DOTTORANDO: **Michele Guindani**

TITOLO: **Bayesian Nonpararametric Analysis of Spatial Data.**

ABSTRACT: Many real life phenomena which arise in such diverse areas as climatology, geology, medicine, real estate marketing, and so on, appear to be geographically referenced.

In this thesis, we consider point referenced data modeling. In the context of Bayesian inference, many of the models proposed to study the behavior of physical systems with point referenced data are in the form of a spatial random effects model. A hierarchical model is often used to specify the distribution of the observables, and a mean zero gaussian stationary random field is chosen as a prior for the pure spatial term at the second level of the hierarchy. However, the stationarity assumption turns out to be unappropriate in many practical situations, e.g. in most environmental studies, where the phenomena can be highly affected by the topology of the region in which they occur. Also the Gaussian assumption can be seen as inadequate to cope with the distribution of real data.

Gelfand, Kottas and MacEachern (2005) have recently introduced a non-parametric specification at the second level of the hierarchy of the Bayesian spatial random effects model which is particularly promising for dealing with nonstationarity and nongaussianity. They develop a Spatial Dirichlet process (SDP), which is essentially a Dirichlet Process (DP) whose base measure is the distribution of a specified random field.

In this thesis, we address two different issues raised by the use of a SDP specification in a random effects model. First, we study the smoothness properties of the random surfaces sampled from a SDP. In many areas of spatial data analysis, it is often relevant to investigate smoothness of process realizations. For example, in the environmental sciences, it is often of interest to study the rates of change in levels of pollutants in the atmosphere.

Banerjee and Gelfand (2003) and Banerjee, Gelfand and Sirmans (2003) have recently studied the issue, formalizing the notions of directional finite difference processes and directional derivative processes for random fields defined on a subset of $\Re^d$ and providing complete distributional theory under the assumptions of a stationary Gaussian Process model.

In this work, we study smoothness properties of samples from a SDP and obtain distributional theory for the associated directional finite differences and derivative processes. In particular, we show that directional finite differences and derivatives are themselves samples

from a SDP. We use the SDP specification to model the pure spatial component in a random spatial effects model. Inference and prediction is sought by means of a Gibbs sampling scheme. This can be implemented in several ways, thanks to both the mentioned results and the tractable scheme described in Gelfand, Kottas and MacEachern (2005).

The second issue we address is related to the ability of the SDP to capture a wide range of spatial behaviors. In fact, the SDP specification implies that we can actually sample one and only one common surface for all locations.

In this thesis, we introduce a random distribution for the spatial effect that allows different surface selection at different sites. Moreover, we can model the parameters so to preserve the property that the marginal distribution at each site still comes from a Dirichlet process. This is done constructively, extending to a multivariate setting the stick-breaking representation of the weights which is known to characterize the usual Dirichlet Process (see Sethuraman (1994)). Hence, we define a new class of random probability measures for random vectors and processes, which includes the customary Dirichlet process specification as a special case.

Alternatively, we can achieve the same behavior proceeding from a different perspective. It is known that we can define finite dimensional priors converging in distribution to the Dirichlet Process (see Muliere and Secchi (1995) and Ishwaran and Zarepour (2002)). Analogously, we can define a multivariate random distribution which marginally is a finite sum approximation to the Dirichlet process. We discuss both the approaches outlined above and describe their properties, together with model fitting and inference.

References

Banerjee, S. and Gelfand, A. (2003) On smoothness properties of spatial processes. *Journal of Multivariate Analysis*, **84**, 85–100.

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

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

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

Muliere, P. and Secchi, P. (1995) A note on a proper Bayesian bootstrap. *Tech. rep.*, Università degli Sudi di Pavia, Dipartimento di Economia Politica e Metodi Quantitativi.

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

UNIVERSITÀ COMMERCIALE "LUIGI BOCCONI" – MILANO

Facoltà di Economia

Dottorato di Ricerca in Statistica

XVI Ciclo

Bayesian Nonparametric Analysis of Spatial Data

Coordinatore:

Ch.mo Prof. Pietro Muliere

Tesi di:

Michele Guindani

UNIVERSITÀ COMMERCIALE "LUIGI BOCCONI"
ISTITUTO DI METODI QUANTITATIVI

The thesis "**Bayesian Nonparametric Analysis of Spatial Data**" by **Michele Guindani** is recommended for acceptance by the members of the delegated committee, as stated by the enclosed reports, in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

Dated: January 2005

Research Supervisor: **Sonia Petrone**
Pietro Veronese
External Examiners: **Alan E. Gelfand**

UNIVERSITÀ COMMERCIALE "LUIGI BOCCONI"

Date: **January 2005**

Author: **Michele Guindani**

Title: **Bayesian Nonparametric Analysis of Spatial Data**

Department: **Istituto di Metodi Quantitativi**

Signature of Author

Alla mia famiglia

Table of Contents

Introduction

Many real life phenomena which arise in such diverse areas as climatology, geology, medicine, real estate marketing, and so on, appear to be geographically referenced.

Statistical analysis of these phenomena of course vary according to the nature of the data collected. Traditionally, spatial data fall into three basic types. The case typical of geostatistical studies is the one where data are collected at known locations, as it happens for example when studying rainfall at monitoring stations, or pollution measured at fixed sampling sites (*point referenced* data). Otherwise, data can refer to an entire region rather than a precise location. This is the case of *areal* (or *lattice*) data, where spatial association is usually modeled through the definition of a neighborhood structure associated with each region. Sometimes, we don't know a priori where an event could occur, although we could be interested in determining its point location. For example, this happens when studying the spread of a particular disease or predicting the earthquake epicenters (*point pattern* data).

In this thesis, we consider point referenced data modeling. We suppose to observe data at locations $\mathbf{s}$ over D, a fixed subset of $\mathbb{R}^d$, $d \geq 1$. Data are the realization of a univariate random field $\{Y(\mathbf{s}), \mathbf{s} \in D\}$. In the context of Bayesian inference, many of the models proposed to study the behavior of physical systems with point referenced data are in the form of a spatial random effects model. Thus, we explicitly introduce in the model a term meant to capture variability of a pure spatial nature, which is the object of inference together with the other parameters of the model (e.g. the

coefficients of a regressive term). A hierarchical model is often used to specify the distribution of the observables, and a mean zero gaussian stationary random field is chosen as a prior for the pure spatial term at the second level of the hierarchy. In fact, in most of applications also the distribution of the observables is chosen to be conditionally gaussian and stationary.

However, the stationarity assumption turns out to be unappropriate in many practical situations, e.g. in most environmental studies, where the phenomena can be highly affected by the topology of the region in which they occur. Many studies have recently appeared which try to remove the stationarity assumption (see Sampson and Guttorp (1992), Damian *et al.* (2001), Schmidt and O'Hagan (2003)). Also the Gaussian assumption can be seen as inadequate to cope with the distribution of real data, which often show heavy tails or skewness (see, for example, Palacios and Steel (2004)).

Gelfand, Kottas and MacEachern (2004) have recently introduced a non-parametric specification at the second level of the hierarchy of the Bayesian spatial random effects model which is particularly promising for dealing with nonstationarity and nongaussianity. They develop a Spatial Dirichlet process (SDP), which is essentially a Dirichlet Process (DP) whose base measure is the distribution of a specified random field. Although the base measure is taken to be a mean zero stationary gaussian process, it turns out that samples from the SDP are neither gaussian nor stationary. With respect to competing strategies, the SDP modeling thus provides a well-known framework to deal with non stationarity. For theoretical and model fitting purposes the statistician can rely on the well-established theory of the Dirichlet process (see Ferguson (1973)). For example, there exists in the literature a number of algorithms devised for sampling from either the exact or an approximate posterior of the model. As an example of the former, we recall the algorithms described by Escobar (1988), MacEachern (1994), Escobar and West (1995), Bush and MacEachern (1996), MacEachern and Müller (1998)

among others. With regard to the approximate inference, we recall the papers by Muliere and Tardella (1998) and Gelfand and Kottas (2002).

In this thesis, we address two different issues raised by the use of a SDP specification in a random effects model like the one described above. First, we study the smoothness properties of the random surfaces sampled from a SDP. In many applications, we can be interested in estimating the rate of change of a spatial surface at a given location in a given direction. For example, in digital terrain models we could explore surface roughness, in meteorology or environmental sciences we could study temperature, rainfall or pollution gradients, in real estate marketing we could examine how home pricing varies according to the distance from a central business district, and so on. Banerjee and Gelfand (2003) and Banerjee, Gelfand and Sirmans (2003) have recently provided new theory to study these issues. In particular, building upon the existing theory for stochastic processes on $\mathbb{R}$, they have formalized the concept of mean square differentiability of a random field defined on a general set $D \subseteq \mathbb{R}^d$, $d \geq 1$, together with the notions of directional finite difference and directional derivative processes.

Here, we provide conditions under which samples from the SDP are either mean square continuous or mean square differentiable. Moreover, we consider the gradient processes associated with these random surfaces and obtain complete distributional theory results. In particular, we show that the gradient processes are themselves samples from a SDP, whose base measure is the distribution of the gradient process associated with the base measure of the original SDP. Therefore, interesting inference can be carried on through the use of standard results on DP and the results in Banerjee, Gelfand and Sirmans (2003).

The second issue we address is related to the ability of the SDP to capture a wide range of spatial behaviors. Since the SDP is essentially a Dirichlet process defined on a space of surfaces, its realizations are a.s. discrete probability measures with

infinite support (see Ferguson (1973) and Sethuraman (1994)). Therefore, whenever the distribution of the observables is assumed to be conditionally gaussian, our model can be characterized as an infinite mixture of normals, and so in principle it is able to capture virtually any distribution of the observables.

However, the way this is achieved can be unsatisfactory for inferential purposes. In fact, the SDP specification implies that we can actually sample one and only one common surface for all locations. In other words, that could yield a smoothing of the mixturing weights, for example whenever two different spatial random effects are in place at two different sites.

In this thesis, we introduce a random distribution for the spatial effect that allows different surface selection at different sites. Moreover, we can model the parameters so to preserve the property that the marginal distribution at each site still comes from a Dirichlet process. This is done constructively, extending to a multivariate setting the stick-breaking representation of the weights which is known to characterize the usual Dirichlet Process (see Sethuraman (1994)). Hence, we define a new class of random probability measures for random vectors and processes, which includes the customary Dirichlet process specification as a special case. Moreover, this class can be seen as an extension of the generic class of priors described in Hjort (2000) and Ishwaran and James (2001), which also take their aim from the stick-breaking representation.

Alternatively, we can achieve the same behavior proceeding from a different perspective. It is known that we can define finite dimensional priors converging in distribution to the Dirichlet Process (see Muliere and Secchi (1995) and Ishwaran and Zarepour (2002b)). Analogously, we can define a multivariate random distribution which marginally is a finite sum approximation to the Dirichlet process. We discuss both the approaches outlined above and describe their properties, together with model fitting and inference.

In chapter 1, we provide some basic notions about the Dirichlet process and its generalizations proposed in the literature. Here, we also discuss the general Bayesian semi-parametric hierarchical modeling by means of the DP, together with the techniques developed for sampling from the posterior of the parameters involved. Since the generalization we propose can be compared with recent proposals considering random probability measures dependent on the values of some underlying covariates, we discuss the issue, outlining the differences among the several approaches proposed.

In chapter 2, we present the basics of the Bayesian modeling of point referenced data. We also define the Spatial Dirichlet process recently introduced by Gelfand, Kottas and MacEachern (2004) and review its properties. The important concepts and definitions referring to the smoothness of spatial processes are also presented. More specifically, we review the literature concerning a.s. continuity, mean square continuity, and mean square differentiability of spatial processes and introduce the related concepts of directional finite difference and directional derivative processes. For stationary random fields, existence of the directional derivative process requires existence and continuity of all the second-order partial and mixed derivatives of the covariance function. Therefore, in the Appendix to this chapter, we discuss the differentiability of the covariance functions belonging to the Matern class (Matern (1986)), which is a very flexible class of covariance functions and turns out to be vey useful in many practical applications (Handcock and Stein (1993)).

In chapter 3, we apply the concepts presented above to determine the smoothness properties of the surfaces sampled from a Spatial Dirichlet process. According to the intuition, we prove that the smoothness of these processes is connected to the smoothness of the base measure. In fact, we provide conditions for mean square continuity and differentiability of a sample from a SDP. As expected, directional finite difference and directional derivative processes are themselves samples from a SDP, whose base

measure is the distribution of the corresponding gradient process associated with the base measure of the original SDP. We build a semi-parametric hierarchical model to get the desired inference. In particular, we focus on the mean of the distribution of the observables at the first level of the hierarchy and predict the values of its gradients at old and new locations. This can be done in several ways, exploiting the distributional theory results that we have obtained before.

In chapter 4, we propose a generalization of the SDP with the properties described above. That is, we define a random probability measure on a space of surfaces over a region D, whose finite dimensional distributions are a.s. discrete and have infinite support. Moreover, the probabilities at any jump point are obtained by means of a multivariate stick-breaking construction. We show how this specification extends the Sethuraman's representation characterizing the Dirichlet process. As with the SDP, this specification can be usefully employed to model the spatial component in a spatial random effects model. In particular, we show how to specify the stick-breaking components in a way that is appealing for modeling purposes and offers a feasible computational strategy for model fitting. We discuss the features of our approach by means of a simulation example.

In chapter 5, we pursue an alternate strategy. It is known that the Dirichlet Process can be approximated by means of a Dirichlet-Multinomial process (see Muliere and Secchi (1995), or more generally a finite dimensional Dirichlet prior (see Ishwaran and Zarepour (2002b). Those are essentially random probability measures which are a.s. discrete, but whose support is finite and whose weights are Dirichlet distributed. Therefore, we can define a Spatial finite dimensional Dirichlet prior in a way similar to what Gelfand, Kottas and MacEachern (2004) have done for the SDP. Moreover, it is possible to proceed analogously to what done in chapter 4 and define a Generalized Spatial finite dimensional Dirichlet prior. We still model the weights by means of a

Dirichlet distribution, but they are spatially varying and we expect them to follow a reasonable spatial behavior. We discuss the properties of this measure and show that it can be viewed as a very peculiar mixture of finite dimensional priors. Thanks to this characterization, we are able to prove some limiting results, which show that this random probability measure converges to particular instances of the DP for an opportune choice of the parameters of the Dirichlet distribution on the weights.

Finally, we draw some conclusions and briefly discuss some further directions of research.

Chapter 1

An overview of the Dirichlet Process.

This chapter is intended to define and describe the main properties of the Dirichlet Process. The Dirichlet process was discovered by Freedman (1963) through the notion of a tail-free measure (see also Fabius (1964)), and its properties and theory were developed by Ferguson (1973) and Blackwell and MacQueen (1973). In the Bayesian setting, it has been widely applied in hierarchical models. In this context, it usually serves as a flexible way to specify the unknown prior for the parameters of the distribution of the observables at the first stage of the hierarchy. We refer to the seminal paper of Ferguson (1983) for review and properties of the arising mixture model specification.

In section 1.1, we provide some basic notions about the Dirichlet process (DP). In particular, we recall the fundamental definition of a sample from a Dirichlet Process. This definition turns out to be very useful in practical applications, e.g. when we need to define a stage with a nonparametric specification in Bayesian hierarchical models. In section 1.2, we discuss an important characterization of the Dirichlet Process due to Sethuraman (1994) by means of a weighted infinite sum of random indicator variables. More generally, we review a general class of random distributions proposed by Ishwaran and Zarepour (2002b) which are expressible as infinite weighted sums. There are many

other generalizations of the DP taking their aim from the Sethuraman's characterization. We discuss some of the most recent ones in section 1.3. Bayesian semiparametric hierarchical modeling by means of the DP is explicitly introduced in section 1.4. There, we also discuss different techniques to sample from the posterior of the parameters of the model. In particular, we discuss the Pólya Urn scheme based Gibbs sampler and the methods based on approximations of the DP by means of truncations of the infinite sum representation. In section 1.5, we examine dependent random probability measures, that is prior distributions whose realizations depend on the values of some underlying covariates. In particular, we address mixtures of products of Dirichlet Processes (MPDP), originally developed to account for partially exchangeable observations whose distribution is unknown (see section 1.5.1). In section 1.5.2, we deal with random probability measures for vectors taking values on $\mathbb{R}^k$. In particular, we notice that it is always possible to define a DP whose base measure is the joint distribution of a vector on $\mathbb{R}^k$. We thoroughly discuss the differences between this latter approach and the former based on MPDP. It is straightforward to extend the two types of modeling above in order to define random probability measures for stochastic processes. Therefore, in section 1.5.3, we study DP whose base measure is the distribution of a stochastic process. Then, in section 1.5.4, we introduce Dependent Dirichlet processes, which can be seen as an extension of MPDP to an infinite collection of distributions (see MacEachern (2000)) and show how they differ from the latter. Finally, in section 1.5.5, we outline the basics of the approach we pursue in chapter 4. There, we consider a collection of random probability measures for stochastic processes, whose finite dimensional distributions are a.s. DP with state space index-varying weights. In other words, we concede the possibility to have different marginal DP at each index set in the state space of the base stochastic process, a feature not allowed in the models presented in section 1.5.3 and that turns useful in the spatial context we will consider

later.

1.1 Basic notions on the Dirichlet Process.

We start giving a definition of what is meant by a random probability measure. Consider a measurable space $(\Omega, \mathcal{F})$ and the set $\mathbb{P}$ of all the probability measures defined on an arbitrary measurable space $(\mathbb{X}, \mathcal{X})$. Then, if we endow $\mathbb{P}$ with the σ-algebra generated by the collection of sets $\{p(A),\, p \in \mathbb{P}\}$, obtained allowing A to vary in $\mathcal{X}$, and denote it with $\mathcal{A}$, we define a random probability measure on $(\mathbb{X}, \mathcal{X})$ as any function $P(\cdot)$ from Ω into $\mathbb{P}$, which is $\mathcal{F}/\mathcal{A}$-measurable. In particular, if we allow $\mathbb{X}$ to be a separable complete metric space and $\mathcal{X}$ to be its Borel sigma algebra, and $\mathbb{P}$ is endowed with the topology of the weak convergence, then $\mathcal{A} = \mathcal{B}(\mathbb{P})$, and $\mathbb{P}$ can also be metrized as a Polish space. We refer to Regazzini (1996) and Gosh and Ramamoorthi (2003) for a review of the results hereby stated from a full measuristic approach. In the following, we will assume $\mathbb{X}$ to be a complete separable metric space. In particular, it will often be the d-dimensional Euclidean space, for some integer $d \geq 1$. In this chapter, we stick to the notation in Ferguson (1973) and denote with $\mathcal{P}$ a probability measure on $(\mathbb{P}, \mathcal{B}(\mathbb{P}))$. Accordingly, P indicates the random probability measure chosen according to $\mathcal{P}$.

The Dirichlet Process is a particular random probability measure that arises from Dirichlet distributed finite dimensional specifications of $\mathcal{P}$, as the following definition shows.

Definition 1.1.1. *Let $\alpha(\cdot)$ be a finite non-null measure on $(\mathbb{X}, \mathcal{X})$. We say P is a Dirichlet Process (DP) on $(\mathbb{X}, \mathcal{X})$ with parameter α if, for every $k = 1, 2, \ldots$ and measurable partition $(B_1, \ldots, B_k)$ of $\mathbb{X}$, the distribution of $P(B_1), \ldots, P(B_k))$ is Dirichlet, $\mathcal{D}(\alpha(B_1), \ldots, \alpha(B_n))$.*

For reasons that will be clear soon, we may consider the normalized probability measure $H(\cdot) = \alpha(\cdot)/\alpha(\mathbb{X})$ and set $\alpha(\mathbb{X}) = \alpha$. In fact, in Bayesian nonparametric modeling it is customary to refer to P as a Dirichlet Process with parameter αH, and denote it as $P \sim DP(\alpha H)$. From the definition above, it follows that for every $B \in \mathcal{X}$, the law of $P(B)$ is Beta($\alpha H(B)$, $\alpha(1 - H(B))$), with mean $E_{\mathcal{P}}(P(B)) = H(B)$ and variance $Var_{\mathcal{P}}(P(B)) = H(B)(1 - H(B))/(1 + \alpha)$, where $E_{\mathcal{P}}$ and $Var_{\mathcal{P}}$ denote, respectively, the expectation and the variance with respect to $\mathcal{P}$. In particular, notice that as α gets bigger, the variance of the process gets smaller. Because of these results, the distribution H is said to be the base measure and α the precision parameter of the process.

Whenever inference is pursued, the concept of a sample from the random probability measure P plays a key role.

Definition 1.1.2. *Let P be a random probability measure on $(\mathbb{X}, \mathcal{X})$. We say that $X_1, \ldots, X_n$ is a sample of size n from P if for any $m = 1, 2, \ldots$ and measurable sets $A_1, \ldots, A_m, C_1, \ldots, C_n$,*

$$\mathcal{P}\{X_1 \in C_1, \ldots, X_n \in C_n | P(A_1), \ldots, P(A_m), P(C_1), \ldots, P(C_n)\} = \prod_{j=1}^{n} P(C_j) \quad a.s. \tag{1.1.1}$$

In other words, we could say that, given a realization of the process P, the X_i's are indipendent and identically distributed according to P. In Bayesian hierarchical modeling, this is usually expressed in symbols as $X_i | P \overset{i.i.d}{\sim} P$ and $P \sim \mathcal{P}$.

One of the main features of the Dirichlet Process is its conjugacy. Let P be a Dirichlet Process on $(\mathbb{X}, \mathcal{X})$ with parameter αH and let $X_1, \ldots, X_n$ be a sample of size n from P. Then, it turns out that the posterior distribution of P given $X_1 \ldots, X_n$ is still a Dirichlet Process with parameter $\alpha H + \sum_{i=1}^{n} \delta_{X_i}$, where δ_x denotes the measure

on $(\mathbb{X}, \mathcal{X})$ giving mass one to the point $x \in \mathbb{X}$, that is

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

Notice that the parameter of the posterior may be rewritten as $\alpha H + n\hat{P}_n$, where $\hat{P}_n$ is the empirical probability. If we allow $\alpha \to 0$, then the posterior distribution of the process is concentrated on the empirical distribution function. It follows that we may replicate many frequentist results allowing α to get smaller, as Ferguson (1973) showed. For example, under a quadratic loss function, the Bayes estimates of P and of the mean $H(B)$ tend, respectively, to the empirical distribution and the sample mean. This has led many people, including Ferguson, to interpret the parameter α as a measure of faith in the prior guess H. However, it has been shown (see Sethuraman and Tiwari (1982)) that when $\alpha \to 0$, the prior process P, instead of being diffuse, tends to be concentrated at a single value, drawn from the base measure H, so that the previous interpretation has no clear meaning.

1.2 Sum representations of the Dirichlet process.

The Dirichlet process can be defined alternatively, as an infinite weighted sum of indicator variables. In fact, it turns out that any probability measure P drawn from $\mathcal{P}$ is almost surely discrete. This was first showed by Ferguson in his 1973 seminal paper. His definition was based on earlier work by Ferguson and Klass (1972), who provided a representation for the gamma process based on arrival times from a homogeneous Poisson Process. In fact, it is possible to see that there exist many alternate definitions of the Dirichlet process based on sum representations. This result has been shown in details by Ishwaran and Zarepour (2002b). Actually, their arguments encompass any random probability measure that can be expressed as an infinite weighted sum, whose

weights are functions of a continuous, positive, Levy measure. Consider a positive infinitely divisible non-Gaussian random variable J whose characteristic function can be expressed as

$$\phi(\theta) = \exp\left[-\int_0^\infty \{\exp(i\theta u) - 1\}\, dN(u)\right], \quad -\infty < \theta < +\infty,$$

where the Levy measure N is a Borel measure defined on $(0, \infty)$ by $N(x) = \int_x^\infty dN(u)$ and satisfies the integrability condition

$$\int_\varepsilon^\infty N^{-1}(u)\, du < \infty \quad \text{for all } \varepsilon > 0, \tag{1.2.1}$$

where $N^{-1}(u) = \sup\{x : N(x) \leq u\}$. Let E_i, $i = 1, 2, \ldots$ be a sequence of i.i.d. exponentially distributed random variables with unitary mean, and define $\Gamma_k = E_1 + \ldots + E_k$. Let also θ_i^*, $i = 1, 2, \ldots$, be i.i.d. random elements with distribution H over $(\mathbb{X}, \mathcal{X})$. If N is positive and continuous, then by Ferguson and Klass (1972) it follows that $J = \sum_{k=1}^\infty J_k$ almost surely, where $J_k = N^{-1}(\Gamma_k)$. Thus, the weighted sum

$$\sum_{k=1}^\infty \frac{J_k}{\sum_{i=1}^\infty J_i}\, \delta_{\theta_k^*}(\cdot) \tag{1.2.2}$$

is a random probability measure with random weights $p_k = \frac{J_k}{\sum_{i=1}^\infty J_i}$ based on an infinitely divisible distribution. Ferguson (1973) first showed that the Dirichlet process is indeed an example of a random measure expressible as (1.2.2). In fact, it can be obtained when $J = \sum_{k=1}^\infty J_k$ is a Gamma random variable with shape parameter α and scale parameter $\beta = 1$, and N is the Gamma Levy measure defined by

$$N(x) = \alpha \int_x^\infty \exp(-u)\, u^{-1}\, du, \quad x > 0. \tag{1.2.3}$$

Here, $N^{-1}(u)$ is decreasing in u; hence, the weights are decreasingly ordered, that is $p_1 \geq p_2 \geq \ldots$. Their distribution has been studied by Kingman (1975) and is usually referred to as Poisson-Dirichlet distribution. Ishwaran and Zarepour (2002b) showed

that whenever it's possible to define a random probability measure like (1.2.2), this can also be represented through a different set of weights, obtained by means of a random permutation of the original ones. In fact, the following theorem holds.

Theorem 1.2.1. *Suppose that N is a positive and continuous Levy measure satisfying* (1.2.1). *If W_k are i.i.d. positive random variables independent of Γ_k such that $E(W_1^{-1}) = 1$, then*

$$\sum_{k=1}^{\infty} \frac{N^{-1}(\Gamma_k)}{\sum_{i=1}^{\infty} N^{-1}(\Gamma_i)} \delta_{\theta_k^*}(\cdot) \stackrel{\mathcal{D}}{=} \sum_{k=1}^{\infty} \frac{N^{-1}(W_k\Gamma_k)}{\sum_{i=1}^{\infty} N^{-1}(W_i\Gamma_i)} \delta_{\theta_k^*}(\cdot). \tag{1.2.4}$$

Sum representations such as (1.2.2) imply that the random probability measure P drawn from $\mathcal{P}$ is a.s. discrete. However, Ferguson (1973) proves that the topological support of P is generally greater than the set of all discrete probability measures on $(\mathbb{X}, \mathcal{X})$. Recall that the topological support of a probability measure μ defined on a measurable space $(S, \mathcal{S})$, where S is a separable metric space and $\mathcal{S}$ is its Borel class, is the set

$$S_\mu = \bigcap \{F \subset S : \text{ F is closed and } \mu(F^c) = 0\},$$

and $x \in S_\mu$ iff $\mu(A) > 0$ for any open set A such that $x \in A$ (see Regazzini (1996)). Then, if $(\mathbb{X}, \mathcal{X})$ is a Borel space we may consider the topology of the weak convergence on the set of σ-additive measures on $(\mathbb{X}, \mathcal{X})$. With this topology, it can be shown that if $\alpha(\cdot)$ is σ-additive, the support of P is indeed the set of all σ-additive probability measures whose support is contained in the support of $\alpha(\cdot)$.

1.2.1 Sethuraman's construction.

One very useful example of alternate representation of the Dirichlet process that may be obtained through a random permutation of the Poisson-Dirichlet weights is due to

Sethuraman (1994). Let $\theta_1^*, \theta_2^*, \ldots$ be i.i.d. random elements independent and identically distributed according to H, as before. Notice that, in practice, H is usually assumed to be non-atomic, although this is not necessary for definition purposes. Let $q_1, q_2, \ldots$ be random variables independent of the θ^*'s and i.i.d. among themselves with common distribution Beta$(1, \alpha)$. Sethuraman (1994) shows that, 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

$$P(\cdot) = \sum_{k=1}^{\infty} p_k \, \delta_{\theta_k^*}(\cdot), \tag{1.2.5}$$

is distributed according to the Dirichlet measure $\mathcal{P}$. The distribution of the set of weights $\mathbf{p} = (p_1, p_2, \ldots,)$ is widely known in literature as GEM distribution, after Griffiths, Engen and McCloskey (see Ewens (1988)). If we let Y be a unit sample from P, it can be shown that the GEM distribution is the distribution of a permutation of the original Poisson-Dirichlet weights obtained in such a way that

$$\begin{aligned} p_1 &= P(Y = \theta_1^*) \\ p_k &= P(Y = \theta_k^*, Y \neq \theta_i^*, i = 1, \ldots, k-1). \end{aligned} \tag{1.2.6}$$

This is also apparent from the construction of the p's through the q's. In fact, any q_k, $k \geq 0$ can be interpreted accordingly as $q_k = P(Y = \theta_k^* | Y \neq \theta_i^*, i = 1, \ldots, k-1)$ (see Patil and Taillie (1977); Perman *et al.* (1992); Pitman and Yor (1997)). Such a permutation is customarily defined as a *size-biased permutation* of the original set of weights. It entails that the p_i's can be thought of as arising from a stick-breaking procedure: at any stage i we break off what remains of a stick of unitary length according to the random size q_i and assign that part to p_i. Hence, constructions like (1.2.5), whose weights are defined through a similar partitioning of the interval $[0, 1]$, are usually referred to as *stick-breaking*. Pitman (1996) calls them *residual allocation schemes*. These are usually simpler to work with than competing sum representations.

Moreover, they provide an easy way to define a.s. discrete random probability measures that comprise and extend the DP, as we show in the next section.

1.3 Generalizations of the Dirichlet Process.

Recently, there has been a growing literature on new classes of nonparametric priors, intended to constitute either an alternative or an extension of the Dirichlet Process. Most of them get their inspiration from the sum representations of the Dirichlet Process that we have discussed in so far.

We can roughly divide this kind of priors into two groups: the ones that directly consider infinite sums like (1.2.2) and (1.2.5) and obtain the Dirichlet process as a particular case for some values of the parameters; and the ones that take into consideration finite sums eventually converging to the Dirichlet Process in the limit.

A recent work of Ongaro and Cattaneo (2004) stands as an unifying approach between the two perspectives. They discuss a general class of non parametric priors that can be represented as an a.s. discrete random probability measure,

$$P(\cdot) = \sum_{k=1}^{K} p_k \delta_{\theta_k^*(\cdot)}. \tag{1.3.1}$$

Here K is an integer random variable allowed to be eventually infinite, the θ_k^* are i.i.d. from some distribution H, as before and the weights p_k are allowed to have any distribution on the simplex $\{\mathbf{p} : \sum_{k=1}^{K} p_k = 1, p_k \geq 0, k = 1, \ldots, K\}$. Ongaro and Cattaneo extend to this general case results previously proved in Ferguson (1973) for the DP, e.g. computation of the moments, determination of posterior and predictive distributions. So they provide a nice theoretical framework for many generalizations of the same type. In fact, it can be seen that all the extensions of the DP we are going to discuss in the next paragraphs can be comprised in this general framework.

For example, the class of priors defined by Muliere and Secchi (1995) and Ishwaran and Zarepour (2002b) can be obtained setting K finite and modeling the weights to come from a symmetric Dirichlet distribution $(p_1, \ldots, p_K) \sim D(\alpha/K, \ldots, \alpha/K\}$, for some positive α. They have been called "Dirichlet-Multinomial" priors and "finite dimensional Dirichlet priors" respectively by Muliere and Secchi (1995) and Ishwaran and Zarepour (2002b). We adopt Ishwaran and James (2001) notation and indicate the random measure (1.3.1) so specified as $DP_K(\alpha H)$. As $K \to \infty$, this prior converges weakly to the DP with parameters α and H. This result has been proven first by Muliere and Secchi (1995). Recently, it has been slightly extended by Ishwaran and Zarepour (2002b), who show that the same result holds for any continuous functional g of $DP_K(\alpha H)$. We will discuss these issues further in chapter 5.

For now, notice that the $DP_K(\alpha H)$ can be expressed as a mixture of Dirichlet processes in the sense of Antoniak (1974). In fact, given a random sample $\theta^* = (\theta_1^*, \theta_2^*, \ldots, \theta_K^*)$ from H, consider the empirical measure

$$\xi_K(\theta^*, \cdot) = \frac{1}{K} \sum_{k=1}^{K} \delta_{\theta_1^*}(\cdot).$$

Then, given θ^*, $DP_K(\alpha H)$ is distributed like a Dirichlet process with precision parameter α and base measure $\xi_K(\cdot)$, i.e. $DP(\alpha\, \xi_K(\theta^*, \cdot))$. Therefore, conditioning on θ^* and then integrating, it follows that

$$DP_K(\alpha H)(\cdot) \stackrel{d}{=} \int DP(\alpha \xi_K(\theta^*, \cdot))\, H^K(d\theta^*),$$

where $H^K(d\theta^*) = H(d\theta_1^*)H(d\theta_2^*) \ldots H(d\theta_K^*)$, which defines a mixture of Dirichlet processes. For a thorough discussion of the Dirichlet-Multinomial process and the finite dimensional Dirichlet priors, we refer to section 5.1, where we introduce an interesting generalization.

As anticipated in the introduction to this section, other works have moved along a different path, taking explicitly into consideration infinite sum representations. In

particular, Hjort (2000) and Ishwaran and James (2001) have defined extensions of the Dirichlet process by means of a generalization of the stick breaking construction. Recall that in the Sethuraman's representation of the DP the weights are built upon a sequence of i.i.d. random variables $q_k \overset{i.i.d.}{\sim} Beta(1, \alpha)$.

Hjort (2000) extends this setting and considers i.i.d. random variables $q_1, q_2, \ldots$ drawn from an arbitrary distribution Q on $[0, 1]$. Note that Q does not need to be Beta, although setting Q to be just Beta(a, b) gives rise to an attractive class of priors. Hjort (2000) discusses this situation thoroughly, in relation to skewness of random means and consistency of posterior means. Ishwaran and James (2001) work in a similar direction. In particular, they relax the identical distribution hypothesis, allowing the q_k's to be Beta(a_k, b_k). As with Ongaro and Cattaneo (2004), they consider a.s. discrete random probability measures P that can be represented as (1.3.1), where the sum can be either finite or infinite, but the weights are obtained through a stick-breaking procedure. Notice that when K is finite it is necessary to set $q_K = 1$ to ensure that the weights sum to one and (1.3.1) is well defined. Moreover, notice that this definition for the q_k's encompasses most of the alternate measures recently defined in the literature. Among the others, we can mention here the finite dimensional Dirichlet priors discussed above and the two-parameter Poisson-Dirichlet process developed by Pitman and Yor (1997). In fact, since Ishwaran and James assume the $q_k \sim Beta(a_k, b_k)$,$k = 1, 2, \ldots$, it follows that the law for the random weights $\mathbf{p} = (p_1, \ldots, p_K)$ is a Generalized Dirichlet Distribution with parameters $\mathbf{a} = (a_1, \ldots, a_{K-1})$ and $\mathbf{b} = (b_1, \ldots, b_{K-1})$, denoted as $\mathbf{p} \sim GD(\mathbf{a}, \mathbf{b})$ (see Connor and Mosimann (1969)). The models by Muliere and Secchi (1995) and Ishwaran and Zarepour (2002b) result once we consider that any Dirichlet distributed random variable $X \sim D(a_1, \ldots, a_N)$, is $GD(\mathbf{a}, \mathbf{b})$, with $\mathbf{a} = (a_1, \ldots, a_N)$ and $\mathbf{b} = \left(\sum_{n=2}^{N} a_n, \sum_{n=3}^{N} a_n, \ldots, a_N\right)$. Instead, the process described by Pitman and Yor (1997) is a two-parameter stick breaking measure that can be obtained setting

$a_k = 1 - a$, $b_k = b + ka$, $k = 1, 2, \ldots$, for some $a \in [0, 1)$ and $b \in (-a, \infty)$. In the DP, $a = 0$, $b = \alpha$. Another important example is given when $a = \gamma$ and $b = 0$. This selection of shape parameters yields a measure whose random weights are based on a stable law with index $\gamma \in (0, 1)$.

We conclude this section pointing to an important result present in Ishwaran and James (2001). When $K = \infty$, the stick-breaking construction does not ensure by itself that the weights actually sum to one. Ishwaran and James provide a necessary and sufficient condition to check if that is indeed the case for the given distributional assumptions on the q'k.

Lemma 1.3.1. *Consider the random measure* (1.3.1), *where the weights are stick-breaking, that is* $p_1 = q_1$, $p_2 = q_2(1 - q_1), \ldots$, $p_k = q_k \prod_{j=1}^{k-1}(1 - q_j), \ldots$, *with* $q_k \sim Q_k$, $k = 1, 2, \ldots$. *Then,*

$$\sum_{k=1}^{\infty} p_k = 1 \quad a.s. \ \textit{iff} \quad \sum_{k=1}^{\infty} E(log(1 - q_k)) = -\infty. \tag{1.3.2}$$

In particular, if $q_k \sim Beta(a_k, b_k)$, $k = 1, 2, \ldots$, *the condition on the right is true if* $\sum_{k=1}^{\infty} log(1 + \frac{a_k}{b_k}) = +\infty$.

The second condition is easy to check. In particular, if $q_k \sim \text{Beta}(1, \alpha_k)$, $\alpha_k > 0$, $k = 1, 2, \ldots$, then $\sum_{k=1}^{\infty} log(1 + \frac{a_k}{b_k}) = \log \prod_{k=1}^{\infty}(1 + \frac{1}{\alpha_k})$. Since $(1 + \frac{1}{\alpha_k}) > 1$ for all k, then $\prod_{k=1}^{\infty}(1 + \frac{1}{\alpha_k}) \to \infty$ and the condition is easily satisfied. These arguments will prove useful later in chapter 4.

1.4 Bayesian semiparametric hierarchical models.

Stick breaking measures have been widely used in Bayesian nonparametric and semiparemetric hierarchical models. In fact, in many situations, it is assumed to observe

data $\mathbf{Y} = (Y_1, \ldots, Y_n)$ that arise from a hierarchical model

$$\begin{aligned} Y_i|\,\theta_i,\,\boldsymbol{\xi} &\overset{ind}{\sim} p\,(Y_i|\,\theta_i, \boldsymbol{\xi}), \quad i = 1, \ldots, n, \\ \theta_i|\,P &\overset{i.i.d.}{\sim} P, \quad i = 1, \ldots, n, \\ P &\sim \mathcal{P} \\ \boldsymbol{\xi} &\sim p\,(\boldsymbol{\xi}), \end{aligned} \tag{1.4.1}$$

where $\boldsymbol{\xi} \in \mathbb{R}^d$, $d \geq 1$ is a (vector valued) parameter and $\theta_i \perp\!\!\!\perp \boldsymbol{\xi}$, for $i = 1, 2, \ldots$. When P is taken to be the Dirichlet process, DP(αH), the model is usually called *DP mixture* model. As realizations of the DP are discrete with probability one, it turns out that one typically observes some clustering of the θ_i's. In fact, the DP mixture can also be seen as a countably infinite mixture. If P is a finite dimensional Dirichlet prior, then we simply call (1.4.1) a finite mixture model. Green and Richardson (2001) prefer to use the term Dirichlet-Multinomial allocation (DMA) model. In particular, in their paper they compare the different clustering structures induced under the two models. It's known that the Dirichlet process prior favors unequally sized groups, meaning that most of the data will typically come from just a few distributions. This unbalancedness of the allocation distribution can persist a posteriori, as shown in Petrone and Raftery (1997). Through entropy related arguments, Green and Richardson (2001) show that this behavior is actually more evident in the DP mixture model than in DMA models.

Notice that the posterior of model (1.4.1) is not available in closed form. Therefore, a number of techniques has been devised in the literature for sampling from either the exact or an approximate posterior of model (1.4.1). These methods will be the subject of the next two sections. In fact, the following ideas will turn useful later in chapter 3 and 4 in order to devise efficient algorithms to get inference on the models there described.

1.4.1 Pólya urn Gibbs samplers.

We recall that, according to the definition in Blackwell and MacQueen (1973), a sequence of random elements $\{X_1, X_2, \ldots\}$ defined on a complete separable metric space $\mathbb{X}$ is called a Pólya urn sequence with parameter μ, if μ is a finite non null measure on $(\mathbb{X}, \mathcal{X})$ and $\forall B \in \mathcal{X}$,

$$P(X_1 \in B) = \frac{\mu(B)}{\mu(\mathbb{X})},$$
$$P(X_{i+1} \in B | X_1, \ldots, X_n) = \frac{\mu_n(B)}{\mu_n(\mathbb{X})},$$

where $\mu_n(\cdot) = \mu(\cdot) + \sum_{j=1}^{n} \delta_{X_n}(\cdot)$. They showed that if $X_1, X_2, \ldots$, is a Pólya sequence, then

(a) $\frac{\mu_n(\cdot)}{\mu_n(\mathbb{X})}$ converges almost surely to a discrete random probability measure μ^*.

(b) μ^* is the Ferguson Dirichlet process with parameter μ.

(c) Given μ^*, $X_1, X_2, \ldots$ are independent with distribution μ^*.

It follows that the Pólya urn scheme provides a way to draw values from the Dirichlet process. This fact has been exploited by several authors to devise Gibbs sampling algorithms for sampling from the posterior of (1.4.1). Among others, we recall the algorithms described by Escobar (1988), MacEachern (1994), Escobar and West (1995), Bush and MacEachern (1996), MacEachern and Müller (1998). Ishwaran and James (2001) have developed similar algorithms for the Pitman and Yor (1997) process, the finite dimensional Dirichlet prior, and, more in general, for any stick breaking prior with a known Pólya urn characterization.

We can summarize such Pólya urn Gibbs samplers as follows. Let $\boldsymbol{\theta}_{-i}$ denote the subvector of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ formed by removing the i-th coordinate. Then to draw values from the posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\xi}|\mathbf{Y})$ of (1.4.1), we iteratively draw values

from the conditional distributions $p(\theta_i|\boldsymbol{\theta}_{-i}, \boldsymbol{\xi}, \mathbf{Y})$, $i = 1, \ldots, n$ and $p(\boldsymbol{\xi}|\boldsymbol{\theta}, \mathbf{Y})$. The latter is obtained by application of usual techniques: it can either be in closed form or it can also require a Gibbs sampling and computation of appropriate full conditionals. The former is a full conditional distribution and is obtained thanks to the Pólya urn characterization. In fact,

$$p(\theta_i|\,\boldsymbol{\theta}_{-i},\,\boldsymbol{\xi},\,\mathbf{Y}) = \frac{q_0\, p(\theta_i|\,\boldsymbol{\xi},\,\mathbf{Y}) + \sum_{\substack{j=1 \\ j\neq i}}^{n} q_j\, \delta_{\theta_j}(\theta_i)}{q_0 + \sum_{\substack{j=1 \\ j\neq i}}^{n} q_j}, \quad i = 1, \ldots, n, \tag{1.4.2}$$

where $p(\theta|\,\boldsymbol{\xi},\,\mathbf{Y})$ is a full conditional for the parametric model

$$\begin{aligned} Y_i|\,\theta,\,\boldsymbol{\xi} &\overset{i.i.d.}{\sim} p\,(Y_i|\,\theta, \boldsymbol{\xi}), \quad i = 1, \ldots, n, \\ \theta &\sim H \\ \boldsymbol{\xi} &\sim p\,(\boldsymbol{\xi}), \\ \theta &\perp\!\!\!\perp \boldsymbol{\xi} \end{aligned}$$

and the coefficients q_0 and q_j are given by

$$\begin{aligned} q_0 &= \alpha \int p(\theta|\,\boldsymbol{\xi},\,\mathbf{Y})\, d\theta \\ q_j &= p(Y_i|\,\theta_j, \boldsymbol{\xi}), \quad j = 1, \ldots, n,\ j \neq i. \end{aligned}$$

However, many authors note that Pólya urn Gibbs samplers can sometime be slow, difficult or impossible to use (see Ishwaran and James (2001). In fact, this methods rely on the specification of the full conditional distribution $p(\theta_i|\,\boldsymbol{\theta}_{-i},\,\boldsymbol{\xi},\,\mathbf{Y})$, so that the θ_i's have to be uptated one at a time. This eventually results in a slowly mixing Markov chain. Moreover, most of the models of type (1.4.1) assume conjugacy between the DP base measure H and the distribution of the observables $[Y_i|\,\theta_i, \xi]$. If that is not the case, then it can be difficult to compute the relevant quantities for the Pólya urn Gibbs sampler, especially the coeffient q_0. See MacEachern and Müller (1998), Walker and

Damien (1998), Neal (2000) for discussions of the subject and proposals of alternate algorithms. Finally, Pólya urn schemes do not allow direct inference on the posterior of the process. For example, this means that it is not possible to sample from the posterior distribution of interesting functionals of the process, like its extreme values or quantiles, unless we turn to some different approach.

1.4.2 Finite truncations of the Dirichlet process.

Whenever interest is on the Dirichlet measure itself, posterior inference has been carried out only approximately, by means of partial sums approximations of the infinite sum (1.2.5). In this section, we will discuss two important examples of such approach, that is the ones developed by Muliere and Tardella (1998) and Gelfand and Kottas (2002). For a review and extensions, we generally refer to Ishwaran and Zarepour (2002b).

Finite truncations can be used either to approximate the prior or the posterior distribution of the process. The work of Muliere and Tardella (1998) is an example of the first case. For any $\varepsilon \in (0,1)$, they define a ε-Dirichlet random probability P_ε as a truncation of the DP prior P such that

$$P_\varepsilon(\cdot) = \sum_{k=1}^{n_\varepsilon} p_k \delta_{\theta_k^*}(\cdot) + r_\varepsilon \delta_{\theta_0^*}(\cdot), \tag{1.4.3}$$

where n_ε and r_ε are random variables defined as

$$n_\varepsilon = \inf\{m \in \mathbb{N} : \sum_{k=1}^{m} p_i > 1 - \varepsilon\}$$

and

$$r_\varepsilon = 1 - \sum_{k=1}^{n_\varepsilon} p_k,$$

while $\theta_0^* \sim H$. This construction ensures that P_ε can be made arbitrarily close to P in the total variation distance, that is, for all $\varepsilon > 0$ it is possible to find a value n_ε such

that

$$d(P_\varepsilon, P) < \varepsilon,$$

where $d(Q_1, Q_2) = \sup\{|Q_1(B) - Q_2(B)|,\, B \in \mathcal{X}\}$.

Moreover, the random n_ε turns out to be a linear transformation of a Poisson distributed random variable . In fact, the authors show that $n_\varepsilon - 1 \sim \text{Poisson}(\lambda)$, where $\lambda = -\alpha \log \varepsilon$. Therefore, it follows that n_ε has first and second moments

$$E(n_\varepsilon) = -\alpha \log \varepsilon + 1$$

and

$$Var(n_\varepsilon) = -\alpha \log \varepsilon.$$

Thus, for any given α, opportune tuning of the parameter ε allows to control how many addends we expect to find in the sum (1.4.3), that is the computational burden of drawing P_ε. Then, sampling from the posterior of models like (1.4.1) can be done in different ways. For a recent method, which also applies to any finite dimensional Dirichlet prior, see the blocked Gibbs sampler developed by Ishwaran and James (2001).

An example of truncation applied to the posterior distribution of the Dirichlet measure, arising from hierarchical models such as (1.4.1), is given by Gelfand and Kottas (2002). Here the partial sum approximation is applied after the model is fitted, so that the procedure does not depend on any random element. In fact, consider a sample $\mathbf{Y} = (Y_1, \ldots, Y_n)$ from (1.4.1) and denote with $[P|\,\mathbf{Y}]$ the posterior distribution of the process, where we use the convenient brackets notation for densities and distributions as in Gelfand and Smith (1990). Then, $[P|\,\mathbf{Y}]$ can be obtained from

$$[P|\,\mathbf{Y}] = \int [P|\,\boldsymbol{\theta}] \times [\boldsymbol{\theta}|\mathbf{Y}]\, d\boldsymbol{\theta}\,, \tag{1.4.4}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$. Notice that (1.4.4) is indeed a mixture of Dirichlet processes according to Antoniak (1974). In fact, $P|\,\boldsymbol{\theta} \sim DP(\alpha_n H_n)$, where $\alpha_n = \alpha + n$ and

$H_n = \alpha H + n\hat{P}_n$, where $\hat{P}_n$ is the empirical probability (see section 1.1). Therefore, it's possible to describe a simple scheme to obtain samples from $[P|\,\mathbf{Y}]$. First, we draw from $[\boldsymbol{\theta}|\mathbf{Y}]$, for example using a conventional Gibbs sampler. At each iteration we get a draw of $\boldsymbol{\theta}$, that we can use to sample $P|\,\boldsymbol{\theta}$ from its distribution. In fact, according to Sethuraman's representation of the Dirichlet process, $P|\,\boldsymbol{\theta}$ has to be a.s. of the form $\sum_{k=1}^{\infty} p_k\,\delta_{\theta_k^*}$, where $\theta_k^* \overset{i.i.d.}{\sim} H_n$ and the weights p_k arise from a stick-breaking construction as described in section 1.3, with $q_k \overset{i.i.d.}{\sim} \text{Beta}(\alpha_n, 1)$. Of course, in practice, it's impossible to construct this infinite sum. Therefore, Gelfand and Kottas suggest to obtain a partial sum approximation, truncating the infinite sum at a $K \in \mathbb{N}$ such that, for any given ε the residual mass assigned by the weights is less than ε, that is

$$E\left(\sum_{k=1}^{K} p_k\right) = 1 - E\left(\prod_{k=1}^{K}(1-q_k)\right) = 1 - \left(\frac{\alpha_n}{\alpha_n+1}\right)^K \geq 1-\varepsilon.$$

Once we have samples from $P|\,\mathbf{Y}$ this way, it is immediate to make arbitrarily accurate inference on any functional of the DP, $T(P)$. We refer to Gelfand and Kottas (2002) for exemplifications, in particular with regard to the distribution of quantiles and extreme values of the process.

As a remark at the end of this section, we just mention the alternative approach recently developed by Teh *et al.* (2004), which also moves from a prior finite sum approximation to the DP, but is based on variational methods, a class of deterministic algorithms that convert inference problems into optimization problems. We refer to their paper for further discussion.

1.5 Dependent random measures.

In section 1.1, we have seen that in principle the DP can be defined on an arbitrary measurable space $(\mathbb{X}, \mathcal{X})$. In particular, this space is usually taken to be a Polish space.

For example, the Blackwell-MacQueen Pólya urn scheme explicitly requires $(\mathbb{X}, \mathcal{X})$ to be Polish.

Many applications in the literature have been limited to nonparametric specifications for probability distributions defined on the real line. However, in some situations we need working with probability distributions defined on $\mathbb{R}^k$. Moreover, there could be an interest in modeling the probability distributions of the observables nonparametrically according to the values of some underlying covariates. For example, the income of the households in a town can come from different distributions according to the age of the people, their gender, the neighborhood they live in, and so on.

In accordance with a habit that has recently taken place in the literature, we refer to these models as Dependent random probability measures models, although we understand that the terminology can be somewhat confusing. In fact, in recent approaches, dependency is often referred more to the samples from the DP than to the random probability measures themselves, as the name instead seems to suggest. We will come back to this point later.

In section 1.5.1, we discuss mixtures of products of Dirichles processes (MPDP), originally introduced by Cifarelli and Regazzini (1978) for partially exchangeable observations. In section 1.5.2, we concentrate on the extension of the univariate Dirichlet prior on $\mathbb{R}$ to a multivariate distribution on $\mathbb{R}^k$. Again with some abuse of terminolohy, we call this the *simple* multivariate DP to distinguish it from other approaches. In section 1.5.3, we extend these ideas from random distributions of random vectors to random distributions of stochastic processes. In section 1.5.4 we examine the idea developed by MacEachern (2000) and discuss Dependent Dirichlet processes (DDP). We show how the DDP could be seen as an extension of a MPDP and how they differ from the simple multivariate DP. We will also present here a recent application of the DDP to random fields, developed by Griffin and Steel (2004). Finally, in section 1.5.5, we

conclude drawing a sketch of the approach that we will follow in chapter 4.

1.5.1 Mixtures of products of Dirichlet Processes.

Consider a vector $\mathbf{X} = (X_1, \ldots, X_n)$ of partially exchangeable observations, i.e. there exists an integer $r \in \{1, \ldots, n\}$ and a permutation $\pi(\cdot)$ of the indexes $\{1, \ldots, n\}$ such that

$$P(X_{\pi(1)} < x_{\pi(1)}, X_{\pi(2)} < x_{\pi(2)}, \ldots, X_{\pi(n)} < x_{\pi(n)} | F_1, F_2) = \prod_{i=1}^{r} F_1(x_{\pi(i)}) \prod_{i=r+1}^{n} F_2(x_{\pi(i)}), \tag{1.5.1}$$

where $F_1(\cdot)$ and $F_2(\cdot)$ indicate two distribution functions. For explanatory purposes, we confine ourself to the case of observations coming from two populations, although the arguments that follow can be repeated for the general n populations case, with $n \geq 2$. If we want to pursue a full nonparametric approach, a prior on (F_1, F_2) is needed. Moreover, if we assume that one subset does not provide any information about the other, we can take F_1 and F_2 as independent. Accordingly, it is possible to assign $p(F_1, F_2) = p(F_1)\, p(F_2)$. For an application of the independent case to change point problems, we refer to Muliere and Scarsini (1985). More usually, we assume that $(X_{\pi(1)}, \ldots, X_{\pi(r)})$, and $(X_{\pi(r)+1}, \ldots, X_{\pi(n)})$ are dependent, so that even F_1 and F_2 must be dependent. A suitable class of priors for (F_1, F_2) is that of mixtures of products of Dirichlet processes. This class has been introduced by Cifarelli and Regazzini (1978), extending the Dirichlet mixture priors developed by Antoniak (1974).

Let Θ_1 and Θ_2 be two d-dimensional random vectors with joint distribution $H(d\theta_1, d\theta_2)$. Then consider functions $\alpha_i(\cdot, \cdot)$, $i = 1, 2$, such that

(a) for every $\boldsymbol{\theta}_i \in \mathbb{R}^d$, $\alpha_i(\cdot, \boldsymbol{\theta}_i)$ is a finite, non-null and non negative measure on the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

(b) for every $A \in \mathcal{B}(\mathbb{R})$, $\alpha_i(A, \cdot)$ is $\mathcal{B}(\mathbb{R}^d)$-measurable.

Then, the vector of random probability measures (F_1, F_2) is distributed as a mixture of products of Dirichlet processes (MPDP), with parameters $\alpha_1(\cdot, \boldsymbol{\theta}_1)$ and $\alpha_2(\cdot, \boldsymbol{\theta}_2)$ and mixing distribution $H(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ respectively. In symbols, we can write

$$(F_1, F_2|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \sim \mathrm{DP}(\alpha_1(\cdot, \boldsymbol{\theta}_1))\ \mathrm{DP}(\alpha_2(\cdot, \boldsymbol{\theta}_2)),$$

and

$$(F_1, F_2) \sim \int \mathrm{DP}(\alpha_1(\cdot, \boldsymbol{\theta}_1))\ \mathrm{DP}(\alpha_2(\cdot, \boldsymbol{\theta}_2))\ dH(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2).$$

Inference on F_1 and F_2 is obtained by means of the Bayes theorem. In fact, it turns out that $F_1, F_2|\mathbf{X}$ is a mixture of products of Dirichlet processes, whose kernel si given by

$$F_1, F_2|\boldsymbol{\theta}_1, \boldsymbol{\theta}_1, \mathbf{X} \sim \mathrm{DP}(\alpha_1(\cdot, \boldsymbol{\theta}_1) + \sum_{i=1}^{r} \delta_{x_{\pi(i)}})\ \mathrm{DP}(\alpha_2(\cdot, \boldsymbol{\theta}_2 + \sum_{i=r+1}^{n} \delta_{x_{\pi(i)}})),$$

and the mixing distribution is the posterior of $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$. For applications of this approach to change point problems, we refer to Muliere and Scarsini (1984), Muliere and Scarsini (1985) , Mira and Petrone (1996), Petrone and Raftery (1997). See Cifarelli, Muliere and Scarsini (1981) and Carota and Parmigiani (1997) for applications to the regression analysis setting.

Although the mixture of products of Dirichlet processes constitues an appealing way to create dependency between random probability distributions, actual specification of the mixing distribution $H(\cdot, \cdot)$ can be difficult. As Walker and Muliere (2003) point out, the kind of dependency is somewhat implicitly defined by the mixturing scheme, and it is not so easy to fix $H(\cdot, \cdot)$ to achieve the needed dependence structure in practical applications. These problems are of course amplified when we need to specify dependence among more than two groups of observations, or when the covariates can even assume a continuum of values.

To overcome at least the first of these problems, Walker and Muliere (2003) have recently introduced a bivariate Dirichlet process model for the joint distribution of F_1

and F_2 such that marginally each F_k, $k = 1, 2$ is a DP with the same parameters and the type of dependence is such that for all sets A,

$$corr(F_1(A), F_2(A)) = \rho \geq 0.$$

This is done using a Dirichlet-multinomial point process N (see Lo (1986)). Let F_0 be a distribution function, $c > 0$ a real number and r a nonnegative integer. Let also $F \sim \text{DP}(c, F_0)$. A point process N on the real line is said to be Dirichlet-Multinomial with parameters (c, F_0, r) if for any k and any partition $(B_1, \ldots, B_k)$ of the real line, the random vector $(N(B_1), \ldots, N(B_k))$ has conditional distribution given F

$$\mathcal{M}(r; F(B_1), \ldots, F(B_k)),$$

where $\mathcal{M}$ denotes a multinomial distribution. Therefore, we can consider the joint distribution of F and N, $p(F, N) = DP(F|c, F_0) \times \mathcal{M}(N|r, F)$. The conditional distribution of F given N can be obtained by sampling $Y_1, \ldots, Y_r$ i.i.d. from F and taking $N(\cdot) = \sum_{j=1}^{r} \delta_{Y_j}(\cdot)$. We can consider the extended model given by

$$p(F_1, N, F_2) = DP(F_1|c, F_0)\mathcal{M}(N|r, F_1)DP(F_2|c + r, F_r, N),$$

where $F_r = \frac{cF_0+N}{c+r}$.

Conditionally, $F_1 \sim DP(c, F_0)$, $N|F_1 \sim \mathcal{M}(r, F_1)$ and $F_2|N \sim DP(c+r, F_r)$. Marginally, $F_2 \sim DP(c, F_0)$ and for all sets A,

$$corr(F_1(A), F_2(A)) = \frac{r}{c + r}$$

. We can conclude that the closeness between the two distributions is reflected by a convenient choice of the parameters c and r. Notice that given N, F_1 and F_2 are independent. It follows that the posterior of $F_1|N$ and $F_2|N$ can be easily obtained by conventional results relating to the DP.

An extension of this work to a multidimensional setting has been recently developed by Muliere, Secchi and Walker (2004). There they extend this methodology to partially exchangeable observations arising by k-trees, that is trees with an infinite number of vertices and such that every vertex in the tree has the same number $k \geq 1$ of children. Consider a stochastic process $X = \{X_\sigma, \sigma \in T\}$ of random variables indexed by the vertices of a k-tree T. X is defined so that it describes a Pólya sequence with parameters c and F_0 along any path of the tree. Therefore, it characterizes a collection of Dirichlet Processes that have same marginal distribution along any path of the tree, but are dependent in that they come from the same father.

For a different way of modeling partially exchangeable observations through particular mixtures of Dirichlet processes in clustering problems, we refer to the hierarchical Dirichlet process recently studied by Teh, Jordan, Beal and Blei (2004).

1.5.2 Random probability measures on $\mathbb{R}^k$.

In section 1.1, we have seen that it is possible to define a DP in an arbitrary measurable space $(\mathbb{X}, \mathcal{X})$. Usually, $\mathbb{X}$ is simply taken to be the real line.

In the previous section, we have discussed mixtures of products of Dirichlet processes in an attempt to provide a formal framework for dealing with partially exchangeable data. We have seen that MPDP arise naturally when we could think of data coming from two or more different clusters, or data generating processes, as it is, for example, in change point problems. However, data are usually dependent just because they come in as samples from dependent random variables $(X_1, \ldots, X_k)$ and their dependency is customarily modeled through a k-dimensional joint distribution. In other words, with MPDP, we assume that a vector of dependent random variables $(X_1, \ldots, X_k)$ is such

that

$$X_1, \ldots, X_k | F_1, \ldots, F_k \sim F_1(x_1) \ldots F_k(x_k),$$

for some distributions $F_1(\cdot), \ldots, F_k(\cdot)$, on which we put a common prior distribution

$$F_1, \ldots, F_k \sim p(F_1, \ldots, F_k).$$

However, we could move along a different path, that is we could assume that $(X_1, \ldots, X_k)$ have common joint distribution

$$X_1, \ldots, X_k \sim F_{X_1,\ldots,X_k}(x_1, \ldots, x_k)$$

and then give a prior probability to the whole $F_{X_1,\ldots,X_k}(x_1, \ldots, x_k)$. Of course, this prior can simply be taken as a DP prior on the k-dimensional euclidean space, that is we take $\mathbb{X} = \mathbb{R}^k$ and $\mathcal{X} = \mathcal{B}(\mathbb{R}^k)$. We will often denote $F_{X_1,\ldots,X_k}(x_1, \ldots, x_k)$ by $F^{(k)}$ or notations alike. Then, we can consider the following hierarchical model

$$\begin{aligned} X_1, \ldots, X_k | F^{(k)} &\sim F^{(k)} \\ F^{(k)} &\sim DP(\alpha F_0^{(k)}), \end{aligned}$$

where $F_0^{(k)}$ shortly denotes a base joint distribution function, $F^0_{X_1,\ldots,X_k}(x_1, \ldots, x_k)$, on $\mathbb{R}^k$ and α is the precision parameter of the DP. Since this is a DP defined on $\mathbb{R}^k$, with some abuse of terminology, we can call it a *multivariate* DP. In particular, notice that if we are interested in the marginal distribution of a subset of random variables $(X_1, \ldots, X_j)$, $j < k$, then

$$\begin{aligned} X_1, \ldots, X_j | F^{(j)} &\sim F^{(j)} \\ F^{(j)} &\sim DP(\alpha F_0^{(j)}), \end{aligned}$$

where $F_0^{(j)}$ is the marginal distribution of $F_0^{(k)}$ after integrating with respect to the variables not included in the subset. Accordingly, if we take $j = 1$, we get back to a

univariate DP. This way, we are led to consider the collection of Dirichlet processes $(F_1, \ldots, F_k)$, where F_i is the random marginal distribution of X_i, $i = 1, \ldots, k$. This random marginal probability measures are of course dependent, and their dependence structure is driven by the joint distribution function $F_0^{(k)}$. In fact, consider

$$\begin{aligned} P(X_1 < x_1, \ldots, X_k < x_k) &= E_{\mathcal{P}} \left(F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) \right) \\ &= F^0_{X_1\ldots,X_k}(x_1, \ldots, x_k). \end{aligned}$$

This marks the difference with the MPDP setting. In fact, notice that if we decide to model the vector $(X_1, \ldots, X_k)$ with a MPDP, the marginal distribution would be

$$P(X_1 < x_1, \ldots, X_k < x_k) = \int \prod_{i=1}^{k} F_0(x_i; \boldsymbol{\theta}_i) \, dH(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k),$$

where $F_0(x; \boldsymbol{\theta}_i) = \frac{\alpha((-\infty, x], \boldsymbol{\theta}_i)}{\alpha(\mathbb{R}, \boldsymbol{\theta}_i)}$, $i = 1, \ldots, k$.

We can see this more clearly if we consider how a simple change point problem could be modeled in the multivariate DP case. For the sake of simplicity, consider a set $(X_1, \ldots, X_n)$ that collects the values of a random variable X_i observed at discrete times $i = 1, \ldots, n$. Let r be a change point, $r \in \{1, \ldots, n\}$. Then, we could assume that

$$\begin{aligned} X_1, \ldots, X_r | F_1^{(r)} &\sim F_1^{(r)} \\ F_1^{(r)} &\sim DP(\alpha F_0^{(r)}(\cdot, \boldsymbol{\theta}_1))) \end{aligned}$$

and

$$\begin{aligned} X_{r+1}, \ldots, X_n | F_2^{(n-r+1)} &\sim F_2^{(n-r+1)} \\ F_2^{(n-r+1)} &\sim DP(\alpha F_0^{(n-r+1)}(\cdot, \boldsymbol{\theta}_2))). \end{aligned}$$

Notice that this model is actually an extension of the MPDP setting to mixtures of products of multivariate DP. Then,

$$P(X_1 < x_1, \ldots, X_n < x_n) = \int F_0^{(r)}(x_1, \ldots, x_r; \boldsymbol{\theta}_1) \, F_0^{(n-r+1)}(x_{r+1}, \ldots, x_n; \boldsymbol{\theta}_2) \, dH(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2),$$

and if we compare this expression with (1.5.1), we can conclude that the two models marginally coincide when we assume that the base joint distribution is the distribution of i.i.d. variables, i.e. for any integer k we assume that $F_0^{(k)}(x_1, \ldots, x_k; \boldsymbol{\theta}_i) = \prod_{j=1}^{k} F_0(x_j, \boldsymbol{\theta}_i)$, $i = 1, 2$. However, notice that the two models still differ in that in the multivariate DP case we seek to model the joint random distribution $F_{X_1,\ldots,X_k}$ and a collection of marginal random ditributions $(F_1, \ldots, F_k)$ is deduced, while in the mixture of products of Dirichlet processes we directly consider a collection of marginal random ditributions $(F_1, \ldots, F_k)$ and we put a suitable prior on it. Therefore, while in the multivariate DP we can define a joint Sethuraman's representation for $F_{X_1,\ldots,X_k}$, this is not generally possible for mixtures of products of Dirichlet processes.

1.5.3 Random probability measures of stochastic processes.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and T any index set. Consider the (possibly vector valued) stochastic process $X(\omega) : \{X_t(\omega),\, t \in T\}$. Then, recalling the discussion in the previous section, it's easy to define a random probability measure F for the stochastic process $X(\omega)$.

In fact, consider a k-tuple $(t_1, \ldots, t_k)$ of distinct elements of T and the corresponding vector $(X_{t_1}(\omega), \ldots, X_{t_k}(\omega))^T$. Then, for any k-tuple $(t_1, \ldots, t_k)$, we can define the finite dimensional distribution of the stochastic process X as a multivariate DP, i.e.

$$\begin{aligned} X_{t_1}, \ldots, X_{t_k} | F^{(k)} &\sim F^{(k)} \\ F^{(k)} &\sim DP(\alpha F^0_{X_{t_1},\ldots,X_{t_k}}), \end{aligned} \tag{1.5.2}$$

where, for simplicity, we denote by $F^{(k)}$ the finite dimensional distribution $F_{X_{t_1},\ldots,X_{t_k}}$, α is the precision parameter and $F^0_{X_{t_1},\ldots,X_{t_k}}$ is the finite dimensional distribution of a base stochastic process $X_0(\omega) = \{X_{0,t}(\omega),\, t \in T\}$. For $k = 1$, we go back to the customary univariate Dirichlet process. Of course, this way we get a well-defined

random probability measure F for the whole process X.

In fact, Kolmogorov's consistency conditions are easily checked. It is sufficient to recall Sethuraman's infinite sum representation and consider

$$P(X_{t_1} \in A_1, \ldots, X_{t_k} \in A_k | F^{(k)}) = \sum_{i=1}^{\infty} p_i \, \delta_{X^i_{0,t_1}(\omega)}(A_1) \cdots \delta_{X^i_{0,t_k}(\omega)}(A_k), \qquad (1.5.3)$$

for any collection of sets $(A_1, \ldots, A_k)$ belonging to the Borel space of the state space of the process X, where the vector $\mathbf{X}^i_0 = \left(X^i_{0,t_1}(\omega), \ldots, X^i_{0,t_k}(\omega)\right)^T$, $i = 1, 2, \ldots$ gathers the values of a realization of the base process $X_0(\omega)$ at points $(t_1, \ldots, t_k)$ for any $i = 1, 2, \ldots$. Then, the set of finite dimensional distributions (1.5.3) defines the law of the stochastic proess X whenever the set of finite dimensional distributions $F^0_{X_{t_1}, \ldots, X_{t_k}}$ does.

Notice that the set of random weights p_i is the same for all finite dimensional distributions. In fact, any p_i defines the probability of picking up a whole realization $\mathbf{X}^i_0(\omega)$ of the base process. We can see this considering for example a stochastic process on the positive real line, i.e. $T = \mathbb{R}^+$. This is the case typical of phenomena evolving with time. Using (1.5.2), for any given set of weights p_i and realizations X^i_0, $i = 1, 2, \ldots$, once we have observed a value, say X^j_{0,t_1}, in a point t_1 we can only observe values from the same realization X^j_0 in any other point $t \neq t_1$. In other words, we are not allowed the possibility to choose different realizations of the base process X_0 at different times. This could also be stated saying that the random weights do not vary with $t \in T$. Therefore, these models have been called, not without some abuse, *single-p* Dependent Dirichlet processes by MacEachern (2000). Recent applications include the ANOVA model of De Iorio *et al.* (2004) and the Spatial Dirichlet Process of Gelfand *et al.* (2004). We will discuss the latter thoroughly in chapter 2.

1.5.4 Dependent Dirichlet processes.

A model recently proposed in the literature aimed to account both for dependency driven by the realizations of a stochastic process and index-varying weights in the context of Bayesian nonparametrics has been described by MacEachern (2000).

Let $\boldsymbol{\theta} = \{\boldsymbol{\theta}_t(\omega),\, t \in T\}$ and $\boldsymbol{\xi} = \{\boldsymbol{\xi}_t(\omega),\, t \in T\}$ be two (possibly vector valued) independent stochastic processes with index set T and state space S. Usually, $S = \mathbb{R}^k$. Now, consider a univariate or multivariate DP whose base c.d.f., $F_{0,\boldsymbol{\theta}_t}$, depends upon the values of $\boldsymbol{\theta}_t$. Moreover, let the random weights vary with t and depend on $\boldsymbol{\xi}$ and denote them by $p_{i,\boldsymbol{\xi}}(t)$. Both $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ are assumed to be separable. Then, at any t in T, we can define a random probability measure F_t, which depends on $\boldsymbol{\theta}_t$ and $\boldsymbol{\xi}_t$. In practical applications, we could think that $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ are related to some process of covariates. Then, Dependent Dirichlet processes (DDP) generalize the customary Dirichlet process to allow for a collection of nonparametric distributions, the realizations of which are dependent, with the level of the covariate governing the degree of dependence. In fact, for any t in T and for any given realization of $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$, the random distribution function F_t is defined to be the discrete distribution characterized by

$$P(X_t \in A) = \sum_{i=1}^{\infty} p_{i,\boldsymbol{\xi}}(t)\, \delta_{X_{0,\boldsymbol{\theta}_t}}(A),$$

for any $Ain\mathcal{X}$, where $X_{0,\boldsymbol{\theta}_t}$ are independent samples from $F_{0,\boldsymbol{\theta}_t}$, the weights $p_{i,\boldsymbol{\xi}}(t)$ are obtained as in Sethuraman's representation for each t. In other words, given $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$, $F_t \sim DP(\alpha\, F_{0,\boldsymbol{\theta}_t})$, where α could eventually depend on $\boldsymbol{\xi}_t$.

For vectors $(\boldsymbol{\theta}_{t_1}, \ldots, \boldsymbol{\theta}_{t_k})$ and $(\boldsymbol{\xi}_{t_1}, \ldots, \boldsymbol{\xi}_{t_k})$ $k = 1, 2, \ldots$, we get a collection of random distributions $(F_{t_1}, \ldots, F_{t_k})$, whose dependence is governed by the law of the stochastic processes $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$.

Notice that if T is finite and the weights do not depend on $\boldsymbol{\xi}$, this model reduces to a MPDP. In fact, we could see the DDP as an extension of the MPDP for more

general T and index-varying weights. Therefore, we could repeat the same arguments outlined in section 1.5.2 to illustrate the difference with respect to random probability measures of stochastic processes. In particular, both the approaches reduce to a simple DP marginally at each t, but, whenever we consider a multivariate sample from a DDP, it's not possible to get any joint Sethuraman's infinite sum representation of associated random joint distribution. In fact, as we already pointed out in section 1.5.3, the so-called single-p Dependent Dirichlet processes are not even in the DDP setting, but they more properly describe a random probability measure of a stochastic process.

An interesting implementation of the DDP idea has been recently provided by Griffin and Steel (2004). They move from the Sethuraman's constructive representation and notice that the distribution of P is unchanged by permuting the atoms $\{X^i_{0,\theta}, q_{i,\xi}\}$, where the $q_{i,\xi}$'s represent the components of the stick-breaking construction defining the weights. Therefore, given any realization of $\boldsymbol{\xi}$, define a process of permutations $\pi(\boldsymbol{\xi}) = \{\pi(\boldsymbol{\xi}_t), t \in T\}$ of the set of the integers $\{1, 2, \ldots\}$, and consider

$$p_{i,\xi} = q_{\pi_i(\xi_t)} \prod_{\{j:\pi_j(\xi_t)<\pi_i(\xi_t)\}} (1 - q_j).$$

An interesting implementation of the DDP in the context of spatial processes has been recently provided by Griffin and Steel (2004). They move from the Sethuraman's constructive representation and notice that the distribution of P is unchanged by permuting the atoms $\{X^i_{0,\theta}, q_{i,\xi}\}$, where the $q_{i,\xi}$'s represent the components of the stick-breaking construction defining the weights. Then, consider any realization of the stochastic process $\boldsymbol{\xi}$ and, for simplicity, denote it simply by $\boldsymbol{\xi}$. Then, it's possible to consider a process of permutations $\pi(\boldsymbol{\xi})$ and define

$$p_{i,\xi} = q_{\pi_i(\xi)} \prod_{\{j:\pi_j(\xi)<\pi_i(\xi)\}} (1 - q_j).$$

It is immediate to see that here the dependence among the marginal random distributions is directly deduced by the permutation chosen at each t. This model is denoted by *Order-Based* Dependent Dirichlet Processes, abbreviated as πDDP and characterised by a mass parameter α, centering distribution F_0 and a stochastic process $\pi(\boldsymbol{\xi})$.

Griffin and Steel concentrate on a specific class of varying orderings that are defined by a driving point process Φ and a sequence of sets $U(\boldsymbol{\xi}_t)$ for all values $\boldsymbol{\xi}_t \in S$. $U(\boldsymbol{\xi}_t)$ defines the region in which points are relevant for determining the ordering corresponding to the realized $\boldsymbol{\xi}_t$. The permutation at any $\boldsymbol{\xi}_t \in S$ is chosen to satisfy

$$||\boldsymbol{\xi}_t - z_{\pi_1(\boldsymbol{\xi}_t)}|| < ||\boldsymbol{\xi}_t - z_{\pi_2(\boldsymbol{\xi}_t)}|| < \dots,$$

where $||\cdot||$ is an opportune distance measure and z_i are realizations of the point process Φ. It follows that a realization of this process will necessarily be the same for some regions of S. Then, they associate each atom $\{X^i_{0,\boldsymbol{\theta}}, q_{i,\boldsymbol{\xi}}\}$ with the corresponding z_i according to the resulting permutation, so that $\{z_i, X^i_{0,\boldsymbol{\theta}}, q_{i,\boldsymbol{\xi}}\}_{i=1}^{\infty}$ can be thought of as a marked point process from which we can define the distribution F_t for any value $\boldsymbol{\xi}_t$ and $\boldsymbol{\theta}_t$, $t \in T$. Therefore, Griffin and Steel define a DDP that in principle allows for different stick-breaking constructions at different t, although those result from a permutation of the same set of stick-breaking components, which depend on the realized values both of $\boldsymbol{\xi}_t$ and the underlying point process Φ. Moreover, here the dependence structure depends on the notion of norm considered and for practical purposes it can be difficult to model the type of dependence induced by the point process mechanism.

In fact, the dependence structure can be summarized by the correlation between the random distributions at points $t_1, t_2 \in T$. Let A_{1i} be the set of points before the i-th atom in the permutation at $\boldsymbol{\xi}_{t_1}$ and let A_{2i} be the set of points before the i-th atom in the permutation at $\boldsymbol{\xi}_{t_2}$. Then, define $S_i = A_{1i} \cap A_{2i}$ as the set of points before the i-th atom for both $\boldsymbol{\xi}_{t_1}$ and $\boldsymbol{\xi}_{t_2}$, and $S'_i = A_{1i} \cup A_{2i} - S_i$ as the complement of S_i

with respect to $A_{1i} \cup A_{2i}$. Notice that this set depends on the realized value of Φ.

Let Φ be a general point process. Then, Griffin and Steel show that the correlation between F_1 and F_2 is

$$Corr(F_1(A), F_2(A)) = \frac{2}{M+2} E_z \left[\sum_{i=1}^{\infty} \left(\frac{M}{M+2} \right)^{\#S_i} \left(\frac{M}{M+1} \right)^{\#S_i'} \right],$$

where $\#$ denotes the counting measure for the elements of the set and A is any measurable set in the Borel space of the domain of the sample X. For stationary point processes this expectation can be explicitly computed so that

$$Corr(F_1(A), F_2(A)) = \frac{2\lambda}{M+2} \int \left(\frac{M}{M+2} \right)^{\#S(z)} \left(\frac{M}{M+1} \right)^{\#S'(z)} P_0(d\phi)\, dz,$$

where we have written $S(z)$ and $S'(z)$ to explicitly account for the dependence on z and $P_0(d\phi)$ is the Palm distribution of Φ at the origin, i.e. the distribution of the point process conditioned on the point process having a point at the origin.

1.5.5 A Generalized Dirichlet Process.

In section 1.5.3, we have seen that it is possible to use the Dirichlet Process to sample a random distribution for a stochastic process $X = \{X_t(\omega),\, t \in T\}$. However, when we sample from any given distribution drawn from this DP, all the values X_t at any point t come from one and only one realization X_0 of the base stochastic process. In fact, in the Sethuraman's representation of the DP, the random weights define the probability to pick a up a whole realization of the base stochastic process and we cannot choose different realizations at different indexes $t \in T$.

The Dependent Dirichlet processes discussed in section 1.5.4 can be conveniently introduced to allow more flexibility in the previous setting. However, one major drawback is that the Sethuraman's representation is preserved only marginally at each $t \in T$, while the joint distribution at any k-uple $(t_1, \ldots, t_k)$ has no simple form.

In this section, we introduce a model aimed at dealing with both the issues raised above. In particular, we extend the DP setting in order to allow index-varying weights. In accordance with the discussion in section 1.5.3, this is done through the specification of all random finite dimensional distributions of the stochastic process X. This is actually the model that we will use and discuss later in chapters 4 and 5, where it will be explicitly applied in a spatial context. However, the model is indeed quite general and capable of being implemented in a variety of situations. Therefore, this section is meant to stress the differences between this approach and the ones presented in the previous sections, while we postpone the discussion of most of its properties and specific implementation to the already mentioned chapters.

Let $X(\omega) : \{X_t(\omega),\, t \in T\}$ be a stochastic process with index set T and state space S and consider a k-tuple $(t_1, \ldots, t_k)$ of distinct elements of T together with the corresponding vector $(X_{t_1}(\omega), \ldots, X_{t_k}(\omega))^T$. Then, we can define a random probability measure F on the space of the realizations of X as that measure whose finite dimensional distributions almost surely have the following representation: for any $k \in \mathbb{N}$, for any k-tuple $(t_1, \ldots, t_k)$ and for any $(A_1, \ldots, A_k)$ in $\mathcal{B}(S)$,

$$P(X_{t_1} \in A_1, \ldots, X_{t_k} \in A_k) = \sum_{i_1,\ldots,i_n=1}^{\infty} p_{i_1,\ldots,i_n}(t_1, \ldots, t_n)\, \delta_{X^{i_1}_{0,t_1}}(A_1) \cdots \delta_{X^{i_k}_{0,t_k}}(A_k), \tag{1.5.4}$$

where each $\mathbf{X}^{i_j}_0 = (X^{i_j}_{0,t_1}, \ldots, X^{i_j}_{0,t_k})$ is a realization of a base process X_0 and the weights (from now on denoted simply with $p_{i_1,\ldots,i_n}$) can arise either as realizations of a stochastic process independent of X_0 or as realizations of a random probability functional of an underlying stochastic process (also independent of X_0).

For example, let $\{Z_t(\omega), t \in T\}$ be a stochastic process with state space $\mathbb{N}$ whose distribution G is random. For any realization of G, we could define

$$p_{i_1,\ldots,i_k} = P(X_{t_1} = X^{i_1}_{0,t_1}, \ldots, X_{t_k} = X^{i_k}_{0,t_k}) = P(Z_{t_1} = i_1, \ldots, Z_{t_k} = i_k).$$

It follows that any realization of the random measure F arising from (1.5.4) defines a consistent sequence of finite dimensional distribution for the stochastic process $X(\omega)$.

In fact, for any k-tuple $(t_1, \ldots, t_k)$, $k \in \mathbb{N}$, and any $l = 1, \ldots, k$,

$$p_{i_1,\ldots,i_{l-1},\cdot,i_{l+1},\ldots,i_k} = \sum_{l=1}^{\infty} p_{i_1,\ldots,i_{l-1},l,i_{l+1},\ldots,i_k}$$
$$= \sum_{l=1}^{\infty} P(Z_{t_1} = i_1, \ldots, Z_{t_{l-1}} = i_{l-1}, Z_{t_l} = l, Z_{t_{l+1}} = i_{l+1}, \ldots, Z_{t_k} = i_k).$$

Then, Kolmogorov's consistency conditions are easily derived from the existence of the stochastic process $Z(\cdot)$. In fact, if we set $A_l = S$ in 1.5.4, we get

$$P(X_{t_1} \in A_1, \ldots, X_{t_{l-1}} \in A_{l-1}, X_{t_l} \in S, X_{t_{l+1}} \in A_{l+1}, \ldots, X_{t_k} \in A_k) =$$
$$= \sum_{i_1,\ldots,i_k=1}^{\infty} p_{i_1,\ldots,i_k} \delta_{X_{0,t_1}^{i_1}}(A_1) \cdots \delta_{X_{0,t_l}^{i_l}}(S) \cdots \delta_{X_{0,t_k}^{i_k}}(A_k)$$
$$= \sum_{i_1,\ldots,i_{l-1},i_{l+1},\ldots,i_k=1}^{\infty} \delta_{X_{0,t_1}^{i_1}}(A_1) \cdots \delta_{X_{0,t_k}^{i_k}}(A_k) \left(\sum_{l=1}^{\infty} p_{i_1,\ldots i_{l-1},l,i_{l+1},\ldots,i_k} \right)$$
$$= \sum_{i_1,\ldots,i_{l-1},i_{l+1},\ldots,i_k=1}^{\infty} p_{i_1,\ldots i_{l-1},i_{l+1},\ldots,i_k} \delta_{X_{0,t_1}^{i_1}}(A_1) \cdots \delta_{X_{0,t_k}^{i_k}}(A_k)$$
$$= P(X_{t_1} \in A_1, \ldots, X_{t_{l-1}} \in A_{l-1}, X_{t_{l+1}} \in A_{l+1}, \ldots, X_{t_k} \in A_k).$$

Notice that the process defining the probability functionals for the weights can be more complicated. For example, if we look at the random distribution at any t, we could allow the possibility to define a process that at least marginally retains the Sethuraman's representation. Therefore, we could choose to model directly the stick-breaking components as follows. Then, for any $t \in T$ and for $l = 1, 2, \ldots$, consider the random events $\Theta_l(t) = \{X = X_{0,t}^l\}$ and their complements $\Theta_l^c(t) = \{X \neq X_{0,t}^l\}$. Let us denote $\Theta_l^u(t)$, with $u = 0$ or 1 according if we consider respectively the event or its complement. Then, let $\{p_1(t), p_2(t), \ldots\}$ denote the sequence of weights arising marginally at a point t in T. We can rewrite the stick-breaking components as

$$q_l = P\left(\Theta_l^1(t) \middle| \Theta_i^0(t),\, i < l\right).$$

Let $\{Z_t^l(\omega), t \in T\}_{l=1}^{\infty}$ be a countable collection of independent stochastic processes defined on T, identically distributed according to a random distribution G and with support at least on a real interval containing 0. The call for a random distribution G has to be intended in a wide sense. For example, in a parametric hierarchical bayes setting, $Z_t(\omega)$ could be a gaussian process with unknown expected value, on which we place a convenient prior distribution.

Let $\{\delta_t^l, t \in T\}$ be the process such that

$$\delta_t^l = 1 \quad \text{if} \quad Z_t^l \geq 0$$
$$\delta_t^l = 0 \quad \text{if} \quad Z_t^l < 0.$$

Then, we can assume that the distribution of the random events $\Theta_l(t)$ is the same as that of the process δ_t^l, so that

$$q_l = P(\delta_t^l = 1 | \, \delta_t^i = 0, i < l) = P(\delta_t^l = 1),$$

where the last equality follows from the independence of the processes Z^l.

Therefore, we can define the marginal weights as random probability functionals of an underlying stochastic process through the stick-breaking mechanism.

Of course, the considerations above do not depend on the particular choice of the latent process we made. The key feature is the existence of a random partition expressed by the event $\Theta_l(t)$ at any point t .

In chapter 4, following arguments similar to those outlined above, we show that it is possible to define a multivariate stick-breaking construction also for the specification of the weights $p_{i_1,\ldots,i_n}$ in the joint distribution (1.5.4). This is done in the context of spatial modeling, but can be seen in a general framework. Of course, the resulting behavior can be quite different from that of the multivariate DP, which stands as a particular case of our generalization. We discuss thoroughly the connections between

the two, taking as a reference the Spatial Dirichlet Process introduced by Gelfand, Kottas and MacEachern (2004).

In chapter 5, we move from the ideas characterizing the finite sum approximations to the Dirichlet process and define a multivariate random distribution which is a generalization of the Dirichlet Multinomial process or, more generally, of the finite dimensional Dirichlet priors. There, (1.5.4) is a finite sum, and the weights $p_{i_1,\ldots,i_n}$ are Dirichlet distributed. Since we apply this model to spatial data analysis, we also require some desirable spatial property to be satisfied. In fact, in accordance with the results in Ishwaran and Zarepour (2002b), the modeling of the weights is of crucial importance to determine the behavior of the process, in particular, its limiting behavior. We discuss thoroughly the properties of this measure, especially with regard to their implication for the analysis of spatial data. This is still work in progress. However, we are able to show that this finite dimensional specification joins easiness of interpretation together with a simple Gibbs sampler implementation.

Chapter 2

Bayesian Hierarchical Spatial Modeling.

In this chapter, we consider models for spatially distributed data. These are relevant in many diverse fields, such as the environmental sciences, epidemiology and public health, urban economics, geophysics, and so on. Since it is impossible to provide here a comprehensive account of the subject, we will only discuss the relevant features in order to introduce the models we will use in the next chapters, and refer the interested reader to the monographies of Cressie (1993), Stein (1999) and Banerjee, Carlin and Gelfand (2004) for further details.

In section 2.1 we address the basics of hierarchical spatial modeling for point-referenced data models. There we introduce the important concepts of stationarity and isotropy and address spatial random effects models and their fitting. In section 2.2, we describe the Spatial Dirichlet Process model described by Gelfand, Kottas and MacEachern (2004). We introduce the hierarchical model that we will use again later in chapter 3. In section 2.3, we will discuss smoothness properties of spatial processes. In particular, we will present the results on a.s. continuity of random field given by Kent (1989). We will put more emphasis on mean square continuity and mean square differentiability (see Stein (1999)). We introduce the directional finite difference and

derivative process studied by Banerjee and Gelfand (2003) and Banerjee, Gelfand and Sirmans (2003). Finally, in the Appendix, we derive some smoothness property of the Matern covariance function, working directly on its derivatives.

2.1 Basics of point-referenced data models.

We assume data come from the observation of a univariate stochastic process $\{Y(\mathbf{s}),\ \mathbf{s} \in D\}$, where D is a fixed region of the Euclidean space $\mathbb{R}^d$, $d \geq 1$, and contains a d-dimensional rectangle of positive measure. In practice, however, actual data will be a partial realization of the process observed at a finite set of locations, say $(\mathbf{s}_1, \ldots, \mathbf{s}_n)$. Based on such data, the statistician is interested in inference on the process $Y(\mathbf{s})$ and prediction at new locations in D.

2.1.1 Stationarity and isotropy.

In order to achieve those goals, it is necessary to make some assumptions on the process itself. The most common one is usually to assume some sort of stationary behavior. In particular, the process is said to be *strictly stationary* if, for all finite n, $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$, $t_1, \ldots, t_n \in \mathbb{R}$, and $\mathbf{h} \in \mathbb{R}^d$

$$P(Y(\mathbf{s}_1) \leq t_1, \ldots, Y(\mathbf{s}_n) \leq t_n) = P(Y(\mathbf{s}_1 + \mathbf{h}) \leq t_1, \ldots, Y(\mathbf{s}_n + \mathbf{h}) \leq t_n).$$

Let $Y(\mathbf{s})$ be a random field with finite first two moments. Then, it is said to be *weakly stationary* if it has constant mean and

$$Cov(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = K(\mathbf{h}), \tag{2.1.1}$$

for all $\mathbf{h} \in \mathbb{R}^d$ such that $\mathbf{s}$ and $\mathbf{s} + \mathbf{h}$ both lie in D. In other words, the covariance of a weakly stationary process depends only on the separation vector $\mathbf{h}$. Notice that a

strictly stationary random field with finite second moments is also weakly stationary. The converse is not generally true, unless for Gaussian processes. Stationarity can be thought of as an invariance property under translation of the coordinates. We can also consider invariance under rotations and reflections. In fact, a random field is said to be *strictly isotropic* if, for any $d \times d$ orthogonal matrix $\mathbf{H}$ and any $\mathbf{h} \in \mathbb{R}^d$,

$$P(Y(\mathbf{s}_1) \leq t_1, \ldots, Y(\mathbf{s}_n) \leq t_n) = P(Y(\mathbf{H}\mathbf{s}_1 + \mathbf{h}) \leq t_1, \ldots, Y(\mathbf{H}\mathbf{s}_n + \mathbf{h}) \leq t_n),$$

for all finite n, $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$, $t_1, \ldots, t_n \in \mathbb{R}$, and $\mathbf{h} \in \mathbb{R}^d$. Notice that if $Y(\mathbf{s})$ is strictly isotropic, then it is also strictly stationary. A weakly stationary random field is said to be also *weakly isotropic* if the covariance function depends upon the separation vector only through its norm $||\cdot||$. In other words, assuming weak isotropy means that $Y(\mathbf{s})$ has the same covariance structure, roughly the same behavior, along any direction $\mathbf{h}$ we decide to consider. Again, strict isotropy implies weak isotropy, but the converse is not generally true. However, the two notions coincide for Gaussian processes.

2.1.2 Two important examples of isotropic covariance functions.

Isotropic processes are convenient to deal with because of their simplicity and of the existence of a number of relatively simple parametric forms for their covariance functions. We introduce here two of the most common among them, since we will need them later.

Consider first the *exponential* covariance function, defined as

$$K_\phi(||\mathbf{h}||) = \begin{cases} \sigma^2 \exp\big(-\phi||\mathbf{h}||\big) & \text{if } ||\mathbf{h}|| > 0 \\ \tau^2 + \sigma^2 & \text{otherwise,} \end{cases} \tag{2.1.2}$$

for some positive parameters ϕ, σ and τ. The parameter ϕ is usually referred to as a *decay* parameter, since it governs the rate at which the covariance drops as a function

of the distance. The value R at which the correlation has decreased to only 0.05 is usually called the *effective range*. Since $\log(0.05) \approx -3$, it follows that $R \approx 3/\phi$, and so the effective range is inversely proportional to the decay parameter, as we would expect. For reason that we won't investigate here (see Banerjee, Carlin and Gelfand (2004), ch. 2), the sum $\tau^2 + \sigma^2$, σ^2 and τ^2 are respectively called *sill, partial sill and nugget* terms. In particular, since it enters in (2.1.2) only for $||\mathbf{h}|| = 0$, the nugget τ^2 is often viewed as a *non spatial effect variance*. Accordingly, the partial sill σ^2 is seen as a *spatial effect variance*.

Another class of covariance functions that has proven to be attractive in various respects was originally devised by Matern (1986), to whom it gives its name. Handcock and Stein (1993) and Handcock and Wallis (1994) then showed its flexibility in dealing with a variety of spatial data sets, and this eventually led to its widespread use. There are several equivalent expressions for the *Matern* covariance function. We choose one from which it's particularly simple to recover the correlation function of the process. For $\phi > 0$, $\nu > 0$, and $\sigma > 0$, consider

$$K_{\nu,\phi}(||\mathbf{h}||) = \begin{cases} \frac{2^{-\nu+1}\sigma^2}{\Gamma(\nu)} \left(\phi||\mathbf{h}||\right)^{\nu} \mathcal{H}_{\nu}\left(\phi||\mathbf{h}||\right) & \text{if } ||\mathbf{h}|| > 0 \\ \tau^2 + \sigma^2 \quad \text{otherwise,} \end{cases} \tag{2.1.3}$$

where $\mathcal{H}_{\nu}(\cdot)$ is the modified Bessel function of order ν (see Abramowitz and Stegun (1965), ch. 9). The parameter ϕ can be viewed again as a decay parameter and ν is a parameter that controls the degree of smoothness of the process. We will discuss this point later in this chapter. If $\nu = m + \frac{1}{2}$, for some integer m, the function $\mathcal{H}_{m+\frac{1}{2}}(\cdot)$ is a modified spherical Bessel function of the third kind and admits a representation of the form $\exp(-\phi||\mathbf{h}||)$ times a polynomial in $||\mathbf{h}||$ of degree m (see Abramowitz and Stegun (1965), 10.2.15). Then, for $\nu = \frac{1}{2}$, we get the exponential covariance function. For $\nu = \frac{3}{2}$, we get another simple and often used closed form expression, that

is $K(||\mathbf{h}||) = \sigma^2 \exp(-\phi||\mathbf{h}||)\,(1 + \phi||\mathbf{h}||)$, for $||\mathbf{h}|| > 0$.

2.1.3 Spatial random effects models.

In the following, we will assume that our observations come from a random field $\{Y(\mathrm{s}),\, \mathrm{s} \in D\}$, $D \in \mathbb{R}^d$, such that

$$Y(\mathrm{s}) = \mu(\mathrm{s}) + \theta(\mathrm{s}) + \varepsilon(\mathrm{s}), \tag{2.1.4}$$

where the mean structure $\mu(\mathrm{s})$ can eventually be a regression term $\mathbf{x}(\mathrm{s})^T\boldsymbol{\beta}$. Here $\mathbf{x}(\mathrm{s})$ is a p-dimensional vector and it can be either a trend surface or a set of explanatory variables used as covariates. $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients. The residual term is partitioned into two pieces. The first $\theta(\mathrm{s})$ accounts for spatial variability that it is not captured by the regressive term. Accordingly, it is customarily modeled as a mean zero stationary gaussian process, with variance σ^2 and covariance decay parameter ϕ. The second term, $\varepsilon(\mathrm{s})$, is intended to capture variability of a non spatial nature. It is usually modeled as a Gaussian white noise process with variance τ^2 and it can be interpreted either as a pure error term or as a term incorporating measurement errors or microscale variability.

In a Bayesian setting, this is usually recast in the form of a hierarchical model. Suppose we have a vector of data $\mathbf{Y} = (Y(\mathrm{s}_1), \ldots, Y(\mathrm{s}_n))^T$. At the first stage, a gaussian specification is usually given to the distribution of the observables conditional on the spatial component and the other parameters, that is

$$\mathbf{Y}|\,\boldsymbol{\beta}, \boldsymbol{\theta}, \tau^2 \sim N(\mathbf{X}^T\boldsymbol{\beta} + \boldsymbol{\theta},\, \tau^2 I_n), \tag{2.1.5}$$

where $\boldsymbol{\theta} = (\theta(\mathrm{s}_1), \ldots, \theta(\mathrm{s}_n))^T$ and $\mathbf{X}$ is a matrix whose i-th column is $\mathbf{x}(s_i)$, $i = 1, \ldots, n$. The second stage specification is for the spatial component $\theta(\mathrm{s})$. Accordingly to what

said above, it is usually taken to be a mean zero gaussian stationary process, such that

$$\boldsymbol{\theta}|\,\sigma^2,\phi \sim N(\mathbf{0},\, \sigma^2 H(\phi)), \tag{2.1.6}$$

where $H(\phi)$ is a correlation matrix with terms $H_{i,j} = \rho_\phi(\mathbf{s}_i - \mathbf{s}_j)$, and ρ is a valid isotropic correlation function indexed by some parameter ϕ. The model is completed placing convenient, usually independent, priors on the remaining parameters

$$\boldsymbol{\beta}, \tau^2, \sigma^2, \phi \sim p(\boldsymbol{\beta})\, p(\tau^2)\; p(\sigma^2)\, p(\phi). \tag{2.1.7}$$

In particular, a multivariate normal is commonly assumed for $\boldsymbol{\beta}$, while inverse gamma distributions are assumed both for σ^2 and τ^2. The prior for ϕ depends upon the particular choice of the correlation function ρ, but common choices are either uniform or gamma priors.

2.1.4 Inference and prediction.

Notice that the posterior of the model is not available in closed form, so that, in practice, it is necessary to simulate from it using some MCMC methods. In order to achieve a better sampling behavior of the algorithm, some sort of marginalization is typically required. For example, it is common to integrate out the second stage, so that at the first stage we get simply

$$\mathbf{Y}|\,\boldsymbol{\beta}, \tau^2 \sim N(\mathbf{X}^T\boldsymbol{\beta},\, \sigma^2\, H(\phi) + \tau^2\, I_n), \tag{2.1.8}$$

In fact, the matrix $\sigma^2\, H(\phi) + \tau^2\, I_n$ is generally better behaved than $\sigma^2\, H(\phi)$ in a MCMC setting. For discussion of this and related issues, we refer again to Banerjee, Carlin and Gelfand (2004).

The statistician is usually interested in inference on the spatial component $\boldsymbol{\theta}|\mathbf{Y}$ as well as posterior prediction at new locations $\tilde{\mathbf{s}} = (\tilde{\mathbf{s}}_1, \ldots, \tilde{\mathbf{s}}_m)$, $m \geq 1$, for either the response variable or the spatial component itself. Let us denote $\mathbf{Y}_0 =$

$(Y(\tilde{\mathrm{s}}_1),\ldots,Y(\tilde{\mathrm{s}}_m))^T$ and $\boldsymbol{\theta_0} = (\theta(\tilde{\mathrm{s}}_1),\ldots,\theta(\tilde{\mathrm{s}}_m))^T$. In order to compute the posterior predictive distribution we need to know the value of the design matrix $\mathbf{X}_0$, whose columns are the vectors $\mathbf{x}(\tilde{\mathrm{s}}_i)$, $i = 1,\ldots,m$. Then,

$$p(\boldsymbol{\theta}_0|\,\mathbf{y}) = \int p(\boldsymbol{\theta}_0|\,\boldsymbol{\theta},\sigma^2,\phi)\ \times\ p(\boldsymbol{\theta}|\,\mathbf{y},\boldsymbol{\beta},\tau^2,\sigma^2,\phi)\times$$
$$\times\ p(\boldsymbol{\beta},\tau^2,\sigma^2,\phi|\,\mathbf{y})\ d\boldsymbol{\theta}\ d\boldsymbol{\beta}\ d\tau^2\ d\sigma^2\ d\phi,$$

and

$$p(\mathbf{y}_0|\mathbf{y}) = \int p(\mathbf{y}_0|\,\mathbf{y},\boldsymbol{\theta}_0,\boldsymbol{\beta},\tau^2)\ \times\ p(\boldsymbol{\theta}_0,\boldsymbol{\beta},\tau^2|\,\mathbf{y})\,d\boldsymbol{\theta}_0\,d\tau^2\,d\sigma^2.$$

Notice that $p(\boldsymbol{\theta}_0|\,\boldsymbol{\theta},\sigma^2,\phi)$ and $p(\mathbf{y}_0|\,\mathbf{y},\boldsymbol{\theta}_0,\boldsymbol{\beta},\tau^2)$ are both conditional multivariate normal, while the other distributions can be sampled through the Gibbs sampler. Then a sample from the posterior predictive distribution is usually obtained via the so called *composition sampling*, that is drawing $\mathbf{y}_0^g$ from $p(\mathbf{y}_0|\,\mathbf{y},\boldsymbol{\theta}_0^{(g)},\boldsymbol{\beta}^{(g)},\tau^{2(g)})$, where $\{\boldsymbol{\theta}_0^{(g)},\boldsymbol{\beta}^{(g)},\tau^{2(g)}\}_{g=1}^{G}$ is a collection of MCMC samples from the posterior distribution $p(\boldsymbol{\theta}_0,\boldsymbol{\beta},\tau^2|\,\mathbf{y})$.

2.2 The Spatial Dirichlet Process Model.

The process arising from the hierarchical model described in the previous section is gaussian and stationary. This assumption can be considered particularly unappropriate in certain contexts, where it is supposed that local spatial characteristics can influence the correlation structure of the process. Most environmental studies fall in this category. For example, landscape or topography affects weather patterns and, as a consequence, many other atmospheric events. Because of that, many works have recently appeared trying to introduce modeling of nonstationary processes, either from a parametric or nonparametric perspective.

Among the latter, it is worthy to mention the "deformation" approach of Sampson and Guttorp (1992). They provide a nonparametric approach to estimation of the covariance function of a random process $Y(\mathbf{s})$ observed repeatedly at times $t = 1, \ldots, T$ at a finite number of locations $(\mathbf{s}_1, \ldots, \mathbf{s}_n)$ in the plane. Only temporal stationarity is assumed, so that at each time t we can think at $Y_t(\mathbf{s})$ as a replicate of the same process observed at time $t - 1$. In all other respects, the process $Y(\mathbf{s})$ is taken to be quite general. In particular, no spatial stationarity is assumed. Their approach requires to transform the actual locations into conceptual ones by means of a nonmetric multidimensional scaling. This changes the problem into one for which the covariance function is stationary and isotropic. Together with a thin-plate splines interpolation in the new locations space, the method leads to a nonparametric estimation of the covariance structure of the process. In a Bayesian context, this approach has been developed by Damian, Sampson and Guttorp (2001) and Schmidt and O'Hagan (2003). The first ones consider a prior on the thin plate splines, while the others model the mapping between the actual and the fictitious locations by means of an unknown function $d(\cdot)$, on which it is defined a Gaussian prior.

Gelfand, Kottas and MacEachern (2004) address nonstationarity in a totally different way. They reformulate the hierarchical model described in the previous section, replacing the Gaussian specification for the spatial component at the second stage with a spatially defined Dirichlet process. In accordance with their notation, from now on we will denote with G the random probability measure whose distribution is the Dirichlet measure $\mathcal{P}$ and with G_0 the base measure of the process. Recall from chapter 1 that we can represent a random probability measure arising from a $\mathrm{DP}(\alpha G_0)$ as a weighted infinite sum $\sum_{k=1}^{\infty} p_k \delta_{\theta_k^*}$, where $\theta_k^* \overset{i.i.d.}{\sim} G_0$ and the p_k's can be obtained by a stick-breaking procedure as in Sethuraman's construction. The θ_k^* are usually assumed to be scalar or vector valued, but recall that, in principle, the space $(\mathbb{X}, \mathcal{X})$ on which the Dirichlet

process is defined can be quite general. If we take that space to be the space of the realizations of a stationary random field G_0 on D, then each θ_k^* can be seen as a surface over D, $\theta_k^* = \{\theta_k^*(\mathrm{s}),\ \mathrm{s} \in D\}$. In other words, we can define a spatial Dirichlet process G (SDP) simply allowing the base measure G_0 to be a stochastic process.

Again, in practice, the process can be observed only at a finite set of locations $(\mathrm{s}_1, \ldots, \mathrm{s}_n)$. Therefore, we need to consider the random distribution $G^{(n)}$ induced from G at those locations. Since the weights do not depend upon the location $\mathrm{s} \in D$, $G^{(n)} \sim DP(\alpha G_0^{(n)})$ where $G_0^{(n)}$ is the n-variate finite dimensional distribution of the process G_0. This is usually taken to be a mean zero stationary gaussian process, so that $G_0^{(n)}$ is in fact an n-dimensional multivariate normal distribution.

It follows that if we consider a random field $\{\theta(\mathrm{s}),\ \mathrm{s} \in D\}$ such that $\theta(\cdot)|G \sim G$, where $G \sim SDP(\alpha G_0)$ as above, we get

$$E(\theta(\mathrm{s})|G) = \sum_{k=1}^{\infty} p_k\, \theta_k^*(\mathrm{s}),$$

$$Var(\theta(\mathrm{s})|G) = \sum_{k=1}^{\infty} p_k\, (\theta_k^*(\mathrm{s}))^2 - \left\{ \sum_{k=1}^{\infty} p_k\, \theta_k^*(\mathrm{s}) \right\}^2,$$

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

$$Cov(\theta(\mathrm{s}_i), \theta(\mathrm{s}_j)|G) = \sum_{k=1}^{\infty} p_k\, \theta_k^*(\mathrm{s}_i)\, \theta_k^*(\mathrm{s}_j) - \left\{ \sum_{k=1}^{\infty} p_k\, \theta_k^*(\mathrm{s}_i) \right\} \left\{ \sum_{k=1}^{\infty} p_k \theta_k^*(\mathrm{s}_j) \right\}. \quad (2.2.1)$$

Hence, for any given G, the process $\theta(\cdot)$ has heterogenous variance and is nonstationary. However, notice that it is centered around a stationary process. In fact, assume G_0 is a mean zero stationary process with finite variance σ^2 and correlation function $\rho_\phi(\mathrm{s}_i - \mathrm{s}_j)$. If we marginalize with respect to the Dirichlet process, in the previous expressions we get $E(\theta(\mathrm{s})) = 0$, $Var(\theta(\mathrm{s})) = \sigma^2$ and $Cov(\theta(\mathrm{s}_i), \theta(\mathrm{s}_j)) = \sigma^2 \rho_\phi(\mathrm{s}_i - \mathrm{s}_j)$.

Hence, we can go back to the hierarchical model described in section 1.3 and replace the second stage with the SDP specification. However, notice that in order to fit the

model and learn about the covariance structure of the process, we need to observe the process repeatedly at times $t = 1, \ldots, T$. In fact, as in Sampson and Guttorp (1992), replications of the process are required to pursue a fully nonparametric approach. See Gelfand, Kottas and MacEachern (2004) for a discussion of the point. Therefore, consider data $\mathbf{Y}_t = (Y_t(\mathbf{s}_1), \ldots, Y_t(\mathbf{s}_n))^T$, $t = 1, \ldots, T$, and associate with each $\mathbf{Y}_t$ 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, for $t = 1, \ldots, T$, the following semiparametric hierarchical model arises

$$\begin{aligned}
\mathbf{Y}_t | \boldsymbol{\beta}, \boldsymbol{\theta}_t, \tau^2 &\overset{ind.}{\sim} N(\mathbf{X}_t^T \boldsymbol{\beta} + \boldsymbol{\theta}_t,\, \tau^2 I_n) \\
\boldsymbol{\theta}_t | G^{(n)} &\overset{i.i.d.}{\sim} G^{(n)} \\
G^{(n)} | \alpha, \sigma^2, \phi &\sim DP(\alpha G_0^{(n)}),\ G_0^{(n)}(\cdot | \sigma^2, \phi) = N_n(\cdot | \mathbf{0}_n, \sigma^2 H_n(\phi)) \\
\boldsymbol{\beta}, \tau^2, \alpha, \sigma^2, \phi &\sim p(\boldsymbol{\beta})\, p(\tau^2)\, p(\alpha)\, p(\sigma^2)\, p(\phi).
\end{aligned} \tag{2.2.2}$$

The matrix X_t can be either a matrix of covariates, whose values change across replicates, or a matrix of geographical coordinates, so that it is fixed and the regressive term accounts for a trend surface.

One of the main objectives in spatial statistics is to predict phenomena at locations where we don't have any observation. Therefore, we could be interested in predicting the values of the process at a new set of locations $(\tilde{\mathbf{s}}_1, \ldots, \tilde{\mathbf{s}}_m)$ for any of the T replicates. However, the setting proper of the Dirichlet process allows prediction of the values of the process both at old and new locations through generation of a new surface $Y_0(\cdot)$ from the posterior predictive distribution $p(Y_0 | \mathbf{Y}_t, t = 1, \ldots, T)$. Central to this approach is the Pólya urn scheme described by Blackwell and MacQueen (1973). Here we will show briefly how to get samples from these distribution via composition sampling. Let us indicate with $\boldsymbol{\theta}_t^{(g)} = (\theta_t^{(g)}(\mathbf{s}_1), \ldots, \theta_t^{(g)}(\mathbf{s}_n))^T$, $t = 1, \ldots, T$, $g = 1, \ldots, G$, the output of the Gibbs sampler used to sample from the posterior of the model (2.2.2) at the n original locations. Notice that, for each g, some of the $\boldsymbol{\theta}_t$'s can be the same across

the replicates, because the DP is a.s. discrete. However, for notational simplicity, we won't distinguish here and regard the $\boldsymbol{\theta}_t$'s as distinct one another, referring to the original paper of Gelfand, Kottas and MacEachern (2004) for the relevant details. Then, for each t and for each g, we can sample a vector $\tilde{\boldsymbol{\theta}}_t^{(g)} = (\theta_t^{(g)}(\tilde{\mathrm{s}}_1), \ldots, \theta_t^{(g)}(\tilde{\mathrm{s}}_m))^T$ from the conditional m-variate normal distribution of $\tilde{\boldsymbol{\theta}}_t^{(g)}$ given $\boldsymbol{\theta}_t^{(g)}$ emerging from the joint $N_{n+m}(\cdot \,|\, \mathbf{0}_{n+m}, \sigma^2 H_{n+m}(\phi))$. Therefore, at any iteration of the Gibbs sampler we obtain joint posterior samples $(\boldsymbol{\theta}_t^{(g)}, \tilde{\boldsymbol{\theta}}_t^{(g)})$ of the spatial component for the old and new locations .

Accordingly, define $\boldsymbol{\theta}_0^{(g)} = (\theta_0^{(g)}(\mathrm{s}_1), \ldots, \theta_0^{(g)}(\mathrm{s}_n))^T$ and $\tilde{\boldsymbol{\theta}}_0^{(g)} = (\theta_0(\tilde{\mathrm{s}}_1), \ldots, \theta_0(\tilde{\mathrm{s}}_m))^T$, $g = 1, \ldots, G$, to be joint samples from the predictive distribution of the spatial component $\theta_0(\cdot)$ given $(\boldsymbol{\theta}_t^{(g)}, \tilde{\boldsymbol{\theta}}_t^{(g)})$, $t = 1, \ldots, T$. Using the Blackwell-MacQueen urn scheme, we obtain these samples from a weighted mixture of the base measure and point masses corresponding to values coming from the Gibbs sampler, that is

$$p(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_0 | \, \boldsymbol{\theta}_t^{(g)}, \tilde{\boldsymbol{\theta}}_t^{(g)}, \, t = 1, \ldots, T) \propto \frac{\alpha}{\alpha + T} G_0^{(n+m)}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_0 \,|\, \sigma^2, \phi) + \frac{T}{\alpha + T} \sum_{t=1}^{T} \delta_{(\boldsymbol{\theta}_t^{(g)}, \tilde{\boldsymbol{\theta}}_t^{(g)})}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_0).$$

Now, it's easy to obtain samples from the predictive distribution of the response variable Y. We need only to know the values of the matrices of covariates at the time and locations considered for prediction. Let us indicate such matrices with $\mathbf{X}_0, \tilde{\mathbf{X}}_0$ for the old and new locations respectively, and with $(\mathbf{X}_0, \tilde{\mathbf{X}}_0)$ the matrix obtained putting them together. Then,

$$\mathbf{Y}_0, \tilde{\mathbf{Y}}_0 | \boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_0 \sim N_{n+m}(\cdot \,|\, (\mathbf{X}_0, \tilde{\mathbf{X}}_0)^T \boldsymbol{\beta} + (\boldsymbol{\theta}_0, \tilde{\boldsymbol{\theta}}_0), \tau^2 I_{n+m}),$$

where $\mathbf{Y}_0 = (Y_0(\mathrm{s}_1), \ldots, Y_0(\mathrm{s}_n))^T$ and $\tilde{\mathbf{Y}}_0 = (Y_0(\tilde{\mathrm{s}}_1), \ldots, Y_0(\tilde{\mathrm{s}}_m))^T$ represent the predicted values of the response variable Y at old and new locations.

The Spatial Dirichlet Process can be extended in several directions. For example, a different model can be considered, allowing for more complex structures for the

residuals. Also, it is possible to relax the temporal stationarity assumption and assume that the spatial component $\theta_t(\cdot)$ follows a linear dynamic evolution, that is $\theta_t(s) = \theta_{t-1}(s) + \eta_t(s)$ where $\eta_t(s)$ are independent replications from a spatial Dirichlet process, $t = 1, \ldots, T$. We won't develop this further, although a dynamic linear model for the spatial residual will be used as an example for the generalized spatial Dirichlet Process we will introduce in chapter 4

2.3 Smoothness properties of spatial processes.

In many areas of application, it is relevant to investigate smoothness of process realizations. For example, in so called digital terrain models, researchers need to explore surface roughness; in the environmental sciences, they are often interested in the rates of change in levels of pollutants in the atmosphere; in meteoroly, they need to recognize temperature or rainfall gradients, and so on.

In this section, we discuss smoothness properties of spatial processes. In doing so, we typically need to assume a topology on the space where the random field is defined. From now on, for simplicity, we can regard it as the one induced by the usual Euclidean norm.

In section 3.1, we introduce almost sure continuity, and discuss such property mainly in the context of stationary processes (see Kent (1989)). In section 3.2, we define mean square continuity and show that it is possible to relate this property to the covariance function of the process (see Stein (1999)). Together with mean square continuity, it is possible to define mean square differentiability. Such a notion was originally defined for processes defined on $\mathbb{R}$. Banerjee and Gelfand (2003) and Banerjee, Gelfand and Sirmans (2003) have extended this notion to processes defined on $\mathbb{R}^d$. We follow their arguments and introduce the related concepts of directional finite difference and

directional derivative processes, together with the relevant ditribution theory. For stationary random fields, existence of the directional derivative process requires existence and continuity of all the second-order partial and mixed derivatives of the covariance function. This is not always true for the covariance functions usually considered in the literature. In the Appendix, we discuss this issue with regard to the Matern covariance class, for varying ν. In particular, we provide expressions for the derivatives of the Matern covariance function, since we weren't able to find any explicit reference for them in the current literature.

2.3.1 Almost sure continuity.

Consider a real valued process $\{Y(\mathbf{s}),\, \mathbf{s} \in \mathbb{R}^d\}$. We start by investigating the existence of a version of the random field in which the realizations are everywhere continuous.

Definition 2.3.1. *A real valued process $\{Y(\mathbf{s}),\, \mathbf{s} \in \mathbb{R}^d\}$ is almost surely continuous at $\mathbf{s}_0$ if $Y(\mathbf{s}) \to Y(\mathbf{s}_0)$ a.s. as $s \to s_0$. If the process is almost surely continuous for every $s_0 \in \mathbb{R}^d$ the the process is said to have continuous realizations.*

Notice that the above definition applies to any stochastic process (possibly non-stationary). However, Kent (1989) has provided sufficient conditions to ensure a.s. continuity of stationary random fields.

In fact, let $\{Y(\mathbf{s}),\, \mathbf{s} \in \mathbb{R}^d\}$ to be a real valued stationary random field on $\mathbb{R}^d$ with mean 0 and finite second moments. If there exists an integer $m \geq 0$, such that the covariance function $K(\mathbf{h})$ (see sections 2.1.1 and 2.1.2) , $\mathbf{h} \in \mathbb{R}^d$, is m-times continuously differentiable with respect to $\mathbf{h}$, we can define

$$K_m(\mathbf{h}) = K(\mathbf{h}) - P_m(\mathbf{h}),$$

where $P_m(\mathbf{h})$ is the Taylor polynomial of degree m for $K(\mathbf{h})$ about $\mathbf{h} = \mathbf{0}$. Then, it is possible to prove the following theorem

Theorem 2.3.2 (Kent (1989)). *If $K(\mathbf{h})$ is d-times continuously differentiable and*

$$|K_d(\mathbf{h})| = O\left(\frac{||\mathbf{h}||^d}{|\log(||\mathbf{h}||)|^{3+\gamma}}\right) \quad as\ ||\mathbf{h}|| \to 0 \tag{2.3.1}$$

for some $\gamma > 0$, then there exists a version of the d-dimensional random field $\{Y(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ with continuous realizations.

In particular, setting $d = 1$, we get $|K_1(\mathbf{h})| = K(0) - K(\mathbf{h})$ and this theorem comprises a classic result (see Cramér and Leadbetter (1967)).

However, condition (2.3.1) can be difficult to verify in practice. Therefore, Kent provides two conditions that imply (2.3.1). First, (2.3.1) is true if

$$|K_d(\mathbf{h})| = O\left(||\mathbf{h}||^{d+\beta}\right) \quad \text{as } ||\mathbf{h}|| \to 0 \tag{2.3.2}$$

for some $\beta > 0$. In particular, this happens whenever $K(\mathbf{h})$ is $d+1$-times differentiable.

Condition (2.3.1) also holds if the mixed dth-order partial derivative of $K(\mathbf{h})$ exists and satisfies

$$\left|\frac{\partial^d K_d(\mathbf{h})}{\partial h_1 \ldots \partial h_d}\right| = O\left(\frac{1}{|\log(||\mathbf{h}||)|^{3+\gamma}}\right) \quad \text{as } ||\mathbf{h}|| \to 0. \tag{2.3.3}$$

We can get milder conditions if $Y(\mathbf{s})$ is a stationary gaussian random field. For example, continuity of the realizations of the process follows if the covariance function is suche that $K(\mathbf{0}) - K(\mathbf{h}) = O\left(\frac{1}{|\log(||\mathbf{h}||)|^{1+\varepsilon}}\right)$, for some $\varepsilon > 0$ (see Adler (1981) and Kent (1989)).

2.3.2 Mean square continuity.

Let $\{Y(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ be a real valued random field on $\mathbb{R}^d$ with mean 0 and finite second moments.

Definition 2.3.3. *A process $\{Y(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ is mean square continuous (also said, L_2-continuous) at $\mathbf{s}_0$ if*

$$\lim_{||\mathbf{s}-\mathbf{s}_0|| \to 0} E[Y(\mathbf{s}) - Y(\mathbf{s}_0)]^2 = 0.$$

We will denote mean square continuity at $\mathbf{s}_0$ *as* $Y(\mathbf{s}) \xrightarrow{L_2} Y(\mathbf{s}_0)$.

In general, mean square continuity does not imply a.s. continuity, nor viceversa. For counterexamples, we refer to Banerjee and Gelfand (2003) and Banerjee, Carlin and Gelfand (2004). However, if $Y(\mathbf{s})$ is a bounded process then a.s. continuity implies L_2 continuity.

Denote with $K(\mathbf{t}, \mathbf{s})$ the covariance function $Cov(Y(\mathbf{t}), Y(\mathbf{s}))$ of a general (not necessarily stationary) process $Y(\mathbf{s})$ on D. Then, the definition of mean square continuity is equivalent to

$$\lim_{||\mathbf{s}-\mathbf{s}_0||\to 0} (K(\mathbf{s},\mathbf{s}) - 2K(\mathbf{t},\mathbf{s}) + K(\mathbf{t},\mathbf{t})) = 0.$$

It immediately follows that a sufficient condition for mean square continuity is the continuity of the covariance function $K(\cdot,\cdot)$ in $\mathbf{t}$ and $\mathbf{s}$. For $Y(\mathbf{s})$ a weakly stationary process with covariance function $K(\mathbf{h})$, where $\mathbf{h} = \mathbf{s} - \mathbf{s}_0$, mean square continuity reduces to ask

$$\lim_{||h||\to 0} 2\left[K(\mathbf{0}) - K(\mathbf{h})\right] = 0,$$

so that $Y(\mathbf{s})$ 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, since for two separation vectors $\mathbf{h}_1, \mathbf{h}_2 \in \mathbb{R}^d$

$$\begin{aligned} |K(\mathbf{h}_1) - K(\mathbf{h}_2)| &= |Cov\left(Y(\mathbf{h}_1) - Y(\mathbf{h}_2), Y(\mathbf{0})\right| \\ &\leq \left[Var\left(Y(\mathbf{h}_1) - Y(\mathbf{h}_2)\right) Var\left(Y(\mathbf{0})\right)\right]^{\frac{1}{2}} \\ &= \left[2\left[K(\mathbf{0}) - K(\mathbf{h}_1 - \mathbf{h}_2)\right] K(\mathbf{0})\right]^{\frac{1}{2}}, \end{aligned}$$

and this goes to 0 as $||\mathbf{h}_1 - \mathbf{h}_2|| \to 0$.

Now, assume that our process $Y(\mathbf{s})$ satisfies Kent's sufficient condition for a.s. continuity. As Banerjee and Gelfand (2003) note, if $K(\mathbf{h})$ is d-times differentiable, then

it is of course continuous at $\mathbf{0}$. It follows that processes satisfying Kent condition are also mean square continuous.

Banerjee and Gelfand (2003) also investigate smoothness of continuous functionals of the process $Y(\mathbf{s})$. Denote with L_2 the Hilbert space of random variables induced by the L_2 metric, and suppose that $f : L_2 \to \mathbb{R}$ is a continuous function. Then, consider the transformed process $\{Z(\mathbf{s}) = f(Y(\mathbf{s})), \mathbf{s} \in \mathbb{R}^d\}$. If $Y(\mathbf{s})$ is a.s. continuous, a.s. continuity of $Z(\mathbf{s})$ follows routinely. In particular, if f is bounded, $Z(\mathbf{s})$ is also mean square continuous, because of what we said under definition 2.3.3. However, if $Y(\mathbf{s})$ is mean square continuous, the mean square continuity of $Z(\mathbf{s})$ is not so immediate. Banerjee and Gelfand (2003) prove that this follows if $f : \mathbb{R} \to \mathbb{R}$ is Lipschitz of order 1, that is if there exists a constant C such that

$$|f(Y(\mathbf{s}+\mathbf{h}) - f(Y(\mathbf{s}))| \leq C\,|Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})|\,,$$

for all $\mathbf{s} \in D$ and $\mathbf{h} \in \mathbb{R}^d$ such that $\mathbf{s}+\mathbf{h} \in D$. In fact, in that case

$$E\left[f(Y(\mathbf{s}+\mathbf{h}) - f(Y(\mathbf{s}))\right]^2 \leq C^2 E\left[Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})\right],$$

and the mean square continuity of $Z(\mathbf{s})$ follows directly from the mean square continuity of $Y(\mathbf{s})$.

2.3.3 Mean square differentiability.

In this section, we will introduce the concept of mean square differentiability. We will follow the discussion in Banerjee and Gelfand (2003), who extended the definition from processes defined on the real line to general processes on $\mathbb{R}^d$. The definition is motivated by the analogous definition of total differentiability of a function in $\mathbb{R}^d$ in multivariate calculus. Let $\mathbf{u}$ be a unit vector in $\mathbb{R}^d$, i.e. such that $||\mathbf{u}|| = 1$.

Definition 2.3.4. *A process* $\{Y(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ *is said to be mean square differentiable at* $\mathbf{s}_0$ *if there exists a vector* $\nabla_Y(\mathbf{s}_0)$ *such that, for any scalar* h *and any unit vector* $\mathbf{u} \in \mathbb{R}^d$,

$$Y(\mathbf{s}_0 + h\mathbf{u}) = Y(\mathbf{s}_0) + h\mathbf{u}^T\nabla_{\mathbf{Y}}(\mathbf{s}_0) + r(\mathbf{s}_0, h\mathbf{u}), \tag{2.3.4}$$

where $r(\mathbf{s}_0, h\mathbf{u}) \to 0$ *in the* L^2 *sense as* $h \to 0$.

In other words, we require that for any unit vector $\mathbf{u}$,

$$\lim_{h\to 0} E\left(\frac{Y(\mathbf{s}_0 + h\mathbf{u}) - Y(\mathbf{s}_0) - h\mathbf{u}^T\nabla_{\mathbf{Y}}(\mathbf{s}_0)}{h}\right)^2 = 0.$$

This definition ensures that if $Y(\mathbf{s})$ is a mean square differentiable process on $\mathbb{R}^d$, then it is mean square continuous as well. In fact, this result follows directly from (2.3.4), since $r(\mathbf{s}_0, h\mathbf{u}) \to 0$ in the L^2 sense as $h \to 0$ and $\mathbf{u}^T\nabla_{\mathbf{Y}}(\mathbf{s}_0)$ is a constant in h.

For $d = 1$ and for $Y(s)$ a stationary process, it's possible to characterize mean square differentiability by means of the second derivative of the covariance function of the process $Y(s)$. In fact, 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.

Finite difference processes.

Correspondingly to the concept of mean square differentiability of a process $Y(\mathbf{s})$, we can define that of gradient processes. We start defining the *directional finite difference process* $Y_{\mathbf{u},h}(\mathbf{s})$ at scale h in direction $\mathbf{u}$ as

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

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 critical

importance, such as in many environmental, ecological or demographical applications. In fact, many phenomena could not be appropriately recognized at smaller scales. Banerjee, Gelfand and Sirmans (2003) cite, for example, digital terrain modeling, where at low resolutions we can recognize global features such as hills and valleys, and we concentrate on more local features as long as the resolution increases.

Note that if $E(Y(\mathbf{s})) = 0$ for all $\mathbf{s} \in \mathbb{R}^d$, then also $E(Y_{\mathbf{u},h}(\mathbf{s})) = 0$. Banerjee, Gelfand and Sirmans (2003) develop the necessary distribution theory for this process. Let $C_{\mathbf{u}}^{(h)}(\mathbf{s}, \mathbf{s}')$ denote the covariance function associated with the process $Y_{\mathbf{u},h}(\mathbf{s})$ and let $\Delta = \mathbf{s} - \mathbf{s}'$ denote the separation vector. Then, they show that if $Y(\mathbf{s})$ is stationary

$$C_{\mathbf{u}}^{(h)}(\mathbf{s}, \mathbf{s}') = \frac{2K(\Delta) - K(\Delta + h\mathbf{u}) - K(\Delta - h\mathbf{u})}{h^2}. \tag{2.3.5}$$

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

If $Y(\mathbf{s})$ is isotropic, we obtain

$$C_{\mathbf{u}}^{(h)}(\mathbf{s}, \mathbf{s}') = \frac{2K(||\Delta||) - K(||\Delta + h\mathbf{u}||) - K(||\Delta - h\mathbf{u}||)}{h^2}. \tag{2.3.6}$$

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

Directional derivative processes.

Let us consider a random field $\{Y(\mathbf{s}),\ \mathbf{s} \in D\}$ and the associated finite difference process $Y_{\mathbf{u},h}(\mathbf{s})$ in a given direction $\mathbf{u}$ and at an arbitrary scale h. We define the *directional derivative process* $D_{\mathbf{u}}Y(\mathbf{s})$ in direction $\mathbf{u}$ as the process obtained as the limit in the L_2 sense of the finite difference process $Y_{\mathbf{u},h}(\mathbf{s})$, as $h \to 0$, that is

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

if the limit exists. The directional derivative processes are, in fact, gradient processes. Gradients are quantities of basic importance in geometry and physics. Researchers in

the physical sciences often formulate relationships in terms of gradients.

Again, the covariance structure and the main distribution theory of the process so defined has been studied thoroughly by Banerjee, Gelfand and Sirmans (2003).

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

$$C_{\mathbf{u}}(\mathbf{s}, \mathbf{s}') = \lim_{h \to 0} C_{\mathbf{u}}^{(h)}(\mathbf{s}, \mathbf{s}'), \tag{2.3.7}$$

that is the covariance function of the process is obtained as the limit of the covariance function of the associated finite difference process, as $h \to 0$. Let $Y(\mathbf{s})$ be a stationary process on $D \subset \mathbb{R}^d$ with covariance function $K(\cdot)$ such that all second order partial and mixed derivatives of K exist and are continuous. Then,

$$C_{\mathbf{u}}(\mathbf{s}, \mathbf{s}') = -\mathbf{u}^T \Omega(\Delta)\mathbf{u}, \tag{2.3.8}$$

where $(\Omega(\Delta))_{i,j} = \frac{\partial^2 K}{\partial \Delta_i \partial \Delta_j}$. Hence, for any $\mathbf{u}$ (2.3.8) is a valid covariance function on $\mathbb{R}^d$. In particular, $Var(D_{\mathbf{u}}Y(s)) = -\mathbf{u}^T\Omega(\mathbf{0})\mathbf{u}$.

If $Y(\mathbf{s})$ is isotropic, the directional derivative process will be stationary but not isotropic. In fact,

$$C_{\mathbf{u}}(\mathbf{s}, \mathbf{s}') = -\left\{\left(1 - \frac{(\mathbf{u}^T\Delta)^2}{||\Delta||^2}\right)\frac{K'(||\Delta||)}{||\Delta||} + \frac{(\mathbf{u}^T\Delta)^2}{||\Delta||^2}K''(||\Delta||)\right\}. \tag{2.3.9}$$

See Banerjee and Gelfand (2003) and Banerjee, Gelfand and Sirmans (2003) for a derivation of these results. Notice that they rely on the existence of the second derivatives of the covariance function $K(\cdot)$. In fact, not all the covariance functions usually considered in the literature admit second order derivatives, so that directional derivatives do not exist in these cases. Consider for example the exponential covariance function $K(||\Delta||) = \sigma^2 \exp(-\phi||\Delta||)$, that is not differentiable at 0. Actually, the

only function that is differentiable at 0 in the so called *power* exponential family, $K(||\Delta||) = \sigma^2 \exp(-\phi||\Delta||^\nu)$, $0 < \nu \leq 2$, is obtained for $\nu = 2$, that is the Gaussian covariance function. However, this 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) for further details. Later, we will make use of the Matern covariance class, introduced in section 1.2. Here, the degree of smoothness is governed by the parameter ν. In the Appendix we will provide expressions for the first and second derivatives of the Matern covariance function, discussing also the issue of smoothness with regard to the values assumed by the parameter ν.

If $K''(\cdot)$ is continuous at 0 and we let $\Delta \to 0$ in (2.3.9), then $\lim_{\Delta \to 0} C_{\mathbf{u}}(\mathbf{s}, \mathbf{s}') = -K''(0)$. It follows that $C_{\mathbf{u}}(\mathbf{s}, \mathbf{s}')$ is a valid covariance function continuous at zero. According to what said in section 3.1, the underlying directional derivative process is mean square continuous.

This is the analogous of the result obtained by Stein (1999) for the case $d = 1$ and already mentioned at the beginning of section 3.3. However, for $d > 1$, the existence of the directional derivatice process in all directions $\mathbf{u}$ does not necessarily imply that the process is mean square differentiable. In fact, we refer to Banerjee and Gelfand (2003) for the following counterexample.

Example 2.3.5. *Let* $\{Y(\mathbf{s}), \mathbf{s} = (s_1, s_2) \in \mathbb{R}^2\}$ *be a process defined as follows.*

$$Y(\mathbf{s}) = \begin{cases} \frac{s_1 s_2^2}{s_1^2 + s_2^4} Z & \text{if } \mathbf{s} \neq \mathbf{0}, \text{ where } Z \sim N(0,1) \\ 0 & \text{if } \mathbf{s} = \mathbf{0} \end{cases}$$

Then, $Y_{\mathbf{u},h}(\mathbf{0}) = \frac{u_1 u_2^2}{u_1^2 + h^2 u_2^4} Z$. *So, if* $D_{\mathbf{u}} Y(\mathbf{0}) = \frac{u_2^2}{u_1} Z$ *for any direction* $\mathbf{u}$ *with* $u_1 \neq 0$ *and* $D_{\mathbf{u}} Y(\mathbf{0}) = 0$ *for any direction* $\mathbf{u}$ *with* $u_1 = 0$, $D_{\mathbf{u}} Y(\mathbf{0}) = \lim_{h \to 0} Y_{\mathbf{u},h}(\mathbf{0})$ *in the* L_2 *sense. However, the above process is not mean square continuous at* 0 *(and hence, it*

is not even mean square differentiable), as can be seen by considering the path $s_1 = s_2^2$ *along which* $E[Y(\mathbf{s}) - Y(\mathbf{0})]^2 = \frac{1}{4}$.

However, if $Y(\mathbf{s})$ is a mean square differentiable process in $\mathbb{R}^d$, that is (2.3.4) holds for every $\mathbf{s}$ in $\mathbb{R}^d$, then its directional derivative exists in all directions and, in fact, it can be easily characterized. For each $\mathbf{u} \in \mathbb{R}^d$,

$$D_{\mathbf{u}}Y(\mathbf{s}) = \lim_{h \to 0} \frac{Y(\mathbf{s} + h\mathbf{u}) - Y(\mathbf{s})}{h} = \lim_{h \to 0} \frac{h\mathbf{u}^T \nabla_{\mathbf{Y}}(\mathbf{s}_0) + r(\mathbf{s}_0, h\mathbf{u})}{h} = \mathbf{u}^T \nabla_{\mathbf{Y}}(\mathbf{s}_0),$$

where the limit has to be intended in the L_2 sense.

Now, consider an an orthonormal basis set for $\mathbb{R}^d$, composed by unit vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_d$. Then, 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

$$D_{\mathbf{u}}Y(\mathbf{s}) = \mathbf{u}^T \nabla_Y(\mathbf{s}) = \sum_{i=1}^d w_i \mathbf{e}_i^T \nabla_Y(\mathbf{s}) = \sum_{i=1}^d w_i D_{\mathbf{e}_i} Y(\mathbf{s}). \tag{2.3.10}$$

Hence, to study directional derivative processes in an arbitrary direction $\mathbf{u}$, we need only to know a basis set of directional derivative processes. This can always be taken to be the one defined by the coordinate axes, so that $\mathbf{e}_1$ is a $d \times 1$ vector with all 0's except for a 1 in the i-th row. In fact, with this basis, $\nabla_Y(\mathbf{s}) = (D_{\mathbf{e}_1} Y(\mathbf{s}), \ldots, D_{\mathbf{e}_n} Y(\mathbf{s}))^T$. Notice that reduction to a basis set is not possible for finite difference processes, as it is evident if one considers (2.3.4). From (2.3.10), it is also clear that

$$D_{-\mathbf{u}}Y(\mathbf{s}) = -D_{\mathbf{u}}(\mathbf{s}).$$

Applying the Cauchy-Schwarz inequality to (2.3.10), for every unit vector $\mathbf{u}$, we get

$$D_{\mathbf{u}}^2 Y(\mathbf{s}) \leq \sum_{i=1}^d D_{\mathbf{e}_1}^2 Y(\mathbf{s}).$$

Hence, $\sum_{i=1}^d D_{\mathbf{e}_1}^2 Y(\mathbf{s})$ is the maximum over all directions of $D_{\mathbf{u}}^2 Y(\mathbf{s})$.

Consider well-behaved functions of a mean square differentiable spatial process. If $Y(\mathbf{s})$ is a mean square differentiable process and $D_{\mathbf{u}}Y(\mathbf{s})$ is the associated directional derivative process in direction $\mathbf{u}$, we form the new process $Z(\mathbf{s}) = f(Y(\mathbf{s}))$, where f is continuously differentiable with bounded derivative. Then, $Z(\mathbf{s})$ is mean square differentiable and has mean square derivative in the direction $\mathbf{u}$, given by

$$D_{\mathbf{u}}Z(\mathbf{s}) = f'(Y(\mathbf{s}))D_{\mathbf{u}}Y(\mathbf{s}).$$

More distribution theory.

In this section, we recall other expressions developed by Banerjee, Gelfand and Sirmans (2003) to describe the distribution theory of finite differences and directional derivatives of a random field $Y(\mathbf{s})$. In particular, here we concentrate on cross-covariance functions for the bivariate processes $Z_{\mathbf{u}}^{(h)}(\mathbf{s}) = \begin{pmatrix} Y(\mathbf{s}) \\ Y_{\mathbf{u},h}(\mathbf{s}) \end{pmatrix}$ and $\mathbf{Z}_{\mathbf{u}}(\mathbf{s}) = \begin{pmatrix} Y(\mathbf{s}) \\ D_{\mathbf{u}}Y(\mathbf{s}) \end{pmatrix}$.

Let $\mathbf{s}, \mathbf{s}'$ be a pair of locations in D. For $Y(\mathbf{s})$ stationary, consider first the cross-covariance

$$Cov(Y(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s}')) = \frac{K(\Delta - h\mathbf{u}) - K(\Delta)}{h},$$

from which it follows that $Cov(Y(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s})) = \frac{K(h\mathbf{u}) - K(\mathbf{0})}{h}$, for $\Delta = 0$. But then, taking the limit as $h \to 0$,

$$Cov(Y(\mathbf{s}), D_{\mathbf{u}}Y(\mathbf{s}')) = \lim_{h \to 0} \frac{K(\Delta - h\mathbf{u}) - K(\Delta)}{h} = -D_{\mathbf{u}}K(\Delta) = D_{\mathbf{u}}K(-\Delta),$$

that is the directional derivative of the covariance function $K(\cdot)$ in the direction $\mathbf{u}$. Notice that the last equality follows because of stationarity, since $K(\Delta) = K(-\Delta)$. In particular, for $\Delta = 0$.

$$Cov(Y(\mathbf{s}), D_{\mathbf{u}}Y(\mathbf{s})) = \lim_{h \to 0} \frac{K(h\mathbf{u}) - K(\mathbf{0})}{h} = D_{\mathbf{u}}K(\mathbf{0}).$$

The existence of the directional derivative process ensures the existence of $D_{\mathbf{u}}K(\mathbf{0})$. Moreover, since $K(h\mathbf{u}) = K(-h\mathbf{u})$, $K(h\mathbf{u})$, viewed as a function of h is even, so $D_{\mathbf{u}}K(\mathbf{0}) = 0$. Thus, $Y(\mathbf{s})$ and $D_{\mathbf{u}}Y(\mathbf{s})$ are uncorrelated.

Under isotropy,

$$Cov(Y(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s}')) = \frac{K(||\Delta - h\mathbf{u}||) - K(||\Delta||}{h}.$$

Now, $CovY(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s})) = \frac{K(h)-K(0)}{h}$ and $Cov(Y(\mathbf{s}), D_{\mathbf{u}}Y(\mathbf{s})) = K'(0) = 0$.

Suppose we consider the bivariate process $Z_{\mathbf{u}}^{(h)}(\mathbf{s})$. It is clear that this process has mean $\mathbf{0}$ and, if $Y(\mathbf{s})$ is stationary, has cross-covariance matrix given by

$$V_{\mathbf{u},h}((\Delta) = \begin{pmatrix} K(\Delta) & \frac{K(\Delta-h\mathbf{u})-K(\Delta)}{h} \\ \frac{K(\Delta+h\mathbf{u})-K(\Delta)}{h} & \frac{2K(\Delta)-K(\Delta+h\mathbf{u})+K(\Delta-h\mathbf{u})}{h^2} \end{pmatrix} \tag{2.3.11}$$

Since $Z_{\mathbf{u}}^{(h)}(\mathbf{s})$ arises by linear transformation of $Y(\mathbf{s})$, (2.3.11) is a valid-cross covariance matrix in $\mathbb{R}^d$. But since this is true for every h, letting $h \to 0$, we get that

$$V_{\mathbf{u}}(\Delta) = \begin{pmatrix} K(\Delta) & -D_{\mathbf{u}}K(\Delta) \\ D_{\mathbf{u}}K(\Delta) & -\mathbf{u}^T\Omega(\Delta)\mathbf{u} \end{pmatrix} \tag{2.3.12}$$

is also a valid cross-covariance matrix in $\mathbb{R}^d$. In fact, $V_{\mathbf{u}}$ is the cross covariance matrix for the bivariate process $\mathbf{Z}_{\mathbf{u}}(\mathbf{s})$. If, in addition, we assume $Y(\mathbf{s})$ is a stationary gaussian process it is clear, again by linearity, that $\mathbf{Z}_{\mathbf{u}}^{(h)}(\mathbf{s})$ is a stationary bivariate gaussian process. Then, by a standard limiting moment generating function argument, $\mathbf{Z}_{\mathbf{u}}(\mathbf{s})$ is also a stationary bivariate gaussian process and thus the directional derivative $D_{\mathbf{u}}Y(\mathbf{s})$ is a stationary univariate gaussian process.

The covariance matrices associated with $\mathbf{Z}_{\mathbf{u},h}(\mathbf{s})$ and $\mathbf{Z}_{\mathbf{u}}(\mathbf{s})$ are obtained setting $\Delta = \mathbf{0}$ in (2.3.11) and (2.3.12). In fact, if $Y(\mathbf{s})$ is isotropic, $\Omega(\mathbf{0})$ is of the form $c_0 I$. That is, $\frac{\partial^2 K(||\Delta||)}{\partial^2 \Delta_i}|_{\Delta=0}$ is constant over i by the symmetry of $K(||\Delta||)$ in its arguments. Let c_0 denote such a constant value. The off-diagonal terms are given by $\frac{\partial^2 K(||\Delta||)}{\partial \Delta_j \partial \Delta_i} =$

$\frac{\Delta_j\Delta_i}{||\Delta||^2}\left(K"(||\Delta||) - \frac{K(||\Delta||)}{||\Delta||}\right)$. Since the parenthetic term on the right side of this equality tends to 0 as $\Delta \to 0$ while the other term is bounded, we have $\frac{\partial^2 K(||\Delta||)}{\partial\Delta_j\partial\Delta_i}|_{\Delta=0} = 0$.

2.4 Appendix.

Smoothness properties of the Matern covariance function.

We already noticed that the existence of the directional derivative process for stationary processes imply the existence and continuity of the second order partial and mixed derivatives of the covariance function $K(\cdot)$. This is apparent from expression (2.3.7) and from the distribution theory outlined in the previous sections. In this section, we concentrate on derivatives of the Matern covariance function (see section 2.1.2). We have already stressed that the parameter ν is usually seen as governing the degree of smoothness of the process. This has been motivated in the literature in different ways. For example, Stein (1999) studies the behaviour of the function through a finite sum aproximation in a neighborhood of the origin. He concludes that a process Z on $\mathbb{R}$ is m-times mean square differentiable if and only if $\nu > m$. According to the discussion above, we concentrate here on the existence of the directional derivative process for an isotropic random field $\{Y(\mathbf{s}), \mathbf{s} \in D \subset \mathbb{R}^d\}$. We will show that the existence of the directional derivative process requires $\nu > 1$.

This discussion will prove useful in two respects. First, at my knowledge, noone has tried to motivate the interpretation of ν explicitly in terms of equation (2.3.9) and the derivatives of the covariance function (2.1.3). Secondly, computation of such derivatives will turn useful later in the next chapter, when we assume an isotropic random field with a Matern covariance function and therefore the following expressions provide the necessary distributional specifications for the simulation purposes arising there.

For the sake of notational simplicity, from now on we will fix $\sigma = 1$. In fact, this quantity enters in the expressions below only as a multiplicative term. In other terms, we can work with the Matern correlation function as well.

Lemma 2.4.1. *If* $\{Y(\mathrm{s}), \mathrm{s} \in D\}$ *is an isotropic random field with Matern covariance function* (2.1.3)*, a necessary condition for the existence of the directional derivative process is* $\nu > 1$.

Proof. Let $\mathcal{H}_\nu$ be the Modified Bessel function of order ν. We start by recalling two results concerning Modified Bessel functions (see formulas 9.6.6 and 9.6.28 in Abramowitz and Stegun (1965)), that is

a) $\mathcal{H}_{-\nu}(z) = \mathcal{H}_\nu(z)$ *(reflection formula)*;

b) $\frac{d^k}{dz^k}\left\{z^\nu \mathcal{H}_\nu(z)\right\} = z^\nu e^{-k\pi i}\mathcal{H}_{\nu-k}(z)$.

We need to determine the first two derivatives of the Matern covariance function. Let us consider equation b) above. For $k = 1$, $e^{k\pi i} = e^{\pi i} = -1$. It follows that $\frac{d}{dz}\left(z^\nu \mathcal{H}_\nu(z)\right) = -z^\nu \mathcal{H}_{\nu-1}(z)$ and the first derivative of the Matern covariance function is given by

$$K'_{\nu,\phi}(||\Delta||) = -\frac{\phi}{2^{\nu-1}\Gamma(\nu)}\left(\phi||\Delta||\right)^\nu \mathcal{H}_{\nu-1}(\phi||\Delta||).$$

Notice that, if $\nu > 1$, the former can be characterized as a function in the Matern class with smoothness parameter $(\nu - 1)$, that is

$$K'_{\nu,\phi}(||\Delta||) = -\frac{\phi^2||\Delta||}{2(\nu-1)} K_{\nu-1,\phi}(||\Delta||).$$

As a remark, notice that from the two expressions above we can conclude that $K_{\nu,\phi}(||\Delta||)$ is a decreasing function of $||\Delta||$ for all $\nu > 0$, as expected, since $\mathcal{H}_{\nu-1}(\phi||\Delta||)$ is a positive function of $||\Delta||$ for all $\nu > 0$ (see Abramowitz and Stegun (1965), 9.6.1).

Moreover, recalling the discussion below (2.1.3) and using the reflection formula, for $\nu = \frac{1}{2}$, we get $K'_{\frac{1}{2},\phi}(||\Delta||) = -\phi\, e^{-\phi||\Delta||}$ for any $||\Delta|| \neq 0$, as expected. Analogously, for $\nu = \frac{3}{2}$, we obtain $K'_{\frac{3}{2},\phi}(||\Delta||) = -\phi^2\, ||\Delta||\, e^{-\phi||\Delta||}$.

Now, we go back to the actual computation of $K''_{\nu,\phi}(\cdot)$. For $k = 2$, $e^{2k\pi i} = e^{2\pi i} = 1$. Therefore,

$$K''_{\nu,\phi}(||\Delta||) = -\frac{\phi^2}{2^{\nu-1}\Gamma(\nu)}\,(\phi||\Delta||)^{\nu-1}\,\mathcal{H}_{\nu-1}(\phi||\Delta||) + \frac{\phi^2}{2^{\nu-1}\Gamma(\nu)}\,(\phi||\Delta||)^{\nu}\,\mathcal{H}_{\nu-2}(\phi||\Delta||). \tag{2.4.1}$$

For the existence of the directional derivative process it is necessary to show that $K''(||\Delta||)$ is well behaved, in particular finite for any $||\Delta||$.

For $d = 1$, we know from chapter 2.7 in Stein (1999) that this is true for any $\nu > 1$. In fact, Stein shows that a process $\{Y(s), s \in \mathbb{R}\}$ with Matern covariance function $K_{\phi,\nu}$, where $\nu > 1$, is m-times mean square differentiable if and only if $\nu > m$, $m \in \mathbb{N}$.

However, we have already discussed that for a process on $\mathbb{R}^d$, differentiability and existence of the derivative processes are not equivalent. Therefore, the condition $\nu > 1$ is now just a necessary condition for 1-time differentiability.

To prove the lemma, we could indeed follow the same reasoning as in Stein. However, we pursue an alternate way, trying to infer the behavior of $K''_{\nu,\phi}(||\Delta||)$ for $||\Delta|| \to 0$ directly from expression (2.4.1).

First, notice that we need only to discuss the behavior of $K''_{\nu,\phi}(||\Delta||)$ as $||\Delta|| \to 0$. In fact, for any $||\Delta|| > 0$ and for any $\nu > 1$, $K''_{\nu,\phi}(||\Delta||)$ is positive and bounded, since $\mathcal{H}_\nu$ is a positive and bounded function for any $\nu > -1$. Then, let us concentrate on the case $||\Delta|| \to 0$.

For $\nu > 2$, it's again quite easy to prove that $K''_{\nu,\phi}(||\Delta||)$ is bounded. In fact, since

$$\mathcal{H}_\nu(z) \sim \frac{1}{2}\Gamma(\nu)(\frac{1}{2}z)^{-\nu}, \quad \nu > 0,$$

(see 9.6.9 in Abramowitz and Stegun (1965)), it follows that both $z^{\nu-1}\,\mathcal{H}_{\nu-1}(z)$ and

$z^\nu \mathcal{H}_{\nu-2}(z)$ are finite and positive for $z \to 0$, and so it is $K''_{\nu,\phi}(||\Delta||)$.

For $\nu \in (0,2]$, we can argue as follows. First, notice that $\mathcal{H}_{\nu-2}(z) = \mathcal{H}_{|\nu-2|}(z)$, by the reflection formula. Therefore, it's immediate to see that $z^\nu \mathcal{H}_{|\nu-2|}(z) \to \frac{1}{2^{\nu-1}}\Gamma(\nu) z^{2(\nu-1)}$ as $z \to 0$ and $\mathcal{H}_{\nu-2}$ is positive and bounded as long as $\nu > 1$. In particular, notice that for $\nu > 1$, $z^\nu \mathcal{H}_{|\nu-2|}(z) \to 0$ as $z \to 0$. □

As a final remark, notice that for $\nu > 2$, (2.4.1) can be characterized as follows

$$K''_{\nu,\phi}(||\Delta||) = -\frac{\phi^2}{2(\nu-1)}\left\{K_{\phi,\nu-1}(||\Delta||) - \frac{\phi^2}{2(\nu-2)}||\Delta||^2 K_{\phi,\nu-2}(||\Delta||)\right\}.$$

From the properties of the correlation functions and from the discussion above on the boundedness of $\mathcal{H}_{\nu-2}(z)$ for $z > 0$ and $\nu > 1$, it follows that as $||\Delta|| \to 0$

$$K''_{\nu,\phi}(||\Delta||) \to -\frac{\phi^2}{2(\nu-1)}K_{\phi,\nu-1}(0) = -\frac{\phi^2}{2(\nu-1)},$$

for all $\nu > 1$. In particular, recalling the discussion below (2.3.12), this means that $Var(D_u Y(s)) = -\frac{\phi^2}{2(\nu-1)}$.

Chapter 3

Directional rates of change under Spatial Dirichlet Process models.

In this chapter we study smoothness properties of samples from the Spatial Dirichlet Process (SDP) developed by Gelfand, Kottas and MacEachern (2004) and described in chapter 2. We start by recalling that a sample $Y(\cdot)$ from a SDP is a process $\{Y(\mathbf{s}), \mathbf{s} \in D\}$ such that $Y(\mathbf{s})|G \sim G$ where $G \sim SDP(\nu G_0)$, for some scale parameter ν and a base measure G_0. Any realization of the random probability measure G is almost surely discrete. In chapter 1 we have discussed the Sethuraman's representation of Dirichlet processes. This representation turns to be very useful to prove most of the result of this chapter. We refer to chapter 2 for the basic definitions and distributional theory concerning gradients of random fields.

The format of this chapter is as follows. In section 3.1.1, we discuss and prove some results on almost sure and mean square continuity of samples from a SDP, while in section 3.1.2 we deal with mean square differentiability of samples from the SDP. Distribution theory for directional finite difference and derivative processes obtained from samples of the SDP is discussed 3.2. In particular, we show that directional finite differences and derivatives are themselves samples from a SDP. In section 3.3, we address model fitting, inference and prediction. We use the SDP to model the

spatial component in a random spatial effect model. Interest is in prediction of finite differences and derivatives at old and new locations. The computations involved are detailed in section 3.4. Finally, in section 3.5 we provide a simulated example aimed at showing the relevant features of our model with respect to competing parametric ones. The following is joint work with A.E. Gelfand.

3.1 Smoothness properties of the Spatial Dirichlet Process.

We now consider random fields arising as a sample from a SDP and discuss their smoothness properties. In particular, in the next section we start by considering the continuity properties, while we discuss differentiability in section 3.1.2.

3.1.1 Almost sure and mean square continuity of samples from a SDP.

Since now the distribution of the field is itself random, we need to specify the definitions given in the previous section to this nonparametric context. In fact, let $\mathcal{P}$ denote the Dirichlet measure, that is the distribution of the random probability measure G defined on $(\Theta, \mathcal{B})$, where Θ is the space of surfaces over D and $\mathcal{B}$ a convenient σ-field. Let $C = \{\theta \in \Theta : \lim_{||s-s_0||\to 0} Y(s) = Y(s_0)\}$. Notice that data come from a given realization of G, which is almost surely discrete. We are interested on smoothness properties of the fields sampled from this realized distribution, analogously to what happens in the usual parametric context, where we are interested in properties of a field whose distribution is indexed by a fixed number of parameters, which we seek to estimate. In other words, we are interested in $Y(s)|G$ more than in $Y(s)$, the expected

sample from $\mathcal{P}$. Therefore, we can say that $Y(\mathbf{s})$ is a.s. continuous at a point $\mathbf{s}_0$ if $G(C) = 1$ for all G in a set of $\mathcal{P}$-measure one. For example, if we denote with S the set of random distribution functions chosen according to the Sethuraman's representation, we know that $\mathcal{P}(S) = 1$ and in the previous definition we can consider all G in S. Therefore, to investigate the a.s. continuity of samples from a SDP, we can limit the study to samples from an element in S. Analogously, we can say that $Y(\mathbf{s})$ is mean square continuous at a point $\mathbf{s}_0$ if $E[Y(\mathbf{s})^2|G] < \infty$ and $E[(Y(\mathbf{s}) - Y(\mathbf{s}_0))^2|G] \to 0$ as $||\mathbf{s} - \mathbf{s}_0|| \to 0$ for all G in a set of $\mathcal{P}$ measure one, e.g. all G in S.

We start discussing a.s. continuity and prove the following result

Theorem 3.1.1. *Any random field $Y(\mathbf{s})$ sampled from a SDP is a.s. continuous iff the random field G_0 is a.s. continuous.*

Proof. The proof follows after noticing that for all $\mathbf{s}, \mathbf{s}_0 \in D$ and for all $A \in \mathcal{B}$

$$G(Y(\mathbf{s}) - Y(\mathbf{s}_0) \in A) = \sum_{j=1}^{\infty} p_j \, \delta_{\theta_j^*(\mathbf{s}) - \theta_j^*(\mathbf{s}_0)}(A), \quad \text{a.s. -}\mathcal{P}, \tag{3.1.1}$$

that is $Y(\mathbf{s}) - Y(\mathbf{s}_0)$ is a sample from a SDP, with smoothness parameter ν and base measure the distribution of $\theta_1^*(\mathbf{s}) - \theta_1^*(\mathbf{s}_0)$. 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 $C \in \mathcal{B}$ defined as before.

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

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

Theorem 3.1.1 is actually an extension of a previous result in Gelfand, Kottas

and MacEachern (2004) and MacEachern (2000), where it is shown that if G_0 is a.s. continuous, then $Y(\mathbf{s})$ converges weakly to $Y(\mathbf{s}_0)$ as $||\mathbf{s} - \mathbf{s}_0|| \to 0$, for all G in S. Nevertheless, in the non-parametric context, their result can be conveniently restated to show the a.s. convergence of $G(Y(s))$ to $G(Y(s_0))$, as it is shown in the following proposition, where the result is formally proved.

Proposition 3.1.2. *In a SDP, if the base measure G_0 is a.s. continuous in $\mathbf{s}_0$, then the random probability measure $G(Y(s))$ converges a.s. to $G(Y(s_0))$ as $||\mathbf{s} - \mathbf{s}_0|| \to 0$.*

Proof. Let $\mathbb{P}$ be the set of all the probability measure defined on Θ and $G \in \mathbb{P}$ be a realization of the SDP random probability measure. For an arbitrary set $A \in \mathcal{B}$, let $P(Y(\mathbf{s}) \in A|G)$ denote the probability under G that a sample takes values in A at a site $\mathbf{s}$. Then, consider the set

$$C = \{G \in \mathbb{P} : \lim_{\|\mathbf{s}-\mathbf{s}_0\|\to 0} P(Y(\mathbf{s}) \in A|G) = P(Y(\mathbf{s}_0) \in A|G)\}.$$

For any fixed G in S, we have

$$\begin{aligned}\lim_{\|\mathbf{s}-\mathbf{s}_0\|\to 0} G(Y(\mathbf{s}) \in A) &= \sum_{j=1}^{\infty} p_j \lim_{\|\mathbf{s}-\mathbf{s}_0\|\to 0} \delta_{\theta_j^*(\mathbf{s})}(A) \\ &= \sum_{j=1}^{\infty} p_j \, \delta_{\theta_j^*(\mathbf{s}_0)}(A) = G(Y(\mathbf{s}_0) \in A), \quad \text{q.c.-}G_0\end{aligned}$$

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 $\mathbf{p} = (p_1, p_2, \ldots)$. Therefore, we can conclude that $S \cap C$ is indeed of the form $S \cap C = \mathbf{P} \times E^{\infty}$, where $\mathbf{P}$ is the support of the Poisson-Dirichlet distribution defined on the weights and $E = \{\theta \in \Theta : \lim_{\|\mathbf{s}-\mathbf{s}_0\|\to 0} \theta^*(\mathbf{s}) = \theta^*(\mathbf{s}_0)\}$. If we denote with $\pi(\mathbf{p})$ the distribution of $\mathbf{p}$ and exploit the independence structure of the SDP measure, we

get

$$\begin{aligned}\mathcal{P}(S \cap C) &= \int_{\mathbf{P}\times E^{\infty}} \pi(\mathbf{p})\, G_0^{\infty}(d\theta) \\ &= \int_{\mathbf{P}} \pi(\mathbf{p}) \int_{E} G_0(d\theta) = 1,\end{aligned}$$

since E has measure 1 w.r.t. G_0 by hypothesis. □

Turning to mean square continuity, note that if we marginalize with respect to $\mathcal{P}$, we get

$$E[(Y(\mathbf{s}) - Y(\mathbf{s}_0))^2] = E_{G_0}[(\theta_1^*(\mathbf{s}) - \theta_1^*(\mathbf{s}_0))^2. \tag{3.1.2}$$

Therefore, mean square continuity of an expected sample from $\mathcal{P}$ is equivalent to mean square continuity of the base measure. Since

$$E[(Y(\mathbf{s}) - Y(\mathbf{s}_0))^2] = E\left[E\left[(Y(\mathbf{s}) - Y(\mathbf{s}_0))^2|G\right]\right], \tag{3.1.3}$$

we are induced to expect that mean square continuity of the base measure implies mean square continuity of $Y(\mathbf{s})|G$. In fact, $E[(Y(\mathbf{s}) - Y(\mathbf{s}_0))^2|G]$ is a non negative random variable when considered as a function of G; therefore, by Fatou's Lemma, it follows that $E\left[\liminf_{||\mathbf{s}-\mathbf{s}_0||\to 0} E[(Y(\mathbf{s}) - Y(\mathbf{s}_0))^2|G]\right]$ exists and

$$E\left[\liminf_{||\mathbf{s}-\mathbf{s}_0||\to 0} E\left[(Y(\mathbf{s}) - Y(\mathbf{s}_0))^2|G\right]\right] \leq \lim_{||\mathbf{s}-\mathbf{s}_0||\to 0} E\left[(Y(\mathbf{s}) - Y(\mathbf{s}_0))^2\right] = 0, \tag{3.1.4}$$

leading to

$$\liminf_{||\mathbf{s}-\mathbf{s}_0||\to 0} E\left[(Y(s) - Y(s_0))^2|G\right] = 0 \quad \text{a.s.-}\mathcal{P}.$$

However, in order to get the desired result, we need some extra-conditions to ensure the existence of the limit. It turns out that a.s. continuity by itself is enough when the process is defined on a compact space. In fact, we can prove the following proposition.

Proposition 3.1.3. *Let G_0 be a separable process a.s. continuous on a compact $K \subset D$, where D is an Hausdorff space. Then, any random field $Y(\mathbf{s})$ sampled from a SDP is mean square continuous on K.*

Proof. 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

$$\lim_{\|s-s_0\|\to 0} E[(Y(s) - Y(s_0))^2|G] = \lim_{\|s-s_0\|\to 0} \sum_{j=1}^{\infty} p_j (\theta_j^*(s) - \theta_j^*(s_0))^2 = 0, \quad \text{a.s.-}\mathcal{P}. \quad (3.1.5)$$

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 (3.1.5) 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 (3.1.5). 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 (3.1.5) is bounded by the integrable function Z^2 and the conditions for the dominated convergence theorem are satisfied. □

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 (3.1.3) 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 bounded convergence theorem to obtain also mean square continuity of the

base measure, i.e. of the marginal samples, since

$$\lim_{||s-s_0||\to 0} E\left((\theta^*(s)-\theta^*(s_0))^2\right) = E\left(\lim_{||s-s_0||\to 0}(\theta^*(s)-\theta^*(s_0))^2\right).$$

The previous discussion shows that, when we fix a distribution G, mean square continuity of the base measure is not sufficient to claim mean square continuity of the samples $Y(s)$ from G. This is not totally unexpected, since any G is a discrete probability measure with probability 1, and therefore we expect that its smoothness properties depend on the smoothness of the realized surfaces $\theta_j(s)$, $j = 1, 2, \ldots$ which define the support of G.

However, we can show that $G(Y(s))$ is a random variable which converges in distribution to $G(Y(s_0))$. Intuitively, we can see this by the following argument. From (3.1.3), it follows that $E[(Y(s)-Y(s_0))^2|G]$ is a random variable which converges to 0 in mean (with respect to $\mathcal{P}$). Since

$$E\left[(Y(s)-Y(s_0))^2|G\right] \geq \left(E\left[(Y(s)-Y(s_0))|G\right]\right)^2,$$

by Jensen's inequality, it follows that $E[Y(s)|G]$ converges in L_2 (a fortiori, in distribution) to $E[Y(s_0)|G]$. But $E[Y(s)|G] = \sum_{j=1}^{\infty} p_j\,\theta_j^*(s)$ q.c.-$\mathcal{P}$, by Sethuraman's representation. Recall that the θ_j's are i.i.d. draws from G_0, chosen independently from p_j, $j = 1, 2, \ldots$. Therefore, we can conclude that $E[Y(s)|G]$ converges in distribution to $E[Y(s_0)|G] = \sum_{j=1}^{\infty} p_j\,\theta_j^*(s_0)$, 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 the distribution of $\theta(s_0)$, say G_{0,s_0}. In symbols, $Y(s_0)|G_{s_0} \sim G_{s_0}$, $G_{s_0} \sim DP(\nu G_{0,s_0})$. In fact, this is a particular case of a more general result, which is proved in the next proposition.

Proposition 3.1.4. *Let $Y(s)|G \sim G$, $G \sim DP(\nu G_{0,s})$, $s \in D$. Let $g(\cdot)$ be a real valued measurable function defined on $(\Theta, \mathcal{B})$, which is integrable with respect to $G_{0,s}$.*

Let $G^g_{0,\mathrm{s}}$ be the distribution of $g(\theta^*(\mathrm{s}))$ induced from $G_{0,\mathrm{s}}$. Moreover, let s_0 be an arbitrary point in D, and suppose that $G^g_{0,\mathrm{s}}$ converges weakly to a distribution H_{0,s_0}, as $||\mathrm{s} - \mathrm{s}_0|| \to 0$. Then, $g(Y(\mathrm{s}))$ 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}.

Proof. First, notice that $g(Y(\mathrm{s}))$ is a sample from a DP, with parameter ν and base measure $G^g_{0,\mathrm{s}}$, i.e. we can define $G_{\mathrm{s}}(g) = \sum_{j=1}^{\infty} p_j \, \delta_{g(\theta^*_j(\mathrm{s}))}(\cdot)$ and $g(Y(\mathrm{s}))|G_{\mathrm{s}}(g) \sim G_{\mathrm{s}}(g)$ and $G_{\mathrm{s}}(g) \sim G^g_{0,\mathrm{s}}$. By the usual properties of the Dirichlet process, it follows that if $G^g_{0,\mathrm{s}}$ converges weakly, then also $g(Y(\mathrm{s}))$ converges in distribution (with respect to the collection of Dirichlet measures defined at each $\mathrm{s} \in D$). 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 $G_{\mathrm{s}}(g)$ converges in distribution to the random probability measure $\tilde{G} = \sum_{j=1}^{\infty} p_j \, \delta_{\xi_i(\mathrm{s}_0)}(\cdot)$, where $\xi_i(\mathrm{s}_0) \overset{i.i.d.}{\sim} H_{0,\mathrm{s}_0}$.
In order to prove this result, we use the characteristic function method illustrated in Ishwaran and Zarepour (2002b).
We start recalling that Sethuraman (1994) proves that $G \sim DP(\nu G_0)$ is the unique solution satisfying the following distributional equation

$$G \overset{\mathcal{D}}{=} q_1 \, \delta_{\theta^*_1(\mathrm{s})} + (1 - q_1) \, G, \tag{3.1.6}$$

where, on the right hand side, q_1 has distribution $Beta(1, \nu)$, $\theta^*_1(\mathrm{s})$ is independent of q_1 and G is independent of $(q_1, \theta^*_1(\mathrm{s}))$.
Now, let $\psi(t, \mathrm{s}) = E\{\exp itG(g, \mathrm{s})\}$ denote the characteristic function of $G_{\mathrm{s}}(g)$. By (3.1.6), it follows that

$$\psi(t, \mathrm{s}) = E\left\{\exp\left\{it\left[q_1 \, g(\theta^*_1(\mathrm{s})) + (1 - q_1) \, G_{\mathrm{s}}(g)\right]\right\}\right\}. \tag{3.1.7}$$

Analogously, let $\phi(t, \mathrm{s})$ denote the characteristic function of $g(\theta^*_1(\mathrm{s}))$. Then, we can

exploit the independence relations in order to obtain

$$\psi(t, \mathbf{s}) = E_{q_1} \left\{ \phi(t\, q_1, \mathbf{s})\, \psi(t\,(1 - q_1), \mathbf{s}) \right\}, \tag{3.1.8}$$

where the expected value is taken with respect to the distribution of q_1. Any characteristic function satisfying (3.1.8) must be the characteristic function for $G_{\mathbf{s}}(g)$.

Now consider $\psi^*(t, \mathbf{s}_0) = \lim_{||\mathbf{s}-\mathbf{s}_0||\to 0} \psi(t, \mathbf{s})$, and $\phi^*(t, \mathbf{s}_0) = \lim_{||\mathbf{s}-\mathbf{s}_0||\to 0} \phi(t, \mathbf{s})$. Since $g(\theta_1^*(\mathbf{s}))$ converges in distribution to $\xi(\mathbf{s}_0)$, $\phi^*(t, \mathbf{s}_0)$ is the characteristic function of $\xi(\mathbf{s}_0)$.

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

$$\psi^*(t, \mathbf{s}_0) = E_{q_1} \left\{ \phi^*(t\, q_1, \mathbf{s}_0)\, \psi^*(t\,(1 - q_1), \mathbf{s}_0) \right\},$$

which is the characteristic function of $\tilde{G}$, by the uniqueness of the solution of (3.1.8). Then, we have proved that $g(Y(\mathbf{s}))$ converges to a random variable, say $Z(s_0)$, such that $Z(s_0)|\tilde{G} \sim \tilde{G}$ and $\tilde{G} \sim DP(\nu H_{0,\mathbf{s}_0})$. □

3.1.2 Mean square differentiability of samples from a SDP.

Now we turn attention to mean square differentiability of a process arising from a SDP. We consider the finite difference process $Y_{\mathbf{u},h}(\mathbf{s})$ and define the directional derivative process $D_{\mathbf{u}}Y(\mathbf{s})$ as that process satisfying

$$\lim_{h\to 0} E[(Y_{\mathbf{u},h}(\mathbf{s}) - D_{\mathbf{u}}Y(\mathbf{s}))^2 | G] = 0, \tag{3.1.9}$$

for all G in set of $\mathcal{P}$-measure 1. Moreover, if $D_{\mathbf{u}}Y(\mathbf{s})$ is a linear function of $\mathbf{u}$, the process $Y(\mathbf{s})$ is mean square differentiable.

As before, if we consider the properties of the process $Y(\mathbf{s})$ marginally with respect to the Dirichlet measure, they are immediately derived from those of the base measure.

In fact, for any scalar $h_n, h_m > 0$, and any $\mathrm{s} \in D$,

$$E\left(Y_{\mathbf{u},h_n}(\mathrm{s}) - Y_{\mathbf{u},h_m}(\mathrm{s})\right)^2 = E_{G_0}\left(\theta^*_{\mathbf{u},h_n}(\mathrm{s}) - \theta^*_{\mathbf{u},h_m}(\mathrm{s})\right)^2. \tag{3.1.10}$$

Therefore, 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, given a realization G from the SDP, mean square differentiability of the base measure is not helpful to decide about mean square differentiability of a sample from G. In fact, the former relies on the analytical properties of the surfaces specifying the support of G, analogously to what we have discussed in proposition 3.1.3 for mean square continuity.

Proposition 3.1.5. *Let G_0 be a.s. continuously differentiable on a compact $K \subset D$, where D is an Hausdorff Space. Then, any random field $Y(\mathrm{s})$ sampled from a SDP is mean square differentiable on $\dot{K} = int(K)$.*

Proof. The proof mimics that of proposition 3.1.3. In fact, let $h_n, h_m > 0$ be arbitrary scalars, and $\mathrm{s}_0 \in \dot{K}$. Then,

$$E\left(Y_{\mathbf{u},h_n}(\mathrm{s}_0) - Y_{\mathbf{u},h_m}(\mathrm{s}_0)|\, G\right)^2 = \sum_{j=1}^{\infty} p_j \left(\theta^*_{\mathbf{u},h_n}(\mathrm{s}_0) - \theta^*_{\mathbf{u},h_m}(\mathrm{s}_0)\right)^2 \quad \text{q.c.-}\mathcal{P}. \tag{3.1.11}$$

Since $\theta^*_j(\mathrm{s}) \in C^1$, $\lim_{h_n,h_m\to 0}(\theta^*_{\mathbf{u},h_n}(\mathrm{s}_0) - \theta^*_{\mathbf{u},h_m}(\mathrm{s}_0))^2 = 0$ and $\theta^*_j(\mathrm{s}_0)$ has directional derivative $D_{\mathbf{u}}\theta^*_j(\mathrm{s}_0) = \mathbf{u}^T\nabla_{\theta^*}(\mathrm{s}_0)$. Therefore, $Y_{\mathbf{u},h_n}(\mathrm{s}_0)|G$ converges in L^2 with respect to G to a random variable $D_{\mathbf{u}}Y(\mathrm{s})|G = \mathbf{u}^T\nabla_Y(\mathrm{s}_0)|G$, whose distribution is $\sum_{j=1}^{\infty} p_j\, \delta_{D_{\mathbf{u}}\theta^*_j(\mathrm{s}_0)}(\cdot)$, since

$$\lim_{h_n\to 0}\sum_{j=1}^{\infty} p_j \left(\theta^*_{\mathbf{u},h_n}(\mathrm{s}_0) - \mathbf{u}^T\nabla_{\theta^*}(\mathrm{s}_0)\right)^2 = \sum_{j=1}^{\infty} p_j \lim_{h_n\to 0}\left(\theta^*_{\mathbf{u},h_n}(\mathrm{s}_0) - \mathbf{u}^T\nabla_{\theta^*}(\mathrm{s}_0)\right)^2,$$

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

Banerjee and Gelfand (2003) have discussed what conditions are required on the covariance function $K(\mathbf{s})$ of a stationary process in order that the directional derivative process $D_{\mathbf{u}}Y(\mathbf{s})$ has a.s. continuous realizations. Suppose $K(\mathbf{s}) \in C^{d+2}$, meaning it is d-times continuously differentiable. Let $P_d(\mathbf{s})$ denote the Taylor polynomial in $\mathbf{s}$ of degree d and $K_d(\mathbf{s})$ denote the remainder term, i.e. $K_d(\mathbf{s}) = K(\mathbf{s}) - P_d(\mathbf{s})$. Then, applying Kent's result to the covariance of the directional derivative process, they found that a sufficient condition for a stationary process to have a.s. continuous derivatives is that $K_d(\mathbf{s}) = O(||\mathbf{s}||^{d+2+\beta})$, for some $\beta > 0$.

In conclusion, even if the base process G_0 is mean square differentiable, it doesn't follow that also $Y(\mathbf{s})|G$ 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 (3.1.4) to show that the limit $D_{\mathbf{u}}Y(\mathbf{s})$ is indeed obtained as a sample from a SDP whose base measure is the distribution of $D_{\mathbf{u}}\theta(\mathbf{s})$, that is the directional derivative of 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.

3.2 Some distribution theory.

Let $Y(\mathbf{s})$, $\mathbf{s} \in D$ be a random field sampled from a $SDP(\nu G_0)$ and $Y_{\mathbf{u},h}(\mathbf{s})$ be the associated directional finite difference process, for some fixed unit vector $\mathbf{u}$ and scalar $h > 0$. Then it is easy to prove that also $Y_{\mathbf{u},h}(\mathbf{s})$ is a sample from a SDP with same precision parameter ν and with base measure the distribution of the finite difference

process $\theta^*_{\mathbf{u},h}(\mathbf{s})$, say $G_0^{\mathbf{u},h}$. In fact, for any pair $(\mathbf{s}, \mathbf{s}+h\mathbf{u}) \in D$ and any real t, consider

$$P\left(Y_{\mathbf{u},h}(\mathbf{s}) \leq t \middle| G\right) = P\left(Y(\mathbf{s}+h\mathbf{u}) \leq Y(\mathbf{s}) + th \middle| G\right)$$
$$= \int P\left(Y(\mathbf{s}+h\mathbf{u}) \leq Y(\mathbf{s}) + th \middle| Y(\cdot), G\right)\ G(dY(\cdot)),$$

where we use $Y(\cdot)$ to denote the whole process and exploit the fact that $Y(\cdot)|G \sim G$. Notice that, since G is given and $Y(\cdot)$ is sampled from G, the integrand is just an indicator function, which takes value one if $Y(\mathbf{s}+h\mathbf{u}) \leq Y(\mathbf{s}) + th$. Therefore,

$$P\left(Y_{\mathbf{u},h}(\mathbf{s}) \leq t \middle| G\right) = \int I_{(-\infty, Y(\mathbf{s})+th]}(Y(\mathbf{s}+h\mathbf{u}))\ G(dY(\cdot))$$
$$= \sum_{j=1}^{\infty} p_j\ I_{(-\infty, \theta^*_j(\mathbf{s})+th]}(\theta^*_j(s+h\mathbf{u})), \quad \text{a.s.-}\mathcal{P}$$

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

$$I_{(-\infty, \theta^*_j(\mathbf{s})+th]}(\theta^*_j(s+h\mathbf{u})) = I_{(-\infty, t]}\left(\theta^{*j}_{\mathbf{u},h}(\mathbf{s})\right),$$

so to conclude that $Y_{\mathbf{u},h}(\mathbf{s})|\,G$ is a sample from $\sum_{j=1}^{\infty} p_j\, \delta_{\theta^{*j}_{\mathbf{u},h}(\mathbf{s})}(\cdot)$, 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

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

and for any pair of locations $(\mathbf{s}, \mathbf{s}')$ in D,

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

while the distribution of the marginal process $Y_{\mathbf{u},h}(\mathbf{s})$ is the same as the distribution of the base process $\theta^*_{\mathbf{u},h}(\mathbf{s})$ (see section 2.3).

Now consider the directional derivative process $D_{\mathbf{u}}Y(\mathbf{s})$ and suppose that G_0 admits directional derivatives in all directions. Let $G'_{0,\mathbf{u}}$ denote the distribution of the process $D_{\mathbf{u}}\theta^*(\mathbf{s})$. Then, we can apply proposition (3.1.4) in order to prove that $D_{\mathbf{u}}Y(\mathbf{s})$ is a sample from a SDP with smooth parameter ν and base measure $G'_{0,\mathbf{u}}$. In symbols, $D_{\mathbf{u}}Y(\mathbf{s})|G'_{\mathbf{u}} \sim G'_{\mathbf{u}}$ and $G'_{\mathbf{u}} \sim SDP(\nu G'_{0,\mathbf{u}})$. In fact, we have showed that for any direction $\mathbf{u}$ and any fixed h, the finite difference process $Y_{\mathbf{u},h}(\mathbf{s})$ is a sample from a SDP, i.e. $Y_{\mathbf{u},h}(\mathbf{s})|G^{\mathbf{u},h} \sim G^{\mathbf{u},h}$, $G^{\mathbf{u},h} \sim SDP(\nu G_0^{\mathbf{u},h})$. Therefore, marginally at any given site $\mathbf{s}$, $Y_{\mathbf{u},h}(\mathbf{s})$ is just a sample from a DP$(\nu\, G^{\mathbf{u},h}_{0,\mathbf{s}})$, where $G^{\mathbf{u},h}_{0,\mathbf{s}}$ is the marginal distribution of $G_0^{\mathbf{u},h}$ in $\mathbf{s}$. Since G_0 admits directional derivatives in all directions, $G^{\mathbf{u},h}_{0,\mathbf{s}}$ converges weakly to $G'_{0,\mathbf{u}}$. Therefore, the conditions of Proposition (3.1.4) are satisfied and the previous assertion is proved.

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

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

from the mean square convergence of $\theta^*_{\mathbf{u},h}(\mathbf{s})$ to $D_{\mathbf{u}}\theta^*(vs)$, it follows that $E\left(Y_{\mathbf{u},h}(\mathbf{s})|G\right)$, that is the mean of the DP for $Y_{\mathbf{u},h}(\mathbf{s})$, converges in L_2 (with respect to the Dirichlet Measure $\mathcal{P}$) to the random variable $E\left(D_{\mathbf{u}}Y(\mathbf{s})|G\right)$. But $E\left(Y_{\mathbf{u},h}(\mathbf{s})|G\right) = \sum_{j=1}^{\infty} p_j\, \theta^*_{\mathbf{u},h}(\mathbf{s})$ q.c.-$\mathcal{P}$ 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, $E\left(Y_{\mathbf{u},h}(\mathbf{s})|G\right)$ converges in distribution to $E(D_{\mathbf{u}}Y(\mathbf{s})|G) = \sum_{j=1}^{\infty} p_j\, D_{\mathbf{u}}\theta^*_j(\mathbf{s})$, that is the mean of the probability mass function $\sum_{j=1}^{\infty} p_j\, \delta_{D_{\mathbf{u}}\theta^*_j(\mathbf{s})}(\cdot)$, which is the almost sure representation of $G'_{\mathbf{u}}$. Therefore, it is immediate to guess that $G(Y_{\mathbf{u},h}(\mathbf{s}))$ converges in distribution to $G'_{\mathbf{u}}(D_{\mathbf{u}}Y(\mathbf{s}))$, and by the uniqueness of the limit, that $D_{\mathbf{u}}Y(\mathbf{s})$ 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 recover the first and second moments of the directional derivative process. In fact,

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

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

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

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

Again, the distribution for the marginal process coincides with that of $D_{\mathbf{u},h}\theta^*(\mathbf{s})$ and has been described by Banerjee, Gelfand and Sirmans (2003).

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

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

according to Sethuraman's representation. Here $\nabla_{\theta_j^*(\mathbf{s})} = (D_{\mathbf{e}_1}\theta^*(\mathbf{s}), \ldots, D_{\mathbf{e}_d}\theta^*(\mathbf{s}))$ 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 fact, if the base measure is mean square differentiable, we can study the behavior of $D_{\mathbf{u}}Y(\mathbf{s})|G'_{\mathbf{u}}$ in arbitrary directions by means of an orthonormal basis $(\mathbf{e}_1, \ldots, \mathbf{e}_d)$, i.e. for any direction $\mathbf{u}$, there exists a set of weights $(w_1, \ldots, w_d)$ such that $\mathbf{u} = \sum_{i=1}^{d} w_i\mathbf{e}_i$ with $w_i = \mathbf{u}^T\mathbf{e}_i$, $\sum_{i=1}^{d} w_i^2 = 1$, and

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

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

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

and

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

Now consider the bivariate process $\mathbf{Z}_{\mathbf{u}}^{(h)}(\mathbf{s}) = \begin{pmatrix} Y(\mathbf{s}) \\ Y_{\mathbf{u},h}(\mathbf{s}) \end{pmatrix}$. We can follow arguments similar to those outlined before for the finite difference process $Y_{\mathbf{u},h}(\mathbf{s})$ and prove that for any $\mathbf{s}$ in D, $\mathbf{Z}_{\mathbf{u}}^{(h)}(\mathbf{s})$ is a sample from a bivariate Dirichlet process. In fact, for any real h, t_1, t_2,

$$\begin{aligned} P\left(Y(\mathbf{s}) \le t_1, Y_{\mathbf{u},h}(\mathbf{s}) \le t_2|G\right) &= E\left(E\left(\delta_{Y(\mathbf{s})}(-\infty, t_1)\, \delta_{Y_{\mathbf{u},h}(\mathbf{s})(-\infty,t_2)}|\, Y(\cdot), G\right)\right) \\ &= \sum_{i=1}^{\infty} p_i\, \delta_{\theta_i^*(\mathbf{s})}(-\infty, t_1)\, \delta_{\theta_{\mathbf{u},h}^{i*}(\mathbf{s})}(-\infty, t_2). \end{aligned}$$

Therefore, $\mathbf{Z}_{\mathbf{u}}^{(h)}|G$ is a sample from a SDP with precision parameter ν and base measure the joint distribution of $\begin{pmatrix} \theta^*(\mathbf{s}) \\ \theta_{\mathbf{u},h}^*(\mathbf{s}) \end{pmatrix}$ (see Banerjee, Gelfand and Sirmans (2003)). Accordingly, the cross-covariance functions are given by

$$Cov(Y(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s})|G) = \sum_{i=1}^{\infty} p_i\, \theta_i^*(\mathbf{s})\, \theta_{\mathbf{u},h}^{i*}(\mathbf{s}),$$

and

$$Cov(Y(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s})) = Cov(\theta^*(\mathbf{s}), \theta_{\mathbf{u},h}^*(\mathbf{s})),$$

respectively, for the conditional and the marginal processes.

We can define the bivariate process $\mathbf{Z}_{\mathbf{u}}(\mathbf{s}) = \begin{pmatrix} Y(\mathbf{s}) \\ D_{\mathbf{u}}Y(\mathbf{s}) \end{pmatrix}$ as the L_2 limit of $\mathbf{Z}_{\mathbf{u}}^{(h)}(\mathbf{s})$ with respect to the Dirichlet measure $\mathcal{P}$. Recall that $\mathbf{Z}_{\mathbf{u}}^{(h)}(\mathbf{s})$ converges in L^2 to $\mathbf{Z}_{\mathbf{u}}(\mathbf{s})$

if

$$\lim_{h\to 0} E\left(\|\mathbf{Z}_{\mathbf{u}}^{(h)}(\mathbf{s}) - \mathbf{Z}_{\mathbf{u}}(\mathbf{s})\|^2\right) = 0.$$

If $||\cdot||$ is the Euclidean norm on $\mathbb{R}^d$, convergence in L_2 is equivalent to convergence of each of the vector components (see Banerjee and Gelfand (2003)). Therefore, we can proceed as in proposition (3.1.4) to show the convergence of the random probability measures and conclude that $\mathbf{Z}_{\mathbf{u}}(\mathbf{s})$ is a sample from a SDP with precision parameter ν and base measure the distribution of $\begin{pmatrix} \theta^*(\mathbf{s}) \\ D_{\mathbf{u}}\theta^*(\mathbf{s}) \end{pmatrix}$. Again, the necessary distribution theory is obtained from the standard distribution theory of the Dirichlet processes and that of directional derivative processes.

Similar arguments can of course be developed for the vectors $\mathbf{Z}_{\mathbf{u}_1,\mathbf{u}_2,h}(\mathbf{s}) = \begin{pmatrix} Y(\mathbf{s}) \\ Y_{\mathbf{u}_1,h}(\mathbf{s}) \\ Y_{\mathbf{u}_2,h}(\mathbf{s}) \end{pmatrix}$ and $\mathbf{Z}_{\mathbf{u},h_1,h_2}(\mathbf{s}) = \begin{pmatrix} Y(\mathbf{s}) \\ Y_{\mathbf{u},h_1}(\mathbf{s}) \\ Y_{\mathbf{u},h_2}(\mathbf{s}) \end{pmatrix}$ and so on, for any pair of directions $\mathbf{u}_1, \mathbf{u}_2$ and for any real h_1, h_2. In particular, we could consider the vector $\begin{pmatrix} Y(\mathbf{s}) \\ Y_{\mathbf{u}_1,h}(\mathbf{s}) \\ D_{\mathbf{u}_2}Y(\mathbf{s}) \end{pmatrix}$, and get the cross-covariance function between the finite difference and the directional derivative processes $Y_{\mathbf{u},h}(\mathbf{s})$ and $D_{\mathbf{u}}Y(\mathbf{s})$.

3.3 Model fitting and inference.

We work in a $d = 2$ dimensional space and suppose to observe T replicates of a random field $\{Y(\mathbf{s}), \mathbf{s} \in D\}$ in n locations $(\mathbf{s}_1, \ldots, \mathbf{s}_n)$. We assume that $Y(\mathbf{s})$ arises from a

random spatial effects model such that at any location s in D and for $t = 1, \ldots, T$,

$$Y_t(\mathbf{s}) = \mu_t(\mathbf{s}) + \theta_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}). \tag{3.3.1}$$

The mean structure component can be either constant or, more frequently, a regressive term $\mathbf{x}(\mathbf{s})^T\beta$. The elements of the p-dimensional vector $\mathbf{x}(\mathbf{s})$ can be diverse. For example, they can be functions of geographical coordinates, so that the regressive term represents indeed a trend surface and therefore it is constant across replicates. More often, $\mathbf{x}(\mathbf{s})$ will be a vector of covariates varying with t and therefore it will be appropriately denoted with $\mathbf{x}_t(\mathbf{s})$. Choice of the appropriate modeling for the mean term can be suggested by the problem at hand. For example, in the study of land values gradients, Majumdar *et al.* (2004) consider $\mathbf{x}(\mathbf{s}) = e^{-\|\mathbf{s}-\mathbf{s}^*\|}$ or $\mathbf{x}(\mathbf{s}) = 1/(a + \|\mathbf{s}-\mathbf{s}^*\|)^b$, 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 discrict located in s^*. As an example of the second choice, in studying selling prices of single family homes, Banerjee, Gelfand and Sirmans (2003) consider a vector of home specific covariates including its age, square feet of living area, other area, and number of bathrooms.

Notice that we don't consider any dynamical evolution of the model for now, so that the same model is assumed accross t's. In other words, let us denote with $\mathbf{Y}_t = (Y_t(\mathbf{s}_1), \ldots, Y_t(\mathbf{s}_n))^T$ the vector of observed values at the n locations for each $t = 1, \ldots, T$. Correspondingly, let $\mathbf{X}_t$ be the $n \times p$ matrix whose i-th column is the vector $\mathbf{x}_t(\mathbf{s}_i) = (\mathbf{x}_{t,1}(\mathbf{s}_i), \ldots, \mathbf{x}_{t,p}(\mathbf{s}_i))^T$, $i = 1, \ldots, n$, and $\boldsymbol{\theta}_t = (\theta_t(\mathbf{s}_1), \ldots, \theta_t(\mathbf{s}_n))^T$ be the vector of the spatial components, $t = 1, \ldots, T$. Then, we assume that $\mathbf{Y}_t$ given β and $\boldsymbol{\theta}_t$ are drawn indipendently from a density $f(\mathbf{Y}_t|\mathbf{X}_t^T\beta + \boldsymbol{\theta}_t, \tau^2)$, usually assumed to be gaussian. The vector of spatial components $\boldsymbol{\theta}_t$ is a sample from a SDP, such that $\boldsymbol{\theta}_t|G^{(n)} \sim G^{(n)}$. 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 $(H_n(\phi))_{i,j} = \rho_\phi(\mathbf{s}_i, \mathbf{s}_j)$ is the correlation function, indexed by some vector of parameters ϕ. Hence, we can specify the following semiparametric hierarchical model,

$$\begin{aligned}
\mathbf{Y}_t|\,\beta, \boldsymbol{\theta}_t, \tau^2 &\overset{ind.}{\sim} N_n(\mathbf{Y}_t|\,\mathbf{X}_t^T\beta + \boldsymbol{\theta}_t, \tau^2 I_n), \quad t = 1, \ldots, T \\
\boldsymbol{\theta}_t|G^{(n)} &\overset{i.i.d.}{\sim} G^{(n)}, \quad t = 1, \ldots, T \\
G^{(n)}|\,\nu, \sigma^2, \phi &\sim DP(\nu G_0^{(n)}), \quad G_0^{(n)}(\cdot|\,\mathbf{0}_n, \sigma^2 H_n(\phi)) \\
\beta, \tau^2 &\sim N_p(\beta|\,\beta_0, \Sigma_\beta) \times IGamma(\tau^2|\,a_\tau, b_\tau) \\
\nu, \sigma^2, \phi &\sim Gamma(\nu|\,a_\nu, b_\nu) \times IGamma(\sigma^2|\,a_\sigma, b_\sigma) \times [\phi],
\end{aligned} \tag{3.3.2}$$

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 matern covariance class, with decay parameter ϕ and smoothness parameter $\nu \in (1, 2)$, i.e. our base process is only once differentiable. Accordingly, a Gamma prior is considered for ϕ and a uniform on $(1, 2)$ for ν.

We assume to be able to observe T replicates of the random field at each location. As noted in Gelfand, Kottas and MacEachern (2004), this 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 to a conventional parametric specification. Well-known results already present in the literature ensure consistency of DP mixtures (see Ghosal, Ghosh and Ramamoorthi (1999)).

Inference is typically sought for directional finite differences and derivatives of the mean process $m(\mathbf{s}) = E(Y(\mathbf{s})|\,\mathbf{X}, \beta, \theta) = \mathbf{x}(\mathbf{s})^T\beta + \theta(\mathbf{s})$. Here $\mathbf{s}$ can denote an element either in the original set of locations $(\mathbf{s}_1, \ldots, \mathbf{s}_n)$ or in a new set $(\tilde{\mathbf{s}}_1, \ldots, \tilde{\mathbf{s}}_m)$. Notice

that, in the second case, we need to know the value of the covariates matrix at these locations. Moreover, the covariates matrix itself can be thought as a realization from a multivariate random field, i.e. the p-dimensional vector process $\{\mathbf{x}(\mathbf{s}), \mathbf{s} \in D\}$. Then, the finite difference and the directional derivative processes of $m(\mathbf{s})$ are, respectively, given by

$$\begin{aligned} m_{\mathbf{u},h}(\mathbf{s}) &= \frac{E(Y(\mathbf{s}+\mathbf{u}h)|\,\mathbf{X},\beta,\theta) - E(Y(\mathbf{s})|\,\mathbf{X},\beta,\theta)}{h} \\ &= \mathbf{x}_{\mathbf{u},h}(\mathbf{s})^T\beta + \theta_{\mathbf{u},h}(\mathbf{s}), \end{aligned} \tag{3.3.3}$$

and

$$D_{\mathbf{u}}m(\mathbf{s}) = D_{\mathbf{u}}E(\mathbf{Y}(\mathbf{s})|\,\mathbf{X},\beta,\theta) = D_{\mathbf{u}}\mathbf{x}(\mathbf{s})^T\beta + D_{\mathbf{u}}\theta(\mathbf{s}), \tag{3.3.4}$$

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 $m_{\mathbf{u},h}(\mathbf{s})$. 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 $X(\mathbf{s})$. Let $Z(\mathbf{s}) = g(X(\mathbf{s}))$ for some arbitrary functional g. Then, Majumdar *et al.* (2004) prove that $Z(\mathbf{s})$ has directional derivative process given by $D_{\mathbf{u}}Z(\mathbf{s}) = g'(X(\mathbf{s}))D_{\mathbf{u}}X(\mathbf{s})$. In other words, this defines a simple chain rule for directional derivative processes. Of course, if we consider a covariate process, we need to extend the semiparametric model described above to take also into account distributional assumptions on $\mathbf{x}(\mathbf{s})$.

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 $[\boldsymbol{\theta}, \boldsymbol{\beta}, \tau^2, \nu, \sigma^2, \phi \mid \text{data}]$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_T)$ and data $=\{\mathbf{Y}_1, ..., \mathbf{Y}_T\}$ 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 (3.3.2)

has been described in Gelfand, Kottas and MacEachern (2004). 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 $\boldsymbol{\theta}$ and with $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_T^*)$ the vector that collects only the distinct $\boldsymbol{\theta}_t$'s. Notice that we can switch back and forth from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}^*$, once we define a vector of labels $\mathbf{w} = (w_1, \ldots, w_T)$, such that $w_t = j$ if and only if $\boldsymbol{\theta}_t = \boldsymbol{\theta}_j^*$, $t = 1, \ldots, T$. Then, $(\boldsymbol{\theta}^*, \mathbf{w})$ is an equivalent representation of $\boldsymbol{\theta}$ and posterior draws from $[\boldsymbol{\theta}, \beta, \tau^2, \sigma^2, \phi, \nu, \mathbf{Y}_t, t = 1, \ldots, T]$ are the same as posterior draws from $[\boldsymbol{\theta}^*, \mathbf{w}, T^*, \beta, \tau^2, \sigma^2, \phi, \nu, \mathbf{Y}_t, t = 1, \ldots, T]$.

As anticipated, interest is in prediction of the directional finite differences and derivatives of the mean process $m(\mathbf{s})$ at locations where the random field $\{Y(\mathbf{s}),\ \mathbf{s} \in D\}$ is not observed. Hereafter, we use $V_{\mathbf{u}}(\mathbf{s})$ to denote either $m_{\mathbf{u},h}(\mathbf{s})$ or $D_{\mathbf{u}}m(\mathbf{s})$. We also use $V_{\mathbf{u}}^X(\mathbf{s})$ and $V_{\mathbf{u}}^\theta(\mathbf{s})$ to indicate specifically the directional finite differences or derivatives of the mean structure and spatial component in (3.3.1). Then, according to (3.3.3) and (3.3.4), $V_{\mathbf{u}}(\mathbf{s}) = V_{\mathbf{u}}^X(\mathbf{s}) + V_{\mathbf{u}}^\theta(\mathbf{s})$ and if we assume independence between the processes for the covariates and the spatial component, we get

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

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

$$f_{V_{\mathbf{u}}}(y|\,\text{data}) = \int f_{V_{\mathbf{u}}^X}(y - z|\,\text{data})\, f_{V_{\mathbf{u}}^\theta}(z|\,\text{data})\, dz,$$

from which it follows that we can get samples from the posterior predictive distribution of $V(\mathbf{s})$ via composition sampling, once we have samples from the posterior predictive of $V_{\mathbf{u}}^\theta(\mathbf{s})$ and knowing the posterior predictive of $V_{\mathbf{u}}^X(\mathbf{s})$. Hereafter we concentrate only on inference over $V_{\mathbf{u}}^\theta(\mathbf{s})$, which is equivalent to assume a spatially constant mean structure in (3.3.2), so that we can simply consider $V_{\mathbf{u}}(\mathbf{s}) = V_{\mathbf{u}}^\theta(\mathbf{s})$ for notational simplicity.

There are several reasons for doing that. First, the posterior predictive of $V_{\mathbf{u}}^{X}(\mathbf{s})$ is usually known in closed form and therefore it can be easily sampled from. Moreover, it doesn't involve the SDP specification. Secondly, we can be interested in isolating the contribution of the pure spatial effect to the pattern observed in the data. In fact, the combined effect of the mean structure and spatial component can be diverse, and one component could even overshadow the other, making the results of the inference over $V_{\mathbf{u}}(\mathbf{s})$ difficult to interpret.

3.4 Computational issues.

According to the discussion in Gelfand, Kottas and MacEachern (2004), 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 we are interested in studying the pattern of gradients for one of the available replicates. Moreover, the new surface is chosen independently from the others given a predicted spatial effect and all the other parameters of the model. Therefore, we are able to catch significant slopes only in case the nature of spatial variability is not purely random, e.g. there exists some unobservable mean structure or trend surface, so that actually the spatial effect is not centered around zero. Even so, the predicted gradients can be smoothed, as it happens when the distribution of the observations is bimodal, because of two different spatial patterns acting across the replicates. For example, this is the case considered in the simulated data example outlined in the next section. On the other hand, prediction for one of the observed replicates 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.

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

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

$$
\begin{aligned}
[\mathbf{V}_{\mathbf{u},t}, \tilde{\mathbf{V}}_{\mathbf{u},t}, t = 1, \ldots, T | \,\text{data}] &= \int [\mathbf{V}_{\mathbf{u},t}, \tilde{\mathbf{V}}_{\mathbf{u},t}, t = 1, \ldots, T | \,\boldsymbol{\theta}_t, \tilde{\boldsymbol{\theta}}_t, t = 1, \ldots, T, \boldsymbol{\psi}] \times \\
&\quad \times [\boldsymbol{\theta}_t, \tilde{\boldsymbol{\theta}}_t, t = 1, \ldots, T, \boldsymbol{\psi} | \,\text{data}] \\
&= [\mathbf{V}^*_{\mathbf{u},t}, \tilde{\mathbf{V}}^*_{\mathbf{u},t}, \mathbf{w}, t = 1, \ldots, T^* | \,\text{data}] = \\
&= \int \prod_{t=1}^{T^*} [\mathbf{V}^*_{\mathbf{u},t}, \tilde{\mathbf{V}}^*_{\mathbf{u},t} | \,\boldsymbol{\theta}^*_t, \tilde{\boldsymbol{\theta}}^*_t, \sigma^2, \phi, \nu] \times \\
&\quad \times \prod_{t=1}^{T^*} [\tilde{\boldsymbol{\theta}}^*_t | \,\boldsymbol{\theta}^*_t, \sigma^2, \phi, \nu] \, [\boldsymbol{\theta}^*, \mathbf{w}, T^*, \boldsymbol{\psi} | \,\text{data}],
\end{aligned}
\tag{3.4.1}
$$

where $\boldsymbol{\psi} = (\beta, \tau^2, \sigma^2, \phi, \nu)$ denotes the vector of parameters of the model. We can sample from (3.4.1) via composition sampling. In fact, $[\boldsymbol{\theta}^*, \mathbf{w}, T^*, \boldsymbol{\psi} | \,\text{data}]$ is the output of the Gibbs sampling procedure described in Gelfand, Kottas and MacEachern (2004) for model (3.3.2). If the base measure G_0 is gaussian, then also $[\mathbf{V}^*_{\mathbf{u},t}, \tilde{\mathbf{V}}^*_{\mathbf{u},t} | \,\boldsymbol{\theta}^*_t, \tilde{\boldsymbol{\theta}}^*_t, \sigma^2, \phi, \nu]$

and $[\tilde{\boldsymbol{\theta}}_t^* | \boldsymbol{\theta}_t^*, \sigma^2, \phi, \nu]$ are gaussian. In fact, for any fixed $t = 1, \ldots, T^*$, the joint distribution $[\boldsymbol{\theta}_t^*, \tilde{\boldsymbol{\theta}}_t^*, \mathbf{V}_{\mathbf{u},t}^*, \tilde{\mathbf{V}}_{\mathbf{u},t}^*]$ is multivariate gaussian and has been thoroughly described by Banerjee, Gelfand and Sirmans (2003). Hence, if we are interested in the predictive distribution only for the new set of locations, we can consider just

$$\begin{aligned}[\tilde{\mathbf{V}}_{\mathbf{u},t}, t = 1, \ldots, T | \text{ data}] &= [\tilde{\mathbf{V}}_{\mathbf{u},t}^*, t = 1, \ldots, T^*, \mathbf{w} | \text{ data}] \\ &= \int \prod_{t=1}^{T^*} [\tilde{\mathbf{V}}_{\mathbf{u},t}^* | \boldsymbol{\theta}_t^*, \sigma^2, \phi, \nu] \, [\boldsymbol{\theta}^*, \mathbf{w}, T^*, \boldsymbol{\psi} | \text{ data}],\end{aligned} \tag{3.4.2}$$

where $[\tilde{\mathbf{V}}_{\mathbf{u},t}^* | \boldsymbol{\theta}_t^*, \sigma^2, \phi, \nu]$ is again conditionally gaussian.

As a remark, we can mention a different method to obtain samples from $[\mathbf{V}_{\mathbf{u},t}, \tilde{\mathbf{V}}_{\mathbf{u},t}, t = 1, \ldots, T | \text{data}]$ when the base measure of the SDP is gaussian. In fact, model (3.3.2) 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 $\mathbf{Y}_t | \boldsymbol{\beta}, \boldsymbol{\theta}_t, \tau^2$ with $\mathbf{Y}_t | \boldsymbol{V}_{\mathbf{u},t}, \boldsymbol{\psi}$, where

$$[\mathbf{Y}_t | \boldsymbol{V}_{\mathbf{u},t}, \boldsymbol{\psi}] = \int [\mathbf{Y}_t | \boldsymbol{\theta}_t, \beta, \tau^2] \, [\boldsymbol{\theta}_t | \mathbf{V}_{\mathbf{u},t}, \sigma, \phi, \nu] \quad t = 1, \ldots, T,$$

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

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

where K_1 is the matrix of regression coefficients of $\boldsymbol{\theta}_t$ on $\mathbf{V}_{\mathbf{u},t}$, 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 $V_{\mathbf{u},t}(\mathbf{s})$ is a sample from a SDP (see section 3.2) and apply the algorithm outlined by Gelfand, Kottas and MacEachern (2004) to obtain samples from the posterior distribution $[\mathbf{V}_{\mathbf{u},t}, t = 1, \ldots, T, \boldsymbol{\psi} | \text{ data}]$ for old locations. Then, for new locations it is sufficient to consider the distinct $\mathbf{V}_{\mathbf{u},t}$'s and the distribution $[\tilde{\mathbf{V}}_{\mathbf{u},t}^* | \mathbf{V}_{\mathbf{u},t}^*, \sigma^2, \phi, \nu]$, for $t = 1, \ldots, T^*$.

We conclude this section addressing the problem of prediction when we are interested in predicting roughness on a totally new predicted replicate, say $Y_0(\mathbf{s})$. Let us denote with $(\mathbf{V}_{\mathbf{u},0}, \tilde{\mathbf{V}}_{\mathbf{u},0})$ the gradient processes for this new replicate. Then, samples from the predictive density of $(\mathbf{V}_{\mathbf{u},0}, \tilde{\mathbf{V}}_{\mathbf{u},0})$ can be obtained via composition sampling from

$$\begin{aligned}[\mathbf{V}_{\mathbf{u},0}, \tilde{\mathbf{V}}_{\mathbf{u},0}|\,\text{data}] = \int [\mathbf{V}_{\mathbf{u},0}, \tilde{\mathbf{V}}_{\mathbf{u},0}|\,\mathbf{V}_{\mathbf{u},t}, \tilde{\mathbf{V}}_{\mathbf{u},t}, t=1,\ldots,T, \sigma^2, \phi, \nu] \times \\ \times [\mathbf{V}_{\mathbf{u},t}, \tilde{\mathbf{V}}_{\mathbf{u},t}, t=1,\ldots,T, \psi|\,\text{data}].\end{aligned} \tag{3.4.3}$$

Notice that the second factor in the integrand is the posterior predictive distribution (3.4.1). The first factor is easily sampled from recalling that $V_{\mathbf{u},t}(\mathbf{s})$ is a sample from a SDP and exploiting the Blackwell-MacQueen Pólya urn scheme. Therefore, it is given by

$$\begin{aligned}[\mathbf{V}_{\mathbf{u},0}, \tilde{\mathbf{V}}_{\mathbf{u},0}|\,\mathbf{V}_{\mathbf{u},t}, \tilde{\mathbf{V}}_{\mathbf{u},t}, t=1,\ldots,T, \sigma^2, \phi, \nu] = \frac{\nu}{\nu+T} G_0^V(\mathbf{V}_{\mathbf{u},0}, \tilde{\mathbf{V}}_{\mathbf{u},0}|\,\sigma^2, \phi, \nu) \\ + \frac{1}{\nu+T} \sum_j^{T^*} T_{\cdot j}^* \delta_{(\mathbf{V}^*_{\mathbf{u},j}, \tilde{\mathbf{V}}^*_{\mathbf{u},j})}(\mathbf{V}_{\mathbf{u},0}, \tilde{\mathbf{V}}_{\mathbf{u},0}),\end{aligned}$$

where $G_0^V(\cdot|\,\sigma^2, \phi, \nu)$ is either $G_0^{\mathbf{u},h}(\cdot|\,\sigma^2, \phi, \nu)$ or $G'_{0,\mathbf{u}}(\cdot|\,\sigma^2, \phi, \nu)$ according to $V_{\mathbf{u}}(\mathbf{s}) = \theta_{\mathbf{u},h}(\mathbf{s})$ or $V_{\mathbf{u}}(\mathbf{s}) = D_{\mathbf{u}}\theta(\mathbf{s})$.

3.5 A simulation example.

Now we consider a simulation example. We generate a dataset with the intended purpose to stress some peculiarity of our model. In particular, we consider a situation where usual gaussian modeling is certainly unappropriate. In fact, modeling through an SDP increases the flexibility of the model and the ability to fit to a wider range of situations (see Gelfand, Kottas and MacEachern (2004)). Moreover, we take advantage of the clustering feature of the SDP and show the possibility to identify the existence of several types of spatial effects through distinct replicates.

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

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

for some $\alpha \in (0,1)$, that is a mixture of two independent gaussian fields $Z_1(\mathbf{s})$ and $Z_2(\mathbf{s})$, with mean $E(Z_i(\mathbf{s})) = \eta_i(\mathbf{s})$ and covariance structure specified by some correlation function, $\sigma^2 \rho_{\phi_i}(\mathbf{s}, \mathbf{s}')$, $i = 1, 2$. Therefore, data are generated so that at each replicate they could come from one of the two spatial processes, $Z_1(\mathbf{s})$ or $Z_2(\mathbf{s})$. For example, usually temperature decreases with altitude. However, during the passage of a cold front or for overnight radiative cooling, temperature inversions occur and the temperature of the atmosphere increases with altitude. Temperature inversions are relevant, for example, in studying pollution phenenomena, sound propagation or thunderstorms. Therefore, it could be appropriate to model the possibility of distinct spatial effects acting at times when we collect data. Another example is when a signal comes from distinct sources at different times, and we don't know both the presence of the sources and their locations. Then, it could be interesting to evaluate the intensity of the signal at each monitoring site and its direction, in order to understand which source is effective at a given time and its location. Notice that these are all situations where a simple gaussian assumption valid for all replicates is not appropriate. Our dataset can be seen as a simplified version of one of the settings described above.

Let

$$\eta_i(\mathbf{s}) = \beta_{0i} + \beta_{1i}\, e^{-\psi_i \|\mathbf{s}-\mathbf{s}_{0i}\|^2}, \quad i = 1, 2 \tag{3.5.1}$$

where $\mathbf{s}_{01}$ and $\mathbf{s}_{02}$ 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,

$$E(Z(\mathbf{s})) = \alpha\beta_{01} + (1-\alpha)\beta_{02} + \alpha\beta_{11}\, e^{-\psi_1(\|\mathbf{s}-\mathbf{s}_{01}\|)} + (1-\alpha)\beta_{12}\, e^{-\psi_2(\|\mathbf{s}-\mathbf{s}_{02}\|)}, \quad (3.5.2)$$

and

$$\begin{aligned} Cov(Z(\mathbf{s}), Z(\mathbf{s}')) = \alpha(1-\alpha)\,[\eta_1(\mathbf{s}') - \eta_2(\mathbf{s}')]\,[\eta_1(\mathbf{s}) - \eta_2(\mathbf{s})] + \\ + \sigma^2\{\alpha\,\rho_{\phi_1}(\mathbf{s},\mathbf{s}') + (1-\alpha)\,\rho_{\phi_2}(\mathbf{s},\mathbf{s}')\}. \end{aligned} \quad (3.5.3)$$

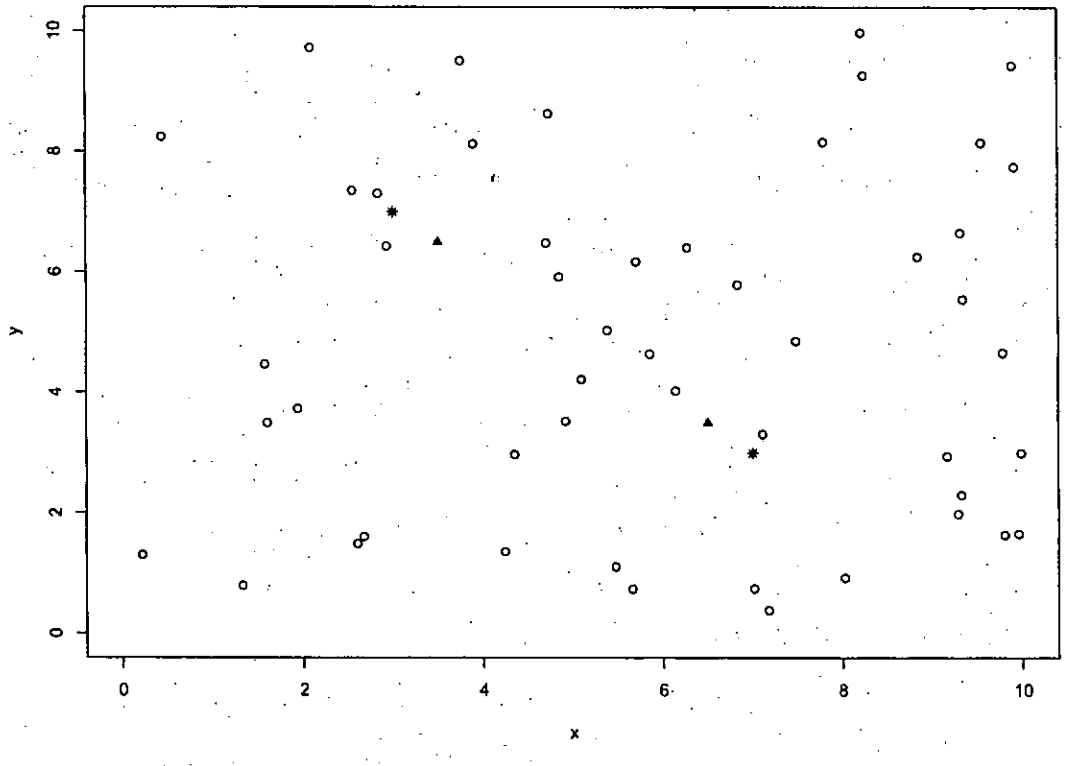

Figure 3.1: Location of the 50 sites where the random field has been observed, together with the indication of the sources (⁕) and prediction sites (▲).

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 showed in figure 3.1, togheter 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 figure 3.1), 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. Then, we consider $T = 100$ independent observations of the random field $Y(\mathbf{s})$ 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 $Z(\mathbf{s})$ 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. $\beta_{01} \exp\{-\psi||\mathbf{s} - \mathbf{s}_{0i}||\} \leq 0.05$) at points further than 2.15 units from the sources. Here $\rho_\phi(\cdot)$ is the Matern 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 figure (3.2), we plot the density of the observations for two sites, one close to a source, one more distant. It's 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 if we sample from Z_1 or Z_2. At points far from the sources, the signal is negligible over its long range mean. Therefore, the density is unimodal and centered around the mean.

Based on observations collected at s_1^* and s_2^* and points around, we can plot the finite differences $m_{\mathbf{u},h}(\mathbf{s}_i^*) = \theta_{\mathbf{u},h}(\mathbf{s}_i^*)$, $i = 1, 2$ for $h = 0.01$. In figure 3.3, we show the distribution of the observed gradients at angles 0 and 135. It is immediate to see that it is bimodal, again reflecting the data generating mechanism. Careful analysis of the

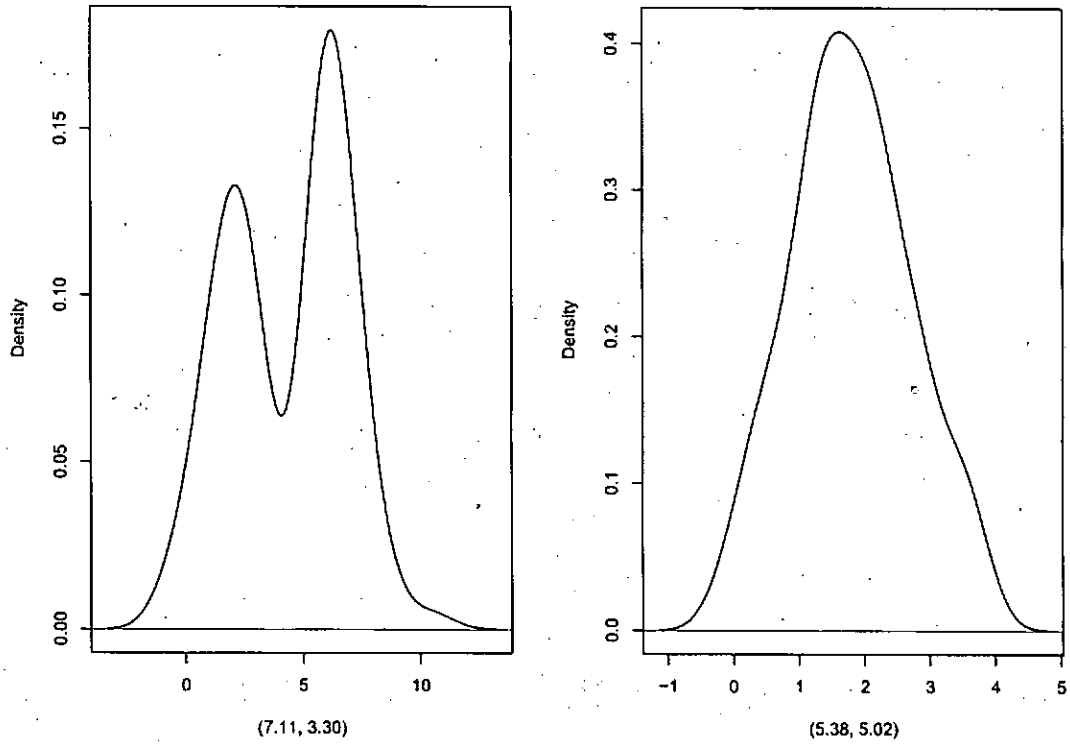

Figure 3.2: Plot of the estimated density in two sites, one close to a source (7.11,3.30), one distant (5.38, 5.02).

modes of the distribution also seems to suggest that most of the data come from Z_2 rather than Z_1.

We suppose that the existence of the sources, hence of the spatial component, is unknown. Therefore, we fit the observations by means of the hierarchical model (3.3.2), that is we assume $Y(\mathbf{s}) = \mu + \theta(\mathbf{s}) + \varepsilon(\mathbf{s})$, where μ is a constant and the distribution of $\theta(\mathbf{s})$ 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 effective

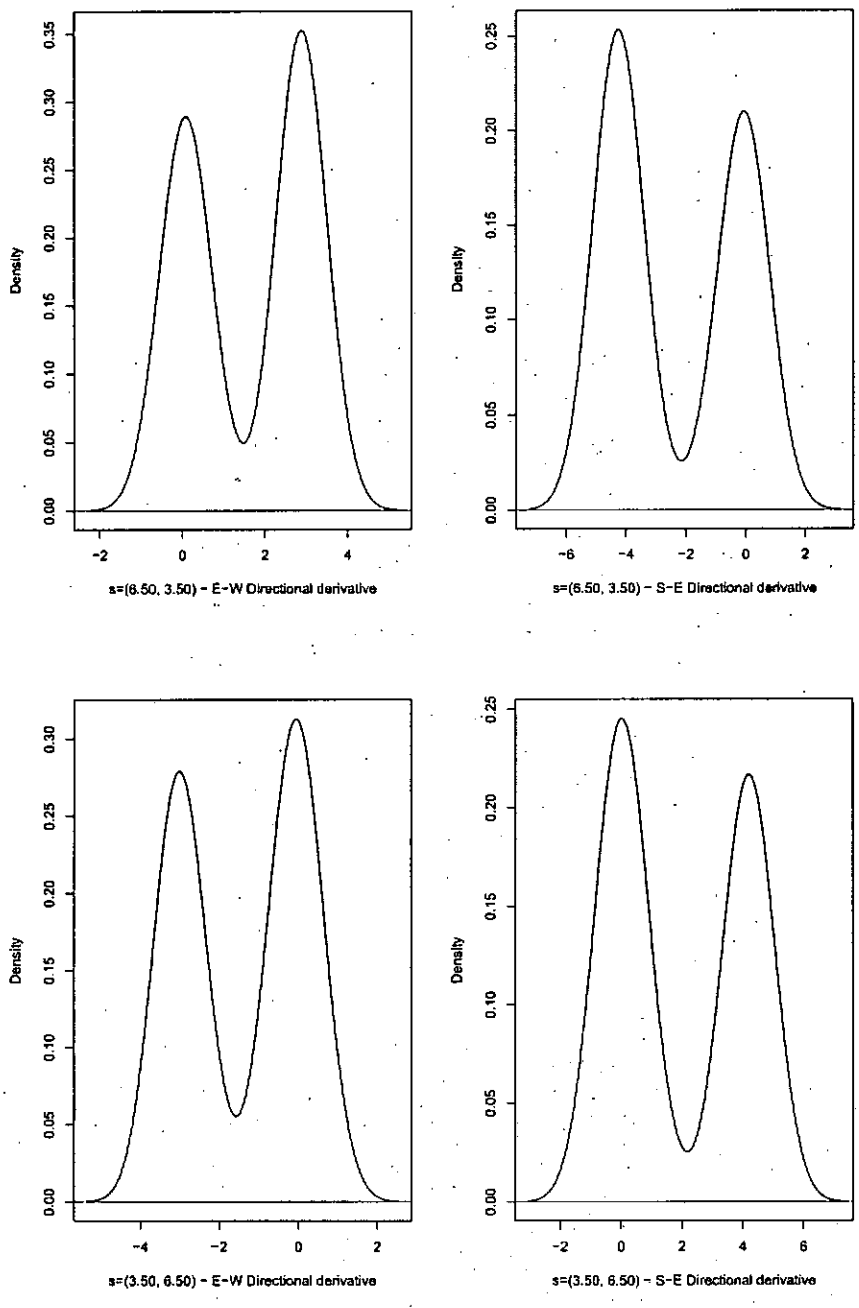

Figure 3.3: Finite differences ($h = 0.01$) at angles 0 and 135, observed at the two locations $s_1^* = (3.5, 6.5)$ (above) and $s_2^* = (6.5, 3.5)$ (below).

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 figure 3.3 or the observed gradients in order to validate our inference.

In figure 3.4 we consider two exploratory image plots of data generated, respectively, at the 99th and 100th replicate. It's immediate to guess that the former is a sample from Z_1, while the latter from Z_2. This conjecture is confirmed by looking at figures 3.5 and 3.6, 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 figure 3.6). We can see that at site s_2^* there are a significant uphill gradient moving along the E-W direction and a significant downhill gradiend toward S-W. On the other hand, the posterior predictive distributions are centered around zero in s_1^*, thus suggeting 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 (figure (3.5)).

In conclusion, by means of the preceding analysis we are able to detect the presence of distinct gradient behaviors across the replicates. Notice that the usual assumption of a single gaussian process for $\theta(\mathbf{s})$ in (3.3.1) would lead to a smoothing ot the estimated gradient processes through all the replicates. Such a smoothing is also present in our estimates, but in a less degree and only for certain directions, the effect probably depending on the number of sites considered and the contribution of the pure spatial effect

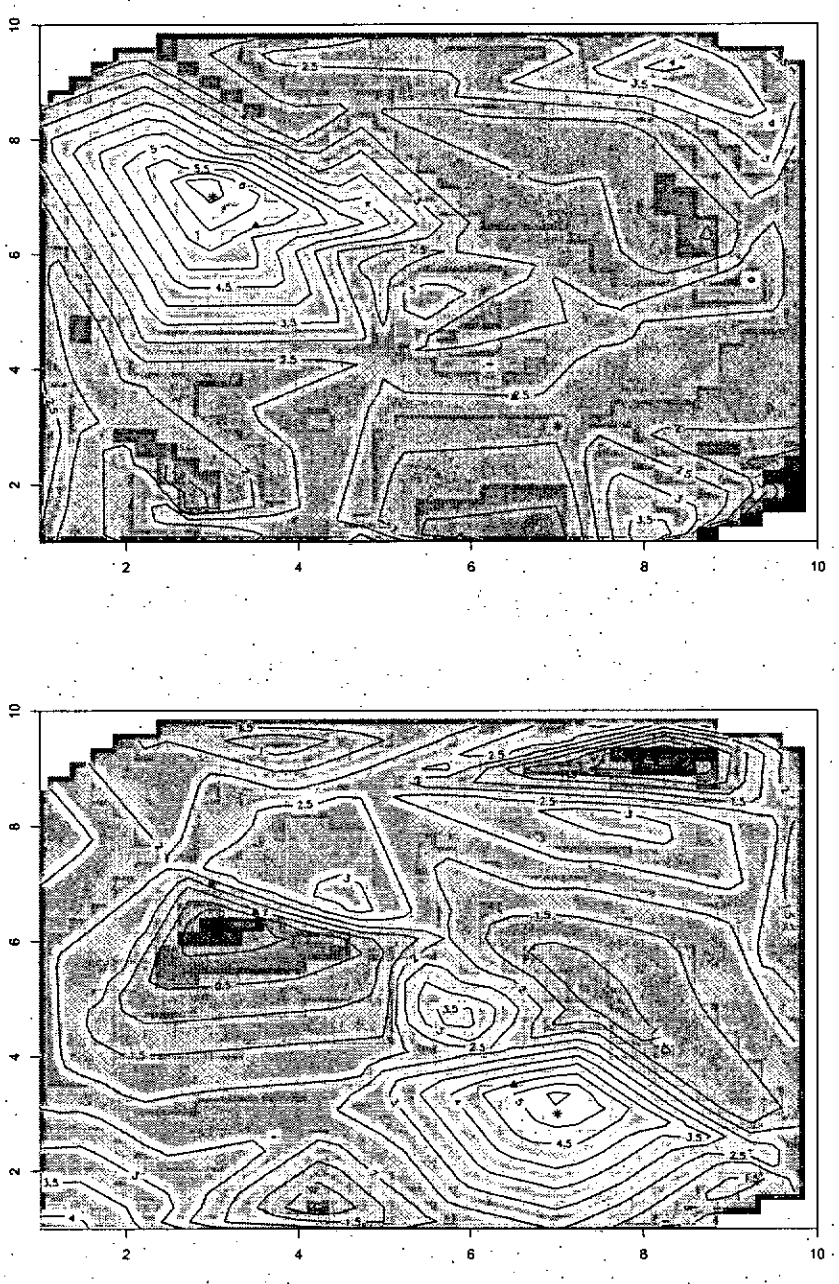

Figure 3.4: Image plots with countour lines of data generated at the 99th (above) and 100th (below) replicates.

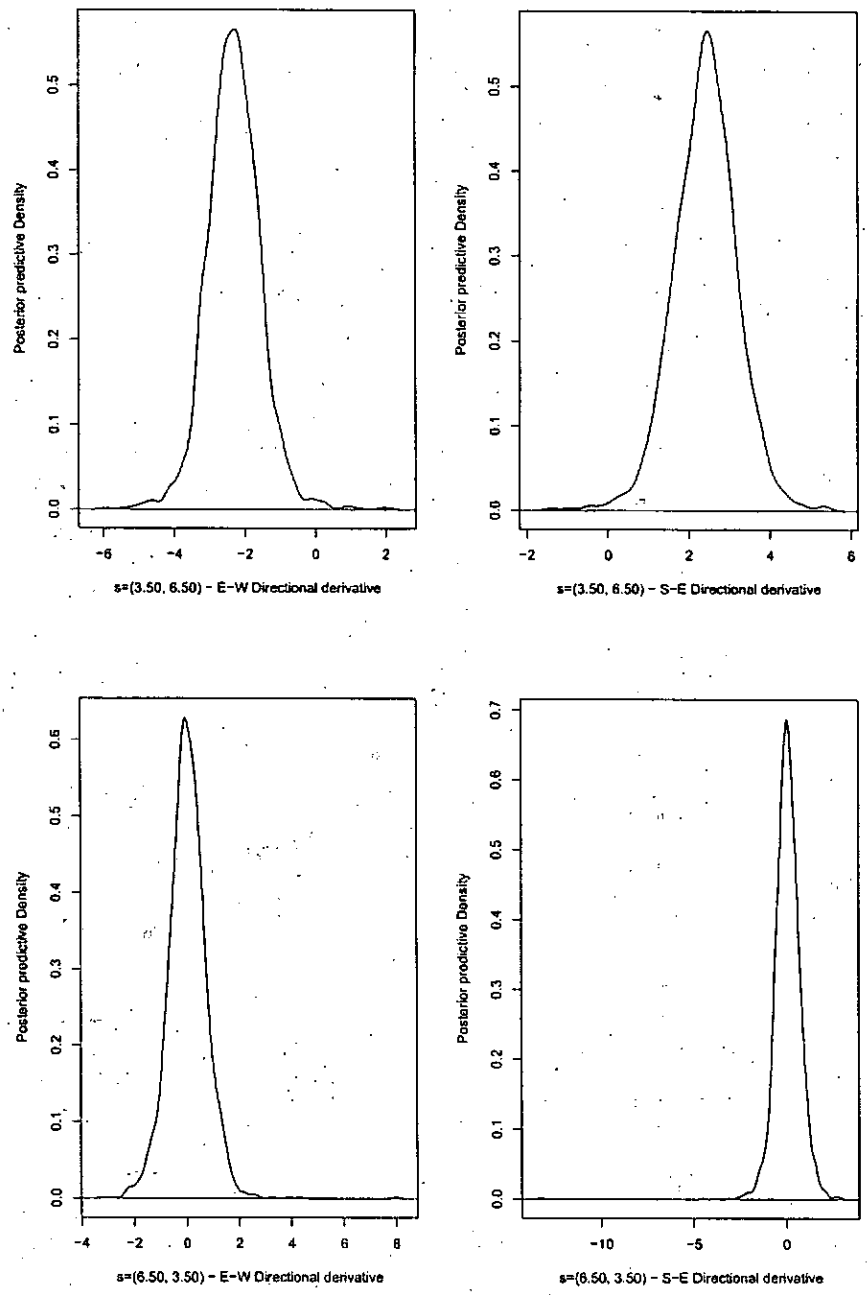

Figure 3.5: Predictive posterior distribution of the directional derivatives at angles 0 and 135 and at sites $s_1^* = (3.5, 6.5)$ and $s_2^* = (6.5, 3.5)$ for the 99th replicate.

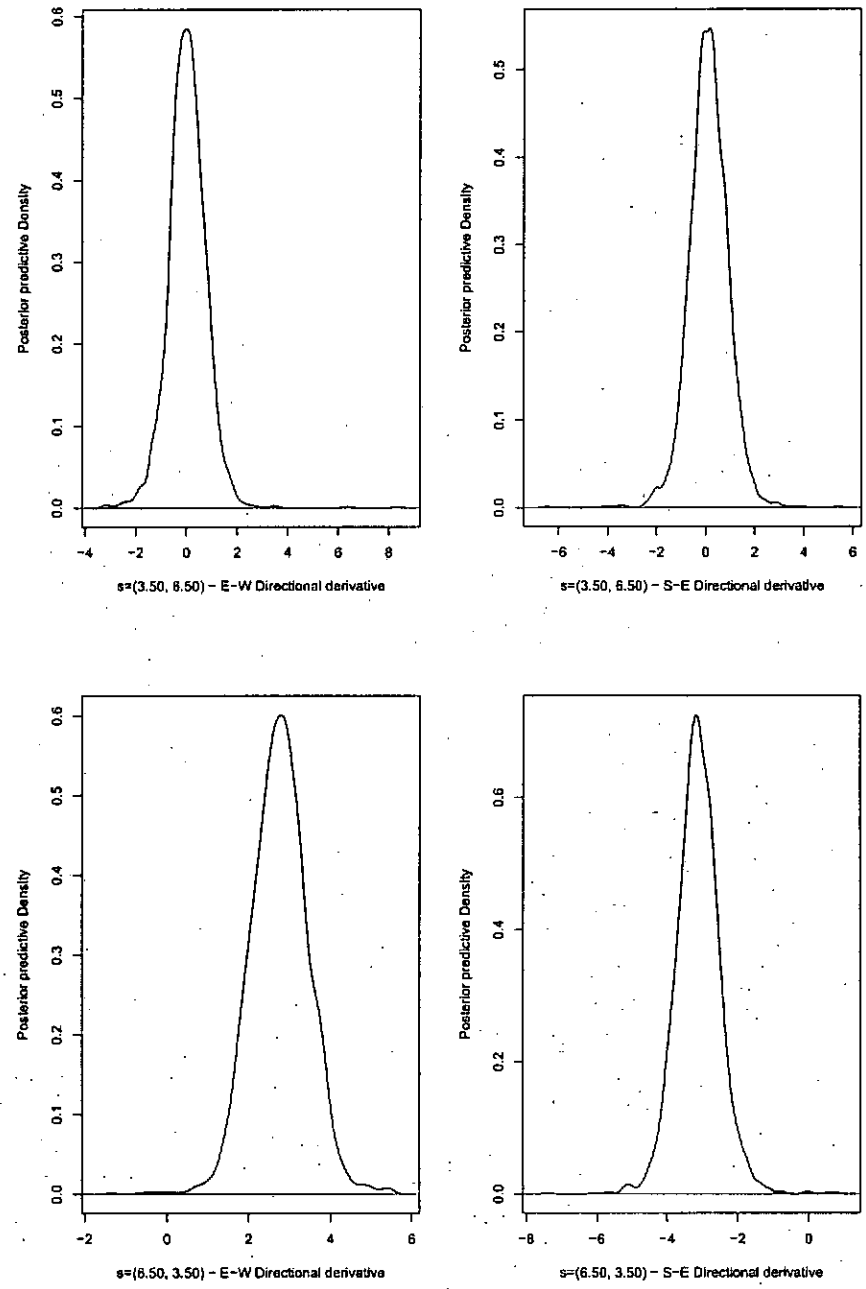

Figure 3.6: Predictive posterior distribution of the directional derivatives at angles 0 and 135 and at sites $s_1^* = (3.5, 6.5)$ and $s_2^* = (6.5, 3.5)$ for the 100th replicate.

ε in partially hiding the gradient behavior of the mean process $m(\mathbf{s})$. Tables 3.1 and 3.2 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 $0, 45, 90, 135, 180, 225, 270$ degrees. Those are compared with the gradients computed on the base of data generated at sites s_1^* and s_2^*, as specified above.

Angle	$m_{\mathbf{u},0.01}(\mathbf{s}_2^*)$ obs. at 100 th	Quant. of $m_{\mathbf{u},0.01}(\mathbf{s}_2^*)$ 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 (-4.08,-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)

Table 3.1: 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 figure 3.3-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	$m_{\mathbf{u},0.01}(\mathbf{s}_1^*)$ obs. at 99 th	Quant. of $m_{\mathbf{u},0.01}(\mathbf{s}_1^*)$ obs. over all replic.	Posterior predictive (2.5%, 97.5%) intervals			
			$D_{\mathbf{u}}m(\mathbf{s}_1^*)$	$m_{\mathbf{u},0.01}(\mathbf{s}_1^*)$	$m_{\mathbf{u},0.1}(\mathbf{s}_1^*)$	$m_{\mathbf{u},1}(\mathbf{s}_1^*)$
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 (0.32,1.83)
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 3.2: 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 (2.5%, 97.5%) predictive intervals for the density in figure 3.3-above; all the other columns provide the posterior medians and (2.5%, 97.5%) predictive intervals for directional derivatives and finite differences ($h = 0.01, 0.1, 1$).

Chapter 4

Generalized Spatial Dirichlet Process Models.

In the preceding chapters, we have used the SDP to model the distribution of the spatial component in a spatial random effects model. Therefore, we pursue a fully Bayesian semiparametric approach that relies for fitting purposes on well-known results and algorithms developed for the Dirichlet process (DP) (see Escobar and West (1995) and MacEachern and Müller (1998)). These methods require that data come as a set of replications at the observed sites. This is not unexpected since replications are typically needed for a full nonparametric approach (see, e.g. Sampson and Guttorp (1992) and Damian, Sampson and Guttorp (2001)). These methods allow the possibility to infer about the (random) distribution function that is operating at any given location in the region. We have seen that nonparametric spatial prediction under such modeling can be pursued not only at a new set of locations for each replicate, but more generally through the generation of an entire new predictive surface.

Since the SDP is essentially a Dirichlet process defined on a space of surfaces, its realizations are a.s. discrete probability measures with infinite support (see Ferguson (1973) and Sethuraman (1994)). Therefore, whenever the (conditional) distribution of the observables is assumed to be Gaussian, such as it is in (2.2.2), the spatial random

effects model (2.1.4) can be characterized as a countable mixture of normals, and so in principle it is able to capture virtually any distribution of the observables.

However, the way this is achieved can be unsatisfactory for inferential purposes. In fact, the SDP specification implies that we can actually sample one and only one common surface for all locations. In this thesis, we introduce a random distribution for the spatial effects that allows different surface selection at different sites. Moreover, we show that it is possible to specify the model so to preserve the property that the marginal distribution of the effect at each site still comes from a Dirichlet process. This is done constructively, extending to a multivariate setting the stick-breaking representation of the weights which is known to characterize the usual Dirichlet process (see section 1.2.1). This way, we define a new class of random probability measures for random vectors and processes, which includes the customary Dirichlet process specification as a special case. We refer to this class as generalized spatial Dirichlet process models (GSDP). Additionally, this class can be seen as an extension of the generic class of priors described in Hjort (2000) and Ishwaran and James (2001), which also take their aim explicitly from the stick-breaking representation.

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

In the context of Bayesian analysis of spatial data, we are aware of only two other recent approaches that also consider mixture models for spatial data where the weights

are allowed to vary across locations. Fernandez and P. (2002) confine their attention to Markov random fields over lattices and Poisson distributed data. However, they consider problems where it is only the weights in the mixture that vary from one location to another. Our model differs from that since it applies to general point referenced data and both the weights and the parameters of the mixed distribution are allowed to vary spatially. Griffin and Steel (2004) present an implementation of the dependent Dirichlet process in the context of spatial processes using Sethuraman's constructive representation, providing a random marginal distribution at each site. The components of the marginal stick breaking are the same at each location, but they are randomly permuted according to the realizations of a latent point process, so that at each site the resulting weights are assigned to different surfaces, inducing spatial dependence. Our approach is different (see section 1.5.5). We define a joint stick-breaking construction for any number and choice of locations, and also allow the marginal components to vary in space. Moreover, in our approach the closeness between the random distribution is ruled directly by the topology of the space, rather than the realizations of an underlying point process.

The format of the chapter is as follows. Section 4.1 formalizes the generalized spatial Dirichlet process (GSDP) and develops its properties. It also treats mixing of Gaussian kernels using this process. Section 4.2 elaborates the spatially varying probabilities model that is a component of the GSDP. Section 4.3 presents the computational strategy for fitting such models while section 4.4 offers an example. This is joint work with A.E. Gelfand and J. Duan. In particular, J. Duan has provided all the data analysis described in section 4.4.

4.1 The Generalized Spatial Dirichlet Process Model

In subsection 4.1.1 we formally develop the GSDP model. In subsection 4.1.1 we employ this model as a mixing distribution, mixing against a Gaussian kernel.

4.1.1 Model details

In the spatial Dirichlet Process developed by Gelfand *et al.* (2004), the random distribution of the pure spatial effect is essentially a Dirichlet Process defined on the space of the random surfaces over D generated by a mean 0 base spatial process. Then the a.s. characterization of the process implies that the random G for s is not the same as that for s' since $\{\theta_l^*(s)\}$'s is not the same as $\{\theta_l^*(s')\}$. However, each has the same stick-breaking probabilities and they come from the same Dirichlet process. Additionally, for any group of n locations, the joint distribution uses the same set of stick-breaking probabilities inducing common surface selection for all locations in the group. Indeed, the uncountable collection of spatial random effects are all selected from the same random surface. The spatial dependence is introduced only through the underlying base measure, and it is not possible to capture the situation in which spatial effects can be selected from different surfaces at different locations. This limitation of the SDP is in common with other recent works relating to the so-called dependent Dirichlet process (see MacEachern (2000), for example the ANOVA construction of De Iorio *et al.* (2004)).

Building upon MacEachern (2000)'s idea, but proceeding along a slightly different direction, we introduce a random distribution for the spatial effect that allows different finite dimensional distributions across locations in the sense that surface selection can vary with location and that the joint surface selection for n locations can vary with the

choice of locations. Moreover, we still preserve the property that the marginal distribution at each location turns out to come from a usual univariate Dirichlet Process. This is achieved constructively, defining a new multivariate stick-breaking prior in which a spatial dependence structure is also introduced in the modeling of the weights.

Accordingly, we start by considering a base random field G_0, which, for convenience, we take to be stationary and gaussian, and indicate with $\theta_i^* = \{\theta_i^*(s), s \in D\}$ a realization from G_0, i.e. a surface over D. Let $G_0^n(\cdot) \equiv G_{0,s_1,\ldots,s_n}(\cdot)$ indicate its finite dimensional distributions at locations $(s_1, \ldots, s_n)$. We say that a random probability measure G on the space of surfaces over D is a *Generalized Spatial Dirichlet Process* if for any set of locations $(s_1, \ldots, s_n)$, the induced finite dimensional distributions have a.s. the following representation,

$$G^{(n)}(A_1 \times \ldots \times A_n) = \sum_{i_1=1}^{\infty} \ldots \sum_{i_n=1}^{\infty} p_{i_1,\ldots,i_n}(s_1, \ldots, s_n) \quad \delta_{\theta^*_{i_1}(s_1)}(A_1) \ldots \delta_{\theta^*_{i_n}(s_n)}(A_n), \tag{4.1.1}$$

where $\theta_j^* \overset{i.i.d}{\sim} G_0$, i_j is an abbreviation for $i(s_j)$, $j = 1, 2, \ldots, n$, and the weights $p_{i_1,\ldots,i_n}(s_1, \ldots, s_n)$, $i_j = 1, \ldots,$, $j = 1, \ldots, n$, have a distribution defined on the infinite dimensional simplex $\mathbb{P} = \{p_{i_1,\ldots,i_n}(s_1, \ldots, s_n) \geq 0 : \sum_{i_1=1}^{\infty} \ldots \sum_{i_n=1}^{\infty} p_{i_i,\ldots,i_n}(s_1, \ldots, s_n) = 1\}$, are independent of the $\theta's$ and arise from a spatial process described below. For notational simplicity, we will often write simply $p_{i_1,\ldots,i_n}$ instead of $p_{i_i,\ldots,i_n}(s_1, \ldots, s_n)$, unless we need to specify the set of locations to avoid confusion. In symbols, then we write $G^{(n)} \sim GSDP(p_{i_1,\ldots,i_n}, G_0^{(n)})$.

Analogously to Ferguson (1973), we say that a vector $\mathbf{Y} = (Y(s_1), \ldots, Y(s_n))^T$ is a sample of size 1 from $G^{(n)}$ if $\mathbf{Y}|G^{(n)} \sim G^{(n)}$ and $G^{(n)} \sim GSDP(p_{i_1,\ldots,i_n}, G_0^{(n)})$.

The generalization of the usual Dirichlet process setting is apparent and it is evident that we allow the possibility to choose different surfaces at different locations. We will return to this point later in the section. For now, it will be enough to notice that the

weights need to satisfy a consistency condition so that the collection of distributions $G^{(n)}$, $n = 1, 2, \ldots$ define a random process $Y(s)$ over D. Specifically, we need that for any set of locations $\{s_1, \ldots, s_n\}$, $n \in \mathbb{N}$ and $\forall k \in \{1, \ldots, n\}$,

$$p_{i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_n} = p_{i_1,\ldots,i_{k-1},\cdot,i_{k+1},\ldots,i_n} \equiv \sum_{j=1}^{\infty} p_{i_1,\ldots,i_{k-1},j,i_{k+1},\ldots,i_n}. \tag{4.1.2}$$

In addition, we insist that the weights satisfy a continuity property. In fact, we want the random laws associated with locations s_1 and s_2 near to each other to be similar. Equivalently, for locations s and s_0, as $s \to s_0$, $p_{i_1,i_2} = P(Y(s) = \theta^*_{i_1}(s), Y(s_0) = \theta^*_{i_2}(s_0))$, tends to the marginal probability $p_{i_2} = P(Y(s_0) = \theta^*_{i_2}(s_0))$ when $i_1 = i_2$, and to 0 otherwise. Analogously, if we consider three locations $\{s_1, s_2, s_3\}$, if s_3 is close to say, s_2, we require p_{i_1,i_2,i_3} to be close to p_{i_1,i_2} if $i_2 = i_3$ and to 0 otherwise. Extension to n locations is clear; we avoid introducing further notation, and from now on refer to this property simply as *a.s. continuity* of the weights. The name is suggested by the a.s. continuity of the paths of a univariate spatial process, as defined in Kent (1989) or Banerjee *et al.* (2003). Recall that a univariate spatial process $\theta(s)$, $s \in D$ is said to be almost surely continuous at a point s_0 if $\theta(s) \to \theta(s_0)$ a.s. as $||s - s_0|| \to 0$. We take up an illustrative construction of a.s. continuous weights in Section 4.2 with associated formal arguments supplied in Appendix I. Now, if we also assume the random field G_0 to be a.s. continuous, we are able to establish the following proposition whose proof is also given in Appendix I.

Proposition 4.1.1. *Let $\{Y(s), s \in D\}$ be a random field, whose random finite dimensional distributions are given by (4.1.1) for all $n \in \mathbb{N}$. If the set of weights $\{p_{i_1,\ldots,i_n}\}$ and the base random field G_0 are almost surely continuous, then for all $s_0 \in D$, $Y(s)$ converges weakly to $Y(s_0)$ a.s. as $||s - s_0|| \to 0$.*

In fact, the proof demonstrates a.s. convergence of the random probability measures. Note that Proposition 1 is an extension to our case of analogous results stated

in MacEachern (2000) and Gelfand *et al.* (2004) for the SDP.

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

$$E(Y(s)|G) = \sum_{l=1}^{\infty} p_l(s)\, \theta_l^*(\mathbf{s}) \tag{4.1.3}$$

$$Var(Y(s)|G) = \sum_{l=1}^{\infty} p_l(s)\, \theta_l^{*^2}(s) - \Big\{ \sum_{l=1}^{\infty} p_l\, \theta_l^*(\mathbf{s}) \Big\}^2, \tag{4.1.4}$$

and, for a pair of sites s_i, s_j,

$$Cov(Y(s_i), Y(s_j)|G) = \sum_{l,m=1}^{\infty} p_{l,m}(s_i, s_j)\theta_l^*(\mathbf{s}_i)\, \theta_m^*(\mathbf{s}_j) - \Big\{ \sum_{l=1}^{\infty} p_l(s_i)\, \theta_l^*(\mathbf{s}_i) \Big\} \Big\{ \sum_{m=1}^{\infty} p_m(s_j)\, \theta_m^*(\mathbf{s}_j) \Big\}. \tag{4.1.5}$$

These results are analogous to the ones in Gelfand, Kottas and MacEachern (2004) for the process without spatially varying weights, so that in both cases we can conclude that the process $Y(s)$ has heterogeneous variance and is nonstationary. However, when we marginalize over G, we can see more clearly the difference with our case. In fact, let G_0 be a mean zero stationary gaussian process with finite variance σ^2 and correlation function $\rho_\phi(s_i - s_j)$, where ϕ is a (possibly vector valued) parameter specifying $\rho_\phi(\cdot)$. Then, $E(Y(s)) = 0$ and $Var(Y(s)) = \sigma^2$ as before, but now

$$Cov(Y(s_i), Y(s_j)) = \sigma^2 \rho_\phi(s_i - s_j) \sum_{l=1}^{\infty} E(p_{ll}(s_i, s_j)). \tag{4.1.6}$$

Notice that $\sum_{l=1}^{\infty} E(p_{ll}(s_i, s_j)) < 1$, unless $p_{ll'}(s_i, s_j) = 0$, $l \neq l'$, as it is in Gelfand, Kottas and MacEachern (2004) or, more generally, in the single-p dependent Dirichlet process discussed by MacEachern (2000). We can call this limiting situation as the one of *maximum concordance* among the surfaces chosen at the two locations. In all other cases, the association structure is diminished by the amount of mass that the process (4.1.1) is expected to place on the not equally indexed θ^*'s. Moreover, from (4.1.6) it follows that, although the base measure G_0 is stationary, the process Y(s) is centered

around a stationary process only when $E(p_{ll}(s_i, s_j))$ is a function of $s_i - s_j$ for all s_i and s_j.

We now turn to the specification of $p_{i_1,\ldots,i_n}$ for any choice of n and $s_1, \ldots, s_n$. We propose a theoretically attractive and computationally feasible approach through a multivariate extension of the stick-breaking construction that usually characterizes the univariate Dirichlet process. For the sake of simplicity, we will show the main features of our approach in a bivariate setting, considering the random measure (4.1.1) for a pair of sites s_i, s_j, and we will provide details on how to extend it to the general multivariate case when necessary. First, we define a convenient process which retains the same Dirichlet process structure marginally at each site and then we move to a more general setting.

We start by recalling that in the Sethuraman's univariate stick-breaking construction the random measure $\sum_{l=1}^{\infty} p_l \delta_{\theta_l^*}$ has weights p_l defined by $p_1 = q_1$, $p_l = q_l \prod_{m=1}^{l-1}(1 - q_m)$, $l \geq 2$, where for all $l \geq 1$, q_l are i.i.d. $Beta(1, \nu)$ random variables independent of θ_l. Any realization of such a measure has evidently support on the set of realized θ_l^*'s, $l = 1, 2, \ldots$ Then, it is immediate to define the random events $\{Y = \theta_l^*\}$, denoted by Θ_l, as elements of the sigma algebra of the space on which the θ_l^*'s take values, together with their complements Θ_l^c, and interpret the sequence of weights $\{p_1, p_2, \ldots\}$ as arising from $q_1 = P(\Theta_1)$, $q_l = P(\Theta_l | \Theta_m^c, m < l) = P(Y = \theta_l^* | Y \neq \theta_m^*, m < l)$, $l = 1, 2, \ldots$. Analogously, turning back to our model, at each location s we can define events $\Theta_l^u(s)$, $u = 0, 1$, such that $\Theta_l^1(s) = \{Y(s) = \theta_l^*(s)\}$ and $\Theta_l^0(s) = \{Y(s) \neq \theta_l^*(s)\}$. Then, for any two locations s_i, s_j, we can consider the probabilities $q_{1,u,v}(s_i, s_j) = P(\Theta_1^u(s_i), \Theta_1^v(s_j))$, $q_{l,u,v}(s_i, s_j) = P(\Theta_l^u(s_i), \Theta_l^v(s_j) | \Theta_m^0(s_i), \Theta_m^0(s_j), m < l)$, $l \geq 2$, $u, v \in \{0, 1\}$. For all $l = 1, 2, \ldots$, we can enter these probabilities in the form of Table 4.1. Note that, formally, e.g., $q_{l,1,1}(s_i, s_j) + q_{l,1,0}(s_i, s_j) = q_{l,1,+}(s_i, s_j)$ and we need to argue that $q_{l,1,+}(s_i, s_j) = q_l(s_i)$. (Similarly for $q_{l,+,1}(s_i, s_j) = q_l(s_j)$.) The argument is supplied in

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

Table 4.1: Relevant probabilities in the multivariate stick-breaking construction in the special case of $n = 2$ locations, for $l = 1, 2, \ldots$.

Appendix I as Lemma 4.5.2. Then, accordingly, we can define the weights in (4.1.1) as

$$\begin{aligned} p_{lm} &= P(Y(s_i) = \theta_l^*(s_i), Y(s_j) = \theta_m^*(s_j)) \\ &= P(\Theta_l^1(s_i), \Theta_m^1(s_j), \Theta_k^0(s_i), k < l, \Theta_r^0(s_j), r < m) \\ &= \begin{cases} \prod_{k=1}^{l-1} q_{k,0,0}\, q_{l,1,0} \prod_{r=l+1}^{m-1}(1 - q_r)\, q_m & \text{if } l < m \\ \prod_{r=1}^{m-1} q_{r,0,0}\, q_{m,0,1} \prod_{k=m+1}^{l-1}(1 - q_k)\, q_l & \text{if } m < l \\ \prod_{r=1}^{l-1} q_{r,00}\, q_{l,11} & \text{if } l = m \end{cases} , \end{aligned} \tag{4.1.7}$$

where we have suppressed s_i and s_j. Although not immediate, close inspection of expression (4.1.7) reveals that the weights are determined through a partition of the unit square similar to the one induced on the unit segment by the usual stick-breaking construction, so that indeed the former can be considered as a bivariate extension of the latter. We can see this clearly from the exemplification in figure 4.1. At the first stage, if both the events $\Theta_1^1(s_i)$ and $\Theta_1^1(s_j)$ are true, we break off a region of the unit square of the same size as the realized value of $q_{1,11}(s_i, s_j)$. This is region A in figure 4.1. If only $\Theta_1^1(s_i)$ (or $\Theta_1^1(s_j)$) is true, we remain only with a piece corresponding to region B (D). In fact, given $\Theta_1^1(s_i)$ ($\Theta_1^1(s_j)$), the next stages we go on with a univariate stick-breaking procedure, so that we break off a part of region B (C) according to the values of $q_l(s_j)$ ($q_l(s_i)$), $l = 2, 3, \ldots$. If neither $\Theta_1^1(s_i)$ nor $\Theta_1^1(s_j)$ are true, then we discard all regions A, B, and D and remain only with region C, whose size is determined by $q_{1,00}(s_i, s_j)$.

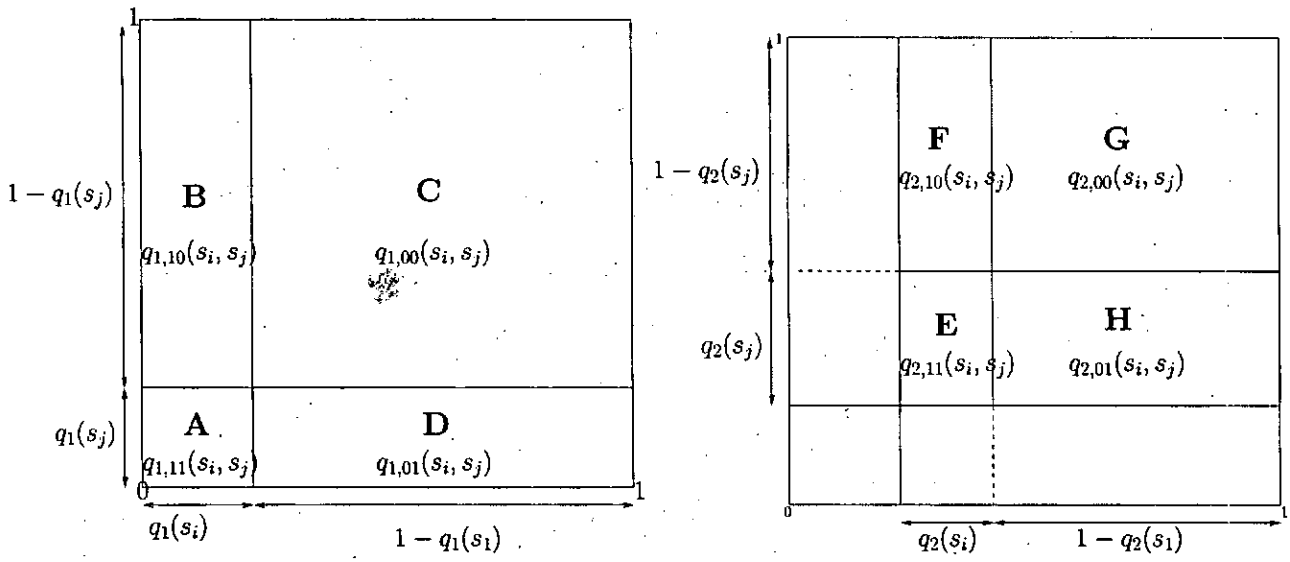

Figure 4.1: An exemplification of the multivariate stick-breaking procedure for the special case of $n = 2$ locations

Then, at stage two, we repeat the same arguments as above for region C, and so on (see Figure 4.1).

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

The construction relies on the specification of probabilities $q_{l,u_1,\ldots,u_n}, u_j \in \{0,1\}, j = 1,2,\ldots,n$, where u_j is an abbreviation for $u(s_j)$ at any set of locations $\{s_1,\ldots,s_n\}$. This is generally difficult, since it entails defining a spatial process which 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 (4.1.2) for all $l = 1,2,\ldots$ and any set of locations

$\{s_1, \ldots, s_n\}, n \in \mathbb{N}$ and for all $k \in \{1, \ldots, n\}$, that is

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

However, in the next section, we offer a flexible construction under which this can be done consistently. For now, as a special case, suppose the process retains the same marginal distribution at each location. Referring to Table 4.1, this can be easily achieved imposing $q_l(s) = q_l$, together with the symmetry condition $q_{l,1,0}(s_i, s_j) = P(Y(s_i) = \theta_l^*(s_i), Y(s_j) \neq \theta_l^*(s_j)) = P(Y(s_i) \neq \theta_l^*(s_i), Y(s_j) = \theta_l^*(s_j)) = q_{l,0,1}(s_i, s_j)$, for all $l = 1, 2 \ldots$ and $s \in D$. Then, given q_l, if we can compute say $q_{l,1,1}(s_i, s_j)$ as a function of q_l, the remainder of the table is determined. Then, according to Sethuraman's construction, if we allow q_l to be Beta$(1, \nu)$, we get a process which marginally is a Dirichlet process with precision parameter ν and base measure G_0. Together with (4.1.6), this illuminates the role of the distribution of the q's in specifying the dependence structure in a multivariate Dirichlet process.

Notice that there are other ways to achieve this particular result. For example, if we consider a process such that each $q_{l,0,1}$ given $q_{l,1,1}$ has a Beta-Stacy distribution with parameters $1, \nu$, $1 - q_{l,1,1}$ and $q_{l,11}$ is assumed to be Beta$(1, \nu + 1)$, then q_l is Beta$(1, \nu)$. The model we present in section 4.2 offers an alternative spatially-explicit way to specify q_l and $q_{l,1,1}$, at the same time preserving the marginal DP behaviour and allowing computational feasibility. Similar arguments could be easily developed to encompass the n-dimensional case; in particular, symmetry conditions similar to the one stated above must be assumed in order to obtain the same marginal behaviour at each site.

Modelling the marginals to be Dirichlet processes allows direct comparison with the models described by Gelfand, Kottas and MacEachern (2004) and De Iorio, Müller, Rosner and MacEachern (2004). However, it is worth noting that, though we employ

a generalized stick-breaking construction and achieve DP marginal distributions, our model doesn't generally describe a joint Dirichlet Process for a collection of locations. In particular, it follows that, given the dependence between the θ^*'s in the sum representation (4.1.1), we are not able to trace a joint urn scheme, but only a marginal one. The SDP model described in Gelfand, Kottas and MacEachern (2004) stands as a particular case of the model described here, where in table (4.1) we set $q_{l,0,1} = q_{l,1,0} = 0$ and $q_{l,1,1} = q_l$ for all locations and for all l.

We can see the difference from the SDP model also by looking at the random conditional distribution associated with $Y(s_i)|Y(s_j)$ for any pair of locations s_i, s_j. In fact, in the SDP this is just a random indicator function. In our model, it turns out to be another random measure. In fact, the random distribution $Y(s_i)|Y(s_j) = \theta^*_m(s_j)$ is a.s. discrete of the form $\sum_{l=1}^{\infty} p_{l|m}(s_i, s_j)\delta_{\theta^*_l(s_i)}$, where

$$\begin{aligned} p_{l|m}(s_i, s_j) &= P(Y(s_i) = \theta^*_l(s_i)|Y(s_j) = \theta^*_m(s_j)) = \\ &= \frac{p_{lm}(s_i, s_j)}{\prod_{k=1}^{m-1}(1 - q_k(s_j))\, q_m(s_j)}, \end{aligned} \tag{4.1.8}$$

since $\sum_l p_{l,m}(s_i, s_j) = p_m(s_j)$ due to marginal stick-breaking. But, substituting the expressions in (4.1.7) and again suppressing s_i and s_j, we get

$$p_{l|m} = \begin{cases} \prod_{k=1}^{l-1} \frac{q_{k,0,0}}{(1-q_k)} \frac{q_{l,1,0}}{1-q_l} & \text{if } l < m \\ \prod_{k=1}^{m-1} \frac{q_{k,0,0}}{(1-q_k)} \frac{q_{m,0,1}}{q_m} \prod_{k=m+1}^{l-1}(1 - q_k)\, q_l & \text{if } m < l \\ \prod_{k=1}^{l-1} \frac{q_{k,0,0}}{(1-q_k)} \frac{q_{l,1,1}}{1-q_l} & \text{if } l = m. \end{cases} \tag{4.1.9}$$

If we proceed along the lines that lead us to (4.1.7), we can show that for any given m, based on conditional reasoning, (4.1.9) defines a stick-breaking partition of the unit segment. However, this is not obtained through the usual Beta$(1, \nu)$ random variables, even if the process is marginally Dirichlet. In fact, the random measure arising from (4.1.9) can be seen as a generalized Dirichlet process, in the spirit of the more general definitions of Hjort (2000) and Ishwaran and James (2001).

As a final remark, notice that defining a stick-breaking construction does not necessarily ensure that the random weights sum to one with probability one. This depends of course on the distribution of the weights. In the context of univariate stick-breaking priors, however, it is possible to provide a necessary and sufficient condition for that to happen (see Lemma 1.3.1 in section 1.3). The same condition also holds for our model, as long as we marginally get a DP prior (or, more in general, a stick-breaking prior). The precise argument is a direct extension of the result of Ishwaran and James and is developed for the bivariate case in Appendix I as Lemma 4.5.2. Extension to the general n-dimensional case is again straightforward.

4.1.2 Mixing using a Generalized Spatial Dirichlet Process.

In Bayesian hierachical modelling, if a semiparametric specification is sought using Dirichlet processes, the DP is introduced at the lower levels of the hierarchy to model uncertainty on the priors for the parameters of the model. In spatial data modeling, the DP has been used [see Gelfand, Kottas and MacEachern (2004) and section 2.2] to model the distribution of the spatial component $\theta(s)$ in a random effect model of the type

$$Y(s) = \mu(s) + \theta(s) + \varepsilon(s), \tag{4.1.10}$$

where $\mu(s)$ is a constant mean term, typically assumed to be a regression term $\mathbf{X}(s)^T\boldsymbol{\beta}$ for some vector of covariates $\mathbf{X}(s)$ and some vector of parameters $\boldsymbol{\beta}$, and $\varepsilon(s)$ is a pure error (nugget) component, usually specified as a white noise with mean zero and variance τ^2. In our framework, $\theta(\cdot)$ follows the GSDP as above. Therefore, if we denote by $G^{(n)}$ the finite dimensional distributions defined by 4.1.1, for any finite set of locations $\mathbf{s} = (s_1, \ldots, s_n)$, $n \in \mathbb{N}$, the joint distribution of $\mathbf{Y} = (Y(s_1), \ldots, Y(s_n))^T$,

given $G^{(n)}$, $\boldsymbol{\mu}(\cdot)$ and τ^2 is given by

$$F\left(\mathbf{y}|\, G^{(n)}, \boldsymbol{\mu}, \tau^2\right) = \int F_{N_n}\left(\mathbf{y}|\, \boldsymbol{\theta} + \boldsymbol{\mu}, \tau^2 I_n\right)\, G^{(n)}(d\boldsymbol{\theta}), \tag{4.1.11}$$

where $\boldsymbol{\theta} = (\theta(s_1), \ldots, \theta(s_n))^T$, $\boldsymbol{\mu} = (\mu(s_1), \ldots, \mu(s_n))^T$, and $F_{N_n}(\cdot|\, \mathbf{m}, \boldsymbol{\Sigma})$ denotes the p-dimensional normal distribution with mean vector $\mathbf{m}$ and covariance matrix $\boldsymbol{\Sigma}$. Differentiating to densities,

$$f\left(\mathbf{y}|G^{(n)}, \boldsymbol{\mu}, \tau^2\right) = \int N_n\left(\mathbf{y}|\, \boldsymbol{\theta} + \boldsymbol{\mu}, \tau^2 I_n\right)\, G^{(n)}(d\boldsymbol{\theta}). \tag{4.1.12}$$

As with the SDP, since $G^{(n)}$ is almost surely discrete, the conditional density (4.1.12) can be rewritten a.s. as a countable location mixture of normals,

$$f\left(\mathbf{Y}|G^{(n)}, \boldsymbol{\mu}, \tau^2\right) = \sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} p_{i_1,\ldots,i_n} N_n(\mathbf{Y}|\, \boldsymbol{\theta}_{i_1,\ldots,i_n} + \boldsymbol{\mu}, \tau^2 I_n), \tag{4.1.13}$$

where, for simplicity, we have suppressed the locations in $p_{i_1,\ldots,i_n}$ and indicated $\boldsymbol{\theta}_{i_1,\ldots,i_n} = (\theta_{i_1}(s_1), \ldots, \theta_{i_n}(s_n))^T$. Computation of the moments of this distribution is immediate. Given $G^{(n)}$, $\mathbf{Y}$ is a random vector which has density a.s. absolutely continuous with respect to the Lebesgue measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, expected value $E(\mathbf{Y}|G^{(n)}, \boldsymbol{\mu}, \tau^2) = \sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} p_{i_1,\ldots,i_n} \boldsymbol{\theta}_{i_1,\ldots,i_n} + \boldsymbol{\mu}$, and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Y}}|G^{(n)}, \boldsymbol{\mu}, \tau^2 = \tau^2 I_n + \boldsymbol{\Sigma}_{\theta}^{\mathrm{s}}$, where $(\boldsymbol{\Sigma}_{\theta}^{\mathrm{s}})_{i,j} = Cov\left(\theta(s_i), \theta(s_j)|G^{(n)}\right)$ is given by (4.1.5).

Under the assumptions of Proposition 4.1.1, if, in addition, the mean vector $\boldsymbol{\mu}$ describes a continuous surface over D, it is easy to prove that an analogous statement holds for the convolved process $\mathbf{Y}$. In fact, the normal density is a bounded continuous function of the mean. Then the bounded convergence theorem applies to (4.1.11), and together with a.s. convergence of the random probability measures $G^{(n)}$ proved in Proposition 4.1.1, this implies that with probability one $Y(s)$ converges weakly to $Y(s_0)$ for any $s, s_0 \in D$, as $||s - s_0|| \to 0$.

4.2 The Spatially Varying Probabilities Model

In this section we discuss how to specify the stick-breaking components in a way that is appealing for modeling purposes and ensures the existence of the processes sampled from G. This discussion also distinguishes our approach from that of Griffin and Steel (2004), based on mixtures of order-based dependent Dirichlet processes. We defer this distinction to the end of the section.

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

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

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

In particular, suppose $\Theta_l^1(s)$ occurs i.f.f. $Z_l(s) \in A_l(s)$, i.e., we can work with the stochastic process $\{\delta^*_{A_l(s)}, s \in D, l = 1, 2, \ldots\}$ defined by

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

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

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

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

It is easy to see that such a characterization guarantees that (4.1.2) is true, hence the existence of the processes sampled from the random distribution (4.1.1).

In the following, we assume that $\{Z_l(s),\, s \in D,\, l = 1, 2, \ldots\}$ is a countable collection of independent stationary Gaussian random fields on D having variance 1 and correlation function $\rho_Z(\cdot, \eta)$. We assume that the mean of the process, say $\mu_l(s)$, is unknown and we put a convenient prior on it, so that the distribution of $Z_l(s)$ (and hence of the q_l's) can be actually viewed as random. We also choose $A_l(s) = \{Z_l(s) \geq 0\}$. With these assumptions, it follows that

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

because of the independence of the processes $\{Z_l(s)\}$ over the index l. For example, for $n = 2$, we get $q_{l,0,1} = P(Z_l(s_1) < 0, Z_l(s_2) \geq 0 | \mu_l(s_1), \mu_l(s_2))$. If the process $\{\mu_l(s),\, s \in D\}$ is the same across $l = 1, 2, \ldots$, then also the $q_{l,u_1,\ldots,u_n}(s_1, \ldots, s_n)$ are i.i.d. because of the independence of the $\{Z_l(s)\}$.

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

$$q_{l,1}(s) = P(Z_l(s) \geq 0) = 1 - \Phi(-\mu_l(s)) = \Phi(\mu_l(s)), \tag{4.2.1}$$

where $\Phi(\cdot)$ denotes the univariate standard normal distribution function. If the $\mu_l(s)$ are such that $\Phi(\mu_l(s)) \overset{i.i.d.}{\sim} Be(1, \nu)$ then for each s, the marginal distribution of $Y(s)$ is a DP with probabilities that vary with location. In the special case that $\mu_l(s) = \mu_l, \forall s$, with $\Phi(\mu_l) \overset{i.i.d.}{\sim} Be(1, \nu)$ then, again, marginally the $Y(s)$ are a DP where the marginal weights are the same for each s but the marginal distributions are not the same since $\theta^*_l(s) \neq \theta^*_l(s')$.

It is worth noting that marginal reduction to a DP is not strictly necessary for the definition of the GSDP (although it can be useful for purposes of comparison with the SDP by Gelfand, Kottas and MacEachern (2004) or other competing approaches). For instance, if we retain the $\mu_l(s)$, then, since we would like to encourage $Z_l(s)$ to resemble $Z_l(s')$ when s is close to s', we could take $\mu_l(s)$ to be a realization of say a Gaussian

spatial process rather than say i.i.d. as above. Now, the stick-breaking components are no longer Beta-distributed.

We have described the construction of a flexible model for the spatial random effects $\theta(s)$ in order that they can come from different random spatial surfaces at different locations. Two properties must be satisfied by this construction. We first demand the random finite dimensional distribution $G^{(n)}$ to satisfy the Kolmogorov consistency condition so that $G^{(n)}$ is the finite dimensional distribution of a spatial process having distribution G. Second the continuity property showed in Proposition 4.1.1 should also be satisfied, that is if location s is near s', we want the probability of picking up the same sample surface for s and s' to be high.

To recapitulate, we will never actually calculate the random weights $p_{i_1,\ldots,i_n}$. Rather, as is frequently done in hierarchical modeling, we introduce latent variables, in this case, a countable collection of independent Gaussian process realizations. That is, $Z_l(s)$ is a Gaussian process with mean $\mu_l(s)$, variance 1, and covariance function $\rho_Z(\cdot,\eta)$ and we let

$$\begin{aligned} p_{i_1,\ldots,i_n} = P\Big(& Z_1(s_1) < 0, \ldots, Z_{i_1-1}(s_1) < 0, Z_{i_1}(s_1) \geq 0; \\ & Z_1(s_2) < 0, \ldots, Z_{i_2-1}(s_2) < 0, Z_{i_2}(s_2) \geq 0; \ldots; \\ & Z_1(s_n) < 0, \ldots, Z_{i_n-1}(s_n) < 0, Z_{i_n}(s_n) \geq 0 | \{\mu_l(s_i)\} \Big), \end{aligned} \tag{4.2.2}$$

In the appendix we prove that the construction above satisfies the Kolmogorov consistency and the continuity conditions.

Finally, spatially varying weights have recently been considered by Griffin and Steel (2004), who work in the framework of dependent Dirichlet processes. They proceed from the assumption that the distribution of a $DP(\nu G_0)$ is unaffected by a permutation of the atoms $\{\theta_l^*(\cdot), q_l(\cdot), l = 1, 2, \ldots\}$ in Sethuraman's constructive representation. Then, if $\{\pi(s),\ s \in D\}$ is a process of permutations of the set of integers $\{1, 2, \ldots\}$, it

is possible to define an *order-based dependent stick-breaking prior* over D, abbreviated πDDP as a process $\{F_\pi(s),\, s \in D\}$, such that at any $s \in D$, given a realization of the process $\pi(s)$,

$$F_\pi(s) = \sum_{l=1}^{\infty} p_l(s)\, \delta_{\theta_l^*(s)},$$

where

$$p_l(s) = q_{\pi_l(s)} \prod_{j<l} (1 - q_{\pi_j(s)}).$$

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

$$P(Y(s_i) = \theta_l^*(s_i), Y(s_j) = \theta_m^*(s_j)) = \int p_l(s_i)\, p_m(s_j)\, dH(\pi(s_i), \pi(s_j)),$$

that is as the expected values of the marginal probabilities with respect to the distribution of the permutation field at the two locations. By the defintion of πDDP, it follows that the dependence among the marginal random distribution functions is directly deduced by the permutation at each s. In particular, this is given by means of an auxiliary latent point process Z. In fact, Griffin and Steel first associate each atom $\{\theta_i^*(s), q_i\}$ with a realization z_i from Z, for $i = 1, 2, \ldots$. Then, at any s, the πDDP is defined permuting the set of $q's$ according to the realizations of the latent point process Z. In fact, $\pi(s)$ is defined so to satisfy

$$||s - z_{\pi_1(s)}|| < ||s - z_{\pi_2(s)}|| < \ldots.$$

It follows that a realization from this process will necessarily be the same for some regions of D, while allowing different stick-breaking constructions for points far apart

from each other. However, the representation they get at any s depends on how the process Z is associated with the atoms of the process, so that the representation does not seem to be invariant to a reordering of the z's. Moreover, for practical purposes it can be difficult to model the type of dependence induced by the point process mechanism, unless we choose simple processes, such as a stationary Poisson process. On the other hand, in our approach the stick-breaking components depend on the distribution of the latent gaussian process Z and can vary across locations if this is true for the mean of Z. In fact, as we have discussed above, we directly model the spatial behavior of the q's, either marginally or jointly for any finite set of locations, and this leads to an easier dependence structure, as outlined in equation (4.1.5) and (4.1.6).

4.3 Simulation Based Model fitting for the GSDP

We now define the spatial model we work with and discuss the techniques used for model fitting and inference. Let $\mathbf{Y}_t = (Y_t(s_1), Y_t(s_2), \ldots, Y_t(s_n))^T$, $t = 1, \ldots, T$ be T groups of observations at the same set of locations $(s_1, \ldots s_n) \in D \subset \mathbb{R}^2$. We consider a random effects model like (4.1.10), where the mean surface $\mu(s), s \in D$ is explicitly modeled by a linear regression, i.e. $\mu(s) = \mathbf{x}^T(s_i)\beta$. The spatial random effect $\theta(s)$, $s \in D$ is specified through a GSDP, whose weights are spatially varying according to the realizations of a collection of stationary gaussian random fields $\{Z_l(s), s \in D\}$, $l = 1, 2, \ldots$ as outlined in section 4.2. Let $\boldsymbol{\theta}_t = (\theta_t(s_1), \theta_t(s_2), \ldots, \theta_t(s_n))^T$ indicate a vector of spatial component for each $t = 1, \ldots, T$.

Therefore, we get the following semiparametric hierarchical model:

$$\begin{aligned}
\mathbf{Y}_t \mid \boldsymbol{\theta}_t, \boldsymbol{\beta}, \tau^2 &\overset{ind.}{\sim} N_n(\mathbf{Y}_t \mid X_t^T\boldsymbol{\beta} + \boldsymbol{\theta}_t, \tau^2 I_n), \quad t = 1, ..., T \\
\boldsymbol{\theta}_t \mid G^{(n)} &\overset{i.i.d.}{\sim} G^{(n)}, \quad t = 1, ..., T \\
G^{(n)} \mid p_{i_1,...,i_n}, \sigma^2, \phi &\sim GSDP(p_{i_1,...,i_n}, G_0^{(n)}), \quad G_0^{(n)} \equiv N_n\left(0, \sigma^2 R_n(\phi)\right) \\
\boldsymbol{\beta}, \tau^2 &\sim N_p(\boldsymbol{\beta} \mid \boldsymbol{\beta}_0, \Sigma_{\boldsymbol{\beta}}) \times \text{IGamma}(\tau^2 \mid a_\tau, b_\tau) \\
\sigma^2, \phi &\sim \text{IGamma}(\sigma^2 \mid a_\sigma, b_\sigma) \times [\phi],
\end{aligned} \tag{4.3.1}$$

where the model is completed once we determine how to model the set of $p_{i_1,...,i_n}$, $i_j = 1, 2, \ldots$, $j = 1, \ldots, n$. According to what discussed in the previous section we choose to model $p_{i_1,...,i_n}$ as in (4.2.2), where the vectors $\mathbf{Z}_{t,l} = (Z_{t,l}(s_1), \ldots, Z_{t,l}(s_n))^T$ are given by

$$\mathbf{Z}_{t,l},\ l = 1, 2, 3, \ldots, t = 1, 2, ...T \sim N_n\left(\mu_l \mathbf{1}_n,\ H_n(\eta),\right)$$

with μ_l, $l = 1, 2, \ldots$ such that

$$\Phi(\mu_l) \overset{i.i.d.}{\sim} Beta(1, \nu)$$

and $\eta \sim [\eta]$. The priors for ϕ and η depend on the specific form of the correlation functions specifying the matrices $R_n(\phi)$ and $H_n(\eta)$. The replications across t enable us to learn about the mean of the latent Z's, μ_l, or the process driving $\mu_l(s)$.

We can obtain samples from the posterior of model (4.3.1) by means of a Gibbs sampler. Notice that, although the marginal random distribution at one location s follows a Dirichlet process, the joint random distribution $G^{(n)}$ does not. In particular, the traditional method of marginalizing $G^{(n)}$ and exploit a Pólya urn scheme cannot be used in this case. Instead, we must approximate $G^{(n)}$ with a finite sum

$$G_K^{(n)}(\cdot) = \sum_{(i_1,...,i_n) \in \{1,2,...,K\}^n} p_{i_1,...,i_n} \quad \delta_{\theta^*_{i_1}(s_1)}(\cdot)\, \delta_{\theta^*_{i_2}(s_2)}(\cdot) \ldots \delta_{\theta^*_{i_n}(s_n)}(\cdot), \tag{4.3.2}$$

for K suitably large. Then, the distribution of the observables at the first level of the hierarchy in (4.1.1) is a finite mixture. $G_K^{(n)}(\cdot)$ is determined once we know $\boldsymbol{\theta}_l^* = (\boldsymbol{\theta}_l^*(s_1), \ldots, \boldsymbol{\theta}_l^*(s_n))^T$, $l = 1, \ldots, K$ and $\boldsymbol{Z}_{t,l}$, $l = 1, \ldots, K-1$, $t = 1, \ldots, T$. In fact, we need a truncating rule for the weights. It's immediate to set

$$p_K(s) = P\left(Z_1(s) < 0, \ldots, Z_{K-1}(s) < 0\right).$$

The actual computation of the weights $p_{i_1,\ldots,i_n}$ is very difficult because it involves the evaluations of multivariate normal cdf's. But that is not necessary to implement a Gibbs sampler, if we sample and use the latent variables $\boldsymbol{Z}_l$'s directly. In fact, in order to sample from the conditional distribution of $[\boldsymbol{\theta}_t | \, G_K^{(n)}]$ we can use the following equivalence. At any location s and at any $t = 1, \ldots, T$ the observed θ is given by

$$\theta_t(s) = \theta_1^*(s) I_{\{Z_{t,1}(s) \geq 0\}} + \theta_2^*(s) I_{\{Z_{t,1}(s) < 0, Z_{t,2}(s) \geq 0\}} + \ldots + \theta_K^*(s) I_{\{Z_{t,1}(s) < 0, \ldots, Z_{t,K-1}(s) < 0\}}, \tag{4.3.3}$$

that is a deterministic function of the $\theta_l^*(s)$'s, $l = 1, \ldots, K$ and $Z_{t,l}(s)$'s, $l = 1, \ldots, K-1$, $t = 1, \ldots, T$. Therefore, we can reparameterize the first stage of the hierarchical model as

$$[\mathbf{Y}_t | \boldsymbol{\theta}_t, \boldsymbol{\beta}, \tau^2] = [\mathbf{Y}_t | \mu, \boldsymbol{\theta}^*, \mathbf{Z}_t, \boldsymbol{\beta}, \tau^2], \tag{4.3.4}$$

where $\boldsymbol{\theta}^* = \{\boldsymbol{\theta}_l, l = 1, \ldots, K\}$ and $\mathbf{Z}_t = \{\mathbf{Z}_{t,l}, l = 1, \ldots, K-1, t = 1, \ldots, T\}$. Notice that this reparametrization is not invertible. Then, the conditional likelihood for $\mathbf{Y}_t$ can be expressed in different ways as

$$\begin{aligned}
[\mathbf{Y}_t |, \boldsymbol{\theta}^*, \mathbf{Z}_t, \boldsymbol{\beta}, \tau^2] &\propto e^{-\frac{1}{2\tau^2} \sum_{i=1}^n \left(Y_t(s_i) - X_t^T(s_i)\beta - \theta_t(s_i)\right)^2} \\
&\propto e^{-\frac{1}{2\tau^2} \sum_{l=1}^K \sum_{i=1}^n \left(Y_t(s_i) - X_t^T(s_i)\beta - \theta_l^*(s_i)\right)^2 I_{\{Z_{t,1}(s) < 0, \ldots, Z_{t,l-1}(s) < 0, Z_{t,l}(s) \geq 0\}}} \\
&\propto \prod_{i=1}^n \left(\sum_{l=1}^K e^{-\frac{1}{2\tau^2} \left(Y_t(s_i) - X_t^T(s_i)\beta - \theta_l^*(s_i)\right)^2} I_{\{Z_{t,1}(s) < 0, \ldots, Z_{t,l-1}(s) < 0, Z_{t,l}(s) \geq 0\}} \right).
\end{aligned}$$

Therefore, the posterior distributions for the latent variables and parameters of model (4.3.1) are proportional to this likelihood function multiplied by the priors, that is

$$\prod_{t=1}^{T}[\mathbf{Y}_t|\theta^*,\mathbf{Z}_t,\tau^2]\times\prod_{l=1}^{K}[\theta_l^*|\sigma^2,\phi]\times\prod_{t=1}^{T}\prod_{l=1}^{K-1}[\mathbf{Z}_{t,l}|\mu_l,\eta]\times \\ \times[\mu_l]\times[\sigma^2]\times[\phi]\times[\tau^2]\times[\eta]. \tag{4.3.5}$$

The previous posterior can sampled from by means of a Gibbs sampler. The details of all the full conditionals are given in the appendix.

4.4 Data Illustration

We illustrate the main features of our model by means of a simulated data set. We generate data from a finite mixture model of gaussian random fields, whose weights are taken to be spatially varying. More precisely, we draw T independent groups of observations from a model with different multi-modal distributions at distinct locations. For example, that could reflect the effect of different topography in the neighborhood of each location. Notice that the resulting dataset shows features that wouldn't be conveniently revealed by models with a simple gaussian specification for the spatial random effect. Also the SDP specification seems to be unappropriate to catch the type of multimodality we consider.

Therefore, let $\mathbf{Y}_t$ be as in section 4.3, $t=1,\ldots,T$. Then, we sample our observations as follows. We consider $\mathbf{Y}_t(s)$ to arise from a mixture of two gaussian processes, so that at each location s,

$$Y_t(s)\sim\alpha(s)\,G^1_{0,s}+(1-\alpha(s))\,G^2_{0,s}, \tag{4.4.1}$$

where $G^1_{0,s}$ and $G^2_{0,s}$ are GP specified, respectively, by a mean ξ_i and a covariance function $\sigma_i^2\,\rho_{\psi_i}(s,s')$, $i=1,2$, $s,s'\in D$. The weight $\alpha(s)$ is given by

$$\alpha(s)=P(Z(s)>0),$$

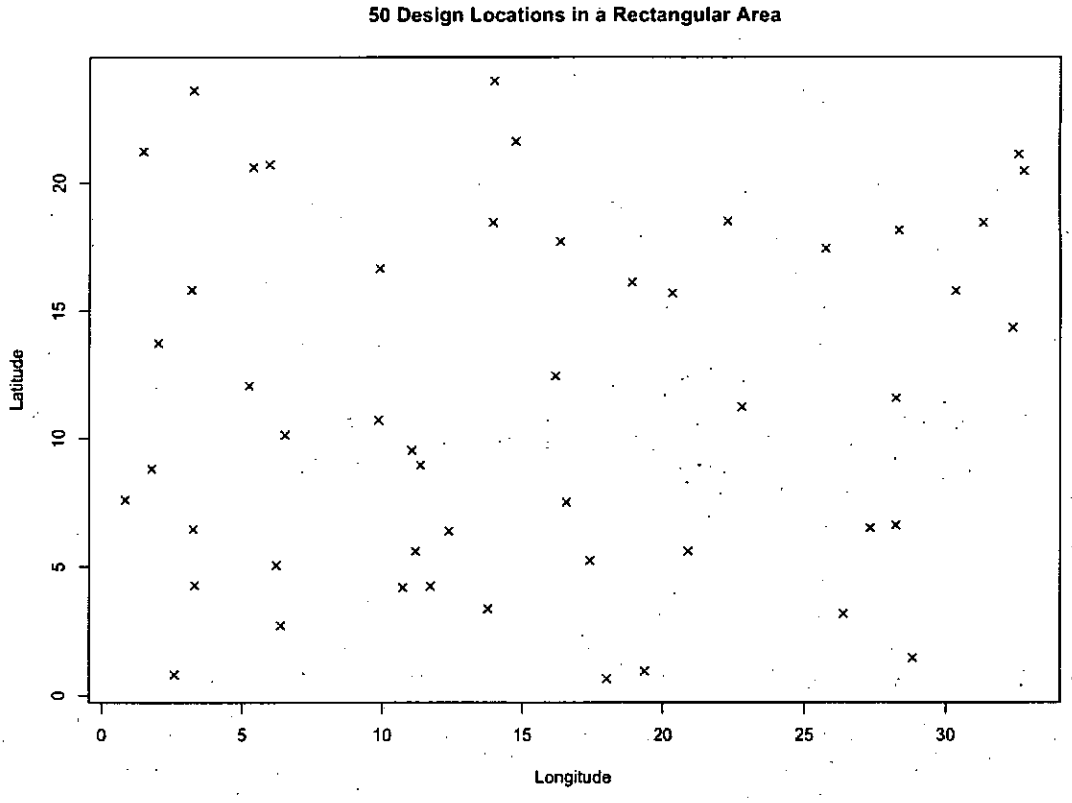

Figure 4.2: Plot of 50 design locations randomly sampled in a 35×25 rectangular area.

where $Z(s)$ is a mean zero stationary GP with covariance function $\rho_\eta(s, s')$. Therefore, for all $s \in D$ and $t = 1, \ldots, T$, we choose

$$Y_t(s) \sim G^1_{0,s} \quad \text{if } Z_t(s) \geq 0$$
$$Y_t(s) \sim G^2_{0,s} \quad \text{if } Z_t(s) < 0.$$

Since $Z(s)$ is centered at zero, it follows that marginally,

$$Y_t(s) \sim \frac{1}{2} N_n\left(\xi_1, \sigma_1^2\right) + \frac{1}{2} N_n\left(\xi_2, \sigma_2^2\right).$$

However, if we consider the joint distribution for a pair of locations s, s' in D, we have

$$\begin{aligned}(Y_t(s), Y_t(s')) \sim & \alpha_{1,1}(s, s')\, G^1_{0,s,s'} + \alpha_{2,2}(s, s')\, G^2_{0,s,s'} + \\ & + \alpha_{1,2}(s, s')\, G^1_{0,s}\, G^2_{0,s'} + \alpha_{2,1}(s, s')\, G^2_{0,s}\, G^1_{0,s'},\end{aligned}$$

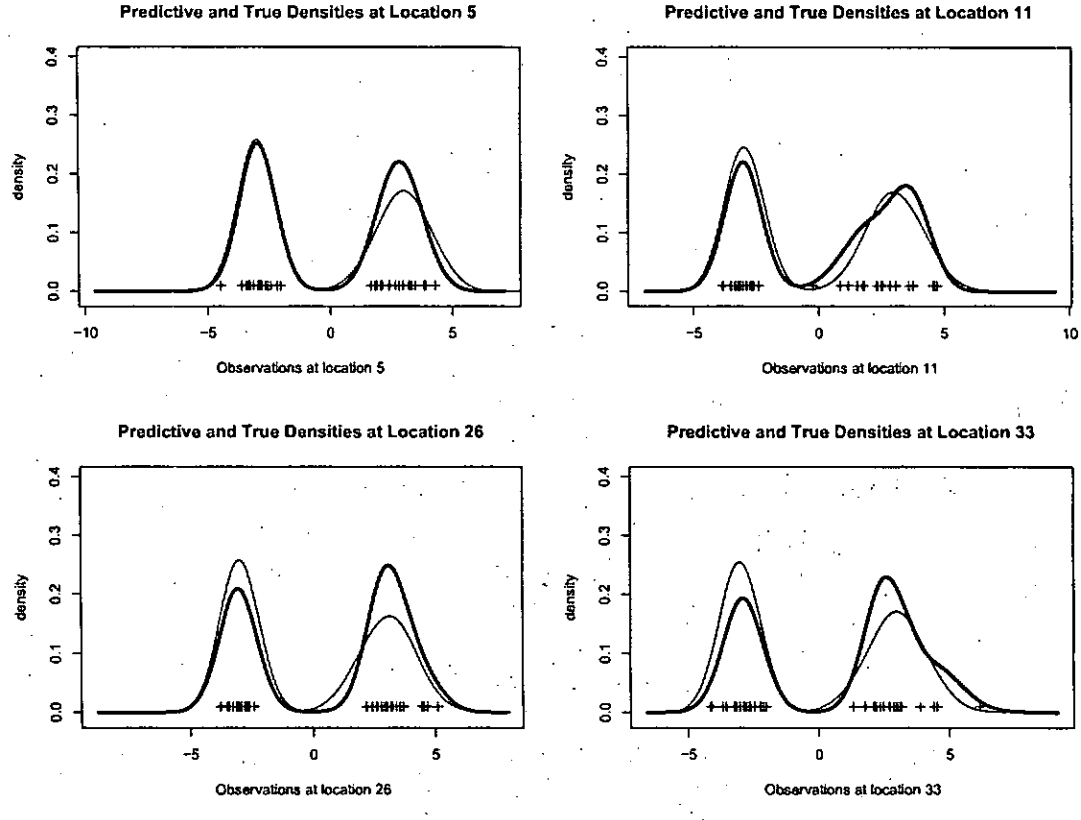

Figure 4.3: Posterior predictive densities (thick) and true densities (thin) for the 4 locations $(s_5, s_{11}, s_{26}, s_{33})$. The +'s mark the values of the 40 observations at those 4locations.

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 randomly sample $n = 50$ design locations in a rectangular area shown in figure 4.2. Then, we sample $T = 40$ independent groups of observations at those locations. In doing so, we have considered the following values of the parameters for the mixture 4.4.1: $\xi_1 = -\xi_2 = 3$, $\sigma_1 = 2\,\sigma_2 = 2$, $\phi_1 = \phi_2 = 0.3$, and $\eta = 0.3$.

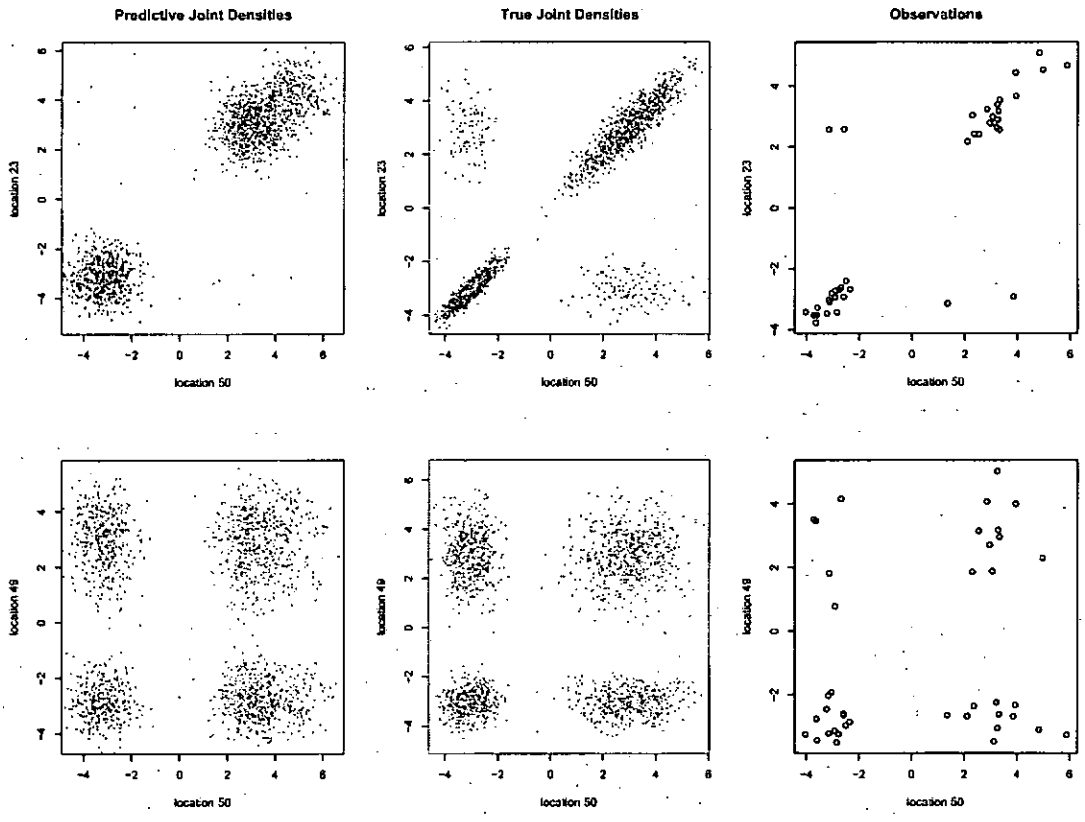

Figure 4.4: Scatter plots of 2000 drawns from the posterior predictive density (left) and from the true density (center), together with a scatter plot of the 40 observed data (right) at sites (s_{50}, s_{23}) and (s_{50}, s_{49})

We fit the data by means of model (4.3.1). In order to focus on the modeling of the spatial association, we set $\mu(s) = 0$, that is our observations don't have any mean structure. We implement a Gibbs sampler following the discussion in section 4.3 and in the appendix. In particular, we approximate $G^{(n)}$ with the finite sum 4.3.2 where $K = 20$.

The Bayesian goodness of fit is illustrated by the posterior predictive densities. In this example, we show not only the marginal posterior predictive density at each location, but also the joint posterior predictive densities for some pairs of locations.

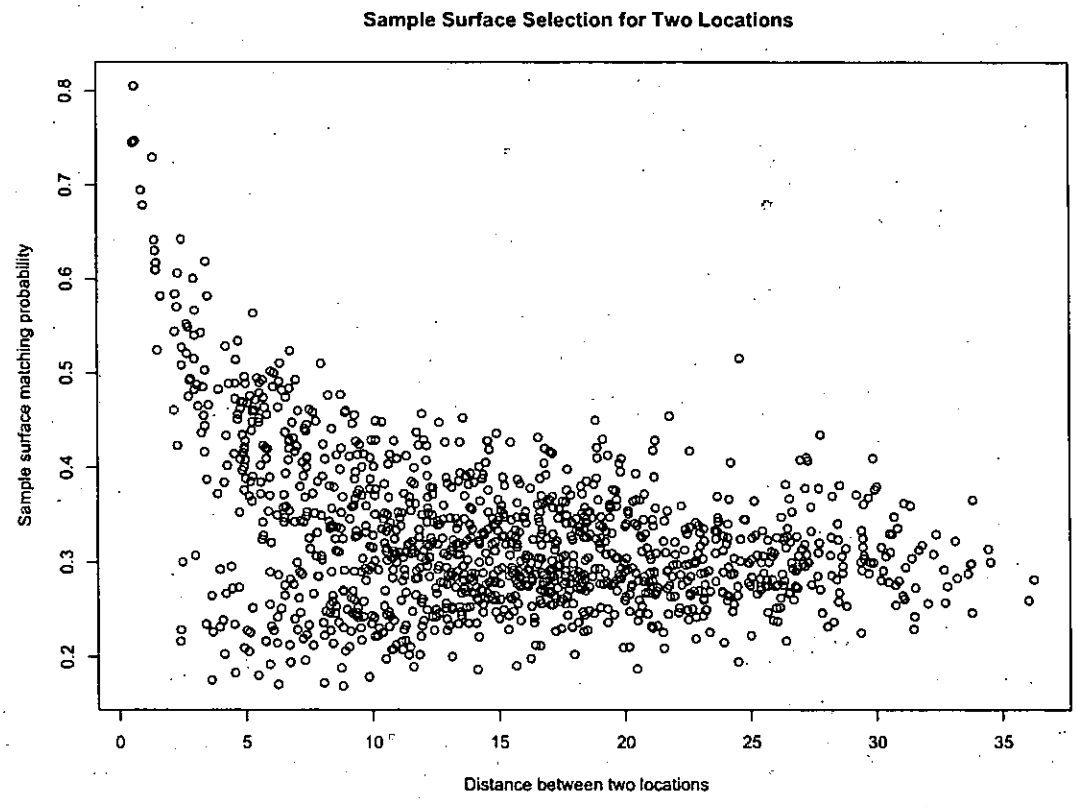

Figure 4.5: Matching probabilities against the location distance.

For example, in Figure 4.3, we plot the posterior predictive density for four randomly selected locations, precisely $(s_5, s_{11}, s_{26}, s_{33})$. The thick density curves represent the posterior predictive densities obtained via composition sampling after fitting the model. The thin density curves are the true densities of the model from which we simulated the data. It is evident that the semiparametric model (4.3.1) is able to capture the bimodality of the data at each location.

Now, we select 2 pairs of sites and consider the posterior predictive joint densities at those sites. In particular, we choose the first pair (s_{50}, s_{23}) so that the locations are close to each other. Instead, the second pair (s_{50}, s_{49}) is chosen so that they are distant. In figure 4.4, we show the scatter plots of 2000 drawns from the predictive and

true joint densities at those sites. It is evident that the predictive sample clouds for $(Y(s_{50}), Y(s_{23}))$ are highly correlated. This feature is in accordance with the clouds of the observations and with the samples from the true density. Instead, the scatter plot for $(Y(s_{50}), Y(s_{49}))$ shows less correlation, as the observed data also do. Therefore, our model is able to detect spatial behaviors characterized by several spatial effects acting over different regions and different replicates. Those are revealed by joint multimodality, although the marginal distribution is only bimodal. Notice that this behavior wouldn't be detected by a simple SDP mixture model. In fact, the former would have shown the same bimodality both marginally and jointly.

Finally, in figure 4.5, we plot an estimate of the matching probabilities, i.e. the probabilities that the same sample surface is selected for two distinct locations, against their distance. These values are obtained from the corresponding matching frequencies calculated in the samples from the posterior distribution when the Gibbs sampler has reached convergence. We can see that the observations are more likely to be on the same sample surface if the locations are close to each other, as expected. However, even if the distance is great, there's always a positive probability to sample observations from the same surface, although it can be small. In fact, it follows from (4.1.6) that the association between the samples in two sites which are distant one from the other is close to zero, according to the rate of decay both of the covariance function $\rho_\phi(s - s')$ and of the matching probabilities.

4.5 Appendices

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

Appendix I: Theoretical arguments

Proof of Proposition 4.1.1.

Proof. Consider two sites s, s_0 in D. According to (4.1.1) the joint distribution of the process is almost surely a realization of the random element

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

for all $A, B \in \mathcal{B}(\mathbb{R})$. If we denote by $\mathbf{p}$ the distribution of the weights and let s tend to s_0,

$$\begin{aligned}
\lim_{||s-s_0||\to 0} P(Y(s) \in A, Y(s_0) \in B) &= \lim_{||s-s_0||\to 0} E_{G_0,\mathbf{p}}\Big[P\Big(Y(s) \in A, Y(s_0) \in B|G\Big)\Big] \\
&= \lim_{||s-s_0||\to 0} E_{G_0,\mathbf{p}}\left[\sum_{l,m=1}^{\infty} p_{l,m}\, \delta_{\theta_l^*(s)}(A)\, \delta_{\theta_m^*(s_0)}(B)\right] \\
&= \sum_{l=1}^{\infty} E_{\mathbf{p}}\left(p_l\right)\, E_{G_0}\left(\delta_{\theta_l^*(s_0)}(A \cap B)\right),
\end{aligned}$$

because of Tonelli's theorem, the independence between $\mathbf{p}$ and G_0, the finiteness of the limits involved and the a.s. continuity hypotheses. Then,

$$\begin{aligned}
\lim_{||s-s_0||\to 0} P(Y(s) \in A, Y(s_0) \in B) &= P_{G_0}(\theta^*(s_0) \in A \cap B) \\
&= P(Y(s_0) \in A \cap B),
\end{aligned}$$

and the result stated is obtained when we consider $B = \mathbb{R}$. □

Lemma 4.5.1. *The probabilities*

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

$u, v \in \{0, 1\}$, *are such that*

$$q_{l,1,+}(s_i, s_j) = q_l(s_i)$$

and

$$q_{l,+,1}(s_i, s_j) = q_l(s_j),$$

for any $l = 1, 2, \ldots$.

Proof. Recall that

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

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

Hence,

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

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

Lemma 4.5.2. *For any given* s_i, s_j *in* D,

$$\sum_{l,m=1}^{\infty} p_{l,m}(s_i, s_j) = 1 \quad \textit{iff} \quad \sum_{l=1}^{\infty} E\left(\log(1 - q_l(s_i))\right) = -\infty. \tag{4.5.1}$$

Proof. First, notice that if we marginalize with respect to s_i, we need

$$\sum_{l,m=1}^{\infty} p_{l,m}(s_i, s_j) = \sum_{m=1}^{\infty} p_m(s_j) = 1,$$

and condition (4.5.1) has been proved by Ishwaran and James (2001).

Now we turn to the proof of sufficiency. This is actually an extension for the bivariate case of the proof in Ishwaran and James (2001). In fact, consider for any $N = 1, 2, \ldots$, the remainder term

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

Assume condition (4.5.1) holds. We need to prove that $R_{N,M}(s_i, s_j) \to 0$ a.s. as $N, M \to \infty$. It's easy to see that

$$R_{N,M}(s_i, s_j) = \sum_{l=1}^{N} \sum_{m=M+1}^{\infty} p_{l,m}(s_i, s_j) + \sum_{m=1}^{M} \sum_{l=N+1}^{\infty} p_{l,m}(s_i, s_j) + \sum_{l=N+1}^{\infty} \sum_{m=M+1}^{\infty} p_{l,m}(s_i, s_j). \tag{4.5.2}$$

Since all the terms in the sums are positive, to show $R_{N,M}(s_i, s_j) \to 0$ it is necessary and sufficient that all the series tend to zero, as $N, M \to \infty$. Then we can work with each of them separately. Consider the first term in the sum and substitute (4.1.7) to all $p_{l,m}(s_i, s_j)$, so that

$$\begin{aligned} \sum_{l=1}^{N} \sum_{m=M+1}^{\infty} p_{l,m}(s_i, s_j) &= \sum_{l=1}^{N} \sum_{m=M+1}^{\infty} \prod_{k=1}^{l-1} q_{k,0,0}(s_i, s_j)\, q_{l,1,0}(s_i, s_j) \prod_{r=l+1}^{m-1} (1 - q_r(s_j))\, q_m(s_j), \\ &= \sum_{l=1}^{N} \prod_{k=1}^{l-1} q_{k,0,0}(s_i, s_j)\, q_{l,1,0}(s_i, s_j) \sum_{m=M+1}^{\infty} \prod_{r=l+1}^{m-1} (1 - q_r(s_j))\, q_m(s_j). \end{aligned} \tag{4.5.3}$$

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

$$\begin{aligned} \sum_{m=M+1}^{\infty} \prod_{r=l+1}^{m-1} (1 - q_r(s_j))\, q_m(s_j) &= \sum_{m=M+1}^{\infty} P(\Theta_m^1(s_j) | \, \Theta_r^0(s_j), r = 1, \ldots, l) \\ &= \sum_{m=M+1}^{\infty} \frac{P(\Theta_m^1(s_j), \Theta_r^0(s_j), r = 1, \ldots, m-1)}{P(\Theta_r^0(s_j), r = 1, \ldots, l)} \\ &= \frac{\sum_{m=M+1}^{\infty} \prod_{r=1}^{m-1} (1 - q_r(s_j))\, q_m(s_j)}{\prod_{r=1}^{l} (1 - q_r(s_j))} \\ &= \frac{\sum_{m=M+1}^{\infty} p_m(s_j)}{1 - \sum_{m=1}^{l} p_m(s_j)}. \end{aligned}$$

Then, if we let $M \to \infty$, the numerator tends to 0, because it is the remainder of the sum of the weights for the marginal model in s_j, and so we can directly apply the result in Ishwaran and James (2001). Then, each term of the series in (4.5.2) tends to 0 as $M \to \infty$. So,

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

We can follow a similar argument for the second remainder term in (4.5.2). Now consider

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

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

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

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

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

Proposition 4.5.3. *Let $\{\theta(s_1),\theta(s_2),\ldots\theta(s_n), s_i \in D, i = 1,\ldots n\}$ have random finite dimensional distribution given by (10). If the set of weights $\{p_{i_1,\ldots,i_n}\}$ is defined by (11), then this random finite dimensional distribution defines a random field $\theta(s)$ on D.*

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

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

Note that if $\theta(s_i) = \theta_k^*(s_i)$, we have the sequence of random variables

$(Z_1(s_i),\ldots,Z_k(s_i),\ldots) \in (-\infty,0)_1 \times \cdots \times (-\infty,0)_{k-1} \times [0,\infty)_k \times \mathbb{R} \times \cdots$

Denote $S_{i,k} = (-\infty,0)_1 \times \cdots \times (-\infty,0)_{k-1} \times [0,\infty)_k \times \mathbb{R} \times \cdots$. Rewrite

$$\begin{aligned} p_{i_1,\ldots,i_{l-1},i_{l+1},\ldots i_n} = P\Big(& (Z_1(s_i),\ldots,Z_{i_1}(s_i),\ldots) \in S_{1,i_1};\ldots; \\ & (Z_1(s_{l-1}),\ldots,Z_{i_{l-1}}(s_{l-1}),\ldots) \in S_{l-1,i_{l-1}}; \\ & (Z_1(s_{l+1}),\ldots,Z_{i_{l+1}}(s_{l+1}),\ldots) \in S_{l+1,i_{l+1}};\ldots; \\ & (Z_1(s_n),\ldots,Z_{i_n}(s_n),\ldots) \in S_{n,i_n}\Big) \end{aligned}$$

and

$$\begin{aligned} p_{i_1,\ldots,i_{l-1},k,i_{l+1},\ldots i_n} = P\Big(& (Z_1(s_i),\ldots,Z_{i_1}(s_i),\ldots) \in S_{1,i_1};\ldots; \\ & \big(Z_1(s_{l-1}),\ldots,Z_{i_{l-1}}(s_{l-1}),\ldots\big) \in S_{l-1,i_{l-1}}; \\ & (Z_1(s_l),\ldots,Z_k(s_l),\ldots) \in S_{l,k}; \\ & \big(Z_1(s_{l+1}),\ldots,Z_{i_{l+1}}(s_{l+1}),\ldots\big) \in S_{l+1,i_{l+1}};\ldots; \\ & (Z_1(s_n),\ldots,Z_{i_n}(s_n),\ldots) \in S_{n,i_n}\Big) \end{aligned}$$

By the continuity of probability measure,

$$\begin{aligned} \sum_{k=1}^{\infty} p_{i_1,\ldots,i_{l-1},k,i_{l+1},\ldots i_n} = P\Big(& (Z_1(s_i),\ldots,Z_{i_1}(s_i),\ldots) \in S_{1,i_1};\ldots; \\ & \big(Z_1(s_{l-1}),\ldots,Z_{i_{l-1}}(s_{l-1}),\ldots\big) \in S_{l-1,i_{l-1}}; \\ & (Z_1(s_l),\ldots,Z_k(s_l),\ldots) \in \bigcup_{k=1}^{\infty} S_{l,k}; \\ & \big(Z_1(s_{l+1}),\ldots,Z_{i_{l+1}}(s_{l+1}),\ldots\big) \in S_{l+1,i_{l+1}};\ldots; \\ & (Z_1(s_n),\ldots,Z_{i_n}(s_n),\ldots) \in S_{n,i_n}\Big) \end{aligned}$$

Note $\bigcup_{k=1}^{\infty} S_{l,k} = \bigotimes_{k=1}^{\infty} \mathbb{R}$, therefore $\sum_{k=1}^{\infty} p_{i_1,\ldots,i_{l-1},k,i_{l+1},\ldots i_n} = p_{i_1,\ldots,i_{l-1},i_{l+1},\ldots i_n}$.
For any $A_i \in \mathcal{B}(\mathbb{R})$, $i = 1,\ldots,k$, we have

$$\begin{aligned} & P(\theta(s_1) \in A_1,\ldots,\theta(s_{l-1}) \in A_{l-1},\theta(s_l) \in \mathbb{R},\theta(s_{l+1}) \in A_{l+1},\ldots,\theta(s_n) \in A_n|G_0) \\ = & \sum_{(i_1,\ldots,i_n)\in\{1,2,\ldots\}^n} p_{i_1,\ldots,i_n}\ \delta_{\theta^*_{i_1}(s_1)}(A_1)\cdots\delta_{\theta^*_{i_l}(s_l)}(\mathbb{R})\cdots\delta_{\theta^*_{i_n}(s_n)}(A_n) \\ = & \sum_{(i_1,\ldots,i_{l-1},i_{l+1},\ldots,i_n)\in\{1,2,\ldots\}^{n-1}} \delta_{\theta^*_{i_1}(s_1)}(A_1)\cdots\delta_{\theta^*_{i_n}(s_n)}(A_n)\left(\sum_{k=1}^{\infty} p_{i_1,\ldots,i_{l-1},k,i_{l+1},\ldots i_n}\right) \\ = & \sum_{(i_1,\ldots,i_{l-1},i_{l+1},\ldots,i_n)\in\{1,2,\ldots\}^{n-1}} p_{i_1,\ldots,i_{l-1},i_{l+1},\ldots,i_n}\delta_{\theta^*_{i_1}(s_1)}(A_1)\cdots\delta_{\theta^*_{i_n}(s_n)}(A_n) \\ = & P(\theta(s_1) \in A_1,\ldots,\theta(s_{l-1}) \in A_{l-1},\theta(s_{l+1}) \in A_{l+1},\ldots,\theta(s_n) \in A_n|G_0) \end{aligned}$$

The Kolmogorov Consistency conditions are satisfied. □

Proposition 4.5.4. *Let $\{\theta(s),\ s \in D\}$ be a random field defined in proposition 2. Then for all $s_0 \in D$, $Y(s)$ converges weakly to $Y(s_0)$ a.s. as $||s - s_0|| \to 0$.*

Proof. One assumption of our model is that $\theta^*(\cdot)$ and $Z(\cdot)$ are a.s. continuous,

$$\lim_{||s_n - s_{n-1}|| \to 0} P(\theta(s_1) \in A_1, \ldots, \theta(s_n) \in A_n) = P(\theta(s_1) \in A_1, \ldots, \theta(s_{n-1}) \in A_{n-1} \cap A_n).$$

In fact,

$$\lim_{||s_n - s_{n-1}|| \to 0} P(\theta(s_1) \in A_1, \ldots, \theta(s_n) \in A_n) =$$

$$= \lim_{||s_n - s_{n-1}|| \to 0} E_{G_0}\left(P(\theta(s_1) \in A_1, \ldots, \theta(s_n) \in A_n | G_0)\right)$$

$$= \lim_{||s_n - s_{n-1}|| \to 0} \sum_{\{i_1,\ldots,i_n\} \in \{1,2,3\ldots\}^n} p_{i_1,\ldots,i_n} \quad E_{G_0}\left(\delta_{\theta^*_{i_1}(s_1)}(A_1)\,\delta_{\theta^*_{i_2}(s_2)}(A_2)\ldots\delta_{\theta^*_{i_n}(s_n)}(A_n)\right)$$

$$= \sum_{(i_1,\ldots,i_n) \in \{1,2,3,\ldots\}^n} p_{i_1,\ldots,i_{n-1}} \quad \lim_{||s_n - s_{n-1}|| \to 0} E_{G_0}\left(\delta_{\theta^*_{i_1}(s_1)}(A_1)\,\delta_{\theta^*_{i_2}(s_2)}(A_2)\ldots\delta_{\theta^*_{i_n}(s_n)}(A_n)\right)$$

because of Tonelli's Theorem and the finiteness of the limits involved,

$$= \sum_{(i_1,\ldots,i_n) \in \{1,2,3,\ldots\}^n} p_{i_1,\ldots,i_{n-1}} \quad E_{G_0}\left(\delta_{\theta^*_{i_1}(s_1)}(A_1)\,\delta_{\theta^*_{i_2}(s_2)}(A_2)\ldots\delta_{\theta^*_{i_{n-1}}(s_{n-1})}(A_{n-1} \cap A_n)\right)$$

$$= E_{G_0}\left(\sum_{(i_1,\ldots,i_n) \in \{1,2,3,\ldots\}^n} p_{i_1,\ldots,i_{n-1}} \quad \delta_{\theta^*_{i_1}(s_1)}(A_1)\,\delta_{\theta^*_{i_2}(s_2)}(A_2)\ldots\delta_{\theta^*_{i_{n-1}}(s_{n-1})}(A_{n-1} \cap A_n)\right)$$

$$= E_{G_0}\Big(P(\theta(s_1) \in A_1, \ldots, \theta(s_n) \in (A_{n-1} \cap A_n) | G_0)\Big)$$

$$= P(\theta(s_1) \in A_1, \ldots, \theta(s_n) \in (A_{n-1} \cap A_n)).$$

Notice that

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

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

and this result is independent of the particular mean around which we center the process Z. □

Appendix II: Full conditionals for the Gibbs sampler

1. Full conditionals for the Z's.

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

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

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

$$\tilde{\mu}_{l,i} = \mu_l - h_i(\eta)^T H_{(-i)}^{-1}(\eta)\, \mathbf{Z}_{t,l}^{(-j)}, \tag{4.5.4}$$

$$\tilde{H}_i(\eta) = 1 - h_i(\eta)^T H_{(-i)}^{-1}(\eta) h_i(\eta), \tag{4.5.5}$$

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

Now consider the full conditionals. We start considering the full conditional of $Z_{t,1}(s_i)$, for some $i = 1, \ldots, n$. Let us indicate with $\psi = (X_t, \beta, \boldsymbol{\theta}^*, \tau^2, \sigma^2, \phi, \mu_l, l > 1, \eta)$ the vector of parameters of the model other than the Z's. Then, the full conditional of $Z_{t,1}(s_i)$ is given by

$$[Z_{t,1}(s_i)|Y_t, \mathbf{Z}_{t,1}^{(-j)}, \mathbf{Z}_{t,l}(s_j), l > 2, \psi] \propto [Z_{m,1}(s_i)|Z_{m,1}(s_j), j \neq i, \boldsymbol{\theta}^*, \mu, \eta, \phi] \times$$
$$\times \left(\sum_{k=1}^{K} e^{-\frac{1}{2\tau^2}\left(Y_t(s_i) - X_t^T(s_i)\beta - \theta_k^*(s_i)\right)^2} I_{\{Z_{t,1}(s_i)<0, \ldots, Z_{t,l-1}(s_i)<0, Z_{t,l}(s_i)\geq 0\}} \right), \tag{4.5.6}$$

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

for the weights we can define the two quantities

$$\omega^- = e^{-\frac{1}{2\tau^2}(Y_t(s_i)-X_t^T(s_i)\beta-\theta_1^*(s_i))^2}$$

$$\omega^+ = e^{-\frac{1}{2\tau^2}(Y_t(s_i)-X_t^T(s_i)\beta-\theta_l^*(s_i))^2}$$

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

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

we can see that (4.5.6) is a mixture of two truncated gaussian. Therefore, with probability π_1, we sample $Z_{t,1}(s_i)$ from the truncated normal distribution $N(\tilde{\mu}_{1,i}^1, \tilde{H}_i(\eta))I_{\{Z_{t,1}(s_i)\geq 0\}}$; with π_l sample $Z_{t,1}(s_i)$ from the truncated normal distribution $N(\tilde{\mu}_{1,i}, \tilde{H}_i(\eta))I_{\{Z_{t,1}(s_i)<0\}}$. We next proceed repeating the same arguments for $Z_{t,2}(s_i)$.

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

Otherwise if $Z_{t,j}(s_i) < 0$, for $j < l$, the full conditional is again a binary mixture of truncated normals as we have seen for $Z_t^1(s_i)$ (see equation 4.5.6). Say $Z_{t,k}(s_i) \geq 0$ for some $k > l$, again let

$$\omega^- = e^{-\frac{1}{2\tau^2}(Y_t(s_i)-X_t^T(s_i)\beta-\theta_k^*(s_i))^2}$$

$$\omega^+ = e^{-\frac{1}{2\tau^2}(Y_t(s_i)-X_t^T(s_i)\beta-\theta_k^*(s_i))^2},$$

and

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

Therefore, the full conditional for $Z_{t,l}(s_i)$ is again a mixture of two truncated normals. With probability π_l, we sample $Z_{t,l}(s_i)$ from the truncated normal distribution

$N(\tilde{\mu}_{l,i}^k, \tilde{H}_i(\eta))I_{\{Z_{t,l}(s_i)\geq 0\}}$; with probability π_l, we sample $Z_{t,l}(s_i)$ from the truncated normal distribution $N(\tilde{\mu}_{l,i}, \tilde{H}_i(\eta))I_{\{Z_{t,l}(s_i)<0\}}$. Next, proceed repeating similar arguments for $Z_{t,l+1}(s_i)$..

2. Full conditional for the θ^*'s.

We can update the θ^*'s all at once for all locations. In fact, in order to keep the notation simple, let us consider at each point $s \in D$ the partition induced on the space of the Z's by the allocative process, that is, for $t = 1, \ldots, T$, consider the sets

$$\mathcal{Z}_t^1(s) = \{s \in D : Z_{t,1}(s) > 0\}$$

$$\mathcal{Z}_t^2(s) = \{s \in D : Z_{t,1}(s) < 0, Z_{t,2}(s) > 0\}$$

$$\ldots$$

$$\mathcal{Z}_m^{K-1}(s) = \{s \in D : Z_{t,1}(s) < 0, \ldots, Z_{t,K-2}(s) < 0, Z_{t,K-1}(s) > 0\}$$

$$\mathcal{Z}_m^K(s) = \{s \in D : Z_{t,1}(s) < 0, \ldots, Z_{t,K-1}(s) < 0\}.$$

Then, $I(\mathcal{Z}_t) = diag(I_{\mathcal{Z}_{t,l}(s_1)}, \ldots, I_{\mathcal{Z}_{t,l}(s_n)})$ is the diagonal matrix whose i-th entry is equal to one when the component l is chosen at location s_i.

It's immediate to see that the full conditional for $\boldsymbol{\theta}_l^* = (\theta_l^*(s_1), \ldots, \theta_l^*(s_n))$ is given by

$$[\boldsymbol{\theta}_l^* | Y_t, \mathbf{Z}_t, t = 1, \ldots, T, \beta, \tau^2, \sigma^2, \phi] \propto e^{-\frac{1}{2\tau^2}\sum_{t=1}^T (Y_t - X_t^T\beta - \theta_k^*)^T I(\mathcal{Z}_t)(Y_t - X_t^T\beta - -\theta_k^*)} \times \times e^{-\frac{1}{2\sigma^2}\theta_k^{*T} R_n^{-1}(\phi)\theta_k^*} \tag{4.5.7}$$

Then,

$$[\boldsymbol{\theta}_l^* | Y_t, \mathbf{Z}_t, t = 1, \ldots, T, \beta, \tau^2, \sigma^2, \phi] \sim N\left(\frac{1}{\tau^2}\Lambda \sum_{t=1}^T I(\mathcal{Z}_t)\left(Y_t - X_t^T\beta\right), \Lambda\right)$$

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

Once we know (θ_l^* and $\mathbf{Z}_t$) for all $l = 1, \ldots, K$ and $t = 1, \ldots, T$, we can compute each θ_t as a function of $(\theta_l^*, \mathbf{Z}_t)$, as shown in equation (4.3.3).

3. Full conditional for β.

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

$$[\beta | X_t, Y_t, \mathbf{Z}_t, \theta_t, \tau^2] \sim N(\hat{\beta}, \hat{\Sigma}_\beta),$$

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

4. Full conditional for τ^2.

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

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

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

5. Full conditional for σ^2.

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

$$[\sigma^2 | \theta_l^*, l = 1, \ldots, K, \phi] \sim IGamma(\tilde{\alpha}_\sigma, \tilde{\beta}_\sigma),$$

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

6. Full conditional for ϕ.

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

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

7. Full conditional for μ.

Generally we must use a Metropolis step for μ_l, $l = 1, \ldots K-1$, unless the α in the $Beta(1, \alpha)$ is equal to 1.

Note that $P(Z_l(s) \geq 0) = \Phi(\mu_l)$ and $P(Z_l(s) \geq 0) \sim Beta(1, \alpha)$ induce a prior for μ_l such that

$$\mu_l \propto [1 - \Phi(\mu_l)]^{\alpha-1} \times e^{-\frac{1}{2}\mu_l^2}$$

. If $\alpha = 1$, the prior for μ_l is just normal, therefore conjugate. The full conditional for μ_l is given by

$$[\mu_l | \mathbf{Z}_{t,l}, \eta] \propto [1 - \Phi(\mu_l)]^{\alpha-1} \times e^{-\frac{1}{2}\mu_l^2} \times e^{-\frac{1}{2}\sum_{t=1}^{T}(Z_{t,l}-\mu_l 1_n)^T H_n^{-1}(\eta)(Z_{t,l}-\mu_l 1_n)}$$

8. Full conditional for η.

Depending on the prior, the full conditional of η can be sampled with a Metropolis within Gibbs step, by

$$[\eta|\mathbf{Z}_t, \mu_k] \sim [\eta] \times e^{-\frac{1}{2}\sum_{t=1}^{T}\sum_{l=1}^{K-1}(Z_{t,l}-\mu_l 1_n)^T H_n^{-1}(\eta)(Z_{t,l}-\mu_k 1_n)},$$

where $[\eta]$ is a short symbol for the prior of η.

Chapter 5

Generalized Spatial finite dimensional Dirichlet Priors.

In chapter 4, we have discussed a generalization of the usual Dirichlet Process, which can be of interest when $\mathcal{X} = \mathbb{R}^k$, $k \in \mathbb{N}$ or $\mathcal{X}$ is the space of realizations of a stochastic process $X(\omega) = \{X_t(\omega), t \in T\}$, for some index set T. In fact, we have introduced it with regard to the analysis of spatial data and the Spatial Dirichlet process (SDP) recently introduced by Gelfand *et al.* (2004). In reference to the Sethuraman's representation of the DP (and SDP), the key feature of our approach was the multivariate stick-breaking construction for the weights, which entailed the possibility to pick up different components across the vectors defining the support of the DP (SDP).

In chapter 1, we have seen that it is possible to define random probability measures through sum representations whose support is discrete and finite (see section 1.3) and the weights are Dirichlet distributed. This is the case of the Dirichlet-Multinomial process introduced by Muliere and Secchi (1995) and, more generally, of the finite dimensional Dirichlet prior (DP_k) defined by Ishwaran and Zarepour (2002b).

In this chapter, we extend those ideas in a way similar to what we have done in chapter 4. As before, this is done in the setting proper of spatial analysis, but the arguments can be easily adapted to several contexts. Therefore, we define a Generalized

Spatial finite dimensional Dirichlet prior ($GSDP_K$), and show how appropriate modeling of the weights allows the existence of different spatial effects in different subregions of the space in which it is defined.

In the first section, we begin with some preliminary considerations on the Dirichlet-Multinomial process developed by Muliere and Secchi (1995) and the finite dimensional Dirichlet prior introduced by Ishwaran and Zarepour (2002b). In particular, we discuss under what conditions a DP_K prior can be seen as an approximation of a DP. Then, we propose our generalization for the spatial setting and discuss its theoretical properties. We also show how to model the parameters of the process in a way feasible for coherent spatial analysis. Since this is still work in progress, conducted jointly with Alan E. Gelfand and Sonia Petrone, we are not yet able to show any relevant data analysis, but we provide the basic ideas for model fitting and inference in section 5.2.2.

5.1 Finite dimensional Dirichlet priors.

In this section, we discuss a class of priors which, under some conditions, can be seen as a finite sum approximation to the DP. These have already been discussed briefly in section 1.3. Here we provide a more thorough discussion that turns useful to give the basic references for the ideas that we develop in the following sections.

We start recalling that a Dirichlet-Multinomial process is defined as a class of priors that can be represented a.s. as a discrete random probability measure

$$P_K(A) = \sum_{i=1}^{K} p_i \delta_{\theta_i^*}(A), \tag{5.1.1}$$

for any $A \in \mathcal{B}(\mathbb{R})$, where K is integer and finite, θ_i^* are i.i.d. from some base distribution G_0, $i = 1, \ldots, K$, and the weights $(p_1, \ldots, p_K) \sim D(\alpha/K, \ldots, \alpha/K)$, for some positive α. We denote it by $P_K \sim DP_K(\alpha\, H)$, according to what already done in section 1.3.

A Dirichlet-Multinomial process is characterized as follows (see Muliere and Secchi (1995)). Given a probability distribution G_0, let $\theta_1^*, \ldots, \theta_K^*$ be an i.i.d. sample of size $K > 0$ from G_0. Set $G_{0,K}$ to be the empirical probability measure of $(\theta_1^*, \ldots, \theta_K^*)$, that is defined by

$$G_{0,K}(A) = \frac{1}{K} \sum_{i=1}^{K} \delta_{\theta_i^*}(A), \tag{5.1.2}$$

for any $A \in \mathcal{B}(\mathbb{R})$. Let $\mathcal{H}_K$ denote the distribution of $G_{0,K}$ induced by (5.1.2) on $(\mathbb{P}, \mathcal{B}(\mathbb{P}))$.

Now we can give the following definition. A random element $P_K \in \mathbb{P}$ is called a *Dirichlet-Multinomial process* with parameters (K, α, G_0) if it is a mixture of Dirichlet processes on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, with mixing distribution $\mathcal{H}_K$.
In symbols, $P_k | G_{0,K} \sim DP(\alpha G_{0,K})$, $G_{0,K} \sim \mathcal{H}_K$.

Then, for any given realization $G_{0,K}$ from $\mathcal{H}_K$, the random probability measure P_K can be a.s. characterized as

$$P_K(A) = \sum_{i=1}^{\infty} w_i \, \delta_{\theta_i^{**}}(A), \tag{5.1.3}$$

by the Sethuraman's representation, where $\theta_i^{**} \overset{i.i.d.}{\sim} G_{0,K}$ and the weights are obtained from the customary stick-breaking procedure, that is $w_1 = q_1$, $w_i = q_i \prod_{j=1}^{i-1}(1 - q_j)$, for $i > 1$, with $q_j \sim Beta(1, \alpha)$, $j = 1, \ldots, i$, $i = 1, \ldots$. However, since $G_{0,K}$ is discrete and has finite support, we can consider the random sums $p_i = \sum_{\{j:\theta_j^{**} = \theta_i^*\}} w_j$ and reduce the infinite sum (5.1.3) to the finite sum

$$P_K(A) = \sum_{i=1}^{K} p_i \delta_{\theta_i^*}(A), \tag{5.1.4}$$

where $\theta_i^* \overset{i.i.d.}{\sim} G_0$. By definition of the DP, for every finite measurable partition $B_1, \ldots, B_m$, $m = 1, 2, \ldots$,

$$(P_K(B_1), \ldots P_K(B_m)) | G_{0,K} \sim Dir(\alpha G_{0,K}(B_1), \ldots, \alpha G_{0,K}(B_m)),$$

where $Dir(\cdot)$ denotes the Dirichlet distribution. Notice that we can always consider a partition $B_1, \ldots, B_K$ such that each realized θ_k^* is contained in one and only one B_k, $k = 1, \ldots, K$. Then,

$$(P_K(B_1), \ldots, P_K(B_K))|G_{0,K} \stackrel{\mathcal{D}}{=} (p_1, \ldots, p_K)|G_{0,K} \sim Dir(\frac{\alpha}{K}, \ldots, \frac{\alpha}{K}).$$

Therefore, (5.1.4) is a finite-sum random probability measure with Dirichlet distributed random weights, which turns to be easy to work with both from a computational and theoretical point of view when used as a prior in Bayesian nonparametric problems. Moreover, Muliere and Secchi (1995) have proved that, as $K \to \infty$, this prior converges weakly to the DP with parameters α and G_0, of which therefore it can be considered a finite-sum approximation.

One could wonder if the same limit result is valid for an arbitrary specification of the parameters of the Dirichlet distribution on the weights $(p_1, \ldots, p_K)$. The answer is negative. In fact, particular care should be taken when defining the distribution of the weights in order that the prior actually converges to a DP in the limit.
Ishwaran and Zarepour (2002a) have explicitly defined the *finite dimensional Dirichlet priors* with parameters (K, α, G_0) as those priors which admit the a.s. discrete representation

$$P_K(A) = \sum_{i=1}^{K} p_i \delta_{\theta_i^*}(A), \tag{5.1.5}$$

for any $A \in \mathcal{B}(\mathbb{R})$, where K is integer and finite, θ_i^* are i.i.d. from some base distribution G_0 (not necessarily nonatomic) and $(p_1, \ldots, p_K) \sim Dir(\alpha_{1,K}, \ldots, \alpha_{K,K})$. As before, we denote (5.1.5) with $P_K \sim DP_K(\alpha, G_0)$.

According to the values assumed by the parameters $(\alpha_{1,K}, \ldots, \alpha_{K,K})$, it is possible to give some limit results for (5.1.5) as $K \to \infty$.

Theorem 5.1.1. *[Ishwaran and Zarepour (2002a)] Let* $P_K \sim DP_K(\alpha, G_0)$ *and*

$$P_K(g) = \int g(x) P_K(dx),$$

denote a random functional of P_K*, where* g *is a non-negative continuous function with compact support. Then,*

1(a) *If* $\alpha_{k,K} = \lambda_k$*, where* $\sum_{k=1}^{\infty} \lambda_k^2/k^2 < \infty$ *and* $\sum_{k=1}^{K} \lambda_k/K \to \lambda_0 > 0$*, then for any* g *as above,* $P_K(g) \xrightarrow{a.s.} \int g(x) G_0(dx)$.

1(b) *If* $\alpha_{k,K} = \lambda_K$*, where* $K\lambda_K \to \infty$*, then for any* g *as above,* $P_K(g) \xrightarrow{p} \int g(x) G_0(dx)$.

2(a) *If* $\alpha_{k,K} = \alpha/K$ *for some* $\alpha > 0$*, then for each real-valued measurable function* g *which is integrable with respect to* G_0*,* $P_K(g) \xrightarrow{\mathcal{D}} P(g)$*, where* P *is the usual Dirichlet process with finite measure* αG_0*, and* $P(g) = \int g(x) P(dx)$.

2(b) *If* $\sum_{k=1}^{K} \alpha_{k,K} \to \alpha > 0$ *and* $\max \alpha_{1,K}, \ldots, \alpha_{K,K} \to 0$ *as* $K \to \infty$*, then for any* g *as above* $P_K(g) \xrightarrow{\mathcal{D}} P(g)$.

3 *If* $\alpha_{k,K} = \lambda_K$*, where* $K\lambda_K \to 0$*, then* $P_K(g) \xrightarrow{\mathcal{D}} \int g(x) \delta_Z(dx)$*, where* Z *has distribution* G_0.

The previous theorem shows that a finite dimensional Dirichlet prior $DP_K(\alpha, G_0)$ does not necessarily converge to a DP, for all values of $\alpha_{i,K}$, $i = 1, \ldots, K$. In particular, notice that case (1b) comprises the common choice of a uniform Dirichlet prior with parameters $\alpha_{i,K} = 1$, $i = 1, \ldots, K$. Then, the theorem shows that in that case the limiting distribution of P_K is simply a parametric distribution. Therefore, models based on this choice are not truly nonparametric once considered in the limit for $K \to \infty$. On the other hand, notice that case (2a) is an extension of the already mentioned result in Muliere and Secchi (1995), since the theorem gives a stronger form of convergence than weak convergence, which only applies to bounded and continuous functions. We

will use these results later in next section when discussing limiting behavior of the generalization we propose.

We conclude this section with a final remark. Notice that (5.1.5) has been defined as a random probability measure on the real line. However, nothing precludes the possibility to extend the definition to a general measurable space $(\mathbb{X}, \mathcal{X})$, similarly to what has been described in sections 1.5.2 and 1.5.3 for the Dirichlet Process. For example, we could consider the n-dimensional Euclidean space, $\mathbb{R}^n$, together with its Borel sigma algebra $\mathcal{B}(\mathbb{R}^n)$. In particular, if we take $\mathbb{X}$ to be the space of all surfaces over a region D and θ_i^* to be realizations of a random field, i.e. $\theta_i^* = \{\theta_i^*(\mathrm{s}) : \mathrm{s} \in D\}$, $i = 1, \ldots, K$, we could define a finite dimensional Spatial Dirichlet prior in a way analogous to the SDP (Gelfand, Kottas and MacEachern (2004)). Let us denote with $SDP_K(\alpha, G_0)$ the random probability measure (5.1.5) arising under such specification. It is immediate to see that Theorem 5.1.1 still applies. Therefore, $P_K \sim SDP_K(\alpha, G_0)$ converges weakly to a random probability measure $P \sim SDP(\alpha, G_0)$ for an appropriate choice of the parameters of the Dirichlet distribution defining the weights $\{p_1, \ldots, p_K\}$.

5.2 Generalized Spatial finite dimensional Dirichlet priors.

At the end of previous section, we have discussed the possibility to define a finite dimensional Spatial Dirichlet prior (SDP_K) analogous to the SDP. A look to (5.1.5) is sufficient to see that the SDP_K exhibits many features typical of the former. In particular, let P_K be a given realization of (5.1.5). Then, any draw from P_K leads to sampling one of the θ_i^*, that is an entire surface over D. As we have discussed in chapter 4 for the SDP, the SDP_K doesn't allow sampling different surfaces at different

(possibly distant) locations.

In this section, we propose a generalization of the SDP_K which is flexible enough to achieve such behavior. As with the GSDP described in chapter 4, the generalization is obtained through appropriate modeling of the weights.

Let G_0 be a base random field, assumed to be stationary and gaussian and $G_0^{(n)} \equiv G_{0,s_1,\ldots,s_n}$ be its finite n-dimensional distribution at locations $(s_1,\ldots,s_n)$. Let $\theta_i^* = \{\theta_i^*(s),\, s \in D\}$ be indipendent realizations from G_0, i.e. surfaces over D, $i = 1,\ldots,K$. We say that G is a *Generalized Spatial finite dimensional Dirichlet prior* with parameters (K, α, G_0) and we denote it with $G \sim GSDP_K(\alpha, G_0)$ if it is a random probability measure defined on the space of surfaces over D such that for any set of locations $(s_1,\ldots,s_n) \in D$ and any collection of sets $(A_1,\ldots,A_n)$, $A_i \in \mathcal{B}(\mathbb{R})$, $i = 1,\ldots,n$,

$$G_{s_1,\ldots,s_n}(A_1 \times A_2 \times \ldots \times A_n) = \sum_{i_1=1}^{K} \ldots \sum_{i_n=1}^{K} p_{i_1,\ldots,i_n}(s_1,\ldots,s_n) \quad \delta_{\theta^*_{i_1}(s_1)}(A_1) \ldots \delta_{\theta^*_{i_n}(s_n)}(A_n), \tag{5.2.1}$$

where the vectors $(\theta_1^*(s_1),\ldots,\theta_n^*(s_n))^T \overset{i.i.d.}{\sim} G_0^{(n)}$, i_j is an abbreviation for $i(s_j)$, $j = 1, 2, \ldots, n$, and the weights $\{p_{i_1,\ldots,i_n}(s_1,\ldots,s_n),\, i_j = 1,\ldots,K, j = 1,\ldots n\}$ are independent of the θ_j^*'s, $j = 1,\ldots,K$ and are assumed to be jointly Dirichlet with parameters $\{\alpha_{i_1,\ldots,i_n}(s_1,\ldots,s_n),\, i_j = 1,\ldots,K, j = 1,\ldots n\}$, such that $\sum_{i_1,i_2,\ldots,i_n=1}^{K} \alpha_{i_1,\ldots,i_n}(s_1,\ldots,s_n) = \alpha$.

For notational simplicity and when there's no possibility of misunderstanding, we denote $p_{i_1,\ldots,i_n}(s_1,\ldots,s_n)$ and $\alpha_{i_1,\ldots,i_n}(s_1,\ldots,s_n)$ simply with $p_{i_1,\ldots,i_n}$ and $\alpha_{i_1,\ldots,i_n}$.

Analogously to Ferguson (1973), we can define the notion of a sample from G. In particular, we say that a vector $\mathbf{Y} = (Y(s_1),\ldots,Y(s_n))$ is a sample of size 1 from $G_{s_1,\ldots,s_n}$ if for any measurable sets $(A_1,\ldots,A_n)$, $A_i \in \mathcal{B}(\mathbb{R})$, $i = 1,\ldots,n$,

$$P(Y(s_1) \in A_1,\ldots,Y(s_n) \in A_n|G) = G_{s_1,\ldots,s_n}(A_1,\ldots,A_n) \quad \text{a.s.} \tag{5.2.2}$$

This definition guarantees the existence of a random field $Y(\mathbf{s})$ having finite dimensional

distributions $G_{s_1,\ldots,s_n}$ for any arbitrary set $(s_1,\ldots,s_n)$. In fact, the weights satisfy conditions of the type (4.1.2), that is for any $(s_1,\ldots,s_n)$, $n \in \mathbb{N}$ and $\forall k \in \{1,\ldots,n\}$,

$$p_{i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_n} = p_{i_1,\ldots,i_{k-1},\cdot,i_{k+1},\ldots,i_n} = \sum_{j=1}^{\infty} p_{i_1,\ldots,i_{k-1},j,i_{k+1},\ldots,i_n}. \tag{5.2.3}$$

From the properties of the Dirichlet distribution, it follows that (5.2.3) is satisfied as long as we require that the set of marginal weights $\{p_{i_1,\ldots,i_{l-1},i_{l+1},\ldots,i_n}\}$ for the $n-1$ locations $(s_1,\ldots,s_{l-1},s_{l+1},\ldots,s_n)$ has a Dirichlet distribution with parameters given by the collection $\{\alpha_{i_1,\ldots,i_{l-1},i_{l+1},\ldots,i_n}\}$, where each $\alpha_{i_1,\ldots,i_{l-1},i_{l+1},\ldots,i_n} = \sum_{j=1}^{K} \alpha_{i_1,\ldots,i_{l-1},j,i_{l+1},\ldots,i_n}$. Then,

$$P(Y(s_1) \in A_1,\ldots,Y(s_{n-1}) \in A_{n-1}, Y(s_n) \in \mathbb{R}|G) = P(Y(s_1) \in A_1,\ldots,Y(s_{n-1}) \in A_{n-1}|G) \quad \text{a.s.}$$

Therefore, G defines a family of mesures satisfying Kolmogorov Consistency conditions and there exists a random field $\{Y(\mathbf{s}),\ \mathbf{s} \in D\}$ with finite dimensional distributions $G_{s_1,\ldots,s_n}$, for any set of locations $(s_1,\ldots,s_n)$, $n = 1,2,\ldots$.

We outline the basic features of our modeling for the case of two locations $s, s' \in D$. Extension to the general n case is straightforward in most cases. However, explicit discussion will be provided whenever necessary. Therefore, let us consider

$$G_{s,s'}(A_1 \times A_2) = \sum_{i=1}^{K}\sum_{j=1}^{K} p_{i,j}\, \delta_{\theta_i^*(s)}(A_1)\, \delta_{\theta_j^*(s')}(A_2), \tag{5.2.4}$$

where the $p_{i,j}$'s have a Dirichlet distribution with parameters $\alpha_{i,j}$, $i,j = 1,\ldots,K$. From (5.2.1) and (5.2.4), the generalization of the usual finite dimensional Dirichlet prior is apparent. In fact, if $\alpha_{i,j} = 0$ for $i \neq j$, then $p_{i,j} = 0$ a.s. and consequently the only non zero elements among the weights are $(p_{1,1}, p_{2,2},\ldots,p_{K,K}) \sim Dir(\alpha_{i,i}, i = 1,\ldots,K)$. Since $p_i = \sum_{i,j=1}^{K} p_{i,j} = p_{i,i}$, $i = 1,\ldots,K$, $G_{s,s'}$ is a simple $DP_K(\alpha, G_0)$. We can call this the situation of *maximum concordance* among the surfaces chosen at the two locations. In all other cases ($p_{i,j} \neq 0$, for some i,j) definition (5.2.4) allows the possibility to

choose different surfaces at different locations. The other extreme situation is obtained when $p_{i,i} = 0$, for all $i = 1, \ldots, K$ (*minimum concordance*).

Now, let us consider the random marginal distribution in an arbitrary location s, that is

$$G_s(A) = \sum_{i=1}^{K} p_i \, \delta_{\theta_i^*(s)}(A),$$

where $A \in \mathcal{B}(\mathbb{R})$. Here, each $p_i = p_i(s) = p_{i+}$ is obtained marginalizing with respect to the other locations, and $(p_1, \ldots, p_K)$ have a Dirichlet distribution with parameters $\alpha_i, i = 1, \ldots, K$. Therefore, we get a model which marginally retains the structure of a finite dimensional Dirichlet prior (5.1.1). In particular, from Theorem 5.1.1, it follows that we can conveniently choose the α_i's so that marginally G can be seen as a finite sum approximation to a DP. For example, using condition (2a) in Theorem 5.1.1, we could require that marginally $\alpha_i = \alpha/K$, for some real constant α and for all $i = 1, \ldots, K$.

From definition (5.2.1), it's immediate to find the first and second moments of the random probability measure G. Again, we consider the case of $n = 2$ for notational simplicity. Then, the expected value of (5.2.1) is given by

$$\begin{aligned} E(G_{s_1,s_2}(A_1 \times A_2)) &= E\left(\sum_{i=1}^{K}\sum_{j=1}^{K} p_{i,j} \, \delta_{\theta_i^*(s)}(A_1) \, \delta_{\theta_j^*(s')}(A_2)\right) \\ &= E\left(\sum_{i=1}^{K} p_{i,i} \, \delta_{\theta_i^*(s)}(A_1) \, \delta_{\theta_i^*(s')}(A_2) + \right. \\ &\quad \left. + \sum_{\substack{i,j=1 \\ i \neq j}}^{K} p_{i,j} \, \delta_{\theta_i^*(s)}(A_1) \, \delta_{\theta_j^*(s')}(A_2)\right) \\ &= \gamma \, G_{0,s_1,s_2}(A_1 \times A_2) + (1-\gamma) \, G_{0,s_1}(A_1) \, G_{0,s_2}(A_2), \end{aligned} \tag{5.2.5}$$

where $\gamma = \sum_{i=1}^{K} E(p_{i,i}) = \sum_{i=1}^{K} a_{i,i}/\alpha$. Equation (5.2.5) has a nice interpretation in terms of deviation from a SDP_K. In fact, if $\gamma = 1$, $p_{i,j} = 0$ a.s. for $i \neq j$. Then,

$E(G_{s_1,s_2}(A_1 \times A_2)) = G_{0,s_1,s_2}(A_1 \times A_2)$ and the discussion below (5.2.4) shows that in this case $G \sim SDP_K(\alpha, G_0)$. On the other hand, as long as more mass is given to the disaligned couples $(\theta_i^*(s), \theta_j^*(s'))$, the process forgets about the couples coming from the same surfaces and therefore, as $\gamma \to 0$, G is centered around the product of the marginal distributions $G_{0,s_1}(A_1)$ and $G_{0,s_2}(A_2)$. For $\gamma = 0$, we are in the situation of minimum concordance among the surfaces chosen at s_1 and s_2. In fact, the resulting process is no more a simple SDP_K. We show later in section 5.2.1 that in this case (5.2.4) reduces to a mixture of DP_K, where the mixing distribution is still a DP_K.

Now, let us consider the association structure between two marginal realizations of the $GSDP_K$. For any two measurable sets $A, B \in \mathbb{R}$, we have

$$E(G_s(A) \times G_{s'}(B)) = E\left(\sum_{i=1}^{K} p_i(s)\, \delta_{\theta_i^*(s)}(A_1) \times \sum_{j=1}^{K} p_j(s')\, \delta_{\theta_j^*(s')}(A_2\right) \tag{5.2.6}$$

$$= (1-\hat{\gamma})\, G_{0,s}(A) \times G_{0,s'}(B) + \hat{\gamma}\, G_{0,s,s'}(A \cap B), \tag{5.2.7}$$

where $\hat{\gamma} = \sum_{i=1}^{K} E\left(p_i(s)\, p_i(s')\right)$. Since

$$\begin{aligned} E\left(p_i(s)\, p_i(s')\right) &= E\left(\sum_{m=1}^{K} p_{i,m}(s,s') \times \sum_{j=1}^{K} p_{j,i}(s,s')\right) \\ &= E\left(p_{i,i}^2(s,s')\right) + \sum_{\substack{m=1 \\ m \neq j}}^{K} \sum_{j=1}^{K} E\left(p_{i,m}(s,s') \times p_{m,i}(s,s')\right) \\ &= \frac{\alpha_{i+}\alpha_{+i} + \alpha_{i,i}}{\alpha(\alpha+1)}, \end{aligned}$$

then $\hat{\gamma} = \frac{1}{\alpha(\alpha+1)} \sum_{i=1}^{K} (\alpha_{i+}\alpha_{+i} + \alpha_{i,i})$ and the covariance between the two realizations is given by

$$Cov(G_s(A), G_{s'}(B)) = \hat{\gamma}\, \left[G_{0,s,s'}(A \cap B) - G_{0,s}(A) \times G_{0,s'}(B)\right]. \tag{5.2.8}$$

Notice that expression (5.2.8) is similar for a SDP_K or a SDP. The only difference is the value of $\hat{\gamma}$. In fact, we get the SDP_K case just imposing $\alpha_{i,j} = 0$ for all $i \neq j$.

Then, $\alpha_{+i} = \alpha_{i+} = \alpha_{i,i}$, and $\hat{\gamma} = \frac{1}{\alpha+1} + \frac{1}{\alpha(\alpha+1)} \sum_{i=1}^{K} \alpha_{i+}^2$, since $\sum_{i=1}^{K} \alpha_{i+} = \alpha$. Notice that the association structure between the marginal random probability measures is usually higher in the SDP_K than in the $GSDP_K$, since in general $\sum_{i=1}^{K} \alpha_{i,i} \leq \alpha$. That is in accordance to what we expected in virtue of definition (5.2.1). In fact, for fixed $i = 1, \ldots, K$, we can say that the $GSDP_K$ reassigns the mass p_i available for the couple $(\theta_i(s), \theta_i(s'))$ in the SDP_K across the independent couples $\{(\theta_i^*(s), \theta_j^*(s')), i \neq j\}$. In particular, notice that if $\alpha_{i,i} = \alpha/K$ as in the Dirichlet-Multinomial process, then $\hat{\gamma} = \frac{1}{\alpha+1} + \frac{\alpha}{K(\alpha+1)}$. Therefore, we can obtain the covariance between marginal SDP's letting $K \to \infty$. It follows that $\hat{\gamma} = \frac{1}{\alpha+1}$, which coincides with the value of $E(p_i^2)$ for a DP (see Theorem 4 in Ferguson (1973)). Therefore, we can conclude that the association structure between marginal random distributions is lower in the SDP than in the $GSDP_K$, which is in turn lower than in the SDP_K.

Turning back to samples from G, the expressions for their conditional and marginal moments are similar to those obtained in the previous chapter for the $GSDP$ (see equations (4.1.3), (4.1.5) and (4.1.6)). Therefore, conditionally on the realized distribution G, any process $Y(s)$ sampled from G has first and second moments given by

$$E(Y(s)|G) = \sum_{i=1}^{K} p_i(s)\, \theta_i^*(s) \tag{5.2.9}$$

$$Var(Y(s)|G) = \sum_{i=1}^{K} p_i(s)\, \theta_i^{*^2}(s) - \Big\{ \sum_{i=1}^{K} p_i(s)\, \theta_i^{*^2}(s) \Big\}^2, \tag{5.2.10}$$

and, for a pair of sites $s, s' \in D$,

$$Cov(Y(s), Y(s')|G) = \sum_{i,j=1}^{K} p_{i,j}(s,s')\, \theta_i^*(s)\, \theta_j^*(s) - \Big\{ \sum_{i=1}^{K} p_i(s)\, \theta_i^*(s) \Big\} \Big\{ \sum_{j=1}^{K} p_j(s')\, \theta_j^*(s') \Big\}. \tag{5.2.11}$$

Marginalizing over G and assuming that G_0 is a mean zero stationary gaussian field with finite variance σ^2 and correlation function $\rho_\phi(s_i - s_j)$, for some parameter ϕ, we

get $E(Y(s)) = 0$ and $Var(Y(s)) = \sigma^2$. Moreover,

$$\begin{aligned} Cov(Y(s), Y(s')) &= \sigma^2 \rho_\phi(s - s') \sum_{i=1}^{K} E(p_{i,i}(s, s')) \\ &= \sigma^2 \rho_\phi(s - s') \sum_{i=1}^{K} \frac{\alpha_{i,i}}{\alpha}. \end{aligned} \tag{5.2.12}$$

Therefore, the process $Y(s)$ is stationary depending on the way we model the parameters $\alpha_{i,i}$, $i = 1, \ldots, K$. For example, $Y(s)$ is stationary if $\sum_{i=1}^{K} \alpha_{i,i} = \alpha$, as it happens in the SDP_K case. That is not the only case in which we get stationarity. In fact, in order to get the same result, it is sufficient to hold the $\alpha_{i,i}$'s constant independently of the locations. However, unless $\sum_{i=1}^{K} \alpha_{i,i} = \alpha$, the covariance function of $Y(s)$ is lower than for processes sampled from a SDP_K or a SDP. Again, that is in accordance with the behavior implied by definition (5.2.1). In fact, in the minimum concordance case, $\alpha_{i,i} = 0$ for all $i = 1, \ldots, K$, and the covariance function of the process is null, that is $Y(s)$ and $Y(s')$ are independent.

5.2.1 Asymptotic behavior of the $GSDP_K$.

From Theorem 5.1.1, we know that for an appropriate choice of the parameters of the Dirichlet Distribution characterizing the weights in the sum representation (5.1.5), the SDP_K weakly converges to a SDP for $K \to \infty$. We have seen that the $GSDP_K$ can be considered as a generalization of the SDP_K, in which we assign positive probability to the event that we sample from independent surfaces in different locations. However, we don't know anything yet about the behavior of the $GSDP_K$ for $K \to \infty$.

In this section, we provide an interesting interpretation of our model. Based on this interpretation we obtain some limit results, characterizing the limit behavior of our generalization for some values of the parameters of the Dirichlet distributions on the weights.

First, we start considering a random distribution F, $F \sim SDP_K(\alpha, G_0)$. Let the random field $\{Y(s),\ s \in D\}$ be a sample from F. As before, we consider two locations $s, s' \in D$. Extending the discussion to the general n locations case only results in more complicate expressions, but it doesn't add much to the interpretation. Therefore, we consider $F_{s,s'}$, that is the random probability measure induced on (s, s') by F; for any measurable sets $A, B \in \mathcal{B}(\mathbb{R})$,

$$F_{s,s'}(A \times B) = \sum_{i=1}^{K} p_{i,i}\, \delta_{\theta_i^*(s)}(A)\, \delta_{\theta_i^*(s')}(B), \tag{5.2.13}$$

that is, $F_{s,s'} \sim DP_K(\alpha, G_{0,s,s'})$. Notice that if G_0 is assumed to be nonatomic, all the couples $(\theta_i^*(s), \theta_i^*(s'))$ are a.s. distinct.

Let $\mathbf{Y} = (Y(s), Y(s'))$ be a sample from $F_{s,s'}$, that is $\mathbf{Y}|F_{s,s'} \sim F_{s,s'}$. Then, we can consider the conditional probability of $Y(s)$ given $Y(s')$, which is given by

$$\begin{aligned} P(Y(s) \in A | Y(s') = y, F) &= \frac{P(Y(s) \in A, Y(s') = y|F)}{P(Y(s') = y|F)} \\ &= \sum_{i=1}^{K} p_{i|j}(s)\, \delta_{\theta_i^*(s)}(A), \end{aligned} \tag{5.2.14}$$

where $p_{i|j} = 1$ if $y = \theta_i^*(s')$ or null otherwise. Since $(\theta_i^*(s), \theta_i^*(s')) \overset{i.i.d.}{\sim} G_{0,s,s'}, i = 1, \ldots, K$, we can consider the random conditional distribution $F_{s|s'}(\cdot|y)$ induced from F by (5.2.14) and conclude that

$$F_{s|s'}(A|y) = \delta_{\zeta^*}(A), \tag{5.2.15}$$

hence degenerate on a point ζ^*, which is a sample from $G_{0,s|s'}(\cdot|y)$, where $G_{0,s|s'}(\cdot|y)$ denotes the distribution of $\theta^*(s)$ conditional on $\theta^*(s') = y$.

Moreover, we can restate (5.2.13) as

$$F_{s,s'}(A \times B) = \sum_{i=1}^{K} \hat{p}_i\, F_{s,s'}(A|\theta_i^*(s'))\, \delta_{\theta_i^*(s')}(B), \tag{5.2.16}$$

where the $\hat{p}_i = p_{i,i}$'s are the weights of the marginal distribution $G_{s'}$ and are equal to those in (5.2.13) because of the structure of the SDP_K. In other words, (5.2.13) has been restated as a mixture of random degenerate measures, whose mixing distribution is a realization from $F_{s'}$, with $F_{s'} \sim DP_K(\alpha G_{0,s'})$.

As a remark, notice that we could repeat similar arguments for a sample from a SDP. In fact, the only remarkable difference between the two is K finite in (5.2.13).

Now, we consider $G \sim GSDP_K(\alpha, G_0)$. Then, for any two locations $(s, s') \in D$, G reduces to $G_{s,s'}(\cdot)$, whose expression is given by (5.2.4). Let $(p_{1,1}, p_{2,2}, \ldots, p_{K,K})$ be the set of *matching* weights, i.e. specifying the probability to pick up the same surface at both sites, and $\eta_K = \sum_{i=1}^{K} p_{i,i}$. Then, we can define

$$\tilde{p}_{i,j} = \begin{cases} \frac{p_{i,i}}{\eta_K}, & i = j, i = 1, \ldots, K; \\ 0, & \text{otherwise}, \end{cases} \tag{5.2.17}$$

as the set of normalized matching weights, where the normalized constant is given by η_K. Analogously, we can set

$$\hat{p}_{i,j} = \begin{cases} \frac{p_{i,j}}{1-\eta_K}, & i, j = 1, \ldots, K, i \neq j; \\ 0, & \text{otherwise}, \end{cases} \tag{5.2.18}$$

in order to define the complemental set of normalized *disentangling* weights. Then, we can rewrite any $p_{i,j}$ in (5.2.4) as a linear combination of the matching and disentangling weights, that is

$$p_{i,j} = \eta_K \, \tilde{p}_{i,j} + (1 - \eta_K)\hat{p}_{i,j}. \tag{5.2.19}$$

Accordingly, we can reexpress (5.2.4) as

$$\begin{aligned} G_{s,s'}(A_1 \times A_2) = \eta_K \sum_{i=1}^{K} \tilde{p}_{i,i} \, \delta_{\theta_i^*(s)}(A_1) \, \delta_{\theta_i^*(s')}(A_2) + \\ + (1 - \eta_K) \sum_{i,j=1, i \neq j}^{K} \hat{p}_{i,j} \, \delta_{\theta_i^*(s)}(A_1) \, \delta_{\theta_j^*(s')}(A_2). \end{aligned} \tag{5.2.20}$$

Notice that the first sum in (5.2.20) is a realization from a $DP_K(\sum_{i=1}^K \alpha_{i,i}, G_{0,s,s'})$. In fact, by writing the variables in terms of their Gamma components, it is possible to prove that $(\tilde{p}_{1,1}, \tilde{p}_{2,2}, \ldots, \tilde{p}_{K,K}) \sim Dir(\alpha_{i,i}, i = 1, \ldots, K)$. The second term in (5.2.20) does not have a similar characterization. In fact, for each fixed $i = 1, \ldots, K$, each $\theta_i^*(s)$ is associated to $(K-1)$ $\theta_j^*(s')$'s. Hence, letting i vary and fixing j, it is possible to see that the same $\theta_j^*(s')$ appears in $(K-1)$ different couples $(\theta_i(s), \theta_j^*(s')), i = 1, \ldots, K$. The former consideration suggests we can proceed in the following way.

First notice that $\hat{p}_{i,j} = \hat{p}_{j|i}\,\hat{p}_{i+}$, where

$$\hat{p}_{i+} = \sum_{i=1}^K \hat{p}_{i,j} = \frac{1}{1-\eta_K} \sum_{j=1, j\neq i}^K p_{i,j},$$

and

$$\hat{p}_{j|i} = \frac{\hat{p}_{i,j}}{\hat{p}_{i+}} = \frac{p_{i,j}}{\sum_{j=1,j\neq i}^K p_{i,j}}, \quad j \neq i.$$

Then,

$$\begin{aligned} G_{s,s'}(A_1 \times A_2) &= \eta_K \sum_{i=1}^K \tilde{p}_{i,i}\, \delta_{\theta_i^*(s)}(A_1)\, \delta_{\theta_i^*(s')}(A_2) + \\ &+ (1-\eta_K) \sum_{i=1}^K \left(\sum_{j=1, j\neq i}^K \hat{p}_{j|i}\, \delta_{\theta_j^*(s')}(A_2) \right) \hat{p}_{i+}\, \delta_{\theta_i^*(s)}(A_1). \end{aligned} \tag{5.2.21}$$

It's immediate to prove that $(\hat{p}_{j|i}, j = 1, \ldots, K) \sim Dir(\alpha_{i,j}, j \neq i, j = 1, \ldots, K)$, and $(\hat{p}_{i+}, i = 1, \ldots, K) \sim Dir(\sum_{j=1,j\neq i}^K \alpha_{i,j}, i = 1, \ldots, K)$. Therefore, the inner sum between brackets in (5.2.21), that is $\hat{G}(i) := \sum_{j=1,j\neq i}^K \hat{p}_{j|i}\, \delta_{\theta_j^*(s')}(A_2)$, can be seen as a realization from a $DP_K(\sum_{j=1,j\neq i}^K \alpha_{i,j}, G_{0,s'})$. It follows that the second term in (5.2.20) can be rewritten as mixture of DP_K's where the mixing distribution is still a DP_K, with parameters $(K, \sum_{i=1}^K \sum_{j=1,j\neq i}^K \alpha_{i,j}, G_{0,s})$. Therefore, (5.2.4) is a linear combination (a mixture) of two components: the first is a simple DP_K with base measure the joint distribution $G_{0,s,s'}$, the second is a mixture of DP_K's with another DP_K as mixing.

The random weights of the mixture are $(\eta_K, 1-\eta_K)$ which are Beta distributed with parameters $\left(\sum_{i=1}^{K}\alpha_{i,i}, \sum_{i,j=1,i\neq j}^{K}\alpha_{i,j}\right)$.

Characterization (5.2.21) is useful for studying the behavior of G as $K \to \infty$. In fact, the main components of the measures are DP_K's. Therefore, all we need is to recall Theorem 5.1.1 and investigate what happens for appropriate choices of the parameters $\alpha_{i,j}$, $i,j = 1,...K$. First of all, consider the case of maximum concordance between the realizations of $\theta^*(s)$ and $\theta^*(s')$ (that is $a_{i,j} = 0$ for $i \neq j$). Then, $\eta_K = 1$ and we already observed that our model reduces to a $SDP_K(\alpha, G_{0,s,s'})$. Therefore, it converges to a SDP when we chose $\alpha_{i,i}$ appropriately according to Theorem (5.1.1).

Now, let us consider $\alpha_{i,j} = \alpha/K^2$, $i,j = 1,\ldots,K$. In other words we expect the same probability to pick up any combination of surfaces at (s,s'). It follows that $\sum_{i=1}^{K}\alpha_{i,i} = \alpha/K \to 0$ as $K \to \infty$. Correspondingly, $\eta_K = \sum_{i=1}^{K} p_{i,i} \to 0$ a.s. Therefore, $G_{s,s'}$ is a.s. a mixture of DP_K as in the second term of (5.2.21).

Then, consider the behavior of $\hat{G}(i)$ for $K \to \infty$. We have seen that $\hat{G}(i)$ is a realization from a $DP_K(\sum_{j=1,j\neq i}^{K}\alpha_{i,j}, G_{0,s'})$. However, notice that $\sum_{j=1,i\neq j}^{K}\alpha_{i,j} = \alpha/K \to 0$ as $K \to \infty$. In fact, $(\hat{p}_{j|i}, j = 1,\ldots,K) \sim Dir(\alpha_{i,j}, j \neq i, j = 1,\ldots,K)$, with $\alpha_{i,j} = \alpha/K^2 = \lambda_K$ and $K\lambda_K \to 0$ as $K \to \infty$. Therefore, we recall result (3) in Theorem 5.1.1 to show that $\hat{G}(i)$ converges to $\delta_{\theta^*(s)}$, where $\theta^*(s) \sim G_0^{(1)}(s)$.

Now consider the remaining mixing distribution. We have seen that this is a realization from a $DP_K(\sum_{i=1}^{K}\sum_{j=1,j\neq i}^{K}\alpha_{i,j}, G_{0,s})$. Since $\sum_{i=1}^{K}\sum_{j=1,i\neq j}^{K}\alpha_{i,j} = \alpha K(K-1)/K^2 \to \alpha$ and $\max(\alpha_{2,1},\ldots,\alpha_{K-1,K}) \to 0$ as $K \to \infty$, we can apply result (2b) in the same Theorem to prove that the second term in (5.2.21) as a whole and hence our process converges to a DP with scale parameter α and base measure the product of the marginals $G_0(s)G_0(s')$ (see also Kingman (1975)).

One could wonder if there exists indeed a choice of $\{a_{i,j}\}_{i,j=1}^{K}$ such that $G_{s,s'}$ converges to a linear combination of DP in the limit as $K \to \infty$. It is easy to see that it is

sufficient to choose $\alpha_{i,i} = \pi\frac{\alpha}{K}$, for $i = j$ and $\alpha_{i,j} = (1-\pi)\frac{\alpha}{K(K-1)}$ otherwise, for some $\pi \in (0,1)$, to get a process which converges to a mixture of two DP, one with base measure $G_0^{(2)}(s,s')$ and the other with base measure $G_0^{(1)}(s)\,G_0^{(1)}(s)$, and with random weights sampled from a Beta$(\alpha\pi, \alpha(1-\pi))$.

We conclude this section providing an expression analogous of (5.2.15) for a $GSDP_K$. In fact, if we set $p_{i|j} = p_{i,j}/p_{j+}$, where $p_{j+} = p_j = \sum_{i=1}^K p_{i,j}$, we obtain that

$$G_{s|s'}(A) = \sum_{i=1}^{K} p_{i|j}\,\delta_{\theta_i^*(s)}(A),$$

where $p_{i|j}, i = 1,\ldots,K \sim Dir(\alpha_{i,j}, i = 1,\ldots,K)$ and $\theta_i^*(s)$ are i.i.d. from $G_{0,s|s'}$. Therefore, $G_{s|s'}$ is a realization from a $DP_K(\sum_{i=1}^K \alpha_{i,j}, G_{0,s|s'})$. As before, we can interpret $G_{s|s'}$ as the random conditional distribution of $Y(s)|Y(s')$ where $\mathbf{Y} = (Y(s), Y(s'))$ is a sample from $G_{s,s'}$.

5.2.2 Assignment of the parameters of the weights.

In the previous section, we have often stressed how the modeling of the parameters of the process is of crucial importance, in particular to characterize its limiting behavior. In the following, we propose a modeling which is flexible enough and tractable for inferential purposes.

First, we choose the parameters so that marginally at an arbitrary site $s \in D$, the random probability distribution G reduces to a $DP_K(\alpha, G_{0,s})$ converging in the limit as $K \to \infty$ to a $DP(\alpha, G_{0,s})$. Therefore, from now on, we set $\alpha_i = \alpha/K$, and the $\alpha_{i,j} = \alpha\, a_{i,j} = \alpha\, E(p_{i,j})$, $i,j = 1,\ldots,K$ are defined so that $a_i = a_{i+} = \sum_{j=1}^K a_{i,j} = 1/K$.

Let $H(\cdot,\cdot;\tau)$ be a distribution function on $[0,1]^2$, with uniform marginal. In other words, let $(\mathcal{U}_1, \mathcal{U}_2)$ 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 any 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 the sets $Q_{i,j} = \left(\frac{i-1}{K}, \frac{i}{K}\right] \times \left(\frac{j-1}{K}, \frac{j}{K}\right]$, $i, j = 1, \ldots, K$.
Therefore, we can set the expected values $E(p_{i,j})$ to be given by the probability that $(\mathcal{U}_1, \mathcal{U}_2)$ belong to $Q_{i,j}$, that is

$$a_{i,j} = P_H\left(\mathcal{U}_1 \in \left(\frac{i-1}{K}, \frac{i}{K}\right], \mathcal{U}_2 \in \left(\frac{j-1}{K}, \frac{j}{K}\right]\right), \quad i, j = 1, \ldots, K. \tag{5.2.22}$$

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

The easiest choice for $H(\cdot, \cdot; \tau)$ is the uniform distribution over the unit square. This corresponds to choose a random probability measure G whose weights have a symmetric Dirichlet distribution, $\alpha_{i,j} = \alpha/K^2$. Notice that since $\sum_{i=1}^{K} a_{i,i} = 1/K$, $Y(s)$ is stationary in virtue of (5.2.12). Moreover, in section 5.2.1 we have seen that for this choice of the parameters G converges in the limit as $K \to \infty$ to a DP with scale parameter α and base measure the product of the marginals $G_0(s)G_0(s')$. In fact, under this choice of the parameters the probabilities $p_{i,j}$ in (5.2.1) have a distribution which is independent of the actual set of locations considered, but depends only on their number n (in general, in fact, $a_{i_1,\ldots,i_n} = 1/K^n$).

We can introduce spatial variability through a copula argument. In fact, let T be a spatial process on D. For example, T could be a gaussian process with correlation function $\rho_\phi(s, s')$ indexed by some parameter ϕ. Then, for any two locations s and s', $(T(s), T(s')) \sim N_2(\boldsymbol{\mu}, \sigma^2 H_2(\phi))$, where as usual $\boldsymbol{\mu}$ is a mean vector, σ^2 is the variance and $H_2(\phi)$ is the correlation matrix, such that $H_{i,j} = \rho_\phi(s, s')$ for $i, j = 1, 2$. Let Φ denote the bivariate normal distribution function, with marginal Φ_i, $i = 1, 2$. Then,

we can set $\mathcal{U}_1 = \Phi_1(T(s))$ and $\mathcal{U}_2 = \Phi_2(T(s'))$ and obtain

$$\begin{aligned} H_\Phi(u_1, u_2; \tau) &= P(\mathcal{U}_1 \leq u_1, \mathcal{U}_2 \leq u_2) \\ &= P(\Phi_1(T(s)) \leq u_1, \Phi_2(T(s')) \leq u_2) \\ &= P(T(s) \leq \Phi_1^{-1}(u_1), T(s') \leq \Phi_2^{-1}(u_2)) \\ &= \Phi\left(\Phi_1^{-1}(u_1), \Phi_2^{-1}(u_2)\right). \end{aligned} \tag{5.2.23}$$

Notice that we have used $H_\Phi(u_1, u_2; \tau)$ to indicate explicitly the dependence of the distribution of $(\mathcal{U}_1, \mathcal{U}_2)$ on the choice of Φ. It is worthy to notice that the gaussian assumption does not play any particular role in the previous arguments and is made only for exemplification purposes. In fact, what we actually need is just a continuous distribution function Φ. Then, for any $i, j = 1, \ldots, K$,

$$\begin{aligned} a_{i,j} &= P_{H_\Phi}\left(\mathcal{U}_1 \in \left(\frac{i-1}{K}, \frac{i}{K}\right], \mathcal{U}_2 \in \left(\frac{j-1}{K}, \frac{j}{K}\right]\right) \\ &= P_\Phi\left(T(s) \in \left(\Phi_1^{-1}\left(\frac{i-1}{K}\right), \Phi_1^{-1}\left(\frac{i}{K}\right)\right], T(s') \in \left(\Phi_2^{-1}\left(\frac{j-1}{K}\right), \Phi_2^{-1}\left(\frac{j}{K}\right)\right]\right). \end{aligned} \tag{5.2.24}$$

Recall from section 4.1.1 that a desirable property for the weights in a spatial setting is continuity, such that for all locations s in an ϵ-neighborhood of a point s', the probability of selecting two surfaces, $p_{i_1,i_2} = P(Y(s) = \theta^*_{i_1}(s), Y(s') = \theta^*_{i_2}(s'))$, is close to the marginal probability $p_{i_2} = P(Y(s) = \theta^*_{i_2}(s'))$ when $i_1 = i_2$, and to 0 otherwise. This property is immediately satisfied for weights defined through 5.2.24, as long as we assume $T(s)$ to be a.s. continuous. In fact, as $s \to s'$, $a_{i,j} \to 0$ for $i \neq j$, while $a_{i,i} \to a_i$. Then, it follows that each $p_{i,j}$ converges to 0 both in mean and a.s. for $i \neq j$. Since $p_i = p_{i+} = p_{i,i} + \sum_{i=1, i \neq j}^{K} p_{i,j}$, it follows also that $p_{i,i} \to p_i$ as $s \to s'$. Then, if the base process G_0 is also almost sure continuous, Proposition 4.1.1 applies and the distribution of $Y(s)$ is close to the distribution of $Y(s_0)$ for any s in an ε-neighborhood of s_0.

Extension to the n locations case is straightforward. It is enough to consider a distribution $H(\cdot)$ on the n-dimensional unit hypercube with uniform marginals. Spatial dependence can be introduced through the coupling of an n-dimensional multivariate normal, that is seen as the finite dimensional distribution of a random field on D.

The discussion at the end of section 5.2.1 can be extended to encompass all the specifications obtained by means of a coupling argument like the one described above, where the distributions $H(\cdot,\cdot;\tau)$ are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$. In fact, consider the random probability measure $G_{s,s'}$ defined by (5.2.21). Then, if we recall expression (5.2.22), it is immediate to see that

$$\sum_{i=1}^{K} a_{i,i} = P_{H_\Phi}\left((\mathcal{U}_1,\mathcal{U}_2) \in \bigcup_{i=1}^{K}\left(\frac{i-1}{K},\frac{i}{K}\right] \times \left(\frac{i-1}{K},\frac{i}{K}\right]\right)$$

is the probability of a set $A_K = \bigcup_{i=1}^{K}\left(\frac{i-1}{K},\frac{i}{K}\right] \times \left(\frac{i-1}{K},\frac{i}{K}\right]$ which tends to a set A of Lebesgue measure zero as $K \to \infty$. Therefore, $\sum_{i=1}^{K} a_{i,i} \to 0$ and $\eta_K \to 0$ as $K \to \infty$. Consider the second term in the mixture (5.2.21). With an argument similar to the one above, we can show that $\sum_{i=1}^{K} \alpha_{i,j} \to 0$ for fixed $j = 1,\ldots,K$, as $K \to \infty$. Then, it's easy to prove that the argument leading to result (3) in Theorem 3 in Ishwaran and Zarepour (2002a) can be extended to this situation. Therefore, we can show that $\hat{G}(i)$ (defined below (5.2.21)) with this assignment converges to $\delta_{\theta^*(s)}$, where $\theta^*(s) \sim G_{0,s}$. Then, turninig to $G_{s,s'}$ as a whole, since $\sum_{i,j=1,i\neq j} \alpha_{i,j} \to \alpha$ and $\max(a_{2,1},\ldots,a_{K-1,K}) \to 0$, we can apply again case (2b) of (5.1.1) to conclude that $G_{s,s'}$ converges to a DP with scale parameter α and base measure the product of the marginals $G_{0,s}\,G_{0,s'}$.

The above result could seem unappropriate in most cases. In fact, the base measure of the DP is the product of independent components regardless of how far s and s' are. However, it is worth noticing that this is a limit result. For any fixed K, arbitrary large K, $\eta_K \neq 0$ and therefore (5.2.21) is actually a mixture of two random probability

measures, the first having base measure the joint distribution $G_{0,s,s'}$, the second the product $G_{0,s}\,G_{0,s'}$.

5.2.3 Model fitting and Inference.

In this section, we address some basic issues about how to make inferences with models embedding a $GSDP_K$ specification. As recalled in the introduction, this is mostly work in progress. Therefore, we won't be exhaustive at all, but provide some reference which will be the subject of further research.

As before, we consider a hierarchical model as (2.2.2), in which at the second level of the hierarchy we substitute the nonparametric step described in the previous sections, that is we set

$$\begin{aligned}
\mathbf{Y}_t|\,\boldsymbol{\beta},\boldsymbol{\theta}_t,\tau^2 &\overset{ind.}{\sim} N(\mathbf{X}_t^T\boldsymbol{\beta}+\boldsymbol{\theta}_t,\,\tau^2\,I_n)\\
\boldsymbol{\theta}_t|\,G_{s_1,\ldots,s_n} &\overset{i.i.d.}{\sim} G_{s_1,\ldots,s_n}\\
G_{s_1,\ldots,s_n}|\,\alpha_{i_1,\ldots,i_n},\sigma^2,\phi &\sim SDP_K(\alpha G_{0,s_1,\ldots,s_n}),\; G_{s_1,\ldots,s_n}(\cdot|\sigma^2,\phi)=N_n(\cdot|\mathbf{0}_n,\sigma^2 H_n(\phi))\\
\boldsymbol{\beta},\tau^2,\,\alpha,\sigma^2,\phi &\sim p(\boldsymbol{\beta})\,p(\tau^2)\,p(\alpha)\,p(\sigma^2)\,p(\phi).
\end{aligned} \tag{5.2.25}$$

Of course, the posterior of the model is not available in closed form. However, we can obtain samples from such posterior through the use of a Gibbs sampler. The implementation of the Gibbs sampler exploits the copula argument that we have used to define the parameters of the Dirichlet distribution for the weights.

For exemplification purposes, we consider again the case with $n=2$. Then, the set of weights $p_{i,j}, i,j=1,\ldots,K$ has a Dirichlet distribution with parameters $\alpha_{i,j}, i,j=1,\ldots,K$, where $\alpha_{i,j}=\alpha\,a_{i,j}$. If we introduce the latent vector $(\mathcal{U}_1,\mathcal{U}_2)$, we can model the $a_{i,j}$ through a copula argument to take into account spatial variability, as showed in section 5.2.2. Then, $a_{i,j}=P_{H_\Phi}((\mathcal{U}_1,\mathcal{U}_2)\in Q_{i,j})$, for some distribution H_Φ, where the

set of blocks $\{Q_{i,j}, i,j = 1,\ldots,K\}$ defines a partition of the unit square.

Now, let (ξ_1,ξ_2) be a sample from a $DP(\alpha H_\Phi)$. In symbols, $(\xi_1,\xi_2)|P \sim P$ and $P \sim DP(\alpha H_{\Phi_0})$. Then, the set of probabilities

$$P\left((\xi_1,\xi_2) \in Q_{i,j}\right) = P(Q_{i,j}),$$

$i,j = 1,\ldots,K$, is such that

$$(P(Q_{1,1}),\ldots,P(Q_{K,K})) \sim D(\alpha H_\Phi(Q_{1,1}),\ldots,\alpha H_\Phi(Q_{K,K})),$$

that is the vector $(P(Q_{1,1}),\ldots,P(Q_{K,K})) \stackrel{D}{=} (p_{1,1},\ldots,p_{K,K})$.

Hence, we can set $P(Q_{i,j}) = E(I_{Q_{i,j}}|P) = p_{i,j}$ for all $i,j = 1,\ldots,K$. Therefore, for any two measurable sets A and B,

$$\begin{aligned} P(\theta^*(s) \in A, \theta^*(s') \in B) &= \sum_{i,j=1}^{K} p_{i,j}\ \delta_{\theta_i^*}(A)\delta_{\theta_j^*}(B) \\ &= \sum_{i,j=1}^{K} E(I_{Q_{i,j}}(\xi_1,\xi_2)|P)\ \delta_{\theta_i^*}(A)\delta_{\theta_j^*}(B) \qquad (5.2.26) \\ &= \sum_{i,j=1}^{K} I_{Q_{i,j}}(\xi_1,\xi_2)\,\delta_{\theta_i^*}(A)\delta_{\theta_j^*}(B). \end{aligned}$$

Equation (5.2.26) enables us to restate model (5.2.25) in a different way, which is easier to deal with for computing the full conditionals required for the Gibbs sampler. In fact, we can characterize the distribution of the spatial component by means of a deeper hierarchical structure, as

$$\theta^*(s),\theta^*(s')|\xi_1,\xi_2,H_\Phi,(\theta_i^*(s),\theta_i^*(s')),i = 1,\ldots,K \sim \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \delta_{\xi_1,\xi_2}(Q_{i,j})\ \delta_{\theta_i^*(s)}(\cdot)\,\delta_{\theta_j^*(s')}(\cdot), \qquad (5.2.27)$$

According to our modeling, the vector $(\theta_i^*(s),\theta_i^*(s')) \sim G_{0,s,s'}$, $i = 1,\ldots,K$ and is independent of (ξ_1,ξ_2). The former is a sample from a bivariate DP. Therefore, if we restate the model as (5.2.25), we can exploit the well-known Gibbs sampling

algorithms for the DP in order to obtain samples from the posterior of the parameters characterizing the model.

Conclusions

In this thesis, we have discussed some issues regarding the use of semi-parametric modeling for the analysis of spatial data in Bayesian inference.

The main tool is represented here by the Spatial Dirichlet process recently introduced by Gelfand, Kottas and MacEachern (2004). We have discussed the smoothness properties of the surfaces sampled from the SDP, and showed how to obtain interesting inference from this specification. Moreover, we have discussed two possible generalizations of the spatial Dirichlet process, which allow more flexibility and more precise inference on the sources of spatial variability available at each location.

All the models could be embedded in a spatio-temporal dynamic linear model, in order to avoid the use of independent replicates. That is, the spatial component at any time $t = 1, \ldots, T$ and at any location $\mathbf{s} \in D$, could be given by

$$\theta_t(\mathbf{s}) = \theta_{t-1}(\mathbf{s}) + \eta_t(\mathbf{s}),$$

where $\eta_t(\mathbf{s})$ are independent samples from a SDP or a Generalized SDP. We could study inference about finite differences and directional finite differences for this spatio-temporal model.

Notice that, although we develop our Generalized Dirichlet processes for analyzing spatial data, the theory is indeed quite general and can be used in several contexts. For example, our model can be used as an alternate method in most of the problems where mixtures of products of Dirichlet processes (see Cifarelli and Regazzini (1978))

and the Dependent Dirichlet Processes recently defined by MacEachern (2000) can be usefully employed. For example, we could show how to apply these models to an ANOVA context and compare the results with other approaches recently proposed in the literature (see, for example, De Iorio *et al.* (2004)). The nature of the problem will suggest which specification is more adeguate for the problem at hand.

The semi-parametric modeling that we have described by means of a SDP has been applied to the modeling of univariate point-referenced data. However, a nonparametric specification can be devised also for the analysis of multivariate data, for example in the Bayesian models of coregionalization (for a review of these models, see chapter 7.2 in Banerjee, Carlin and Gelfand (2004)). Moreover, we could extend this type of modeling to the case of variables available only as block-level summaries (areal unit data).

The issues here discussed will be the object of further research.

Bibliography

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

Adler, R. (1981) *The Geometry of Random Fields.* New York: Wiley.

Adler, R. and Taylor, J. (2003) Random fields and their geometry. Unpublished.

Anderson, T. (2003) *An Introduction to Multivariate Statistical Analysis.* Wiley, third edn.

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

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

Banerjee, S. and Gelfand, A. (2003) On smoothness properties of spatial processes. *Journal of Multivariate Analysis*, **84**, 85–100.

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

Blackwell, D. and MacQueen, J. (1973) Ferguson distributions via Pólya urn scheme. *The Annals of Statistics.*, **1**, 353–355.

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

Carota, C. and Parmigiani, G. (1997) Semiparametric regression for count data. *Tech. rep.*, Institute of Statistics and Decision Sciences, Duke University.

Cifarelli, D., Muliere, P. and Scarsini, M. (1981) Il modello lineare nell'approccio Bayesiano non parametrico. *Pub. Ist. Mat. "G. Castelnuovo"*.

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, n. 12**, 1–13.

Connor, R. J. and Mosimann, J. E. (1969) Concepts of independence for proportions with a generalization of the Dirichlet distribution. *Journal of the American Statistical Association*, **64**, 194–206.

Cramér, H. and Leadbetter, M. (1967) *Stationary and Related Stochastic Processes.* New York: Wiley.

Cressie, N. (1993) *Statistics for Spatial Data.* New York: Wiley, 2nd edn.

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. *Journal of the American Statistical Association*, **99**, 205–215.

Escobar, M. D. (1988) *Estimating the Means of Several Normal Populations by Nonparametric Estimation of the Distribution of the Means.* Ph.D. thesis, Yale University, Dept. of Statistics. Unpublished.

— (1994) Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association*, **89**, 268–277.

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

Ewens, W. (1988) *Mathematical and Statistical Problems in Evolution*, chap. Population genetics theory. University of Montreal Press, Montreal.

Fabius, J. (1964) Asymptotic behavior of Bayes estimates. ann. statist. *The Annals of Statiscs.*, **35**, 846–856.

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

— (1983) Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics* (eds. H. Rizvi and J. Rustagi), 287–302. New York: Academic Press.

Ferguson, T. S. and Klass, M. J. (1972) A representation of independent increment processes without gaussian components. *Ann. Math. Statist*, **43**, 1634–1643.

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

Gelfand, A. and Kottas, A. (2002) A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, **11**, 289–305.

Gelfand, A., Kottas, A. and MacEachern, S. (2004) Bayesian nonparametric spatial modeling with Dirichlet processes mixing. *Submitted to the Journal of the American Statistical Association.*

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

Ghosal, S., Ghosh, J. and Ramamoorthi, R. (1999) Posterior consistency of Dirichlet mixture in density estimation. *The Annals of Statistics*, **27**, 143–158.

Gosh, J. and Ramamoorthi, R. (2003) *Bayesian Non-Parametrics*. Springer.

Green, P. and Richardson, S. (2001) Modeling heterogeneity with and without the Dirichlet process. *Scandinavian Journal of Statiscs.*, **28**, 355–375.

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

Handcock, M. and Stein, M. (1993) A Bayesian analysis of kriging. *Technometrics*, **35**, 403–410.

Handcock, M. and Wallis, J. (1994) An approach to statistical spatio-temporal modeling of meteorological fields (with discussion). *Journal of the American Statistical Association*, **89**, 368–390.

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

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

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

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

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

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

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

Lo, A. (1986) Bayesian statistical inference for sampling a finite population. *The Annals of Statistics*, **14**, 1226–1233.

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

MacEachern, S. N. (1994) Estimating normal means with a conjugate style Dirichlet process prior. *Communications in Statistics - Simulations*, **23**, 727–741.

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

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

Matern, B. (1986) *Spatial Variation.* Berlin: Springer Verlag, 2nd edn.

Mira, A. and Petrone, S. (1996) Bayesian hierarchical nonparametric inference for changepoint problems. In *Bayesian Statiscs, vol. 5* (eds. J. Bernardo, J. Berger, A. Dawid and A. Smith), 693–703. Oxford University Press, Oxford.

Muliere, P. and Scarsini, M. (1984) Bayesian inference for change point problems. *Rivista di Statistica Applicata*, **17**, 93–102.

— (1985) Change point problems: a Bayesian nonparametric approach. *Applied Mathematics*, **30**, 397–402.

Muliere, P. and Secchi, P. (1995) A note on a proper Bayesian bootstrap. *Tech. rep.*, Università degli Sudi di Pavia, Dipartimento di Economia Politica e Metodi Quantitativi.

Muliere, P., Secchi, P. and Walker, S. (2004) Partially exchangeable processes indexed by the vertices of a k-tree constructed via reinforcement. *Tech. rep.*, Unpublished.

Muliere, P. and Tardella, L. (1998) Approximating distributions of random functionals of fergusonDirichlet priors. *Canadian Journal of Statistics*, **26**, 283–297.

Neal, R. M. (2000) Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, **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. (2004) Non-gaussian Bayesian geostatistical modeling. *Warwick Statistics Research report 426*, University of Warwick.

Patil, G. P. and Taillie, C. (1977) Diversity as a concept and its implications for random communities. *Bull. Int. Stat. Inst.*, **XLVII**, 497–515.

Perman, M., Pitman, J. and Yor, M. (1992) Size-biased sampling of poisson point processes and excursions. *Prob. Theory and Related Fields*, **92**, 21–39.

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

Pitman, J. (1996) Random discrete distributions invariant under permutation. *Advances in Applied Probability*, **28**, 525539.

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

Regazzini, E. (1996) Impostazione non parametrica di problemi di inferenza statistica bayesiana. *Tech. rep.*, CNR-IMATI Milano.

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

Schmidt, A. and O'Hagan, A. (2003) Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society: Series 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. In *Proceeding of the Third Purdue Symposium on Statistical Decision Theory and Related Topics* (eds. S. Gupta and J. Berger), 305–315. Academic Press, New York.

Stein, M. (1999) *Interpolation of Spatial Data - Some theory of Kriging*. Springer.

Teh, Y., Jordan, M., Beal, M. and Blei, D. (2004) Hierarchical Dirichlet processes. *Tech. rep.*, Department of Statistics. University of California, Berkeley.

Walker, S. and Damien, P. (1998) *Practical Nonparametric and Semiparametric Bayesian Statistics*, chap. Sampling Methods for Bayesian Nonparametric Inference Involving Stochastic Processes, 243–254. New York: Springer-Verlag.

Walker, S. and Muliere, P. (2003) A bivariate Dirichlet process. *Statistics and probability letters*, **64**, 1–7.

Aknowledgments

Stranger, if you passing meet me and desire to speak to me, why should you not speak to me? And why should I not speak to you?

(To you, by Walt Whitman, *Leaves of Grass*, 1891)

Friends give you the support you need. They understand why you are in a bad mood when nothing seems to work, and then make you smile. Therefore, I thank all my friends. In particular, I thank all the people I met at ISDS at Duke. Because of them, I had a great time over there, and I will always value their friendship. A big hug to all the other friends I met in Durham, in particular, Rohan, Joe, Liping, Monica, Maria Pia, and Fulvio. I'm sure I forgot someone, but oh well, they are my friends and they know how bad I am!

My Duke deskmates Karim and Ananda, and Rebecca and Antonio in Italy deserve special mentioning. I really appreciated their forgiveness for ignoring all the uncomfortable noises and complaints I made while studying.

I'd also like to thank all the people that guided me through my learning. There were many, and I should say that sometimes I had been a difficult student. Namely, I'd like to thank Pietro Muliere, Piero Veronese, Sonia Petrone, and Alan Gelfand. Their contribution to this work has gone far beyond the "Math and Stat". They shared their humanity and so much more. However, I don't want to seem too romantic, so I will stop at that. Thank you.

Last, but not least, I thank my family: mum, dad, my sister Evelina, my brother in law Massimo, and my nieces Jessica and Debora. Well, what more can I say than

Grazie!

Lo logramos!!