SPACE-TIME MODELLING USING BAYESIAN NONPARAMETRIC AND DIFFERENTIAL EQUATION APPROACHES

by

Jun Duan

Institute of Statistics and Decision Sciences
Duke University

Date: _____

Approved:

_____
Alan E. Gelfand, Supervisor

_____
David L. Banks

_____
Michael L. Lavine

_____
C. F. Sirmans

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Institute of Statistics and Decision Sciences
in the Graduate School of
Duke University

2006

ABSTRACT

(STATISTICS)

SPACE-TIME MODELLING USING BAYESIAN NONPARAMETRIC AND DIFFERENTIAL EQUATION APPROACHES

by

Jun Duan

Institute of Statistics and Decision Sciences
Duke University

Date: _____

Approved:

_____
Alan E. Gelfand, Supervisor

_____
David L. Banks

_____
Michael L. Lavine

_____
C. F. Sirmans

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the
Institute of Statistics and Decision Sciences in the Graduate School of
Duke University

2006

Abstract

The present thesis addresses three important issues in modelling spatio-temporal data: (i) develop a flexible nonparametric Bayesian methodology for spatial random effect models; (ii) extend the current Bayesian nonparametric approach to model discrete spatial data; and (iii) construct a spatio-temporal point process that incorporates established scientific models into a Bayesian hierarchical model.

The spatial Dirichlet process(SDP) is the first attempt to introduce a nonparametric model for a neither Gaussian nor stationary spatial process. The SDP arises as a probability weighted collection of random surfaces. This can be unattractive for modelling, hence inferential purposes since it insists that a process realization is one of these surfaces. In Chapter 2, we introduce a generalized spatial Dirichlet process(GSDP) model for the spatial effects that allows different surface selection at different sites. Moreover, we can specify the model to preserve the property that the marginal distribution of the effect at each site still comes from a Dirichlet process. The development is offered constructively, providing a multivariate extension of the stick-breaking representation of the weights. We then introduce mixing using this generalized spatial Dirichlet process (GSDP). We illustrate the fitting of this novel model with a simulated data set and demonstrate how to embed the GSDP within a dynamic linear model.

In Chapter 3, we extend the SDP to a generalized linear model(GLM) setting by proposing a Bayesian nonparametric spatial approach to analyze disease mapping data. We develop a hierarchical specification using random effects modelled with a spatial Dirichlet process prior. We introduce a dynamic formulation for the spatial random effects to apply the model to spatio-temporal settings.

Chapter 4 introduces a novel structured model for spatio-temporal point pro-

cesses. We formulate a dynamic Cox process model where the evolution of latent intensity is governed by deterministic and stochastic differential equations describing the population growth mechanisms. We construct a Bayesian hierarchical model based on this point process and propose a process convolution approximation for statistical inference. We address the Bayesian estimation and space-time prediction issues and illustrate with simulated and real house construction data examples.

Acknowledgements

I would like to take this opportunity to express my heartfelt gratitude to a number of mentors and colleagues who have guided and helped me over the past five and a half years as a PhD student in statistics. My primary thank goes to my advisor Dr. Alan E. Gelfand, for his inspirational advising and encouragement throughout my research experience. Dr. Gelfand has always encouraged his student to be an independent and original thinker, instead of an assistant or implementor for his professor. His thought-provoking pedagogy teaches me how to discover and formulate unsolved problems, and develop scientifically rigorous and practical solutions.

I would also like to thank Dr.Merlise Clyde and Dr. Michael Lavine for their instructions during my early career in statistical research. They help me build a firm foundation for my later work. Dr. Carl F. Mela at Fuqua School of Business and I had many delightful discussions on empirical research in business. He teaches me to think intuitively and be tenacious in difficulties. I would also like to thank Dr. David Banks and Dr. C. F. Sirmans for providing me data and comments as my supervisory committee members.

Special thanks are reserved for my parents and late grandmother for their sacrifice, patience and support. My friends at Duke University have given me numerous helps. Among them I am especially grateful to Dr. Jimmy Teng for his great friendship.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Spatial data are observations associated with a set of geographical locations in a certain domain. For example, (i) in environmental science, we may observe pollutant at fixed stations; (ii) in public health, researchers are interested in geographical dispersion of cancer rates in different states and counties; (iii) in real estate market, people are interested in how house prices are related to locations. Spatial data can be modelled as realizations of vector valued random fields at a set of fixed locations. The spatial correlation structure is introduced through the form of spatial random effects, where a term capturing residual spatial correlation is explicitly introduced. In this thesis, we will propose a class of very flexible nonparametric models for the spatial random effects.

There are cases where the spatial domain, the set of geographical locations and the moments of observations are also random. These constitute a space-time *point pattern*, which arise in many different settings, e.g., (i) ecology where we might seek the evolution of the range of a species over time, (ii) disease incidence examining say the pattern of cancer cases over time, (iii) astronomy where the goal is the ascertainment of newly discovered stars, and (iv) urban development explained using say the pattern of single family homes constructed over time. The random locations

and moments of these random events are customarily modelled as a spatio-temporal point process with an inhomogeneous intensity surface. The classical Cox model views the intensity surface as a realization of another spatial random process. The space-time point process, in the Cox model framework, can be modelled with the dynamics of the intensity surface. In this thesis, we introduce a class of stochastic differential equations to model the dynamics of the evolving intensity surface.

1.1 Bayesian Nonparametric Spatial Statistics

Spatial random effects are traditionally modelled as a mean-zero stationary Gaussian process (GP). The stationarity or the Gaussian assumption is inappropriate in many cases. Flexible and computationally tractable modelling to relax the stationarity assumption includes the spatially varying kernel approach of Higdon et al. (1999) and the local stationarity approach of Fuentes and Smith (2001) but both are still within the setting of GP's. The fundamental paper of Sampson and Guttorp (1992) introduces a nonparametric specification for the covariance function, as does followup work by Damian et al. (2001) and Schmidt and O'Hagan (2003) but all still employ a GP in the likelihood.

The Gaussian assumption is criticized when the spatial variability is attributable to more than one latent processes so that, for example, a mixture of Gaussian processes would be more appropriate. See Brown et al. (2003) for a recent example or Palacios and Steel (2004) for the development of a class of models that can accommodate heavy tail behaviors. Recently, Gelfand et al. (2005) proposed a spatial Dirichlet process (SDP) mixture model which adopts a stationary and Gaussian base measure. However, the resulting random stochastic process is nonstationary and its joint finite dimensional distributions are not normal. The use of the SDP specification to model the distribution of the spatial component in a spatial random effect model

leads to a fully Bayesian semiparametric approach that, for fitting purposes, relies on well-known results and algorithms developed for Dirichlet process (DP) mixing. See, among others, Escobar and West (1995) and MacEachern and Müller (1998). Because our generalized spatial Dirichlet process models are a generalization of the SDP, we will first briefly review the SDP models in the following section.

1.1.1 Spatial Dirichlet Process Models

Denote the stochastic process by $\{Y(s) : s \in D\}$, $D \subseteq R^d$. Within this domain D, let $s^{(n)} = (s_1, ..., s_n)$ be a specific set of distinct locations at which the observations are collected. Assume that we have replicate observations available at each location and therefore that the full data consist of the collection of vectors $Y_t = \{Y_t(s_1), ..., Y_t(s_n)\}^T$, $t = 1, ..., T$. In fact, imbalance or missingness can be accommodated in $Y_t(s_i)$ through customary latent variable methods.

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

To model $Y_D \equiv \{Y(s) : s \in D\}$, following Gelfand et al. (2005), one can conceptually extend θ_l^* to a realization of a random field by replacing it with $\theta_{l,D}^* = \{\theta_l^*(s) : s \in D\}$. For instance, G_0 can be a stationary GP, from which each realization $\theta_{l,D}^*$ is a surface over D. The resulting random process or distribution, G,

for Y_D is denoted by $\sum_{l=1}^{\infty} p_l \delta_{\theta^*_{l,D}}$ and the construction is referred to as a spatial Dirichlet process (SDP) model. The interpretation is that for the $s^{(n)}$ above, G induces a random probability measure $G^{(s^{(n)})}$ on the space of distribution functions for $\{Y(s_1), ..., Y(s_n)\}$. To simplify our notation, we will use $G^{(n)}$ instead of $G^{(s^{(n)})}$ in as follows. Thus, we have that $G^{(n)} \sim DP(\nu G_0^{(n)})$, where $G_0^{(n)} \triangleq G_0^{(s^{(n)})}$ is the n-variate distribution for $\{Y(s_1), ..., Y(s_n)\}$ induced by G_0. E.g., $G_0^{(n)}$ is an n-variate normal distribution if G_0 is taken to be a GP.

Gelfand et al. (2005) notice a connection between the spatial DP above and the concept of a dependent Dirichlet process (DDP) developed by MacEachern (2000). The DDP provides a formal framework that describes a stochastic process of random distributions. These distributions are dependent, such that, at each index value, the distribution is a univariate DP. In the above setting, G induces a random distribution $G(Y(s))$ for each s, hence the set $\mathcal{G}_D \equiv \{G(Y(s)) : s \in D\}$ which, under sufficient conditions (MacEachern 2000) will be a DDP.

For a stationary G_0 (i.e., $cov\{\theta^*_l(s_i), \theta^*_l(s_j)\}$ depends upon s_i and s_j only through $s_i - s_j$), the choice of the covariance function determines how smooth the process realizations are. Kent (1989), for instance, shows that, if the covariance function admits a second order Taylor-series expansion with a remainder that goes to 0 at the rate of $2 + \delta$ for some $\delta > 0$ then $\theta^*(s_i) - \theta^*(s_j) \to 0$, almost surely, as $||s_i - s_j|| \to 0$. But then, in the representation of G as $\sum p_l \delta_{\theta^*_{l,D}}$, the continuity of $\theta^*_{l,D}$ implies that the random marginal distributions $G(Y(s_i))$ and $G(Y(s_j))$, are such that the difference between them tends to 0 almost surely, as $||s_i - s_j|| \to 0$. This continuity property implies that we can learn about $G(Y(s))$ more from data at neighboring locations than from data at locations further away, as in usual spatial prediction.

For G arising from G_0 and ν, note that given G, $E\{Y(s) \mid G\} = \sum p_l \theta^*_l(s)$ and

$var\{Y(s) \mid G\} = \sum p_l \theta_l^{*2}(s) - \{\sum p_l \theta_l^*(s)\}^2$. Moreover for a pair of sites s_i and s_j,

$$cov\{Y(s_i), Y(s_j) \mid G\} = \sum p_l \theta_l^*(s_i)\, \theta_l^*(s_j) - \left\{\sum p_l \theta_l^*(s_i)\right\}\left\{\sum p_l \theta_l^*(s_j)\right\}. \quad (1.1)$$

Hence, the random process G has heterogeneous variance and is nonstationary. If G_0 is a mean zero stationary GP with variance σ^2 and correlation function $\rho_\phi(s_i - s_j)$, where the (possibly vector valued) parameter ϕ specifies $\rho_\phi(\cdot)$, then, marginalizing over G, $E\{Y(s)\}=0$, $var\{Y(s)\}=\sigma^2$ and $cov\{Y(s_i), Y(s_j)\}=\sigma^2\rho_\phi(s_i - s_j)$. That is, G is centered around a stationary process with constant variance but it has non-constant variance and is nonstationary. Also, with almost surely continuous process realizations, (1.1) makes it clear that the SDP is mean square continuous. That is, given G, $\lim_{||s-s'||\to 0} E[\{Y(s) - Y(s')\}^2 |G] = 0$.

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

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

By *differentiating* with respect to Y_D, we obtain the density,

$$f\left(Y_D \mid G, \tau^2\right) = \int k\left(Y_D - \theta_D \mid \tau^2\right) G\left(d\theta_D\right). \quad (1.2)$$

Here $\mathcal{K}$ and k denote the joint distribution function and density function, respectively, of the pure error process over D. k might denote a $N(0, \tau^2)$ or $t_r(0, \tau^2)$ density. Hence for any s, we have

$$f\left(Y(s) \mid G, \tau^2\right) = \int k\left(Y(s) - \theta(s) \mid \tau^2\right) G\left(d\theta(s)\right).$$

In other words, by introducing the normal mixing kernel, we decompose $Y(s)$ into the sum of $\theta(s) + \epsilon(s)$, where $\theta(s)$ arises from the above spatial DP prior model and $\epsilon(s)$ is white noise with $N(0, \tau^2)$ distribution: similar to the customary partitioning the spatially correlated residual into a spatial random effect and a nugget component. The process model is created by convolving distributions rather than convolving process variables as in Higdon et al. (1999) or Fuentes and Smith (2001).

For the finite set of locations $s^{(n)} = (s_1, ..., s_n)$, (1.2) implies that the joint density for $Y = \{Y(s_1), ..., Y(s_n)\}^T$, given $G^{(n)}$ (where $G^{(n)} \sim DP(\nu G_0^{(n)})$) and τ^2, is

$$f\left(Y \mid G^{(n)}, \tau^2\right) = \int N_n\left(Y \mid \theta, \tau^2 I_n\right) G^{(n)}\left(d\theta\right), \tag{1.3}$$

where, to simplify the notations, $\theta \triangleq \theta^{(s^{(n)})} = \{\theta(s_1), ..., \theta(s_n)\}^T$ and $N_n(\cdot \mid \mu, \Sigma)$ denotes the n-variate normal density/distribution (depending on the context) with mean vector μ and covariance matrix Σ. Again, the almost sure representation of $G^{(n)}$ as $\sum p_l \delta_{\theta_l^*}$, where θ_l^* is the vector $\{\theta_l^*(s_1), ..., \theta_l^*(s_n)\}^T$, yields that $f(Y \mid G^{(n)}, \tau^2)$ is almost surely of the form $\sum_{l=1}^{\infty} p_l N_n(Y \mid \theta_l^*, \tau^2 I_n)$, i.e. a countable location mixture of normals. In fact, assuming the existence of expectations given $G^{(n)}$ and τ^2, one can obtain that $E(Y \mid G^{(n)}, \tau^2) = \sum p_l \theta_l^*$ and the covariance matrix $\Sigma_{Y|G^{(n)},\tau^2} = \tau^2 I_n + \Sigma_\theta^{(s^{(n)})}$, where $(\Sigma_\theta^{(s^{(n)})})_{i,j} = \mathrm{cov}\{\theta(s_i), \theta(s_j) \mid G^{(n)}\}$ is the covariance specified in (1.1).

A regression term, $X^T\beta$, could typically be added to the kernel of the mixture model in (1.3) leading to

$$f\left(Y \mid G^{(n)}, \beta, \tau^2\right) = \int N_n\left(Y \mid X^T\beta + \theta, \tau^2 I_n\right) G^{(n)}\left(d\theta\right). \tag{1.4}$$

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

Consider the data $Y_t = \{Y_t(s_1), ..., Y_t(s_n)\}^T$ with associated X_t, $t = 1, ..., T$. Given X_t, the Y_t are assumed independent from $f(Y_t \mid G^{(n)}, \beta, \tau^2)$ as in (1.4). A DP prior

is placed on $G^{(n)}$, i.e., $G^{(n)} \sim DP(\nu G_0^{(n)})$ (induced by the spatial DP prior for G in (1.2)), with $G_0^{(n)}$ being a multivariate normal with mean zero and covariance matrix $\sigma^2 H_n(\phi)$. The full Bayesian model is completed by placing (independent) priors on β, τ^2, ν, σ^2 and ϕ. Associating with each Y_t a $\theta_t = \{\theta_t(s_1), ..., \theta_t(s_n)\}^T$ where the θ_t, t = 1,...,T are independent realizations from $G^{(n)}$, the following semiparametric hierarchical model is established

$$\begin{aligned}
Y_t \mid \theta_t, \beta, \tau^2 &\sim N_n(Y_t \mid X_t^T\beta + \theta_t, \tau^2 I_n), t = 1, ..., T \\
\theta_t \mid G^{(n)} &\sim G^{(n)}, \quad t = 1, ..., T \\
G^{(n)} \mid \nu, \sigma^2, \phi &\sim DP(\nu G_0^{(n)}); \; G_0^{(n)}(\cdot \mid \sigma^2, \phi) = N_n(\cdot \mid 0_n, \sigma^2 H_n(\phi)) \\
\beta, \tau^2 &\sim N_p(\beta \mid \beta_0, \Sigma_\beta) \times \text{IGamma}(\tau^2 \mid a_\tau, b_\tau) \\
\nu, \sigma^2, \phi &\sim \text{Gamma}(\nu \mid a_\nu, b_\nu) \times \text{IGamma}(\sigma^2 \mid a_\sigma, b_\sigma) \times [\phi],
\end{aligned} \tag{1.5}$$

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

The hierarchical nature of this modelling framework enables extensions by replacing the Gaussian distribution (the kernel for the DP mixture) in the first hierarchy with other distributions, such as $t_r(0, \tau^2)$ aforementioned. If $Y_t(s)$ happens to be a discrete random variable, its distribution may belong to any of the exponential-dispersion family, hence we can formulate a semiparametric spatial generalized linear model. This specification extends the work in Diggle et al. (1998) where a stationary GP was used for the spatial random effects (see also, e.g., Heagerty and Lele 1998, Diggle et al. 1998, and Christensen and Waagepetersen 2002). In this spirit, we will apply the spatial Dirichlet process in the context of GLM to model discrete disease incidence data in Chapter 3.

1.1.2 Generalized Spatial Dirichlet Process Models

We give a brief introduction to the research object of the generalized spatial Dirichlet process in this section. The SDP is essentially a Dirichlet process defined on a space of surfaces, with probability one, its realizations being discrete probability measures with countable support (Ferguson 1973; Sethuraman 1994). Mixing against a Gaussian kernel yields an error specification that can be characterized as a countable location mixture of normals. The SDP insists that $Y_t = \{Y_t(s_1), ..., Y_t(s_n)\}^T$ at any t are sampled on only one realization of the random surface. We introduce a random distribution for the spatial effects that allows different surface selection at different sites. Moreover, we can specify the model to preserve the property that the marginal distribution of the effect at each site still comes from a Dirichlet process. This generalization of the SDP is done constructively by developing a multivariate extension of the stick-breaking weights that characterize the usual Dirichlet process (Sethuraman 1994). A new class of random probability measures for random vectors and processes arise thereof, which is referred as the generalized spatial Dirichlet process models (GSDP), including the customary Dirichlet process specification as a special case. Other extensions, also motivated by the stick-breaking representation are described in Hjort (2000) and Ishwaran and James (2001).

By relaxing the restriction that two locations in spatial sample are on the same random surface, we introduce a random distribution for the spatial effects such that surface selection can vary from location to location and the joint selection of surfaces for the n locations can vary with the choice of locations. Moreover, we can still preserve the property that the marginal distribution at each location comes from a usual univariate Dirichlet Process. This is achieved constructively by defining a new multivariate stick-breaking prior in which spatial dependence structure is also introduced in the modelling of the weights.

Accordingly, we start by considering a base random field G_0, which, for convenience, we assume to be stationary and Gaussian, and define $\theta_l^* = \{\theta_l^*(s), s \in D\}$ as a realization from G_0, i.e., a sample surface over D. Then, we define a random probability measure G on the space of surfaces over D, whose finite dimensional distributions almost surely have the following representation: for any set of locations $s^{(n)}=(s_1, \ldots, s_n) \in D$, and any collection of sets $\{A_1, \ldots, A_n\}$ in $\mathcal{B}(\mathbb{R})$,

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

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

The focus of our development is the construction of the weight $p_{i_1,\ldots,i_n}$, which has to satisfy two conditions. First the weights need to satisfy the Kolmogorov consistency condition in order that we can properly define a random process for $Y(\cdot)$. Specifically, we need that for any set of locations $(s_1, \ldots, s_n)$, $n \in \mathbb{N}$ and for all $k \in \{1, \ldots, n\}$,

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

In addition, for the purpose of spatial modelling, we insist that the weights must satisfy a continuity property: we desire the random laws associated with locations s_1 and s_2 near to each other to be similar. That is, for locations s and s_0, as $s \to s_0$, $p_{i_1,i_2} = pr\{Y(s) = \theta^*_{i_1}(s), Y(s_0) = \theta^*_{i_2}(s_0)\}$, tends to the marginal probability $p_{i_2} = pr\{Y(s_0) = \theta^*_{i_2}(s_0)\}$ when $i_1 = i_2$, and to 0 otherwise. If we also assume the random field G_0 to be almost surely continuous (a univariate spatial process $\theta(s)$, $s \in D$ is said to be almost surely continuous at a point s_0 if $\theta(s) \to \theta(s_0)$

with probability one as $||s - s_0|| \to 0$), we want establish the almost sure continuity property. We shall show in Chapter 2 that the consistency and continuity properties are satisfied by our multivariate stick-breaking construction of $p_{i_1,\ldots,i_n}$ using latent Gaussian random processes.

The goal of our formulation is to establish a setting where very few random surfaces are needed to achieve an adequate random spatial effects model; the novel modelling ingredient is a latent stochastic process that determines surface selection. Therefore, the resultant realizations are functions of these surfaces. As a motivating example, consider a study of performance of plant species (presence/absence, abundance, growth, etc.) over a specified region. While some location-specific environmental covariates would be available to explain performance, there will be unobserved local covariate information that affects *suitability* of the location for the species. A latent covariate can be conceptualized to indicate the selection among suitability surfaces. As a second example, we can consider modelling of selling prices for single family homes. After adjusting for house characteristics, unobserved *neighborhood* and *preference* features remain as random effects in the desirability of a house location. A latent local indicator that selects among desirability surfaces naturally captures the random effects. We demonstrate in Chapter 2 by simulation examples the advantage of the GSDP over the SDP in a setting where there are only a small number of locations and replicated observations at each location.

Replications are typically needed for a full nonparametric approach (see, e.g. Sampson and Guttorp 1992) and so in our GSDP case as well. However, with replications that are discretized across time, we can shed the independence assumption by embedding our methodology within a dynamic model, retaining the temporal dependence. These methods allow the possibility to infer about the (random) distribution function that is operating at any given location, at any time, in the region. Non-

parametric spatial prediction under such modelling can be made not only at new locations for each replicate, but more generally through the generation of an entire new predictive surface at a future time. The GSDP model embedded in a dynamic linear model and its Bayesian inference is developed in Chapter 2. Also, though we develop our model in the context of spatial data, the theory is general and can be used when our responses are indexed by covariates in usual regression settings. Hence, we offer an alternative for most of the problems where mixtures of products of Dirichlet processes (Cifarelli and Regazzini 1978) and/or the dependent Dirichlet processes (MacEachern 2000) have been employed. See, for example, De Iorio et al. (2004).

We are aware of only two other recent approaches that also consider mixture models for spatial data where the weights are allowed to vary across locations. Fernandez and Green (2002) confine their attention to Markov random fields over lattices and Poisson distributed data where only the weights in the mixture vary from one location to another. We work with general point referenced data allowing both the weights and the parameters of the mixed distribution to vary spatially. Griffin and Steel (2004) present an implementation of the dependent Dirichlet process using Sethuraman's constructive representation, providing a random marginal distribution at each site. The components of the marginal stick breaking are the same at each location, but they are randomly permuted according to the realizations of a latent point process, so that at each site the resulting weights are assigned to different surfaces, inducing spatial dependence. Instead, we define a multivariate stick-breaking construction for any number and choice of locations, and also allow the marginal components to vary in space. Moreover, in this approach the closeness between the random distributions is ruled directly by the topology of the space, rather than by realizations of an underlying point process.

1.2 Spatio-temporal Point Process Models

The theory of spatial point processes provides convenient tools for the study of random spatial point patterns. The most commonly used and easily interpretable model is the spatial Poisson process: for any region in the area under study, the total number of observed points is a Poisson random variable with mean equal to the integrated intensity over that region. Then, the locations of these points, conditioned on the total number, is selected by using the intensity surface as a (non-normalized) density function. Again, if the points are emerging dynamically and the exact moments of their occurrence are viewed as continuous variables, we turn to the spatio-temporal version of the Poisson process which, in essence, is a three-dimensional generalization of the spatial Poisson processes.

The mathematical theory of point process on a general carrying space has been well established in the literature (Daley and Vere-Jones 1988; Karr 1991). Cressie (1993) and Møller and Waagepetersen (2004) discuss a larger variety of spatial point processes in practice, but most of their applications are restricted to two-dimensional spatial point processes. Recent developments in spatio-temporal point process modelling include Ogata (1998) with application to statistical seismology and Brix and Møller (2001) with application in modelling weeds. Brix and Diggle (2001), in modelling a plant disease, extend the log Gaussian Cox process (Møller et al. 1998) to a space-time version by using a stochastic differential equation model for the spatially varying relative risk. See Diggle (2005b) for a comprehensive review of the current methods.

The motivating problem for our research is to model the constructions of new residential houses. One attractive method of modelling space-time point process is the Cox process (Cox 1955), which is a Poisson process with inhomogeneous intensity arising as a realization of another stochastic process. The focus of our research is

to develop a statistically viable, physically insightful and computationally feasible stochastic process model for this intensity process. When we model urban growth by the construction of new residential houses, it is natural to assume these new houses would be quickly occupied by people, or vice versa the constructions themselves were driven by the population growth. That is, as a measure of urban development, the construction of new residential houses is a suitable surrogate for population growth. Moreover, the former is collected as spatial point pattern while the latter is not so easily available at such spatio-temporal resolution. Mathematically the conceptual connection between urban and population growth suggest adapting the population growth models in mathematical ecology (Kot 2001) to model the intensity process. In Chapter 4, spatial stochastic differential equation models for this intensity process are formulated from three types of population growth mechanisms.

Here we give a brief overview of the structure of our model. Let D be the study region, which would be some metropolitan area or a portion thereof. Depending upon the window of time, it might include primarily urban area or with a later (or longer) window, the suburban and rural areas surrounding it. Let $N_T(D)$ be the number of houses constructed in the period from $t = 0$ to T and $X_T = \left\{x_{1,t_1}, \ldots, x_{N_T,t_{N_T}}\right\}$ be the set of locations and times of these new constructions. The intensity of this space-time point process X_T is $\Omega(t,s), s \in D, t \in [0,T]$, which is a positive-valued function. $\Omega(t,s)$ could be viewed as essentially being a nonparametric specification which, for instance, could be a realization of a space-time process over $D \times [0,T]$ (Nonparametric functions using basis representations would also be possible.) However, in the present work, we choose to view $\Omega(t,s)$ as having a parametric form that is motivated by mechanistic considerations. We want to introduce specific parametric choices where the parameters quantify relationships and inference is sought about these parameters. Indeed, these parameters themselves will be associated with spatial locations and so

will themselves be viewed as realizations of spatial processes; thus, we can see how they vary over D.

More specifically, let

$$\Omega(t,s) = f(t, \theta_l(t,s); l = 1, \ldots, p)$$

where $\theta_l(t,s), s \in D, l = 1, \ldots, p$ are p possibly interdependent space-time processes. Again, this general formulation of the spatio-temporal Cox process allows incorporation of relevant subject matter mechanistic and theoretical behaviors into the statistical model. Again, the parameters and latent variables in the structured model, the $\theta_l(t)$'s, can be shown to capture flexibly spatial variation and correlation since they are realizations of spatial processes.

In the urban development problem, the parametric function $f(t, \theta_l(t); l = 1, \ldots, p)$ can be determined by a variety of differential equations that describe different mechanisms of development, such as the logistic growth equation:

$$\Omega(t,s) = r(t,s) \int_0^t \Omega(\tau, s)\, d\tau \left[1 - \frac{\int_0^t \Omega(\tau, s)}{K(s)}\right],$$

where $r(t,s)$ and $K(s)$, $s \in D$ are space-time processes representing local growth rate and carrying capacity.

The Bayesian inference for this space-time point process with dynamic intensity poses a difficult problem due to the nonlinearity in the model and the large number of observations. We will use a process convolution technique proposed in Xia and Gelfand (2006) to handle the inference problem. Our house construction data are discrete in time, therefore it requires discretizing the stochastic differential equation model and hence formulate a "transition model" (Diggle 2005b) between successive periods.

Our modelling approach is innovative in a number of respects. First, our model is

structured in the time dimension, reflecting physical limitation in density of residential houses, hence belongs to the "mechanistic modelling" category in Diggle (2005b), in contrast to the empirical approaches in Brix and Møller (2001) and Brix and Diggle (2001). The advantage of our approach to the urban development problem is that the population growth models have been well studied and proved "good" by ecologists. The model estimation and prediction are more easily interpretable and can provide insight into the real mechanism of growth. On the other hand, lacking theory to describe the nature of spatial pattern, we choose to employ a more empirical model for it reflecting only structured dependence based upon distance. Second, we employ a Bayesian hierarchical model to retain the conditional structure from the urban development mechanism. We do not have to resort to a partial likelihood as in Diggle (2005a). The estimation and computational difficulties stated in Diggle (2005b) can be overcome by implementing a full Bayesian model fitting. Third, the number of points in our example is very large, requiring likelihood approximation. We introduce a kernel convolution approximation (Xia and Gelfand 2006) to achieve such approximation. To our knowledge, this is the first time that this approximation has been applied to space-time point process models.

It is also worth reiterating that we intentionally specify our latent space-time intensity through a differential equation rather than a spatio-temporal process (see, e.g. Banerjee et al. 2004 and references therein). We intentionally seek to introduce a mechanistic modelling component; we are directly interested in parameters in our differential equation, such as spatially varying growth rates and carrying capacities. Although space-time process realizations are flexible, they do not offer the physical interpretation and insight we seek.

Chapter 2

Generalized Spatial Dirichlet Process Models

In this chapter, we formally present the construction and properties of the generalized Dirichlet process models. We then employ this model as a mixing distribution against a Gaussian kernel. We develop a hierarchical Bayesian model and its statistical inference based on the GSDP. The computational issues in model fitting are carefully presented. We then show how to embed the GSDP within a dynamic linear model. We demonstrate the flexibility of the GSDP model in data analysis by simulation examples.

2.1 GSDP and Its Mixture Models

2.1.1 Constructing the GSDP

In the spatial Dirichlet process developed by Gelfand et al. (2005), the random distribution of the pure spatial effect is essentially a Dirichlet Process defined on the space of the random surfaces over D generated by a mean 0 base spatial process. Then the almost sure characterization of the process implies that the random G for s is not the same as that for s' since $\theta_l^*(s)$ is not the same as $\theta_l^*(s')$. However, each

distribution has the same set of random stick-breaking probabilities. Indeed, for any group of n locations, the joint distribution uses the same set of stick-breaking probabilities inducing common surface selection for all locations in the group. The spatial dependence is introduced only through the underlying base measure, and it is not possible to capture the situation in which spatial effects can be selected from different surfaces at different locations. This limitation of the SDP is common to other recent work relating to the so-called dependent Dirichlet process (MacEachern 2000). See, for example, De Iorio et al. (2004).

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

Accordingly, we start by considering a base random field G_0, which, for convenience, we take to be stationary and Gaussian, and indicate its sample with $\theta_l^* = \{\theta_l^*(s), s \in D\}$, i.e., a surface over D. Then, we define a random probability measure G on the space of surfaces over D, whose finite dimensional distributions almost surely have the following representation: for any set of locations $(s_1, \ldots, s_n) \in D$, and any collection of sets $\{A_1, \ldots, A_n\}$ in $\mathcal{B}(\mathbb{R})$,

$$pr\{Y(s_1) \in A_1, \ldots, Y(s_n) \in A_n\} = \sum_{i_1=1}^{\infty} \ldots \sum_{i_n=1}^{\infty} p_{i_1,\ldots,i_n} \quad \delta_{\theta^*_{i_1}(s_1)}(A_1) \ldots \delta_{\theta^*_{i_n}(s_n)}(A_n), \tag{2.1}$$

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

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

$$p_{i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_n} = p_{i_1,\ldots,i_{k-1},\cdot,i_{k+1},\ldots,i_n} \equiv \sum_{j=1}^{\infty} p_{i_1,\ldots,i_{k-1},j,i_{k+1},\ldots,i_n}. \tag{2.2}$$

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

Proposition 1 *Let* $\{Y(s),\, s \in D\}$ *be a random field, whose random finite dimensional distributions are given by (2.1) 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* $Y(s_0)$ *with probability one as* $||s - s_0|| \to 0$.

Proof. Consider two sites s, s_0 in D. According to (2.1) the joint distribution of the process is almost surely a realization of the random element

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

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

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

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

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

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

In fact, the proof demonstrates almost sure convergence of the random probability measures. Note that Proposition 1 is an extension to our case of analogous results stated in MacEachern (2000) and Gelfand et al. (2005).

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

$$E\{Y(s)|G\} = \sum_{l=1}^{\infty} p_l(s)\,\theta_l^*(s) \tag{2.3}$$

$$var\{Y(s)|G\} = \sum_{l=1}^{\infty} p_l(s)\,\theta_l^{*^2}(s) - \left\{\sum_{l=1}^{\infty} p_l\,\theta_l^*(s)\right\}^2, \tag{2.4}$$

and, for a pair of sites s_i, s_j,

$$\begin{aligned} cov\{Y(s_i), Y(s_j)|G\} = &\sum_{l=1}^{\infty}\sum_{m=1}^{\infty} p_{l,m}(s_i, s_j)\,\theta_l^*(s_i)\,\theta_m^*(s_j) + \\ &- \left\{\sum_{l=1}^{\infty} p_l(s_i)\,\theta_l^*(s_i)\right\}\left\{\sum_{m=1}^{\infty} p_m(s_j)\,\theta_m^*(s_j)\right\}. \end{aligned} \tag{2.5}$$

Equation (2.5) shows that with almost surely continuous realizations from the base process and of the weights, the GSDP is mean square continuous. Again, the process $Y(s)$ has heterogeneous variance and is nonstationary. However, when we marginalize over G, we can see more clearly the difference between the GSDP and SDP models. Suppose G_0 is a mean zero stationary Gaussian process with finite variance σ^2 and correlation function $\rho_\phi(s_i - s_j)$. Then, $E\{Y(s)\} = 0$ and $var\{Y(s)\} = \sigma^2$ as before, but now

$$cov\{Y(s_i), Y(s_j)\} = \sigma^2\rho_\phi(s_i - s_j)\sum_{l=1}^{\infty} E\{p_{ll}(s_i, s_j)\}. \tag{2.6}$$

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 et al. (2005) or, more generally, in the single-p dependent Dirichlet process discussed by MacEachern (2000). We can interpret this limiting situation as the one of maximum concordance among the surfaces chosen at the two locations. In all other cases, the association structure is diminished by the amount of mass that the process (2.1) is expected to place on the not equally indexed θ^*'s. Moreover, from (2.6) 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 $p_{i_1,...,i_n}$ for any choice of n and $s_1, ..., s_n$ through a multivariate stick-breaking construction. For the sake of simplicity, we present our approach in a bivariate setting, considering the random measure

$$pr\{Y(s_i) \in A_i, Y(s_j) \in A_j\} = \sum_{l=1}^{\infty}\sum_{m=1}^{\infty} p_{l,m} \cdot \delta_{\theta_l^*(s_i)}(A_i)\delta_{\theta_m^*(s_j)}(A_j), \qquad (2.7)$$

for a pair of sites s_i, s_j, providing details on extension to the general multivariate case when necessary. First, we define a convenient process which retains the same Dirichlet process structure marginally at each site and then we move to a more general setting.

Sethuraman's univariate stick-breaking construction has weights p_l defined by $p_1 = q_1$, $p_l = q_l \prod_{m=1}^{l}(1 - q_m), l \geq 2$ where, for all $l \geq 1$, q_l are independent $Beta(1, \nu)$ random variables. Denote the random events $\{Y = \theta_l^*\}$ by Θ_l^1 (with their complements Θ_l^0) and interpret the sequence of weights $\{p_1, p_2, \ldots\}$ as arising from $q_1 = pr\{\Theta_l^1\}$, $q_l = pr\{\Theta_l^1|\Theta_m^0, m < l\} = pr\{Y = \theta_l^*|Y \neq \theta_m^*, m < l\}$, $l = 1, 2, \ldots$. Turning back to our model, at each location s we can define events $\Theta_l^u(s)$, $u = 0, 1$, such that $\Theta_l^1(s)$=$\{Y(s) = \theta_l^*(s)\}$ and $\Theta_l^0(s)$=$\{Y(s) \neq \theta_l^*(s)\}$. Then, for any two locations s_i, s_j, we can consider the probabilities $q_{1,u,v}(s_i, s_j) = pr\{\Theta_1^u(s_i), \Theta_1^v(s_j)\}$, $q_{l,u,v}(s_i, s_j) = pr\{\Theta_l^u(s_i), \Theta_l^v(s_j)|\Theta_m^0(s_i), \Theta_m^0(s_j), m < l\}$, $l \geq 2$, $u, v \in \{0, 1\}$. For all $l = 1, 2, \ldots$, we can enter these probabilities in the form of Table 2.1. Note that, formally, e.g., $q_{l,1,1}(s_i, s_j) + q_{l,1,0}(s_i, s_j) = q_{l,1,+}(s_i, s_j)$ and we need to argue that $q_{l,1,+}(s_i, s_j) = q_l(s_i)$. Similarly, $q_{l,+,1}(s_i, s_j) = q_l(s_j)$. The argument is supplied in Lemma 1.

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

Table 2.1: Relevant probabilities in the multivariate stick-breaking construction in the special case of $n = 2$ locations, for $l = 1, 2, \ldots$.

Lemma 1 *The probabilities* $q_{1,u,v}(s_i, s_j) = pr\{\Theta_1^u(s_i), \Theta_1^v(s_j)\}$ *and*

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

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. By definition of the q's,

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

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

Then, accordingly, we can define the weights $p_{l,m}$ in (2.7) as

$$\begin{aligned} p_{l,m} &= pr\{Y(s_i) = \theta_l^*(s_i), Y(s_j) = \theta_m^*(s_j)\} \\ &= pr\{\Theta_l^1(s_i), \Theta_m^1(s_j), \Theta_k^0(s_i), k < l, \Theta_r^0(s_j), r < m\} \\ &= \begin{cases} \prod_{k=1}^{l-1} q_{k,0,0}\ q_{l,1,0} \prod_{r=l+1}^{m-1}(1 - q_r)\ q_m & \text{if } l < m \\ \prod_{r=1}^{m-1} q_{r,0,0}\ q_{m,0,1} \prod_{k=m+1}^{l-1}(1 - q_k)\ q_l & \text{if } m < l \\ \prod_{r=1}^{l-1} q_{r,00}\ q_{l,11} & \text{if } l = m \end{cases}, \end{aligned} \tag{2.8}$$

where we have suppressed s_i and s_j.

Inspection on expression (2.8) reveals that the weights are determined through a partition of the unit square similar to the one induced on the unit segment by

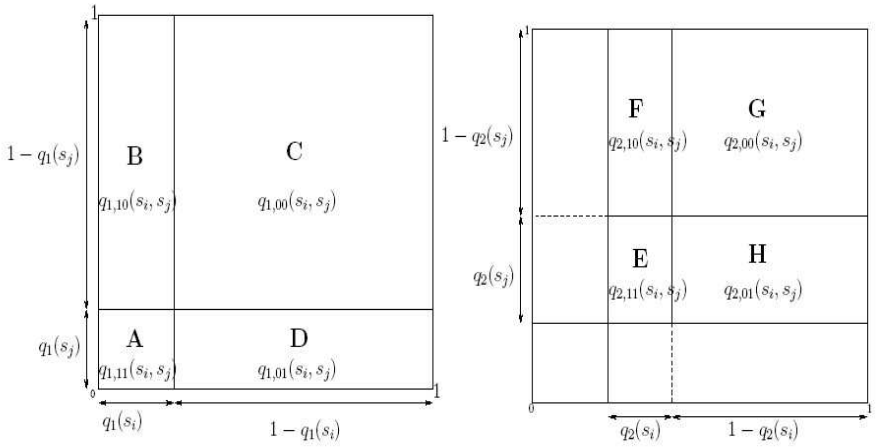

Figure 2.1: An exemplification of the multivariate stick-breaking procedure for the special case of $n = 2$ locations

the usual stick-breaking construction. At the first stage, if both the events $\Theta_1^1(s_i)$ and $\Theta_1^1(s_j)$ are true, we break off a region of the unit square of the same size as the realized value of $q_{1,1,1}(s_i, s_j)$. This is region A in Figure 2.1. If only $\Theta_1^1(s_i)$ (or $\Theta_1^1(s_j)$) is true, we remain only with a piece corresponding to region B (D). In fact, given $\Theta_1^1(s_i)$ ($\Theta_1^1(s_j)$), we go on with a univariate stick-breaking procedure so that we break off a part of region B (C) according to the values of $q_l(s_j)$ ($q_l(s_i)$), $l = 2, 3, \ldots$. If neither $\Theta_1^1(s_i)$ nor $\Theta_1^1(s_j)$ are true, then we discard all regions A, B, and D and remain only with region C, whose size is determined by $q_{1,0,0}(s_i, s_j)$. Then, at stage two, we repeat the same arguments as above for region C, and so on.

For n locations we require an n-dimensional stick breaking construction on the unit n-dimensional hypercube, i.e., we require the specification of probabilities $q_{l,u_1,\ldots,u_n}$, $u_j \in \{0, 1\}$, $j = 1, 2, \ldots, n$, where u_j is an abbreviation for $u(s_j)$, at any set of locations $(s_1, \ldots, s_n)$. This entails defining a spatial process which, conditionally on the locations, has values on the simplex $\mathbb{Q} = \{q_{l,u_1,\ldots,u_n} \geq 0 : \sum_{u_1,\ldots,u_n=0}^{1} q_{l,u_1,\ldots,u_n} = 1\}$, and also satisfies consistency conditions of the type (2.2) for all $l = 1, 2, \ldots$ and any

set of locations $(s_1, \ldots, s_n), n \in \mathbb{N}$ and for all $k = 1, \ldots, n$, that is

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

In the next section, we offer a flexible construction under which this can be done consistently. For the remainder of this section, as a special case, suppose the process retains the same marginal distribution at any location. This can be achieved by imposing $q_l(s) = q_l$, together with the symmetry condition $q_{l,1,0}(s_i, s_j) = pr\{Y(s_i) = \theta_l^*(s_i), Y(s_j) \neq \theta_l^*(s_j)\} = pr\{Y(s_i) \neq \theta_l^*(s_i), Y(s_j) = \theta_l^*(s_j)\} = q_{l,0,1}(s_i, s_j)$, for all $l = 1, 2 \ldots$ and $s \in D$. But, given q_l, if we can compute say $q_{l,1,1}(s_i, s_j)$ as a function of q_l, the remainder of Table 2.1 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 (2.8), 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 of achieving this particular result. For example, we might consider a process such that each q_l given $q_{l,0,0}$ has a Beta-Stacy distribution with parameters $1, \nu - 1,\ 1 - q_{l,0,0}$. If $q_{l,0,0}$ is assumed to be Beta$(1, \nu)$, then q_l is Beta$(1, \nu)$. The model we present in section 4 offers an alternative spatially-explicit way to specify q_l and $q_{l,1,1}$. For the n-dimensional case, symmetry conditions similar to the one stated above must be assumed in order to obtain the same marginal behavior at each site.

Modelling the marginals to be Dirichlet processes allows direct comparison with the models described by Gelfand et al. (2005) and De Iorio et al. (2004). However, it is worth noting that, though we employ a generalized stick-breaking construction and achieve DP marginal distributions, our model doesn't generally describe a joint Dirichlet process for a collection of locations. In particular, it follows that, given the

dependence between the θ^*'s in the sum representation (2.1), we are not able to trace a joint urn scheme, but only a marginal one. The SDP model described in Gelfand et al. (2005) stands as a particular case of the model described here, where we set $q_{l,0,1} = q_{l,1,0} = 0$ and $q_{l,1,1} = q_l$ for all locations and for all l.

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

$$p_{l|m}(s_i, s_j) = pr\{Y(s_i) = \theta^*_l(s_i)|Y(s_j) = \theta^*_m(s_j)\} =$$

$$= \frac{p_{lm}(s_i, s_j)}{\prod_{k=1}^{m-1}\{1 - q_k(s_j)\}\, q_m(s_j)},$$

since $\sum_l p_{l,m}(s_i, s_j) = p_m(s_j)$ due to marginal stick-breaking. But, substituting the expressions in (2.8),

$$p_{l|m} = \begin{cases} \prod_{k=1}^{l-1} \frac{q_{k,0,0}}{(1-q_k)} \frac{q_{l,1,0}}{1-q_l} & \text{if } l < m \\ \prod_{k=1}^{m-1} \frac{q_{k,0,0}}{(1-q_k)} \frac{q_{m,0,1}}{q_m} \prod_{k=m+1}^{l-1}(1 - q_k)\, q_l & \text{if } m < l \\ \prod_{k=1}^{l-1} \frac{q_{k,0,0}}{(1-q_k)} \frac{q_{l,1,1}}{q_l} & \text{if } l = m. \end{cases} \tag{2.9}$$

If we proceed along the lines that lead us to (2.8), we can show that for any given m, based on conditional reasoning, (2.9) 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 (2.9) 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 guarantee that the random weights sum to one with probability one.

This depends on the distribution of the weights. In the context of univariate stick-breaking priors, however, it is possible to provide a necessary and sufficient condition for that to happen (see Lemma 1 in Ishwaran and James 2001). We can expect that this condition holds for our model too, as long as we marginally get a DP prior (or, more in general, a stick-breaking prior). The precise argument is a direct extension of the result of Ishwaran and James (2001) and is given for the bivariate case in Lemma 2.

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

$$\sum_{l=1}^{\infty}\sum_{m=1}^{\infty} p_{l,m}(s_i, s_j) = 1 \quad \textit{if and only if} \quad \sum_{l=1}^{\infty} E\left[\log\{1 - q_l(s_i)\}\right] = -\infty. \tag{2.10}$$

Proof. Necessity follows after noticing that, if we marginalize with respect to s_i, condition (2.10) reduces to condition (5) in Ishwaran and James (2001). Now, consider for any $N, M = 1, 2, \ldots$, the remainder term

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

and assume (2.10) holds. We need to prove that $R_{N,M}(s_i, s_j) \to 0$ with probability one as $N, M \to \infty$. Write $R_{N,M}(s_i, s_j) = R_1 + R_2 + R_3$, where

$$R_1 = \sum_{l=1}^{N}\sum_{m=M+1}^{\infty} p_{l,m}(s_i, s_j),$$

$$R_2 = \sum_{m=1}^{M}\sum_{l=N+1}^{\infty} p_{l,m}(s_i, s_j),$$

$$R_3 = \sum_{l=N+1}^{\infty}\sum_{m=M+1}^{\infty} p_{l,m}(s_i, s_j).$$

Since all the terms in the sums are positive, it is necessary and sufficient that all the series tend to zero, as $N, M \to \infty$. Consider first R_1 and substitute (2.8) into all

$p_{l,m}(s_i, s_j)$, so that

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

Since $\sum_{m=M+1}^{\infty}\prod_{r=l+1}^{m-1}\{1-q_r(s_j)\}\; q_m(s_j) = \left(\sum_{m=M+1}^{\infty} p_m(s_j)\right) / \left(1 - \sum_{m=1}^{l} p_m(s_j)\right)$, for any $l = 1, 2, \ldots, N$, it tends to 0 as $M \to \infty$ again because of condition (5) in Ishwaran and James (2001). Since the result is true for all N, $\sum_{l=1}^{N}\sum_{m=M+1}^{\infty} p_{l,m}(s_i, s_j) \to 0$, with a similar argument for R_2. Now consider R_3. Let $\tau = \min(N, M)$. Then, $R_3 \leq \sum_{l=\tau+1}^{\infty}\sum_{m=\tau+1}^{\infty} p_{l,m}(s_i, s_j) = \prod_{k=1}^{\tau} q_{k,00}(s_i, s_j) \leq \prod_{k=1}^{\tau}\{1 - q_k(s_i)\}$, 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. ■

2.1.2 Mixing Using a Generalized Spatial Dirichlet Process.

The GSDP will be used to model the distribution of the spatial component $\theta(s)$ in a random effect model of the type

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

where $\mu(s)$ is a mean term, typically assumed to be a regression term $X(s)^T\beta$ for some vector of covariates $X(s)$ and some vector of parameters β, and $\varepsilon(s)$ is a white noise (nugget) component with mean zero and variance τ^2. Again, if we denote by $G^{(n)}$ the finite dimensional distributions defined by (2.1), for any finite set of locations $s^{(n)} = (s_1, \ldots, s_n)$, $n \in \mathbb{N}$, the joint distribution for $Y = \{Y(s_1), \ldots, Y(s_n)\}^T$, given $G^{(n)}$, μ and τ^2 is given by

$$f\left(y|G^{(n)}, \mu, \tau^2\right) = \int N_n\left(y|\,\theta + \mu, \tau^2 I_n\right)\; G^{(n)}(d\theta). \tag{2.11}$$

where $\theta = \{\theta(s_1), \ldots, \theta(s_n)\}^T$, $\mu = \{\mu(s_1), \ldots, \mu(s_n)\}^T$. As with the SDP, since $G^{(n)}$ is almost surely discrete, with probability one the conditional density (2.1) can be

rewritten as a countable location mixture of normals,

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

where, for simplicity, we have suppressed $i_j\,(s_j)$, $j=1,\ldots,n$ in $p_{i_1,...,i_n}$ and set the vector $\theta_{i_1,...,i_n}=\{\theta_{i_1}(s_1),\ldots,\theta_{i_n}(s_n)\}^T$. Immediately we know Y is a random vector which has a density absolutely continuous with respect to the Lebesgue measure on $(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$ with probability one. Its expected value is

$$E(Y|G^{(n)},\mu,\tau^2)=\sum_{i_1=1}^{\infty}...\sum_{i_n=1}^{\infty}p_{i_1,...,i_n}\theta_{i_1,...,i_n}+\mu,$$

and covariance matrix

$$\Sigma_{Y|G^{(n)},\mu,\tau^2}\;=\tau^2\,I_n+\Sigma_\theta^s,$$

where $(\Sigma_\theta^s)_{i,j}$=$cov\left\{\theta(s_i),\theta(s_j)|G^{(n)}\right\}$ is given by (2.5).

Under the assumptions of Proposition 1, if, in addition, the mean vector μ describes a continuous surface over D, it is easy to prove that an analogous statement holds for the convoluted process Y. In fact, the normal density is a bounded continuous function of the mean. Then the bounded convergence theorem applies to (2.11). Together with almost sure convergence of the random probability measures $G^{(n)}$ proved in Proposition 1, this implies that, with probability one, $Y(s)$ converges weakly to $Y(s_0)$ for any $s,s_0\in D$, as $||s-s_0||\to 0$.

2.2 The Spatially Varying Probabilities Model

Using latent variables we provide a constructive approach to specify the stick-breaking components in a way that is appealing for modelling purposes and ensures the existence of the processes sampled from G. In its implementation, using MCMC, we never need to sample or even estimate the q_l's or p_l's.

For any $n = 1, 2, \ldots$ and any $l = 1, 2, \ldots$ the stick-breaking components $q_{l,u_1,\ldots,u_n}(s_1, \ldots, s_n)$, $u_j \in \{0, 1\}, j = 1, 2, \ldots, n$ arise through probabilities associated with the events $\Theta_l^{u_j}(s_j)$, $l = 1, 2, \ldots$. So, we can induce a distribution to the stick-breaking components by directly 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 $l = 1, 2, \ldots$, $\delta_{\Theta_l^1(s)} = 1$ if $\Theta_l^1(s)$, $\delta_{\Theta_l^1(s)} = 0$ if $\Theta_l^1(s)$ does not occur. Suppose $\Theta_l^1(s)$ occurs if and only if $Z_l(s) \in A_l(s)$. Then, we can work with the equivalent stochastic process $\{\delta^*_{A_l(s)},\ s \in D, l = 1, 2, \ldots\}$ defined by

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

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

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

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

It is easy to see that such a characterization guarantees that (2.2) holds, hence the existence of the processes sampled from the random distribution (2.1).

We employ Gaussian thresholding to provide binary outcomes. This is routinely done in binary regression modelling (Albert and Chib 1993), is computationally convenient and, as a model for second stage random effects, there will be little posterior sensitivity to this choice. In fact, we assume that $\{Z_l(s),\ s \in D,\ l = 1, 2, \ldots\}$ is a countable collection of independent stationary Gaussian random fields on D having variance 1 and correlation function $\rho_Z(\cdot, \eta)$. We further assume that the mean of the l-th process, say $\mu_l(s)$, is unknown and we put a convenient prior on it, so that the distribution of $Z_l(s)$ (and hence of the q_l's) can be viewed as random. We also choose $A_l(s) = \{Z_l(s) \geq 0\}$. With these assumptions, it follows that

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

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

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

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

where $\Phi(\cdot)$ denotes the univariate standard normal cumulative distribution function(cdf). If the $\mu_l(s)$ are such that the $\Phi\{\mu_l(s)\}$ are independent Beta$(1, \nu)$, $l = 1, 2, \ldots$, then for each s, the marginal distribution of $\theta(s)$ is a DP with probabilities that vary with location. In the special case that $\mu_l(s) = \mu_l$, for all s, with $\Phi(\mu_l)$ independent Beta$(1, \nu)$ then, again marginally, the $\theta(s)$ follow a DP where the marginal weights are same for each s but the marginal distributions are not the same since $\theta_l^*(s) \neq \theta_l^*(s')$.

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

To summarize, we let

$$\begin{aligned} p_{i_1,\ldots,i_n} = pr\Big[& Z_1(s_1) < 0, \ldots, Z_{i_1-1}(s_1) < 0, Z_{i_1}(s_1) \geq 0; \\ & Z_1(s_2) < 0, \ldots, Z_{i_2-1}(s_2) < 0, Z_{i_2}(s_2) \geq 0; \ldots; \\ & Z_1(s_n) < 0, \ldots, Z_{i_n-1}(s_n) < 0, Z_{i_n}(s_n) \geq 0 | \{\mu_l(s_i)\}\Big], \end{aligned} \tag{2.12}$$

Again, we will never actually calculate the random weights $p_{i_1,\ldots,i_n}$ in practice. Finally, following the discussion above Proposition 1, we require two properties for this construction: (i) the random finite dimensional distribution $G^{(n)}$ satisfies the Kolmogorov consistency condition and (ii) the continuity property should be satisfied,

that is, if location s is near s', the probability of choosing the same sample surface for s and s' is high. In Propositions 2 and 3 we prove that these conditions are satisfied.

Proposition 2 *Let* $\{Y(s_1), Y(s_2), \ldots Y(s_n), s_i \in D, i = 1, \ldots n\}$ *have random finite dimensional distribution given by* (2.1)*, for* $n = 1, 2, \ldots$. *If the set of weights* $\{p_{i_1,\ldots,i_n}\}$ *is defined by means of a latent process as in* (2.12)*, then the collection of random finite dimensional distributions define a random field* $Y(s)$ *on* D.

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

$$p_{i_1,\ldots,i_{l-1},i_{l+1},\ldots i_n} = p_{i_1,\ldots,i_{l-1},\cdot,i_{l+1},\ldots i_n} = \sum_{k=1}^{\infty} p_{i_1,\ldots,i_{l-1},k,i_{l+1},\ldots i_n}. \tag{2.13}$$

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

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

$$Z(s_{l+1}) \in S_{l+1,i_{l+1}}, \ldots, Z(s_n) \in S_{n,i_n}\Big\},$$

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

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

have

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

■

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

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

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

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

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

that is, as the expected value of the marginal probabilities with respect to the distribution of the permutation field at the two locations. By the definition of πDDP, it follows that the dependence among the marginal random distribution functions is directly deduced by the permutation at each s. In particular, this is given by means of an auxiliary latent point process Z. In fact, Griffin and Steel first associate each atom $\{\theta_i^*(s), q_i\}$ with a realization z_i from Z, for $i = 1, 2, \ldots$. Then, at any s, the πDDP is defined permuting the set of $q's$ according to the realizations of the latent point process Z. In fact, $\pi(s)$ is defined to satisfy $||s - z_{\pi_1(s)}|| < ||s - z_{\pi_2(s)}|| < \ldots$. It follows that a realization from this process will necessarily be the same for some regions of D, while allowing different stick-breaking constructions for points far apart from each other. However, the representation at any s depends on how the process Z is associated with the atoms of the process, so that the representation does not seem to be invariant to a reordering of the z's. Moreover, for practical purposes it can be difficult to model the type of dependence induced by the point process mechanism, unless we choose simple processes, such as a stationary Poisson process. On the other hand, in our approach the spatial behavior of the stick-breaking components depends on the distribution of the latent Gaussian process Z and can vary across locations if this is true for the mean of Z.

2.3 Simulation Based Model Fitting for the GSDP

Assembling Section 2.1 and 2.2, we work with the following spatial model. Let the vectors Y_t=$\{Y_t(s_1), Y_t(s_2), \ldots, Y_t(s_n)\}^T$, $t = 1, \ldots, T$ denote T groups of independent observations collected at the same set of locations $(s_1, \ldots s_n) \in D \subset \mathbb{R}^2$. The mean surface $u(s), s \in D$ is modelled by a linear regression $u(s) = X(s)^T\beta$. The spatial random effect $\theta(s), s \in D$ has the GSDP rule defined in Section 2.2. The overall model has the following hierarchical structure

$$\begin{aligned}
Y_t \mid \theta_t, \beta, \tau^2 &\sim N_n(y_t \mid X_t^T\beta + \theta_t, \tau^2 I_n), \quad t = 1, ..., T \\
\theta_t \mid G^{(n)} &\sim G^{(n)}, \quad t = 1, ..., T \\
G^{(n)} \mid p_{i_1,\ldots,i_n}, \theta_l^* &= \sum_{i_1,\ldots,i_n=1}^{\infty} p_{i_1,\ldots,i_n}\, \delta_{\theta^*_{i_1}(s_1)}(\cdot) \cdots \delta_{\theta^*_{i_n}(s_n)}(\cdot), \quad l = 1, 2, \ldots \\
p_{i_1,\ldots,i_n} &= pr\left\{Z_{i_l}(s_l) < 0, \ldots, Z_{i_l}(s_l) \geq 0,\, l = 1, \ldots, n\right\}, i_l = 1, 2, \ldots \\
\left\{\theta_l^*(s_1), \ldots \theta_l^*(s_n)\right\}^T &\sim N_n\left(0, \sigma^2 R_n(\phi)\right), \quad l = 1, 2, \ldots \\
\left\{Z_{t,l}(s_1), \ldots, Z_{t,l}(s_n)\right\}^T &\sim N_n\left(\mu_l 1_n, H_n(\eta)\right), \quad l = 1, 2, \ldots,\ t = 1, 2, ...T \\
\mu_l \text{ s.t. } \Phi(\mu_l) &\sim Beta(1, \nu), \quad l = 1, 2, \ldots \\
\beta, \tau^2 &\sim N_p(\beta \mid \beta_0, \Sigma_\beta) \times \text{IGamma}(\tau^2 \mid a_\tau, b_\tau) \\
\sigma^2,\, \phi,\, \eta &\sim \text{IGamma}(\sigma^2 \mid a_\sigma, b_\sigma) \times [\phi] \times [\eta],
\end{aligned} \tag{2.14}$$

The priors for ϕ and η depend on the specific form of covariance function in $R_n(\phi)$ and $H_n(\eta)$. For convenience, in our examples we have set $\nu = 1$. In the version with $\mu_l(s)$ replacing μ_l, for each l, we obtain a realization from a Gaussian process with mean 0 and stationary covariance function $C(\cdot, \psi)$. In either case, the replications across t enable us to learn about the μ_l or the process driving the $\mu_l(s)$.

Although in (2.14) the marginal random distribution at an individual location s follows a Dirichlet process, the joint random distribution $G^{(n)}$ does not; we can not

marginalize over $G^{(n)}$. Instead, we approximate $G^{(n)}$ with a finite sum

$$G_K^{(n)} = \sum_{(i_1,\ldots,i_n)\in\{1,2,\ldots,K\}^n} p_{i_1,\ldots,i_n} \quad \delta_{\theta^*_{i_1}(s_1)}(\cdot)\,\delta_{\theta^*_{i_2}(s_2)}(\cdot)\ldots\delta_{\theta^*_{i_n}(s_n)}(\cdot), \tag{2.15}$$

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

Again, we sample the latent variables Z_l's directly to avoid computation of the weights $p_{i_1,\ldots,i_n}$ in (2.15). We eliminate the sampling of the conditional distribution $[\theta_t|G_K^{(n)}]$ by referring to the following equivalent structure:

$$\theta_t(s) = \theta^*_1(s)\,I_{\mathcal{Z}_{t,1}(s)} + \theta^*_2(s)\,I_{\mathcal{Z}_{t,2}(s)} + \ldots + \theta^*_K(s)\,I_{\mathcal{Z}_{t,K}(s)}. \tag{2.16}$$

In equation (2.16), $\theta_t(s)$ is a deterministic function of $\theta^*_l(s)$, $l = 1,\ldots,K$ and $Z^l_t(s)$, $l = 1,\ldots,K-1$. We rewrite the first stage of the hierarchical model as $[Y_t|\mu,\theta_t] = [Y_t|\mu,\theta^*,Z_t]$. Then, the likelihood function for Y_t can be expressed as

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

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

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

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

2.4 An Illustrative Example

We illustrate the fitting of (2.14) with a simulated data set from a finite mixture model of Gaussian processes that allows different joint multi-modal distributions for different pairs of locations. We first draw a specified number of locations in a given region. They are denoted as $(s_1, \ldots, s_n)$. Suppose there are T independent replicates $\{y_t(s_1), \ldots, y_t(s_n)\}, t = 1, \ldots, T$. We proceed as follows. For $t = 1, \ldots, T$, let

$$\begin{aligned} \left\{\theta_t^1(s_1), \ldots, \theta_t^1(s_n)\right\}^T &\sim N_n^{(1)}\left(-\mu 1_n, \sigma_1^2 R_n(\phi_1)\right) \\ \text{and } \left\{\theta_t^2(s_1), \ldots, \theta_t^2(s_n)\right\}^T &\sim N_n^{(2)}\left(\mu 1_n, \sigma_2^2 R_n(\phi_2)\right) \end{aligned}$$

Also, let $\{Z_t(s_1), \ldots, Z_t(s_n)\}^T \sim N_n(0, H_n(\eta))$. Then, for $i = 1, \ldots, n$, if $Z_t(s_i) \geq 0$, we set $y_t(s_i) = \theta_t^1(s_i)$; if $Z_t(s_i) < 0$, we set $y_t(s_i) = \theta_t^2(s_i)$.

Each $y_t(s_i)$ has a bimodal distribution of the form

$$\frac{1}{2}N^{(1)}\left(-\mu, \sigma_1^2\right) + \frac{1}{2}N^{(2)}\left(\mu, \sigma_2^2\right).$$

For two locations s_i and s_j near each other, the strong association between $Z_t(s_i)$ and $Z_t(s_j)$ makes $y_t(s_i)$ and $y_t(s_j)$ very likely to be from the same component $N^{(k)}(\mu_k, \sigma_k^2), k = 1, 2$. If s_i and s_j are distant, the linkage between $Z_t(s_i)$ and $Z_t(s_j)$ is weak, therefore the component choices for $y_t(s_i)$ and $y_t(s_j)$ are almost independent.

We simulate at 50 design locations in a rectangular region shown in Figure 2.2. Then, 40 independent replicates are sampled for these 50 locations. We choose the values of the parameters as $\mu_1 = -\mu_2 = 3$, $\sigma_1 = 2\sigma_2 = 2$, $\phi_1 = \phi_2 = 0.3$ and $\eta = 0.3$ in the mixture model above. We fit the model in (2.14) to this data set. We approximate $G^{(n)}$ with a finite sum of $K = 20$ components. To focus on the modelling of spatial dependence, we fixed the mean structure of $\{y_t(s_1), \ldots, y_t(s_n)\}$

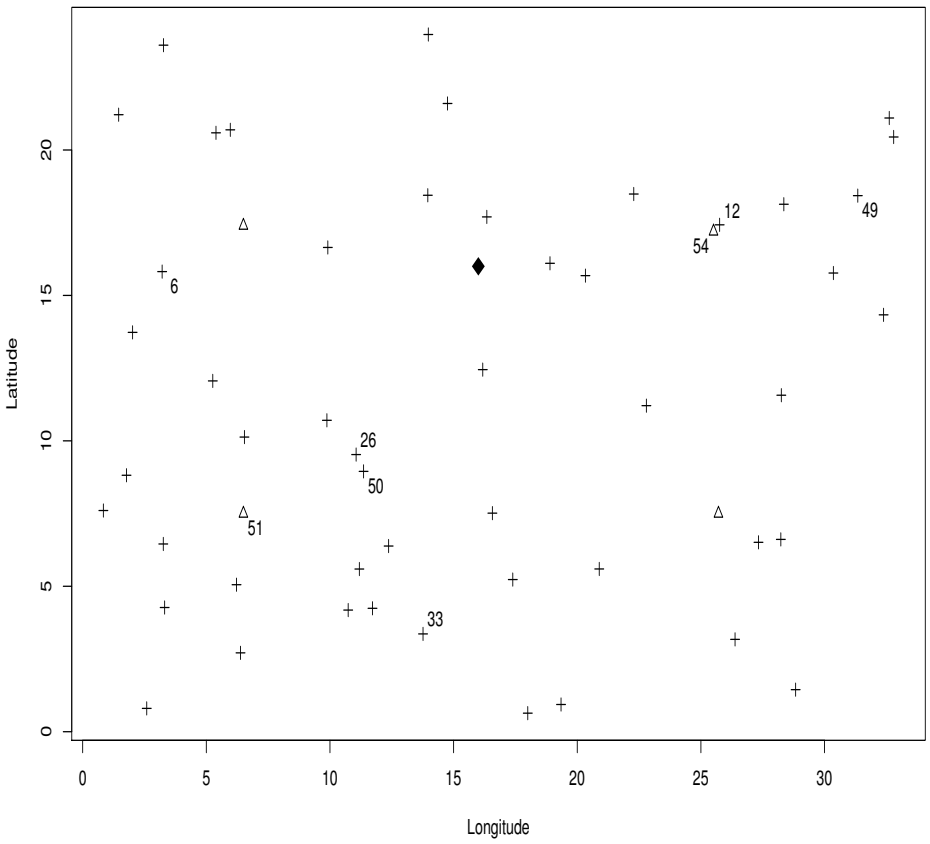

Figure 2.2: The Design Locations for the Simulation Example.

to be zero. For comparison, we considered the analogous SDP, using the same base measure and the same prior for all the parameters.

Performance is examined through the posterior predictive densities, both marginal and joint. In Figure 2.3, for four selected locations $(s_{26}, s_{33}, s_{49}, s_{50})$, we plot the true posterior predictive density, the predictive density estimated under the SDP and the posterior predictive density estimated under the GSDP. The values of the 40 observations at each of these 4 locations are shown along the x-axis. It is evident that the GSDP estimates more closely agree with the true densities of the model. Next, we select 3 pairs of sites and for each pair we show three predictive joint

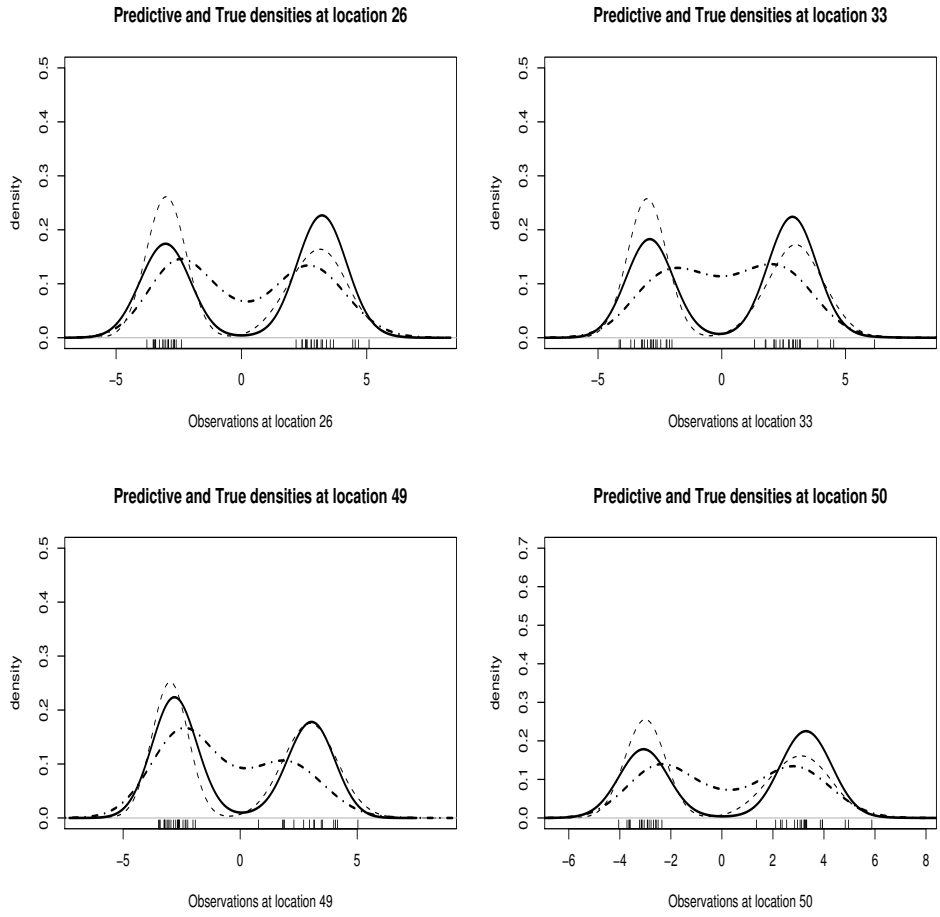

Figure 2.3: Posterior Predictive Densities for the GSDP (thick line $-$) and SDP (dashed line $\cdot-$) model for the locations indicated in Figure 2.2. The dotted line $(-)$ is the estimated true density, and the observed sample is reported on the x-axis.

densities. In Figure 2.4, the first pair $\{y(s_{50}), y(s_{26})\}$ are very close to each other, the second pair $\{y(s_{50}), y(s_{33})\}$ slightly distant and the third pair $\{y(s_{50}), y(s_{49})\}$ very distant (see Figure 2.2). Again, the GSDP is evidently much better than the SDP in capturing the local details and in particular the heights of the local modes.

Figure 2.5 plots the probability that a common sample surface is selected for a pair of locations against the distance between the two locations. We can see the decay in this probability as locations become further apart. Finally, for the SDP, we see no clustering; essentially a separate surface is needed for each replication. For

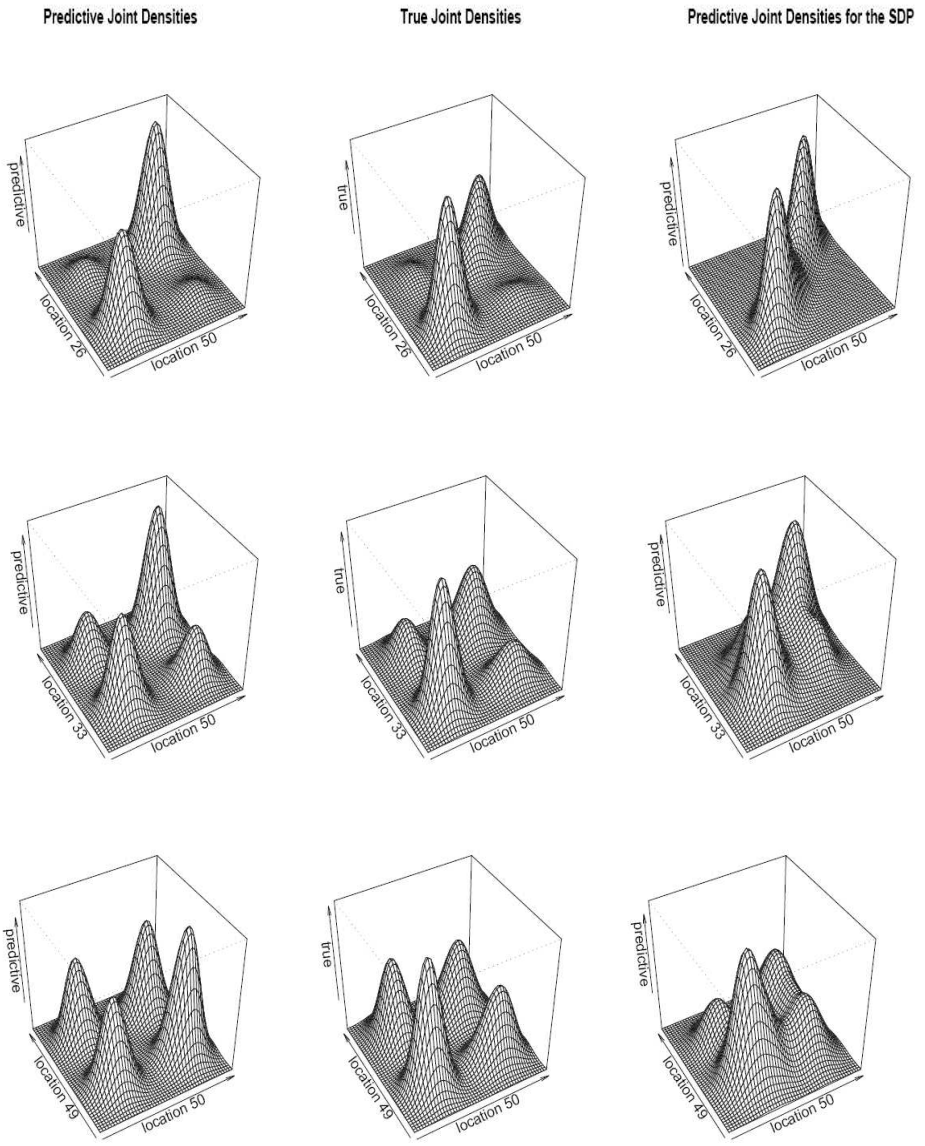

Figure 2.4: Predictive Bivariate Posterior Distributions for the GSDP (on the left) and SDP (on the right) models and the data. The middle column shows the true bivariate posterior distribution from a sample from the true model.

the GSDP, the modal number of surfaces is 3 and the maximum number of surfaces is 5. Clearly, the GSDP is able to respond to the local surface selection, while the SDP is not.

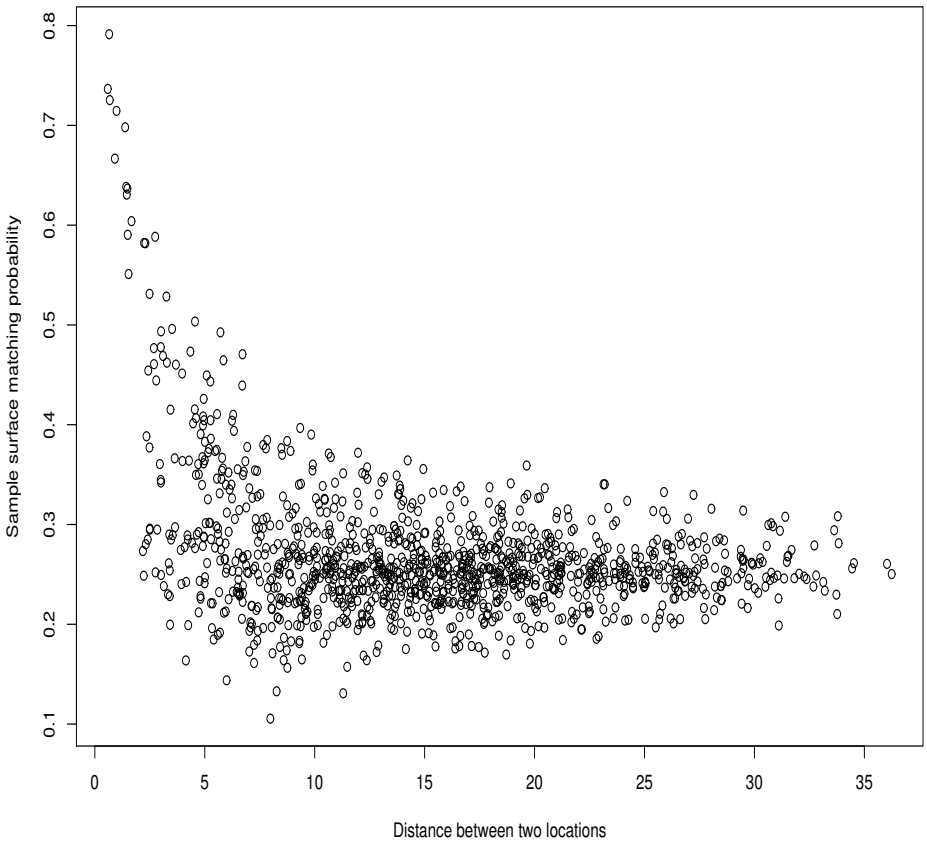

Figure 2.5: Decay in Probability of Common Surface Selection as a Function of Distance (See Section 2.4 for details)

2.5 A Spatio-temporal Dynamic Model Version

In Section 2.3 and 2.4, we assumed the $Y_t = \{y_t(s_1), \ldots, y_t(s_n)\}^T, t = 1, \ldots, T$ to be independent replicates. In practice, these observations are usually made in T consecutive time periods, so it is more realistic to model the evolution of the spatial process over time. In this section we present a version of the spatio-temporal model by embedding the GSDP in a dynamic linear model. We illustrate this spatio-temporal model also by fitting it to a simulated data set.

Preserving the notation in (2.12), the observations at time t can be modelled by the following dynamic linear model structure:

$$\begin{aligned} Y_t &= X_t^T \beta + \theta_t + \varepsilon_t; \varepsilon_t \sim N\left(0, \tau^2 I_n\right) \\ \theta_t &= \gamma \theta_{t-1} + \omega_t; \omega_t \sim GSDP\left(\nu G_0\right) \end{aligned} \tag{2.17}$$

These dynamics yield spatial random effects θ_t that evolve auto-regressively over time with autocorrelation coefficient, $\gamma\,(|\gamma| \leq 1)$. Only the second hierarchical specification in (2.14) changes to reflect (2.17). Updating of the full conditional distributions and the associated MCMC algorithm for the dynamic version is straightforward but careful attention to bookkeeping is required. We detail it in Appendix A.

We illustrate the model above by with a simulated data set. We still use the region and the 50 locations given in Figure 2.2. However, we add 4 new locations (with no observations) labelled 51-54 where we seek to predict. Also, a simple linear regression of $\beta_0 + \beta_1 X\left(s_i\right)$ is added to the model. $X\left(s_i\right)$ denotes the distance from location s_i to a fixed point source represented by the diamond in Figure 2.2.

The simulated observations $\left\{y_t\left(s_1\right), \ldots, y_t\left(s_n\right)\right\}^T$, $t = 1, \ldots, T$ are sampled again from a mixture of two distributions as follows. Following the specifications from the previous section, now consider

$$\left\{\omega_t^1\left(s_1\right), \ldots, \omega_t^1\left(s_n\right)\right\} \sim N_n^{(1)}\left(-\mu 1_n, \sigma_1^2 R_n\left(\phi_1\right)\right)$$

$$\text{and } \left\{\omega_t^2\left(s_1\right), \ldots, \omega_t^2\left(s_n\right)\right\} \sim N_n^{(2)}\left(\mu 1_n, \sigma_2^2 R_n\left(\phi_2\right)\right).$$

Also, let $\left\{Z_t\left(s_1\right), \ldots, Z_t\left(s_n\right)\right\} \sim N_n\left(0, H_n\left(\eta\right)\right)$. Then, for $i = 1, \ldots, n$, if $Z_t\left(s_i\right) \geq 0$, we set $\theta_t\left(s_i\right) = \gamma\,\theta_{t-1}\left(s_i\right) + \omega_t^1\left(s_i\right)$; if $Z_t\left(s_i\right) < 0$, $\theta_t\left(s_i\right) = \gamma\,\theta_{t-1}\left(s_i\right) + \omega_t^2\left(s_i\right)$ for $i = 1, \ldots, n$. Then we obtain $y_t\left(s_i\right) = \beta_0 + \beta_1 X\left(s_i\right) + \varepsilon_t\left(s_i\right)$, where $\varepsilon_t\left(s_i\right)$ is sampled independently from the normal distribution $N\left(0, \tau^2\right)$.

We choose the same values for the parameters of $\mu_1 = -\mu_2 = 3$. $\sigma_1 = 2\sigma_2 = 2$, $\phi_1 = \phi_2 = 0.3$ and $\eta = 0.3$ as in Section 2.4. Also, γ is chosen as 0.7, τ^2 is 9. β_0 and β_1 are 2 and -1 respectively, and T is equal to 40.

In fitting a model to the data, we use the same truncated GSDP for ω_t as given in Section 2.4 with $K = 20$ components. The Bayesian posterior mean of γ is found to be 0.8 in our experiment. The Bayesian goodness of fit is again illustrated by the posterior predictive densities at $T + 1$. We show not only the marginal posterior predictive density at each location, but also the joint posterior predictive densities for two locations.

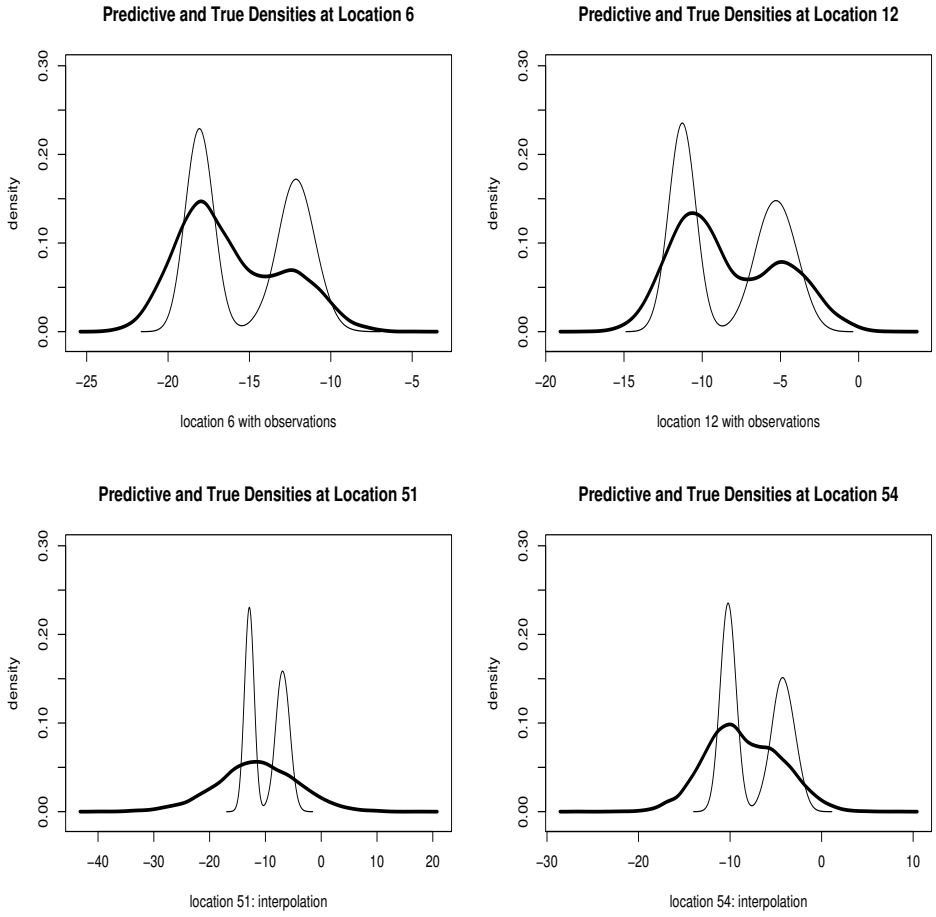

Figure 2.6: Posterior Predictive Densities, True Densities for Two Locations with Observations and Two New Locations at T+1 (See text for details)

In Figure 2.6 we plot the posterior predictive density at $T + 1(= 41)$ for two locations with observation (labelled 6 and 12 on Figure 2.2) and two new locations

(labelled 51 and 54 on Figure 2.2). The thick density curves are the predictive densities estimated from our model. The thin density curves represent the true densities of the model from which we simulated the data. The results are interesting in that, despite the small sample size and the introduction of bimodality only through the innovations at the second stage, we find bimodal behavior at 6 and 12. Location 51 is not very close to any of the sampled locations and, in the absence of data, yields a uni-modal predictive density. However, location 54 is very close to sampled locations and, reflecting the mean square continuity of the GSDP, an indication of two modes emerges.

Turning to bivariate predictive densities, we select 2 pairs of sites to show the predictive joint density at $T+1$. The first pair s_{50}, s_{26} are close to each other. The second pair s_{50}, s_{49} are much farther apart. In Figure 2.7, we provide perspective plots of the predictive and true joint densities. The first pair reveals a bimodal joint density while the second pair shows a density with four modes. If one were interested in developing simultaneous highest posterior density (HPD) confidence sets, one needs to identify the "footprint" associated with a level surface of the joint density. In particular, one must choose the level to provide a specified posterior probability. Of course, these bivariate densities are unavailable analytically so, using the posterior samples, we obtain a bivariate kernel density estimator. However, since level surfaces associated with this density estimate are still difficult to obtain, we evaluate the density estimate at the observed samples, providing an ordering for the samples. Then, according to the desired probability, we choose the density ordinate such that the proportion of the sample with ordinate above this value is the probability we seek. Figure 2.8 provides illustrative 80 % (inner curve) and 95 % (outer curve) HPD's for the site pairs in Figure 2.7.

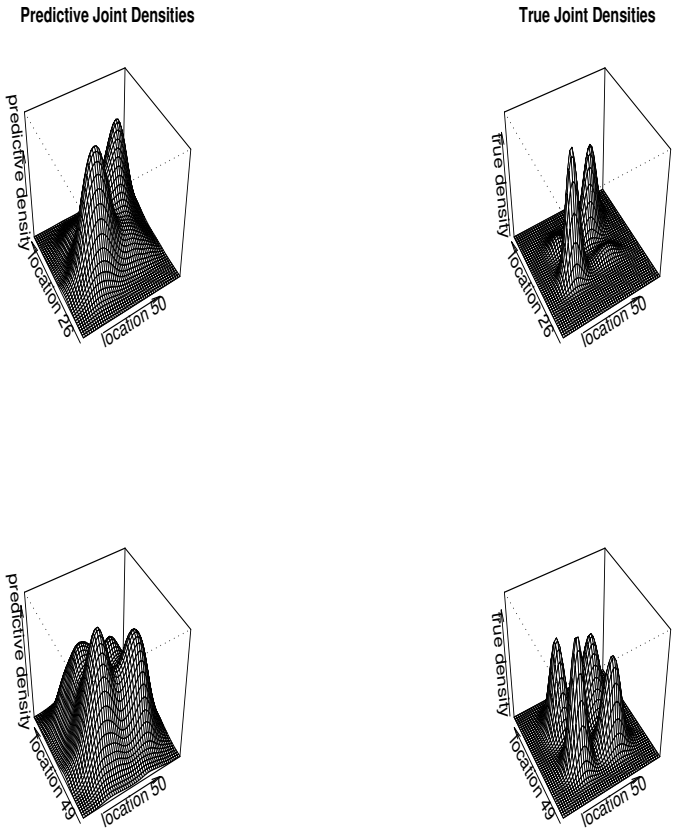

Figure 2.7: Predictive and True Bivariate Predictive Distribution at T+1 for the Simulated Data Example of Section 2.5 (See text for details)

2.6 Discussion and Summary

We have introduced the GSDP as a more flexible successor to the SDP proposed by Gelfand et al. (2005). However, any multivariate density can be approximated by a suitable countable mixture of multivariate normal densities. Since, for any finite set of locations, with probability one the SDP mixture model is such a countable mixture model, what practical advantages can the GSDP offer over the SDP? Why would we take the trouble to implement the much more computationally demanding GSDP? For example, with a bivariate distribution that is the product of two bimodal univariate

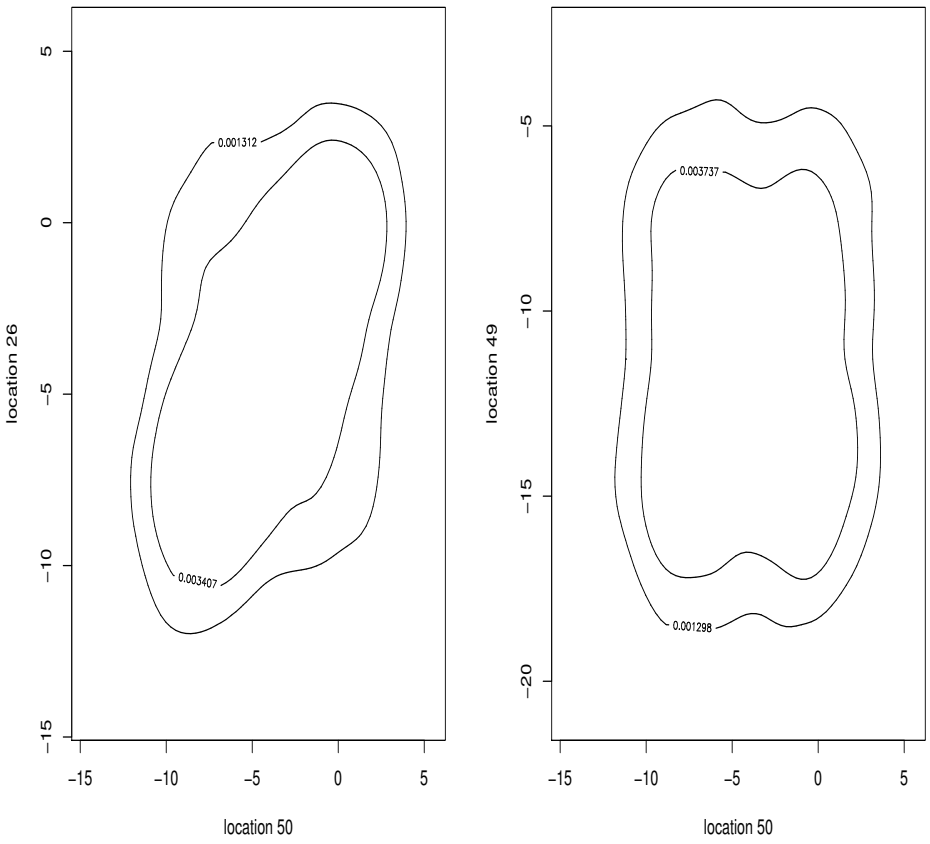

Figure 2.8: 95% and 80% Simultaneous Bivariate Posterior Density Confidence Sets (See Section 2.5 for details)

distributions, while the GSDP might capture such a distribution using essentially two components, wouldn't the SDP be able to do it with four components?

In fact, while, in principle, the SDP can equally well find multiple modes in say a bivariate distribution, it will have a more difficult time distinguishing the joint distribution for points close to each other than from the joint distribution for points far apart. In other words, in practice, the normal mixture model in (2.14) can more quickly adapt to the data. Expressed in different terms, in requiring additional components, the SDP will run into the, a priori, geometrically decaying weights, so it may

struggle to properly allocate mass to the modes. Furthermore, consider the situation where we might have varying numbers of mixing components and these might vary with spatial location. The GSDP, which allows different marginal distributions at each s, is better suited to handle this situation.

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

Chapter 3

Modelling Disease Incidence Data with Spatial and Spatio-temporal Dirichlet Process Mixtures

In this chapter, we will extend the spatial Dirichlet process to model discrete spatial observations using the example of disease incidence data. As stated in Chapter 1, we can replace the Gaussian assumption in the first hierarchy of the SDP model with any exponential family and still model the spatial random effect by the spatial Dirichlet process. We will first review the disease mapping problem, and then formulate the spatial and spatio-temporal Dirichlet process models for it. This model is motivated by the lung cancer incidence data in 88 counties of Ohio.

3.1 Disease Mapping Problem

Data on disease incidence (or mortality) are typically available as rates or summary counts for contiguous geographical regions, e.g., census tracts, post or zip codes, districts, or counties, and collected over time. Hence, though cases occur at point locations (residences), the available responses are associated with entire subregions in the study region. We denote the disease incidence counts (number of cases) by

y_{it}, where $i = 1, ..., n$ indexes the regions B_i, and $t = 1, ..., T$ indexes the time periods. In practice, we may have covariate information associated with the region, e.g., percent African American, median family income, percent with some college education. In some cases, though we only know the areal unit into which a case falls, we may have covariate information associated with the case, e.g., sex, race, age, previous comorbidities. Moreover, any of this covariate information could be time dependent. We devote Section 3.2.3 below to a discussion of how to accommodate such information in our modelling framework. However, the focus here is on flexible modelling of areal unit spatial random effects and so, to avoid obscuring our primary contribution, we do not consider covariates elsewhere.

A primary inferential objective in the analysis of disease incidence data is summarization and explanation of spatial and spatio-temporal patterns of disease (*disease mapping*); also of interest is spatial smoothing and temporal prediction (forecasting) of disease risk. The field of *spatial epidemiology* has grown rapidly in the past fifteen years with the introduction of spatial and spatio-temporal hierarchical (parametric) models; see, e.g., Elliott et al. (2000), and Banerjee et al. (2004) for reviews and further references.

The typical assumption (for rare diseases) is that the disease count y_{it}, conditionally on parameters R_{it}, are independent $\text{Po}(y_{it} \mid E_{it}R_{it})$ (we will write $\text{Po}(\cdot \mid m)$ for the Poisson probability mass function/distribution with mean m). Here, E_{it} is the expected disease count, and R_{it} is the relative risk, for region i at time period t.[1] E_{it} is obtained as $R^* n_{it}$, with R^* an overall disease rate, using either *external* or *internal* standardization, the former developing R^* from reference tables (available for certain types of cancer), the latter computed from the given data set, e.g., $R^* = \sum_{it} y_{it} / \sum_{it} n_{it}$. Next, the relative risks R_{it} are explained through different types of

[1]Below we will use an alternative and, we assert, preferable, specification, writing $n_{it}p_{it}$ for the Poisson mean, where n_{it} is the specified number of individuals at risk in region i at time t and p_{it} is the corresponding disease rate.

random effects. For instance, a specification with random effects additive in space and time is $\log R_{it} = \mu_{it} + u_i + v_i + \delta_t$, where μ_{it} is a component for the regional covariates (e.g., $\mu_{it} = x'_{it}\beta$ for regression coefficients β), u_i are regional random effects (typically, the u_i are assumed i.i.d. $\mathrm{N}(0, \sigma_u^2)$), v_i are spatial random effects, and δ_t are temporal effects (say, with an autoregressive prior).

The most commonly used prior model for the v_i is based on some form of a conditional autoregressive (CAR) structure (see, e.g., Clayton and Kaldor 1987; Cressie and Chan 1989; Besag et al. 1991; Bernardinelli et al. 1995; Besag et al. 1995; Waller et al. 1997; Pascutto et al. 2000). For instance, the widely-used specification suggested by Besag et al. (1991) is characterized through local dependence structure by considering for each region i a set, ϑ_i, of *neighbors*, which, for example, can be defined as all regions contiguous to region i. Then the (improper) joint prior *density* for the v_i is built from the prior full conditionals $[v_i \mid v_j; j \neq i]$. These are normal distributions with mean $m_i^{-1} \sum_{j \in \vartheta_i} v_j$ and variance λm_i^{-1}, where λ is a precision hyperparameter and m_i is the number of neighbors of region i. Alternatively, a multivariate normal distribution for the v_i, with correlations driven by the distances between region centroids, has been used (see, e.g., Wakefield and Morris 1999; Banerjee et al. 2003b). A different hierarchical formulation, discussed in Böhning et al. (2000), involves replacing the normal mixing distribution with a discrete distribution taking values φ_j, $j = 1, ..., k$ (that represent the relative risks for k underlying time-space clusters) with corresponding probabilities p_j, $j = 1, ..., k$. Hence, marginalizing over the random effects, the distribution for each region i and time period t emerges as a discrete Poisson mixture, $\sum_{j=1}^{k} p_j \mathrm{Po}(y_{it} \mid E_{it}\varphi_j)$. See, also, Schlattmann and Böhning (1993) and Militino et al. (2001) for use of such discrete Poisson mixtures in the simpler setting without a temporal component. In this setting, related is the Bayesian work of Knorr-Held and Rasser (2000) and Giudici et al. (2000) based on spatial partition

structures, which divide the study region into a number of clusters (i.e., sets of contiguous regions) with constant relative risk, assuming, in the prior model, random number, size, and location for the clusters. Further related Bayesian work is that of Green and Richardson (2002).

When spatio-temporal interaction is sought, the additive form $v_i + \delta_t$ is replaced by v_{it}. The latter has been modelled using independent CAR models over time, dynamically with independent CAR innovations, or as a CAR in space and time (see Banerjee et al. 2004).

Rather than modelling the spatial dependence through the finite set of spatial random effects, one for each region, an alternative prior specification arises by modelling the underlying continuous-space relative risk (or rate) surface and obtaining the induced prior models for the relative risks (or rates) through aggregation of the continuous surface. This approach is less commonly used in modelling for disease incidence data (among the exceptions are Best et al. 2000; Kelsall and Wakefield 2002). However, it arguably offers a more coherent modelling framework, since by modelling the underlying continuous surfaces, it avoids the dependence of the prior model on the data collection procedure, i.e., the number, shapes, and sizes of the regions chosen in the particular study. It replaces the specification of a proximity matrix, which spatially connects the subregions, with a covariance function, which directly models dependence between arbitrary pairs of locations (and induces a covariance between arbitrary subregions using block averaging).

In this research, we follow this latter approach, our main objective being to develop a flexible nonparametric model for the needed risk (or rate) surfaces. In particular, denote by D the union of all regions in the study area and let $z_{t,D} = \{z_t(s) : s \in D\}$ be the latent disease rate surface for time period t, on the logarithmic scale. Hence, $z_t(s) = \log p_t(s)$, where $p_t(s)$ is the probability of disease at time t and

spatial location s. (With rare diseases, the logarithmic transformation is practically equivalent to the logit transformation). We propose spatial and spatio-temporal nonparametric prior models for the vectors of log-rates $z_t = (z_{1t}, ..., z_{nt})$, which we define by block averaging the surfaces $z_{t,D}$ over the regions B_i, i.e., $z_{it} = |B_i|^{-1} \int_{B_i} z_t(s)ds$, where $|B_i|$ is the area for region B_i. We develop the spatial prior model by block averaging a Gaussian process (GP) to the areal units determined by the regions B_i, and then centering a Dirichlet process (DP) prior (Ferguson 1973; Antoniak 1974) around the resulting n-variate normal distribution. We show that the model is equivalent to the prior model that is built by block averaging a spatial DP (Gelfand et al. 2005). To model the z_t, we can specify them to be independent replications under the DP or we can add a further dynamic level to the model with z_t evolving from z_{t-1} through independent DP innovations. We use the former in our simulation example in Section 3.4.1; we use the latter with our real data example in Section 3.4.2.

With regard to the existing literature, our approach is, in spirit, similar to that of Kelsall and Wakefield (2002) where an isotropic GP was used for the log-relative risk surface. However, as exemplified in Section 3.2.2, we relax both the isotropy and the Gaussianity assumptions. In addition, we develop modelling for disease incidence data collected over space and time. Moreover, as we show in Section 3.2.1, our nonparametric model has a mixture representation, which is more general than that of Böhning et al. (2000) as it incorporates spatial dependence and it allows model-based identification of the extent of clustering through the structure of the DP prior.

The plan of the chapter is as follows. Section 3.2 develops the methodology for the spatial and spatio-temporal modelling approaches. Section 3.3 discusses methods for posterior inference with more details given in Appendix B. Section 3.4 includes illustrations motivated by a previously analyzed data set involving lung cancers for

the 88 counties in Ohio over a period of 21 years. In fact, in Section 3.4.1 we develop a simulated data set for these counties which is analyzed using both our modelling specification as well as a GP model, revealing the benefit of our approach. We also reanalyze the original data in Section 3.4.2. Finally, Section 3.5 provides a summary and discussion of possible extensions.

3.2 Bayesian Nonparametric Models For Disease Incidence Data

The spatial prior model is discussed in Section 3.2.1. Section 3.2.2 demonstrates how the use of the SDP provides foundation for the modelling approach presented in Section 3.2.1. Section 3.2.3 discusses how to include different types of covariate information. Lastly, Section 3.2.4 develops a nonparametric spatio-temporal modelling framework.

3.2.1 The Spatial Prior Probability Model

Here, we treat the log-rate surfaces $z_{t,D}$ as independent realizations (over time) from a stochastic process over D. We build the model by viewing the counts y_{it} and the log-rates z_{it} as aggregated versions of underlying (continuous-space) stochastic processes. The finite-dimensional distributional specifications for the y_{it} and the z_{it} are induced through block averaging of the corresponding spatial surfaces.

For the first stage of our hierarchical model, we use the standard Poisson specification working with the $n_{it}p_{it}$ form for the mean, following the footnote in Section 3.1. We note that this parametrization seems preferable to the $E_{it}R_{it}$ form, since it avoids the need to develop the E_{it} through standardization; the overall log-rate emerges as the intercept in our model. Thus, the y_{it} are assumed conditionally independent, given $z_{it} = \log p_{it}$, from $\text{Po}(y_{it} \mid n_{it} \exp(z_{it}))$.

This specification can be derived through aggregation of an underlying Poisson process under assumptions and approximations as follow. For the time period t, assume that the disease incidence cases, over region D, are distributed according to a non-homogeneous Poisson process with intensity function $n_t(s)p_t(s)$, where $\{n_t(s) : s \in D\}$ is the population density surface and $p_t(s)$ is the disease rate at time t and location s. If we assume a uniform population density over each region at each time period (this assumption is, implicitly, present in standard modelling approaches for disease mapping), we can write $n_t(s) = n_{it}|B_i|^{-1}$ for $s \in B_i$. Hence, aggregating the Poisson process over the regions B_i, we obtain, conditionally on $z_{t,D}$, that the y_{it} are independent, and each y_{it} follows a Poisson distribution with mean $\int_{B_i} n_t(s)p_t(s)ds = n_{it}p^*_{it}$, where $p^*_{it} = |B_i|^{-1}\int_{B_i} p_t(s)ds$. If we approximate the distribution of the p^*_{it} with the distribution of the $\exp(z_{it})$, we can write $y_{it} \mid z_{it} \overset{ind.}{\sim} \text{Po}(y_{it} \mid n_{it}\exp(z_{it}))$ for the first stage distribution. We note that the stochastic integral for p^*_{it} is not accessible analytically. Moreover, using Monte Carlo integration to approximate the p^*_{it} is computationally infeasible (Short et al. 2005). Also, Kelsall and Wakefield (2002) use a similar approximation working with relative risk surfaces. Brix and Diggle (2001) do so as well, using a stochastic differential equation to model $p_t(s)$.

To build the prior model for the log-rates z_t, we begin with the familiar form, $z_t(s) = \mu_t(s) + \theta_t(s)$, for the log-rate surfaces $z_{t,D}$. Here, $\mu_t(s)$ is the mean structure and $\theta_{t,D} = \{\theta_t(s) : s \in D\}$ are spatial random effects surfaces. As discussed in Section 3.2.3, the surfaces $\{\mu_t(s) : s \in D\}$ can be elaborated through covariate surfaces over D. In the absence of such covariate information, we might set $\mu_t(s) = \mu$, for all t, and use a normal prior for μ. Alternatively, we could set $\mu_t(s) = \mu_t$, where the μ_t are i.i.d. $\text{N}(0, \sigma^2_\mu)$ with random hyperparameter σ^2_μ. In what follows for the spatial prior model, we illustrate with the common μ specification.

To develop the model for the spatial random effects, first, let the $\theta_{t,D}$, $t = 1, ..., T$,

given σ^2 and ϕ, be independent realizations from a mean-zero isotropic GP with variance σ^2 and correlation function $\rho(||s-s'||;\phi)$ (say, $\rho(||s-s'||;\phi) = \exp(-\phi||s-s'||)$ as in the examples in Section 3.4). Hence by aggregating over the regions B_i, we obtain $z_{it} = \mu + \theta_{it}$, where $\theta_{it} = |B_i|^{-1} \int_{B_i} \theta_t(s)ds$ is the block average of the surface $\theta_{t,D}$ over region B_i. The induced distribution for $\theta_t = (\theta_{1t}, ..., \theta_{nt})$ is a mean-zero n-variate normal with covariance matrix $\sigma^2 R_n(\phi)$, where the (i,j)-th element of $R_n(\phi)$ is given by

$$|B_i|^{-1}|B_j|^{-1} \int_{B_i} \int_{B_j} \rho(||s-s'||;\phi)\, dsds'.$$

Next, consider a DP prior for the spatial random effects θ_t with precision parameter $\alpha > 0$ and centering (base) distribution $\mathrm{N}_n(\cdot \mid 0, \sigma^2 R_n(\phi))$ (we will write $N_p(\cdot \mid \lambda, \Sigma)$ for the p-variate normal density/distribution with mean vector λ and covariance matrix Σ). We denote this DP prior by $\mathrm{DP}(\alpha, \mathrm{N}_n(\cdot \mid 0, \sigma^2 R_n(\phi)))$. The choice of the DP in this context yields data-driven deviations from the normality assumption for the spatial random effects; at the same time, it allows relatively simple implementation of simulation-based model fitting.

Note that the above structure implies for the vector of counts $y_t = (y_{1t}, ..., y_{nt})$ a nonparametric Poisson mixture model given by $\int \prod_{i=1}^n \mathrm{Po}(y_{it} \mid n_{it} \exp(\mu+\theta_{it}))dG(\theta_t)$, where the mixing distribution $G \sim \mathrm{DP}(\alpha, \mathrm{N}_n(\cdot \mid 0, \sigma^2 R_n(\phi)))$. Under this mixture specification, the distribution for the vectors of log-rates, $z_t = \mu 1_n + \theta_t$, is discrete (a property induced by the discreteness of DP realizations), a feature of the model that could be criticized. Moreover, although posterior simulation is feasible, it requires more complex MCMC algorithms (e.g., the methods suggested by MacEachern and Müller 1998 and Neal 2000) than the standard Gibbs sampler for DP based hierarchical models (e.g., West et al. 1994; Bush and MacEachern 1996). Thus, to overcome

both concerns above, we replace the DP prior for the z_t with a DP mixture prior,

$$z_t \mid \mu, \tau^2, G \overset{ind.}{\sim} \int \mathrm{N}_n(z_t \mid \mu 1_n + \theta_t, \tau^2 I_n) dG(\theta_t),$$

where, again, $G \sim \mathrm{DP}(\alpha, \mathrm{N}_n(\cdot \mid 0, \sigma^2 R_n(\phi)))$. That is, we now write $z_{it} = \mu + \theta_{it} + u_{it}$, with u_{it} i.i.d. $\mathrm{N}(0, \tau^2)$. Introduction of a heterogeneity effect in addition to the spatial effect is widely employed in the disease mapping literature dating to Besag et al. (1991) and Bernardinelli et al. (1995), though with concerns about balancing priors for the effects (see, e.g., Banerjee et al. 2004 and references therein). Here, in responding to the above concerns, we serendipitously achieve this benefit.

Hence, the mixture model for the y_t now assumes the form

$$f(y_t \mid \mu, \tau^2, G) = \int \prod_{i=1}^{n} p(y_{it} \mid \mu, \tau^2, \theta_{it}) dG(\theta_t),$$

where $p(y_{it} \mid \mu, \tau^2, \theta_{it}) = \int \mathrm{Po}(y_{it} \mid n_{it} \exp(z_{it})) \mathrm{N}(z_{it} \mid \mu + \theta_{it}, \tau^2) dz_{it}$ is a Poisson-log-normal mixture. Equivalently, the model can be written in the following semiparametric hierarchical form

$$\begin{aligned}
y_{it} \mid z_{it} &\overset{ind.}{\sim} \mathrm{Po}(y_{it} | n_{it} \exp(z_{it})), \;\; i = 1, ..., n, \;\; t = 1, ..., T \\
z_{it} \mid \mu, \theta_{it}, \tau^2 &\overset{ind}{\sim} \mathrm{N}(z_{it} \mid \mu + \theta_{it}, \tau^2), \;\; i = 1, ..., n, \;\; t = 1, ..., T \\
\theta_t \mid G &\overset{i.i.d.}{\sim} G, \;\; t = 1, ..., T \\
G \mid \sigma^2, \phi &\sim DP(\alpha, N_n(\cdot \mid 0, \sigma^2 R_n(\phi))).
\end{aligned} \tag{3.1}$$

The model is completed with independent priors $p(\mu)$, $p(\tau^2)$ and $p(\sigma^2)$, $p(\phi)$ for μ, τ^2, and for the hyperparameters σ^2, ϕ of the DP prior. In particular, we use a normal prior for μ, inverse gamma priors for τ^2 and σ^2, and a discrete uniform prior for ϕ. Although not implemented for the examples of Section 3.4, a prior for α can be added, without increasing the complexity of the posterior simulation method (Escobar 1994).

In practice, we work with a marginalized version of model (3.1),

$$p(\mu)p(\tau^2)p(\sigma^2)p(\phi)p(\theta_1,...,\theta_T \mid \sigma^2,\phi)\prod_{i=1}^{n}\prod_{t=1}^{T}\mathrm{Po}(y_{it}|n_{it}\exp(z_{it}))\mathrm{N}(z_{it} \mid \mu+\theta_{it},\tau^2), \tag{3.2}$$

which is obtained by integrating the random mixing distribution G over its DP prior (Blackwell and MacQueen 1973). The resulting joint prior distribution for the θ_t, $p(\theta_1,...,\theta_T \mid \sigma^2,\phi)$, is given by

$$\mathrm{N}_n(\theta_1 \mid 0,\sigma^2 R_n(\phi))\prod_{t=2}^{T}\left\{\frac{\alpha}{\alpha+t-1}N_n(\theta_t \mid 0,\sigma^2 R_n(\phi))+\frac{1}{\alpha+t-1}\sum\nolimits_{j=1}^{t-1}\delta_{\theta_j}(\theta_t)\right\}, \tag{3.3}$$

where δ_a denotes a point mass at a. Hence, the θ_t are generated according to a Pólya urn scheme; θ_1 arises from the base distribution, and then for each $t=2,...,T$, θ_t is either set equal to θ_j, $j=1,...,t-1$, with probability $(\alpha+t-1)^{-1}$ or is drawn from the base distribution with the remaining probability.

Note that we have defined the prior model for the spatial random effects θ_t by starting with a GP prior for the surfaces $\theta_{t,D}$, block averaging the associated GP realizations over the regions to obtain the $N_n(0,\sigma^2 R_n(\phi))$ distribution, and, finally, centering a DP prior for the θ_t around this n-variate normal distribution. This approach might suggest that the DP prior is dependent, in an undesirable fashion, on the specific choice of the regions (e.g., their number and size). The next section addresses this potential criticism by connecting the model in (3.1) with the spatial DP (SDP) from Gelfand et al. (2005).

3.2.2 Formulation of the Model through Spatial Dirichlet Processes

The SDP can be proposed as the prior for the spatial random effects surfaces $\theta_{t,D}$ to replace the isotropic GP prior that we used to build the DP model in Section 3.2.1. Therefore, now the model is developed by assuming that the $\theta_{t,D}$, $t = 1, ..., T$, given G_D, are independent from G_D, where G_D, given σ^2 and ϕ, follows a SDP prior with precision parameter α and base process $G_{0D} = \mathrm{GP}(0, \sigma^2\rho(||s - s'||; \phi))$ (i.e., the same isotropic GP used in Section 3.2.1).

Next, we block average the $\theta_{t,D}$ over the regions B_i with respect to their distribution that results by marginalizing G_D over its SDP prior. Recall that for any set of spatial locations s_r, $r = 1, ..., M$, over D, the random distribution $G^{(M)}$ induced by G_D follows a DP with base distribution $G_0^{(M)}$ induced by G_{0D}. Because we can choose M arbitrarily large and the set of locations s_r to be arbitrarily dense over D, using the Pólya urn characterization for the DP, we obtain that, marginally, the $\theta_{t,D}$ arise according to the following Pólya urn scheme. First, $\theta_{1,D}$ is a realization from G_{0D}, and then, for each $t = 2, ..., T$, $\theta_{t,D}$ is identical to $\theta_{j,D}$, $j = 1, ..., t-1$, with probability $(\alpha + t - 1)^{-1}$ or is a new realization from G_{0D} with probability $\alpha(\alpha + t - 1)^{-1}$.

Hence, if we block average $\theta_{1,D}$, we obtain the $\mathrm{N}_n(0, \sigma^2 R_n(\phi))$ distribution for θ_1. Then, working with the conditional specification for $\theta_{2,D}$ given $\theta_{1,D}$, if we block average $\theta_{2,D}$, θ_2 arises from $\mathrm{N}_n(0, \sigma^2 R_n(\phi))$ with probability $\alpha(\alpha+1)^{-1}$ or $\theta_2 = \theta_1$ with probability $(\alpha+1)^{-1}$. Analogously, for any $t = 2, ..., T$, the induced conditional prior $p(\theta_t \mid \theta_1, ..., \theta_{t-1}, \sigma^2, \phi)$ is a mixed distribution with point masses at θ_j, $j = 1, ..., t-1$, and continuous piece given by the $\mathrm{N}_n(0, \sigma^2 R_n(\phi))$ distribution; the corresponding weights are $(\alpha + t - 1)^{-1}$, $j = 1, ..., t - 1$, and $\alpha(\alpha + t - 1)^{-1}$. Thus, the prior distribution for the θ_t in (3.3) can be obtained by starting with a SDP prior for the $\theta_{t,D}$ (centered around the same isotropic GP prior used in Section 3.2.1 for the $\theta_{t,D}$),

and then block averaging the (marginal) realizations from the SDP prior over the regions.

As in Section 3.2.1, we extend $z_t = \mu 1_n + \theta_t$ to $z_t = \mu 1_n + \theta_t + u_t$, where the u_t are independent $\mathrm{N}_n(0, \tau^2 I_n)$. Hence, model (3.2) is equivalent to the marginal version of the model above, i.e., with G_D marginalized over its SDP prior.

The argument above, based on SDPs, provides formal justification for model (3.1) – (3.3). The SDP is a nonparametric prior for the continuous-space stochastic process of spatial random effects; regardless of the number and geometry of regions chosen to partition D, it induces the appropriate corresponding version of the model in (3.2).

3.2.3 Introducing Covariates

As noted in Section 3.1, often in looking at disease incidence/mapping data we will seek to bring in covariate information. Here, we indicate how we would do this in the context of the model given in (3.1). Our approach is to consider how we would handle the idealized situation of point-referenced case/non-case data and then propagate the effect of the assumptions and approximations in Section 3.2.1. Our approach is similar in spirit to that of Wakefield and Shaddick (2006). In particular, illustrating with a single covariate surface $\{X_t(s) : s \in D\}$, suppose $z_t(s) = \beta_{0t} + \beta_{1t} X_t(s) + \theta_t(s)$.

If $X_t(s)$ is an areal unit level covariate, i.e., $X_t(s) = X_{it}$, for all $s \in B_i$, then $p^*_{it} = \exp(\beta_{0t} + \beta_{1t} X_{it})\ |B_i|^{-1} \int_{B_i} \exp(\theta_t(s)) ds$. So, for such covariates, no approximation beyond that of Section 3.2.1 is required.

Next, associate with each of the n_{it} individuals at risk in areal unit i at time t an (unknown) location s_{ij}, $j = 1, 2, ..., n_{it}$, and covariate level $X_t(s_{ij})$ (suppressing time t in the notation for s_{ij}). At each location there is a Bernoulli trial with probability $p_t(X_t(s_{ij}))$. (Here, we write $p_t(X_t(s_{ij}))$, instead of $p_t(s_{ij})$, to emphasize the depen-

dence on the covariate.) Since incidence rates are usually very small, we can envision a Poisson approximation to the sum of the n_{it} Bernoulli trials in areal unit i at time t with expectation equal to $p_{it} = \sum_{j=1}^{n_{it}} p_t(X_t(s_{ij}))$.

Suppose that $X_t(s)$ is categorical, in fact, for convenience, binary. Then, though we do not know where they occur, we do know that n_{0it} of the $X_t(s_{ij})$ are 0 and n_{1it} of the $X_t(s_{ij})$ are 1. So, in the absence of spatial effects, $\sum_{j=1}^{n_{it}} p_t(X_t(s_{ij})) = n_{0it}p_t(0) + n_{1it}p_t(1) = n_{it}p_{it}^*$ where $p_{it}^* = n_{it}^{-1}(n_{0it}p_t(0) + n_{1it}p_t(1))$. With spatial effects and with locations assigned at random, we obtain

$$\sum_{j=1}^{n_{it}} p_t(X_t(s_{ij})) = \sum_{\{s_{ij}:X_t(s_{ij})=0\}} \exp(\beta_{0t} + \theta_t(s_{ij})) + \sum_{\{s_{ij}:X_t(s_{ij})=1\}} \exp(\beta_{0t} + \beta_{1t} + \theta_t(s_{ij})).$$

Again, we know the number of 0s and 1s but can only assume they are randomly assigned to the s_{ij}. Hence, for $\ell = 0, 1$,

$$\sum_{\{s_{ij}:X_t(s_{ij})=\ell\}} \exp(\theta_t(s_{ij})) \approx \frac{n_{\ell it}}{n_{it}} \sum_{j=1}^{n_{it}} \exp(\theta_t(s_{ij})) \approx n_{\ell it}|B_i|^{-1} \int_{B_i} \exp(\theta_t(s))ds,$$

and, thus, $\sum_{j=1}^{n_{it}} p_t(X_t(s_{ij})) \approx n_{it}p_{it}^*$, with

$$p_{it}^* = \frac{n_{0it}}{n_{it}} \exp(\beta_{0t})|B_i|^{-1} \int_{B_i} \exp(\theta_t(s))ds + \frac{n_{1it}}{n_{it}} \exp(\beta_{0t} + \beta_{1t})|B_i|^{-1} \int_{B_i} \exp(\theta_t(s))ds.$$

Finally, making the same integral approximation (i.e., $\exp(\theta_{it}) \approx |B_i|^{-1} \int_{B_i} \exp(\theta_t(s))ds$), we can write $p_{it}^* \approx \exp(\beta_{0t} + \theta_{it})\{1 + n_{it}^{-1}n_{1it}[\exp(\beta_{1t}) - 1]\} \approx \exp(\beta_{0t} + \theta_{it})[1 + n_{it}^{-1}n_{1it}\beta_{1t}] \approx \exp(\beta_{0t} + n_{it}^{-1}n_{1it}\beta_{1t} + \theta_{it})$.

Lastly, with a continuous covariate, we may envision two scenarios – (i) that it is available for each of the n_{it} individuals at risk in areal unit i at time t or (ii) more generally, that it is available as a surface known over the entire study region. Again, the quantity of interest is $\sum_{j=1}^{n_{it}} p_t(X_t(s_{ij})) = \sum_{j=1}^{n_{it}} \exp(\beta_{0t} + \beta_{1t}X_t(s_{ij}) +$

$\theta_t(s_{ij})) = n_{it}p^*_{it}$ where $p^*_{it} = n_{it}^{-1}\exp(\beta_{0t})\sum_{j=1}^{n_{it}}\exp(\beta_{1t}X_t(s_{ij}) + \theta_t(s_{ij}))$. In case (i), let $V_{it} = n_{it}^{-1}\sum_{j=1}^{n_{it}} X_t(s_{ij})$ while in case (ii) let $V_{it} = |B_i|^{-1}\int_{B_i} X_t(s)ds$; under our assumptions, in either case, V_{it} can be calculated. Then, as earlier, we approximate the distribution of p^*_{it} by the distribution of $\exp(z_{it})$. In either case, we obtain $p^*_{it} \approx \exp(\beta_{0t} + \beta_{1t}V_{it} + \theta_{it})$.

3.2.4 A Spatio-temporal Modelling Framework

To extend the spatial model (3.1) of Section 3.2.1 to a spatio-temporal setting, we cast our modelling in the form of a dynamic spatial process model (see Banerjee et al. 2004 for parametric hierarchical modelling in this context, and for related references). We now view the log-rate process $z_{t,D} = \{z_t(s) : s \in D\}$ as a temporally evolving spatial process.

To develop a dynamic formulation, we begin, as in Section 3.2.1, by writing $z_t(s) = \mu_t + \theta_t(s)$ and add temporal structure to the model through *transition equations* for the $\theta_t(s)$, say,

$$\theta_t(s) = \nu\theta_{t-1}(s) + \eta_t(s), \tag{3.4}$$

where, in general, $|\nu| < 1$, and the innovations $\eta_{t,D} = \{\eta_t(s) : s \in D\}$ are independent realizations from a spatial stochastic process. We can now define the nonparametric prior for the block averages $\eta_{it} = |B_i|^{-1}\int_{B_i}\eta_t(s)ds$ of the $\eta_{t,D}$ surfaces following the approach of Section 3.2.1 or, equivalently, of Section 3.2.2. Proceeding with the latter, we assume that the $\eta_{t,D}$, given G_D, are independent from G_D, and assign a SDP prior to G_D with parameters α and $G_{0D} = \mathrm{GP}(0, \sigma^2\rho(||s - s'||; \phi))$. Marginalizing G_D over its prior, the induced prior, $p(\eta_1, ..., \eta_T \mid \sigma^2, \phi)$, for the $\eta_t = (\eta_{1t}, ..., \eta_{nt})$ is given by (3.3) (with η_t replacing θ_t). Block averaging the surfaces in the transition equations (3.4), we obtain $\theta_t = \nu\theta_{t-1} + \eta_t$, where $\theta_{t-1} = (\theta_{1,t-1}, ..., \theta_{n,t-1})$. Adding, as before, the i.i.d. $\mathrm{N}(0, \tau^2)$ terms to the z_{it}, we obtain the following general form for the

spatio-temporal hierarchical model

$$\begin{aligned} y_{it} \mid z_{it} &\overset{ind.}{\sim} \text{Po}(y_{it}|n_{it}\exp(z_{it})), \;\; i=1,...,n, \;\; t=1,...,T \\ z_{it} \mid \mu, \theta_{it}, \tau^2 &\overset{ind}{\sim} \text{N}(z_{it} \mid \mu+\theta_{it}, \tau^2), \;\; i=1,...,n, \;\; t=1,...,T \\ \theta_t =& \nu\theta_{t-1} + \eta_t \\ \eta_1,...,\eta_T \mid \sigma^2, \phi \sim& p(\eta_1,...,\eta_T \mid \sigma^2, \phi). \end{aligned} \tag{3.5}$$

The specification for the μ_t will depend on the particular application. For instance, the μ_t could be i.i.d., say, from a $\text{N}(0, \sigma_\mu^2)$ distribution (with random σ_μ^2), or they could be explained through a parametric function $h(t;\beta)$, say, a polynomial trend, $h(t;\beta) = \beta_0 + \sum_{j=1}^m \beta_j t^j$, or the autoregressive structure could be extended to include the μ_t, say, $\mu_t = \nu_\mu \mu_{t-1} + \gamma_t$, with $|\nu_\mu| < 1$, and γ_t i.i.d. $\text{N}(0, \sigma_\mu^2)$. For the Ohio state lung cancer data (discussed in Section 3.4.2), we work with a linear trend function $\mu_t = \beta_0 + \beta_1 t$. We set $\theta_1 = \eta_1$, i.e., $\theta_0 = 0$ (alternatively, an informative prior for θ_0 can be used). We choose priors for τ^2, σ^2 and ϕ as in model (3.2); we take independent normal priors for the components of β; and a discrete uniform prior for ν.

3.3 Posterior Inference and Prediction

We discuss here the types of posterior inference that can be obtained based on the models in Section 3.2. In particular, Section 3.3.1 comments on the (smoothed) inference for the disease rates while, under the dynamic model, Section 3.3.2 discusses forecasting of disease rates using the extension of Section 3.2.4.

3.3.1 Spatial Model

As is evident from expression (3.3), the DP prior induces a clustering in the θ_t (in their prior and hence also in the posterior for model (3.2). Let T^* be the number of distinct θ_t in $(\theta_1,...,\theta_T)$ and denote by $\theta^* = \{\theta_j^* : j = 1,...,T^*\}$ the vector of distinct

values. Defining the vector of configuration indicators, $w = (w_1, ..., w_T)$, such that $w_t = j$ if and only if $\theta_t = \theta_j^*$, (θ^*, w, T^*) yields an equivalent representation for $(\theta_1, ..., \theta_T)$. Denote by ψ the vector that includes (θ^*, w, T^*) and all other parameters of model (3.2). Draws from the posterior $p(\psi \mid \text{data})$, where data $= \{(y_{it}, n_{it}) : i = 1, ..., n, t = 1, ..., T\}$, can be obtained using the Gibbs sampler discussed in Appendix B.

The multivariate density estimate for the vector of log-rates associated with the subregions B_i is given by the posterior predictive density for a *new* $z_0 = (z_{10}, ..., z_{n0})$,

$$p(z_0 \mid \text{data}) = \int \int p(z_0 \mid \theta_0, \mu, \tau^2) p(\theta_0 \mid \theta^*, w, T^*, \sigma^2, \phi) p(\psi \mid \text{data}). \tag{3.6}$$

Here, $p(z_0 \mid \theta_0, \mu, \tau^2)$ is a $\mathrm{N}_n(\mu 1_n + \theta_0, \tau^2 I_n)$ density, $\theta_0 = (\theta_{10}, ..., \theta_{n0})$ is the vector of spatial random effects corresponding to z_0, and

$$p(\theta_0 \mid \theta^*, w, T^*, \sigma^2, \phi) = \frac{\alpha}{\alpha + T} \mathrm{N}_n(\theta_0 \mid 0, \sigma^2 R_n(\phi)) + \frac{1}{\alpha + T} \sum_{j=1}^{T^*} T_j \delta_{\theta_j^*}(\theta_0), \tag{3.7}$$

where T_j is the size of the j-th cluster θ_j^*. Therefore, $p(z_0 \mid \text{data})$ arises by averaging the mixture

$$\frac{\alpha}{\alpha + T} \mathrm{N}_n(z_0 \mid \mu 1_n, \tau^2 I_n + \sigma^2 R_n(\phi)) + \frac{1}{\alpha + T} \sum_{j=1}^{T^*} T_j \mathrm{N}_n(z_0 \mid \mu 1_n + \theta_j^*, \tau^2 I_n)$$

with respect to the posterior of ψ. Hence, the model has the capacity to capture, through the mixing in the θ_j^*, non-standard features in the distribution of log-rates over the regions.

3.3.2 Spatio-temporal Model

Turning to the spatio-temporal model of Section 3.2.4, let $\mu_t = \beta_0 + \beta_1 t$ (as in the example of Section 3.4.2). Denoting by $\psi = (\beta_0, \beta_1, \tau^2, \nu, \sigma^2, \phi, \{(z_t, \eta_t) : t =$

$1, ..., T\}$) the parameter vector corresponding to model (3.5), the posterior $p(\psi|\text{data})$ is proportional to

$$p(\beta_0)p(\beta_1)p(\nu)p(\tau^2)p(\sigma^2)p(\phi)\times$$
$$p(\eta_1, ..., \eta_T \mid \sigma^2, \phi) \prod_{t=1}^{T} N_n(z_t|\lambda_t, \tau^2 I_n) \prod_{i=1}^{n}\prod_{t=1}^{T} Po(y_{it}|n_{it}\exp(z_{it})), \tag{3.8}$$

where $\lambda_t = (\beta_0 + \beta_1 t)1_n + \sum_{\ell=1}^{t} \nu^{t-\ell}\eta_\ell$. The Gibbs sampler given in Appendix B can be used to obtain draws from $p(\psi|\text{data})$. For instance, of interest might be inference for z_t, the vector of log-rates corresponding to specific time periods t. Moreover, given the temporal structure of model (3.5), our interest is temporal forecasting for disease rates at future time points. In particular, the posterior forecast distribution for the vector of log-rates z_{T+1} at time $T+1$,

$$p(z_{T+1}|\text{data}) =$$
$$\int p(z_{T+1}|\eta_1, ..., \eta_T, \eta_{T+1}, \beta_0, \beta_1, \nu, \tau^2)p(\eta_{T+1}|\eta_1, ..., \eta_T, \sigma^2, \phi)p(\psi|\text{data})$$

where $p(z_{T+1}|\eta_1, ..., \eta_T, \eta_{T+1}, \beta_0, \beta_1, \nu, \tau^2)$ is an n-variate normal distribution with mean vector $(\beta_0 + \beta_1(T+1))1_n + \sum_{\ell=1}^{T+1} \nu^{T+1-\ell}\eta_\ell$ and covariance matrix $\tau^2 I_n$, and $p(\eta_{T+1}|\eta_1, ..., \eta_T, \sigma^2, \phi)$ can be expressed as in (3.7) by replacing θ_0 with η_{T+1} and using the, analogous to (θ^*, w, T^*), clustering structure in the $(\eta_1, ..., \eta_T)$.

3.4 Data Illustrations

Our data consists of the number of annual lung cancer deaths in each of the 88 counties of Ohio from 1968 to 1988. The population of each county is also recorded. Figure 3.1 depicts the geographical locations and neighborhood structure of the 88 counties in Ohio. The county location, area, and polygons are obtained from the "map" package in R.

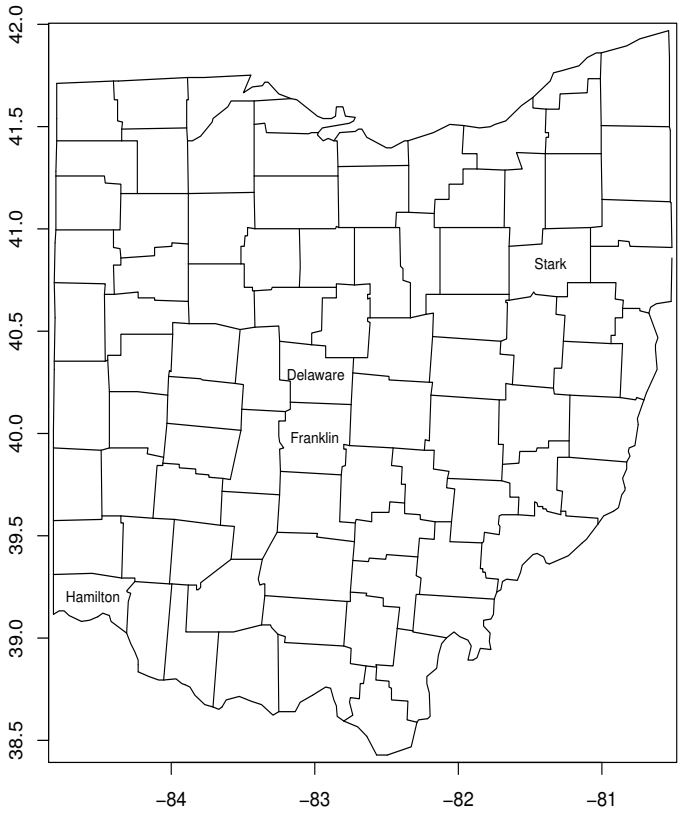

Figure 3.1: Map of the 88 counties in the state of Ohio.

Regarding prior specification, for both models (3.1) and (3.5) we work with an exponential correlation function, $\rho\left(||s-s'||;\phi\right) = \exp\left(-\phi||s-s'||\right)$. For both data examples, the discrete uniform prior for ϕ takes values in $[0.001, 1]$, corresponding to the range from 3 to 3000 miles; σ^{-2} and τ^{-2} have gamma$(0.1, 0.1)$ priors (with mean 1); and α is set equal to 1 (results were practically identical under $\alpha = 5$ and $\alpha = 10$). Finally, the normal priors for μ (Section 3.3.1) and for β_0 and β_1 (Section 3.3.2) have mean 0 and large variance (there was very little sensitivity to choices between 10^2 and 10^8 for the prior variance).

We observed very good mixing and fast convergence in the implementation of the Gibbs samplers discussed in Appendix B. In both of our simulation and Ohio lung cancer example below, we obtain 15,000 samples from the Gibbs sampler, and discard the first 3,000 samples as burn-in. We use 3,000 subsamples from the remaining 12,000 samples, with thinning equal to 4, for our posterior inference.

3.4.1 Simulation Example

We illustrate the fitting of our spatial model in (3.1) – (3.3) with a simulated data set for the 88 counties of Ohio. We simulate the areal incidence rate from a two-component mixture of multivariate normal distributions whose correlation matrix is calculated by block averaging isotropic GPs. The GPs cover the entire area of Ohio. The induced correlation matrix of the 88 blocks is computed by Monte Carlo integration.

In particular, we proceed as follows. For $i = 1, ..., 88$ and $t = 1, ..., T$ (with $T = 40$), we first generate z_{it} independent $\mathrm{N}(\mu + \theta_{it}, \tau^2)$ and, then, y_{it} independent $\mathrm{Po}(n_i \exp(z_{it}))$, where n_i is the population of county i in 1988. The distribution of the spatial random effects $\theta_t = (\theta_{1t}, \ldots, \theta_{nt})$ arises through a mixture of two block-averaged GP's. In particular, for $\ell = 1, 2$, let $\theta^{(\ell)} = \left(\theta_1^{(\ell)}, \ldots, \theta_n^{(\ell)}\right) \sim \mathrm{N}_n((-1)^\ell \mu_\theta 1_n, \sigma_\ell^2 R)$, with the (i, j)-th element of the correlation matrix R given by $|B_i|^{-1}|B_j|^{-1} \int_{B_i} \int_{B_j} \exp\left(-\phi||s - s'||\right) dsds'$. Then, each θ_t is independently sampled from $0.5\theta^{(1)} + 0.5\theta^{(2)}$. The values of the parameters are $\mu = -6.5$, $\mu_\theta = 0.5$, $\sigma_1^2 = \sigma_2^2 = 1/32$, $\tau^2 = 1/256$, and $\phi = 0.6$. Under these choices, marginally, each θ_{it} has a bimodal distribution of the form $0.5\mathrm{N}\left(-\mu_\theta, \sigma_1^2\right) + 0.5\mathrm{N}\left(\mu_\theta, \sigma_2^2\right)$.

We fit model (3.1) to this data set. The Bayesian goodness of fit is illustrated with univariate and bivariate posterior predictive densities for the log-rates, which are estimated using (3.6). In Figure 3.2 we compare the true densities of the model from

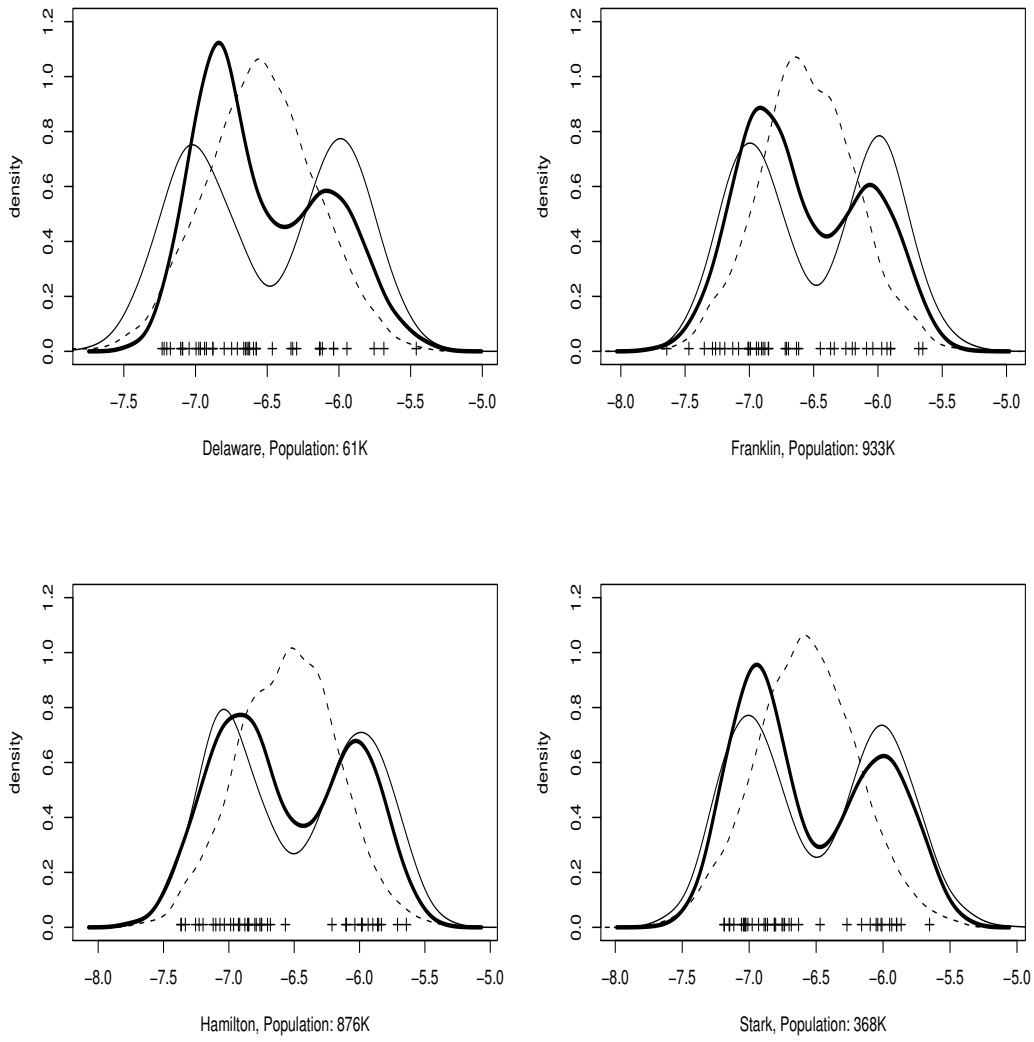

Figure 3.2: For the simulation example in Section 3.4.1, posterior predictive densities for the log-rates, corresponding to four counties, based on the SDP model (thick curves) and the GP model (dashed curves). The true densities are denoted by the thin curves, and the observed log-rates by "+".

which we simulated the data with the SDP model posterior predictive densities for four selected counties. They are "Delaware" and "Franklin" in central Ohio, "Hamilton" in southwest, and "Stark" in northeast. "Franklin" includes Columbus and "Hamilton" includes Cincinnati so these are highly populated counties. "Delaware"

is more suburban and "Stark" is very rural (see Figure 3.1).

The "+" mark the values of the 40 observed log-rates $\log(y_{it}/n_i)$ in each of these four counties. In addition, Figure 3.2 includes posterior predictive densities from a parametric model based on a $\mathrm{GP}(0, \sigma^2 \exp(-\phi||s - s'||))$ for the spatial random effects surfaces. This specification results in a limiting version of model (3.1) (for $\alpha \to \infty$) where the θ_t, given σ^2 and ϕ, are i.i.d. $\mathrm{N}_n(0, \sigma^2 R_n(\phi))$. The SDP model clearly outperforms the GP model with regard to posterior predictive inference.

Next, we pair the four counties above to show in Figure 3.3 the predictive joint densities, based on the SDP model, and, again, to compare with the true joint densities (using samples in both cases). The first pair "Delaware" and "Franklin" are next to each other. The second pair "Hamilton" and "Stark" are distant. We note that, with only 40 replications, our model captures quite well both marginal and joint densities for the log-rates.

3.4.2 Ohio Lung Cancer Data

The exploratory study of the Ohio lung cancer mortality data reveals a spatio-temporal varying structure in the incidence rates. We display the observed log-rates $\log(y_{it}/n_{it})$ for the aforementioned four counties in Figure 3.4. This plot shows clear evidence of an increasing, roughly linear, trend in the log-rate. Therefore we apply the dynamic SDP model (3.5) with a linear trend over time, setting $\mu_t = \beta_0 + \beta_1 t$. Moreover, because negative values for ν do not appear plausible, we use a discrete uniform prior on $[0, 1)$ for ν.

The time t is normalized to be from year $t = 1$ to 21. In order to validate our model, we leave year 21 (year 1988) out in our model fitting and predict the log-rates for all 88 counties in that year, using the posterior forecast distribution developed in Section 3.3.2. Posterior point (posterior medians) and 95% equal-

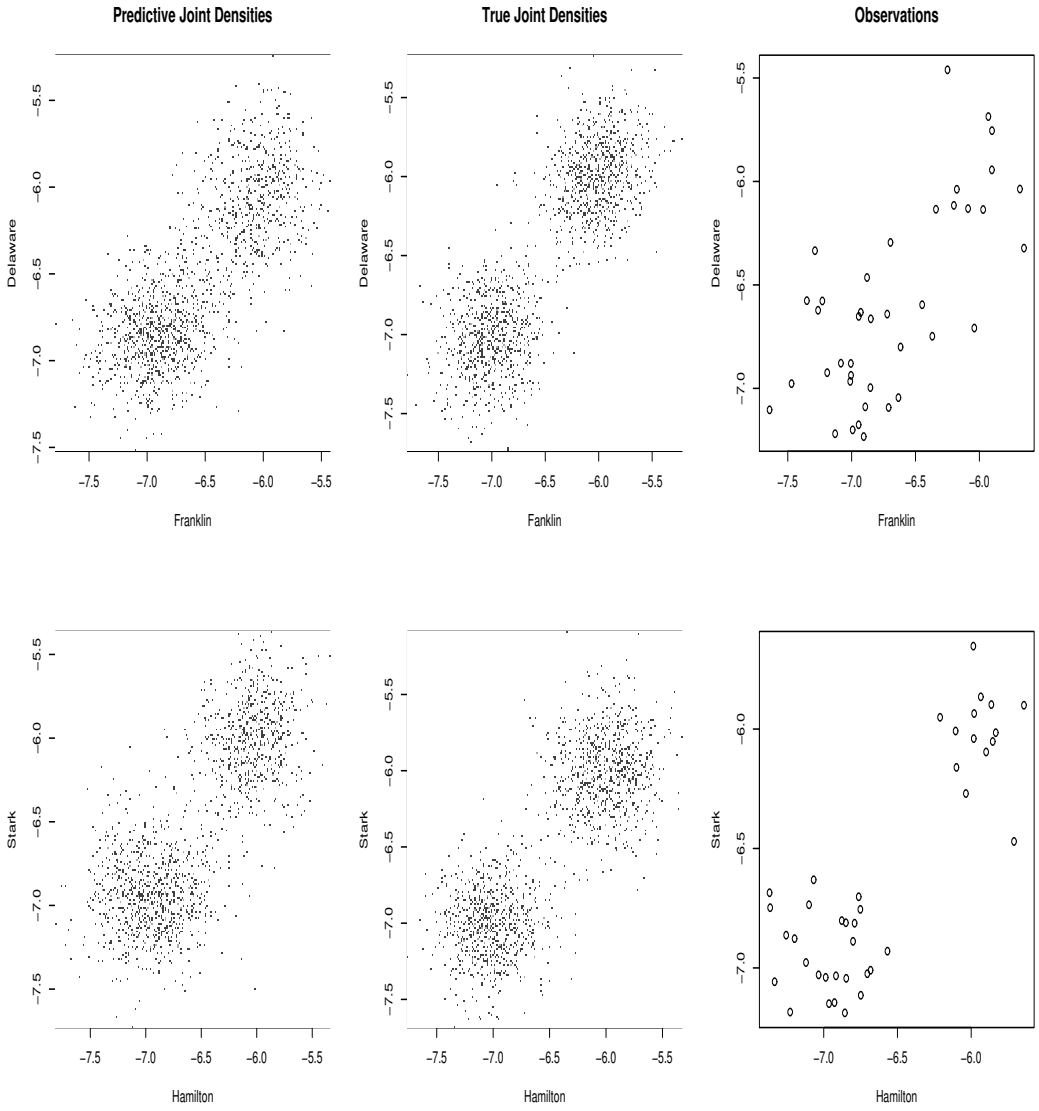

Figure 3.3: For the simulation example of Section 3.4.1, posterior predictive densities (left column) and true bivariate densities (middle column) for log-rates associated with two pairs of counties. The right column includes plots of the corresponding observed log-rates.

tail interval estimates for β_0, β_1 and for ν are given by -8.208 $(-8.319, -8.100)$, 0.0367 $(0.0292, 0.0448)$ and 0.7 $(0.6, 0.8)$, respectively. There was also prior to posterior learning for the other hyperparameters, in particular, point and interval estimates were 0.0586 $(0.0552, 0.0656)$ for ϕ; 0.104 $(0.0855, 0.113)$ for τ^2; and 0.133

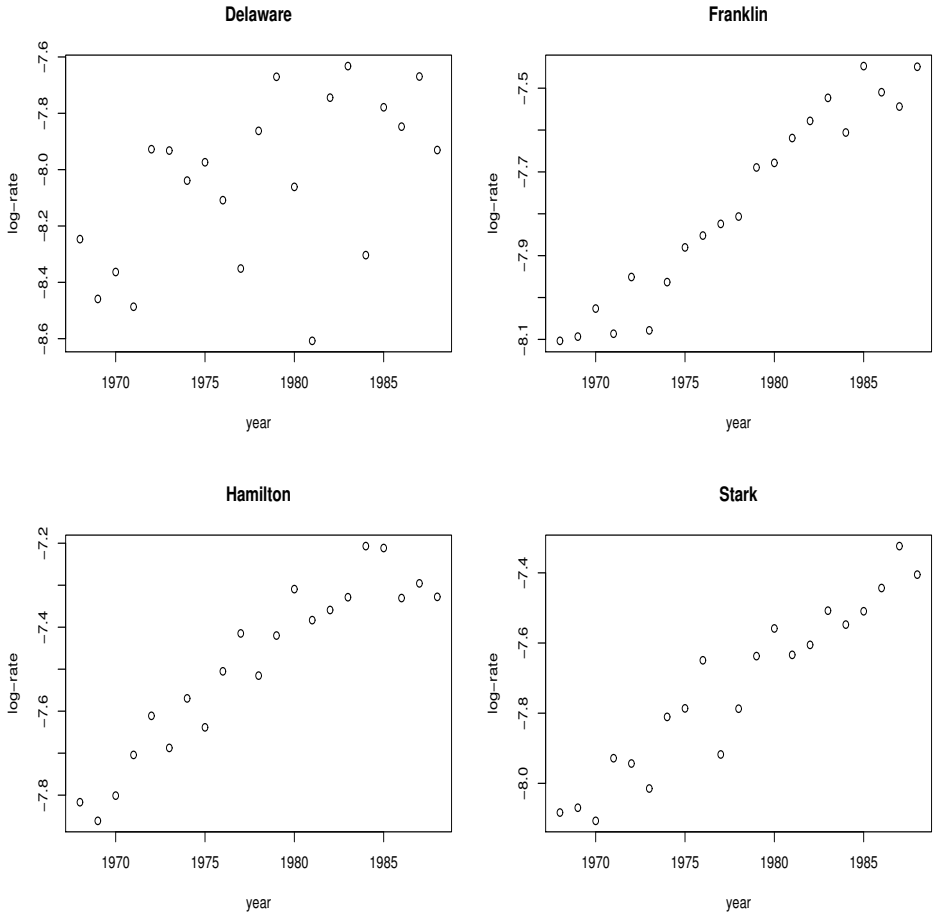

Figure 3.4: Observed log-rates for four counties from 1968 to 1988 for the Ohio data example of Section 3.4.2.

$(0.101, 0.152)$ for σ^2.

In Figure 3.5 we display the marginal posterior forecast density of the log-rate for the earlier four counties in the hold-out year 1988. We also calculated 95% marginal predictive intervals for all 88 counties in 1988 and found that 83 out of 88 observed log-rates (94.3 %) are within their 95% interval; we do not seem to be overfitting or underfitting. In Figure 3.6 we provide the contour plot of the predictive log-rate surface for 1988, using medians from the posterior forecast distribution for each county.

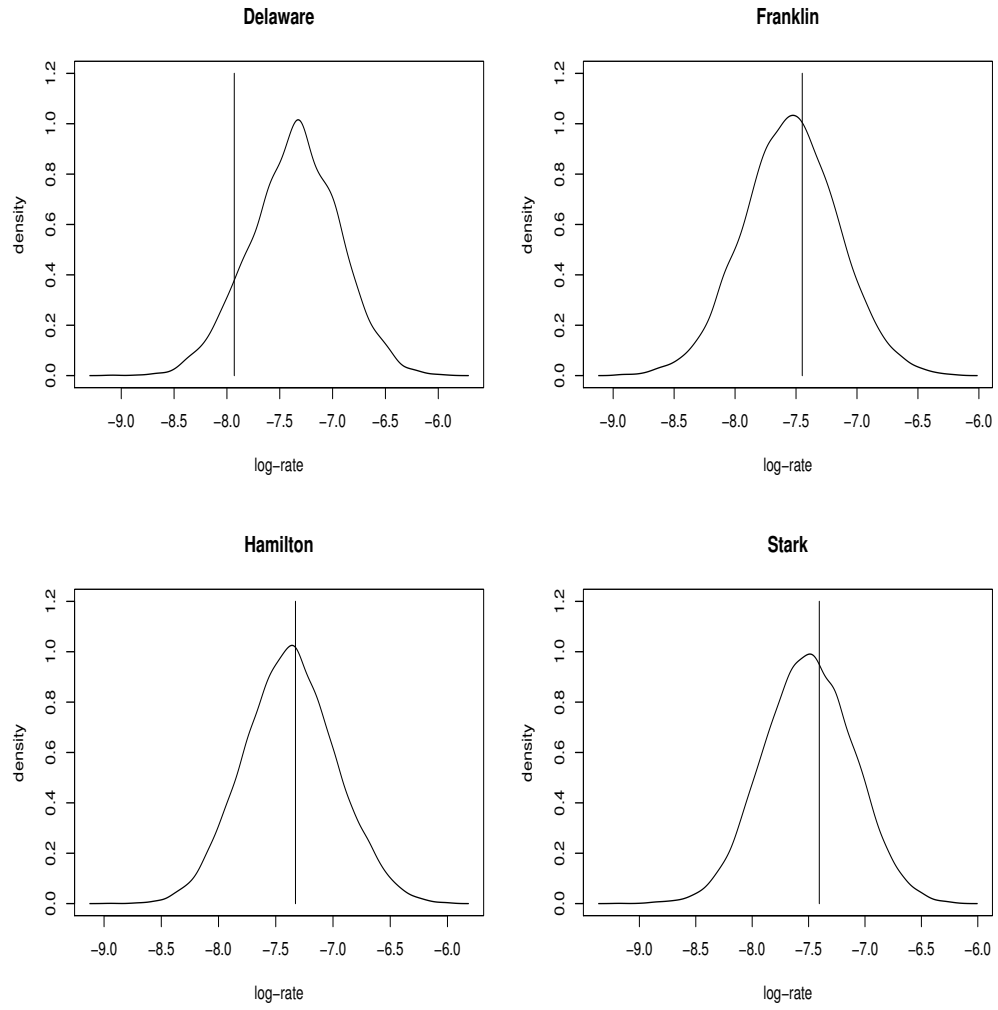

Figure 3.5: Posterior forecast densities for the log-rate of four counties in the hold-out year (year 1988) for the Ohio data example of Section 3.4.2. The vertical line in each plot is the observed log-rate.

3.5 Discussion

We have argued that, with regard to disease mapping, it may be advantageous to conceptualize the model as a spatial point process rather than through more customary areal unit spatial dependence specifications. Aggregation of the point process to suitable spatial units enables us to use it for the observable data. Specifying a non-homogeneous point process requires a model for the latent risk surface. Here,

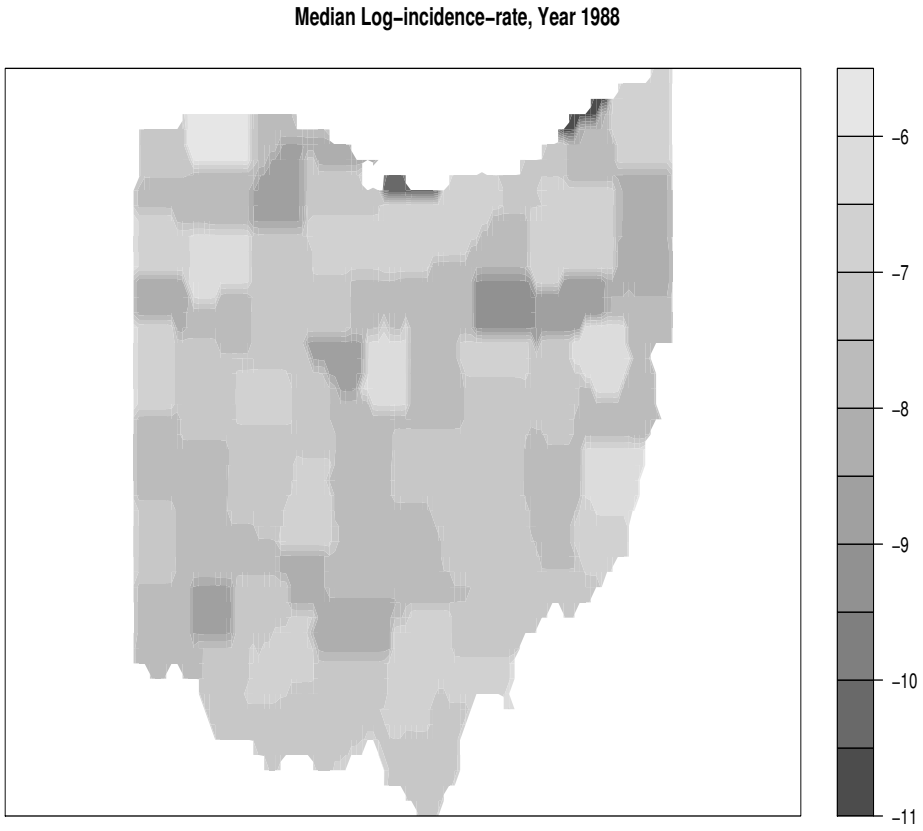

Figure 3.6: For the Ohio data example of Section 3.4.2, medians of the posterior forecast distribution for the log-rate in each county for year 1988.

we have argued that there are advantages to viewing this surface as a process realization rather than through parametric modelling. But then, the flexibility of a nonparametric process model as opposed to the limitations of a stationary GP model becomes attractive. The choice of a spatial DP finally yields our proposed approach. We applied the modelling to both real and simulated data. With the simulated data we clearly demonstrated the advantage of such flexibility.

Extensions in several directions may be envisioned. Three examples are the following. In treating the specification for the μ_t we could provide a nonparametric

model as well through i.i.d. realizations obtained under DP mixing or the associated dynamic version with independent innovations under such a model. Next, we often study concurrent disease maps to try to understand the pattern of joint incidence of diseases. In our setting, for a pair of diseases, this would take us to a pair of dependent surfaces from a bivariate spatial process. We could envision modelling based upon a bivariate SDP centered around a bivariate GP. Finally, how would we handle misalignment issues in this nonparametric setting? That is, what should we do if disease counts are observed for one set of areal units while covariate information is supplied for a different set of units? Banerjee et al. (2004) suggest strategies for treating misalignment but exclusively in the context of GP. Extensions to our SDP setting would be useful.

Chapter 4

Space-time Modelling Using Differential Equations with Application to Urban Development

In this chapter, we develop our spatio-temporal point process whose dynamic intensity is modelled with stochastic differential equations. In Section 4.1, we first review the Cox process model, then propose a structured model for the evolution of latent intensity surfaces over time. Deterministic and stochastic differential equation versions of the growth mechanism are introduced and discussed. In Section 4.2, we formulate a Bayesian hierarchical model based on the theory in Section 4.1 and propose a process convolution approximation. Section 4.3 addresses Bayesian estimation and space-time prediction issues. Section 4.4 provides two illustrative examples, one of which is the real urban growth data of Irving, TX. We conclude with a discussion on future extensions of the current model in Section 4.5.

4.1 Spatio-temporal Point Processes Models with Dynamic Intensity

4.1.1 Space-time Cox Process Model

The motivating problem for our research is to model urban development by the construction of new residential houses. Figure 4.1 shows the residential buildings in Irving, TX from 1951–1968 in four different years. The exact locations and times of new constructions form a point pattern over space and time.

Our fundamental approach to the urban development problem is the space-time Cox process (Cox 1955), i.e,. conditioning on a realization of a positive-valued stochastic process, in a fixed region, the emergence of newly observed points over a certain period of time behaves according to an inhomogeneous Poisson process. If we only model the locations, we can use the univariate Cox Process. However, we might have additional information at the locations, a so-called marked point process. For instance, we might note whether the construction was a single family home, an apartment building, a commercial site, etc. We might expect that commercial buildings have a different distributional pattern over space and time from residential buildings. These associated overlapping and correlated point patterns can be modelled by the multivariate Cox processes. Here, we will only consider the univariate case. The multivariate version is a conceptually straightforward extension though computation will be much more demanding (and fitting our univariate model is already quite challenging).

Focusing on the urban development setting, let D be the study region, which would be some metropolitan area or a portion thereof. Depending upon the window of time, it might include primarily urban area or with a later (or longer) window, the suburban and rural areas surrounding it. Let $N_T(D)$ be the number of houses

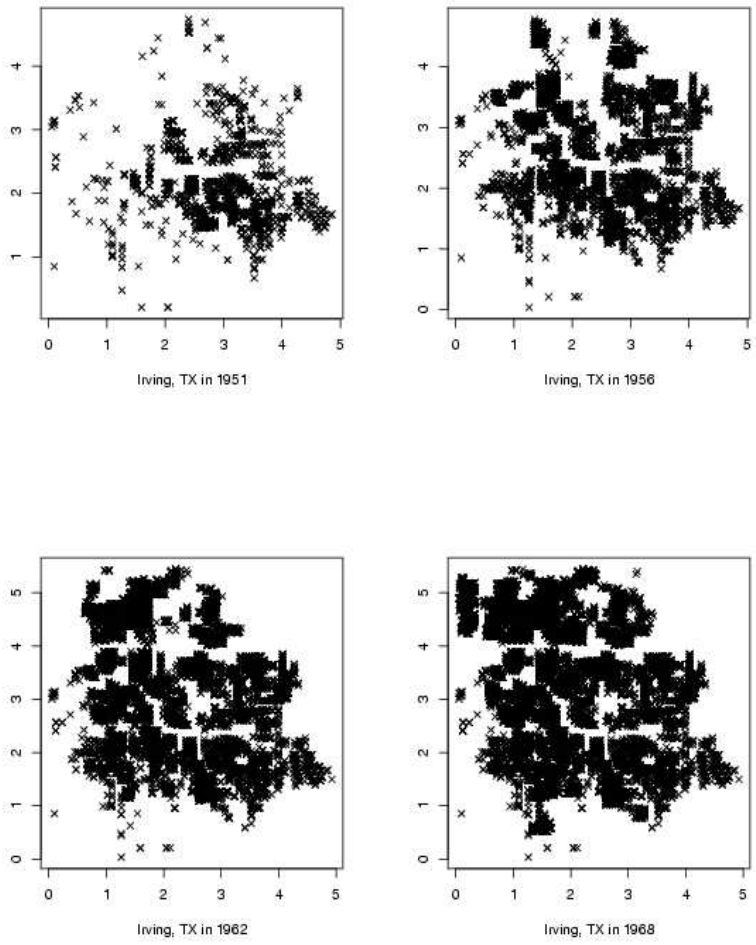

Figure 4.1: Residential houses in Irving, TX.

constructed in the period from $t = 0$ to T and $X_T = \left\{x_{1,t_1}, \ldots, x_{N_T,t_{N_T}}\right\}$ be the set of locations and times of these new constructions. The intensity of the space-time point process model for X_T is $\Omega(t,s), s \in D, t \in [0,T]$, which is a positive-valued function. $\Omega(t,s)$ could be viewed as essentially being a nonparametric specification which, for instance, could be a realization of a space-time process over $D \times [0,T]$ (Nonparametric

functions using basis representations would also be possible.) However, in the present work, we choose to view $\Omega(t,s)$ as having a parametric form that is motivated by mechanistic or theoretical considerations. We want to introduce specific parametric choices where the parameters quantify relationships and inference is sought about these parameters. Indeed, these parameters themselves will be associated with spatial locations and so will themselves be viewed as realizations of spatial processes; thus, we can see how they vary over D.

More specifically, let

$$\Omega(t,s) = f(t, \theta_l(t,s); l = 1, \ldots, p) \tag{4.1}$$

where $\theta_l(t,s); l = 1, \ldots, p$ are p possibly dependent space-time processes. Again, at any location s, the $\theta_l(t,s)$ are the values of the parameters or latent variables which determine $\Omega(t,s)$. Since they are realizations of spatial processes, the $\theta_l(t,s)$ can flexibly capture spatial variation and correlation .

The emergence of new houses is conceptually a problem of continuous-time spatio-temporal point process, because a house is *virtually* there when the land is acquired and the blueprint drawn. However, as buildings are not constructed instantaneously, we only observe new construction over an interval of time. Therefore when using space-time point processes to model house construction, we discretize time and presume that there are only a finite number of periods. In each period there is a finite random set of observed locations, which, altogether, can be treated as a spatial point process. The intensity surface governing that process can still be considered as evolving over continuous time. Expressed in mathematical terms, the chance of an event occurring at any specified time is 0; in order to *observe* a point pattern we have to consider an interval of time.

If we discretize the space-time point process in time, the intensity for any spatial point process, say $X_{[t_1,t_2]}$ consisting of locations emerging in a sub-period $t \in [t_1, t_2)$

($t_1 \geq 0$ and $t_2 \leq T$) is

$$\int_{t_1}^{t_2} \Omega\left(t,s\right) dt = \int_{t_1}^{t_2} f\left(t, \theta_l(t,s); l = 1, \ldots, p\right) dt \tag{4.2}$$

Temporally dependent spatial point processes $X_{[t_1=0,t_2]}, X_{[t_2,t_3]}, \ldots, X_{[t_{J-1},t_J=T]}$ will provide a good approximation to the spatio-temporal point process X_T, when the time intervals are sufficiently small. Moreover, this also allows us to approximate the intensity (4.2) by

$$\int_{t_1}^{t_2} \Omega\left(t,s\right) dt \approx \left(t_2 - t_1\right) \Omega\left(t_1,s\right) = \left(t_2 - t_1\right) f\left(t_1, \theta_l(t_1,s); l = 1, \ldots, p\right) \tag{4.3}$$

In the next subsection we introduce the dynamics for the cumulative spatial intensity $\Lambda\left(t,s\right) = \int_0^t \Omega\left(\tau,s\right) d\tau$ and use the approximation (4.3) to deduce the dynamics for the discrete-time spatial point process.

4.1.2 Modelling Intensity Surface Dynamics

Theory from mathematical ecology and sociology argues that the growth of human population, despite the transcendent intricacies of its infrastructure, bears resemblance in the macro scale to that of any other biological species. We do not attempt to model arrival of people to a metropolitan area using our space-time point process. Rather, we use the construction of single family homes as a surrogate process. Moreover, urban growth is customarily described in aggregate, e.g., at the city-level (e.g. Glaeser 2003; Rossi-Hansberg and Wright 2005 and the references therein). We would argue that providing point referenced modelling for such building enables assessment of urban development at previously unachievable spatial resolution. Additionally, although we do not connect the intensity surface for our "house-building" Cox process to population growth, it is plausible to assume that the spatial intensity evolves with dynamics similar to those of population growth.

Three typical growth models used in ecology and described through differential equations (indexed by location) are

- Exponential growth

$$\frac{d\Lambda(t,s)}{dt} = r(s)\,\Lambda(t,s) \tag{4.4}$$

- Logistic growth

$$\frac{d\Lambda(t,s)}{dt} = r(s)\Lambda(t,s)\left[1 - \frac{\Lambda(t,s)}{K(s)}\right] \tag{4.5}$$

- Gompertz growth

$$\frac{d\Lambda(t,s)}{dt} = r(s)e^{-\alpha(s)t}\Lambda(t,s) \tag{4.6}$$

By introducing a point-referenced spatial component into these equations, there is a differential equation at every spatial location. Moreover, each of these differential equations is random. Our modelling will ensure that the differential equations for locations close to each other will tend to be more similar than those for locations farther apart. Note further that these differential equation models for the cumulative spatial intensity $\Lambda(t,s)$ imply integral equation models for the space-time intensity $\Omega(t,s)$. For example, the logistic growth can be viewed as a nonlinear integral equation:

$$\Omega(t,s) = r(s)\int_0^t \Omega(\tau,s)\,d\tau\left[1 - \frac{\int_0^t \Omega(\tau,s)}{K(s)}\right]. \tag{4.7}$$

In all three of the models above, $r(s)$ denotes the local growth rate across space. The apparent shortcoming of exponential growth with regard to house construction is that the cumulative intensity is not bounded. The logistic growth model uses a carrying capacity $K(s)$ to bound this intensity by making the growth rate decrease to zero when the intensity approaches the capacity. For residential house construction

such a model seems to be plausible and the notion of a carrying capacity has a natural interpretation so we adopt it as our illustrative focus. The Gompertz growth instead uses an intrinsic decay parameter α to control for the infinite growth problem and could be considered as well (see Section 4.1.3 for further discussion).

So, in using the logistic growth model for the intensity, we propose a local growth rate $r(s)$ and carrying capacity $K(s)$ and impose spatial process models on these parameters. In reality, the local growth rate $r(s)$, the local carrying capacity $K(s)$ and the intensity $\Lambda(t,s)$ only have physical interpretations when associated with a region. If we consider the regional growth rate $r(D)$ as the average of the local rate

$$r(D) = \frac{1}{|D|}\int_D r(s)\,ds$$

where $|D|$ denotes the area of D; and the regional capacity is the aggregate of the local capacity

$$K(D) = \int_D K(s)\,ds$$

the global growth in the regional D should have the following dynamics:

$$\frac{d\Lambda(t,D)}{dt} = r(D)\Lambda(t,D)\left[1 - \frac{\Lambda(t,D)}{K(D)}\right] \tag{4.8}$$

where

$$\Lambda(t,D) = \int_D \Lambda(t,s)\,ds.$$

Our local model for one location s can be considered the limit of this global model when the area surrounding the location s goes to zero. Let δ_s be the neighborhood

of s. The limits of the terms in the equation are

$$\begin{aligned}
\lim_{|\delta_s|\to 0} \frac{\Lambda(t,\delta_s)}{|\delta_s|} &= \lim_{|\delta_s|\to 0} \frac{1}{|\delta_s|} \int_{\delta_s} \Lambda(t,s')\, ds' = \Lambda(t,s)\,; \\
\lim_{|\delta_s|\to 0} \frac{K(\delta_s)}{|\delta_s|} &= \lim_{|\delta_s|\to 0} \frac{1}{|\delta_s|} \int_{\delta_s} K(s')\, ds' = K(s)\,; \\
\lim_{|\delta_s|\to 0} r(\delta_s) &= \lim_{|\delta_s|\to 0} \frac{1}{|\delta_s|} \int_{\delta_s} r(s')\, ds' = r(s)\,.
\end{aligned}$$

Therefore the limit of the global growth equation (4.8) is

$$\begin{aligned}
\lim_{|\delta_s|\to 0} \frac{d\Lambda(t,\delta_s)/|\delta_s|}{dt} &= \lim_{|\delta_s|\to 0} r(\delta_s) \frac{\Lambda(t,\delta_s)}{|\delta_s|} \left[1 - \frac{\Lambda(t,\delta_s)/|\delta_s|}{K(\delta_s)/|\delta_s|}\right] \Rightarrow \\
\frac{d\Lambda(t,s)}{dt} &= r(s)\Lambda(t,s)\left[1 - \frac{\Lambda(t,s)}{K(s)}\right]
\end{aligned}$$

which is exactly our local model. In other words, our interpretations for $r(s)$, $K(s)$ and $\Lambda(t,s)$ provide coherent behavior when integrated across D or any portion of D.

In order to capture the spatial variation and suitable dependence for inhomogeneous growth, we assume the local growth rate and capacity have very flexible spatial process models:

$$r(s) = \exp\left(\mu_r(s;\beta_r) + \theta_r(s)\right) \tag{4.9}$$

$$K(s) = \exp\left(\mu_K(s;\beta_K) + \theta_K(s)\right) \tag{4.10}$$

where $\mu_r(s;\beta_r)$ and $\mu_K(s;\beta_K)$ are trend surfaces and $\theta_r(s)$ and $\theta_K(s)$ are mean-zero spatial random effects modelled as realizations of certain spatial processes, such as stationary Gaussian random fields (Cressie 1993; Banerjee et al. 2004) or nonparametric spatial Dirichlet processes (Gelfand et al. 2005).

Returning to (4.4), (4.5) and (4.6) above, we assume the intensity at any location follows the same type of dynamics with spatially varying parameters, such as

$r(s)$ and $K(s)$. The resulting models for the entire study region emerge as infinite-dimensional dynamic systems (Itô 1984). In practice, we approximate them with spatially discretized versions, hence reduce them to multivariate dynamic systems.

Computationally, for the large number of houses we will be working with, it will not be tractable to work with the point process likelihood; rather, we will have to introduce a partition of D and employ a Poisson likelihood for the number of points observed in a given cell of the partition in a given time period. So, consider a subdivision of the study region D into M cells. For each cell $m \in \{1, \ldots, M\}$, the average intensity in this cell can be modelled with the average spatial parameters. For example the logistic model can be approximated by

$$\frac{d\Lambda(t,m)}{dt} = r(m)\Lambda(t,m)\left[1 - \frac{\Lambda(t,m)}{K(m)}\right] \tag{4.11}$$

where $r(m)$ and $K(m)$ are average growth rate and carrying capacity in each cell. The spatial variation and correlation of $r(m)$ and $K(m)$ are inherited from their spatial processes. Conversely, the continuous space model (4.5) can be viewed as the infinitesimal version of (4.11).

4.1.3 Diffusion Equation Models for the Growth Rate

Given $r(s)$ and $K(s)$, model (4.5) yields a deterministic space-time model for the evolution of the spatial and space-time intensity. That is, though we have a spatial process of curves $\Lambda(t,s)$, at any location, given the rate and the carrying capacity, the resulting growth trajectory is fixed. To enrich this behavior, we could try to introduce uncertainty into these trajectories. However, for the house construction problem, we insist that the cumulative intensity must always be nonnegative. A diffusion equation for the intensity achieved by adding a stochastic component to the differential equation need not preserve nonnegativity. Furthermore, not only must

$\Lambda(t,s)$ be nonnegative, but $\Omega(t,s)$ must be nonnegative as well. A typical diffusion equation, such as Feller's branching process:

$$d\Lambda(t,s) = r(s)\Lambda(t,s)\left[1 - \frac{\Lambda(t,s)}{K(s)}\right]dt + \sigma_\Lambda\sqrt{\Lambda(t,s)}dW_t(s) \tag{4.12}$$

where $W_t(s)$ is a spatial Brownian motion, can not guarantee this for the space-time intensity $\Omega(t,s)$. Therefore we choose not to apply a stochastic differential equation to the intensity itself, but instead to add temporal structure to the growth rate and thus formulate a spatial diffusion equation model for $r(s)$.

Let $r(t,s)$ be the time-varying growth rate. Because $r(t,s)$ is also positive, we will model the log growth rate $Z(t,s) = \log r(t,s)$. The Gompertz growth model can be viewed as an extension of the simplest exponential growth with the following differential equation for $Z(t,s)$:

$$dZ(t,s) = -\alpha(s)\,dt, \text{ with } z(0,s) = \log r(0,s). \tag{4.13}$$

If we add a stochastic component with spatial Brownian motion to (4.13), we have the following diffusion equation for $Z(t,s)$:

$$dZ(t,s) = -\alpha(s)\,dt + \sigma_z dW_t(s), \text{ with } Z(0,s) = \log r(0,s).$$

This resulting model for the growth rate $r(t,s)$ is the *geometric Brownian motion*:

$$r(t,s) = r(0,s)\exp\left\{\left[-a(s) + \frac{1}{2}\sigma_z^2\right]t + \sigma_z W(t,s)\right\},$$

which is not stationary.

For the logistic growth, because the crowding effect $[1 - \Lambda(s)/K(s)]$ already controls for the decreasing growth, it is much easier to specify a stationary model for the growth rate. We do so by letting $Z(t,s) = \log r(t,s)$ be a stationary mean-reverting

Ornstein-Uhlenbeck process (See Rossi-Hansberg and Wright 2005 for economic justifications.):

$$dZ(t,s) = [\alpha(s) - \zeta Z(t,s)]\,dt + \sigma_Z dW_t(s). \tag{4.14}$$

The physical interpretation of this model is that the log growth rate will eventually fluctuate about the deterministic level $\alpha(s)/\zeta$. As a result, $Z(t,s)$ is stationary space-time process with the following separable covariance function

$$cov(Z(t_1,s_1), Z(t_2,s_2)) = \sigma_Z^2 \exp(-\beta|t_1 - t_2|)\,\rho(s_1 - s_2;\phi) \tag{4.15}$$

where $\rho(s_1 - s_2;\phi)$ is the correlation function of the spatial Brownian motion $W_t(s)$.

For the spatially discrete model (4.11), the corresponding discrete version for the log growth rate over a subdivision of M cells is

$$dZ(t,m) = \left[\alpha^M - \zeta I_M Z(t,m)\right]dt + \sigma_Z dW_t(m) \tag{4.16}$$

where α^M is an M-dimensional vector with each element α_m^M being the average of $\alpha(s)$ in cell m and I_M is an $M \times M$ identity matrix. For more properties of the multivariate Ornstein-Uhlenbeck process, see Schach (1971). For a general treatment of multivariate stochastic differential equations, see Gard (1988), Itô (1984), Karatzas and Shreve (1991) and Oksendal (2002).

If after a sufficiently long time, $Z(t,s)$ will have converged and will only fluctuate slightly around $\alpha(s)/\zeta$, our deterministic model (4.5) with

$$Z(t,s) \approx \alpha(s)/\zeta \equiv \mu_r(s) + \theta_r(s)$$

is a good approximation to the space-time model. Note this approximation reduces the space-time model for the growth rate to the more parsimonious model (4.9). In the examples that we will present in Section 4.4, we only have very few periods (less than 20) of growth data that is insufficient to fit the model in (4.14), therefore the more parsimonious model (4.9) is deliberately selected. If the data came with many

periods (say more than 350 periods as in the simulated data example in Brix and Diggle 2001), we could attempt to fit the full model with dynamic growth rate. The following sections will focus on the logistic growth model (4.5) with only spatially varying growth rate.

4.2 Statistical Model Fitting and Inference

The dynamic Cox Process models proposed in Section 4.1.2 imply the following statistical hierarchical model: At the first stage, the space-time point patten x_T is a realization of a spatio-temporal Poisson Process X_T. The intensity $\Omega(t, s)$ is a function of the initial intensity

$$\Lambda(0, s) = \int_{-\infty}^{0} \Omega(\tau, s)\, d\tau \tag{4.17}$$

and the latent parameter processes describing the growth rate, such as $r(t, s)$, $K(s)$ etc. The latent parameter processes are realizations of exponential Gaussian processes.

The multi-period spatial Cox Process version of this dynamic model can be viewed as a temporally discretized approximation to the space-time Cox Process above. If we assume there are J evenly spaced periods $\{t_1, \ldots, t_J\}$ in $t \in [0, T]$, the corresponding sequential spatial point pattern x_{t_j} is a realization of the spatial Poisson Process with the intensity $\Delta\Lambda(t_j, s) = \int_{t_{j,1}}^{t_{j,2}} \Omega(\tau, s)\, d\tau$. In the following subsections, we discuss the approximation and discretization of the dynamic spatial intensity $\Delta\Lambda(t_j, s)$ and the statistical inference for the latent parameter processes that define $\Delta\Lambda(t_j, s)$.

4.2.1 The Discretized Cox Process Model

As noted in Section 4.1.2, if we observe the spatial point processes in evenly spaced time intervals Δt and the time intervals are small, the spatial intensity in period t_j

can be approximated as

$$\Delta\Lambda\left(t_j, s\right) = \int_{t_{j,1}}^{t_{j,2}} \Omega\left(\tau, s\right) d\tau \approx \Omega\left(t_{j,1}, s\right) \Delta t.$$

From (4.5), the dynamics of the discretized spatial intensity can be derived as a difference equation

$$\Delta\Lambda\left(t_j, s\right) = r(s)\Lambda\left(t_{j-1}, s\right)\left[1 - \frac{\Lambda\left(t_{j-1}, s\right)}{K\left(s\right)}\right]\Delta t, \tag{4.18}$$

where

$$\Lambda\left(t_j, s\right) = \int_0^{t_j} \Omega\left(\tau, s\right) d\tau \approx \Lambda\left(0, s\right) + \sum_{l=1}^{j} \Delta\Lambda\left(t_l, s\right). \tag{4.19}$$

$\Delta\Lambda\left(t_j, s\right)$ defines the intensity surface for the conditionally independent point processes $X\left(t_j\right), j = 1, \ldots, J$ in the disjoint periods $\{t_1, \ldots, t_J\}$. Equation (4.18) offers an explicit transition model for the intensity over time. The realization of $X\left(t_j\right)$ is a set $x\left(t_j\right)$ with the total number of points $n_j = \left|x\left(t_j\right)\right|$. We will now suppress the index t to use $x_j = \left\{x_{j1}, \ldots, x_{jn_j}\right\}$ as the points in period t_j and let $\Delta\Lambda_j = \left\{\Delta\Lambda\left(t_j, s\right), s \in D\right\}$ be its intensity surface. Note that we begin with x_0, the initial point pattern (i.e., we begin our investigation of growth at some time point after the city was founded), which provides necessary information for the initial intensity surface $\Lambda_0 = \left\{\Lambda\left(0, s\right), s \in D\right\}$.

Suppose we model the initial intensity Λ_0 as a realization of an log Gaussian spatial process,

$$\Lambda_0\left(s\right) = \exp\left(\mu_\Lambda\left(s\right) + \theta_\Lambda\left(s\right)\right), \theta_\Lambda\left(s\right) \sim \mathrm{GP}\left(0, C_\Lambda\left(s - s'; \phi_\Lambda\right)\right). \tag{4.20}$$

Then, we can construct a Bayesian hierarchical model for $x_0, \ldots, x_J$ with (4.9) and

(4.10) as follows:

$$
\begin{aligned}
x_j|\Delta\Lambda_j &\sim \text{Poisson}\left(D, \Delta\Lambda_j\right);\ \ j = 1, \ldots, J \\
x_0|\Lambda_0 &\sim \text{Poisson}\left(D, \Lambda_0\right) \\
\Delta\Lambda_j\left(s\right) &= r(s)\Lambda_{j-1}\left(s\right)\left[1 - \frac{\Lambda_{j-1}\left(s\right)}{K\left(s\right)}\right]\Delta t; s \in D;\ \ j = 1, \ldots, J \\
\Lambda_j\left(s\right) &= \Lambda_0\left(s\right) + \sum_{l=1}^{j-1}\Delta\Lambda_l\left(s\right) \\
\log\Lambda_0\left(s\right) &= \mu_\Lambda\left(s; \beta_\Lambda\right) + \theta_\Lambda\left(s\right) \\
\theta_\Lambda\left(s\right) &\sim \text{GP}\left(0, C_\Lambda\left(s - s'; \phi_\Lambda\right)\right); s, s' \in D \\
\log r\left(s\right) &= \mu_r\left(s; \beta_r\right) + \theta_r\left(s\right) \\
\theta_r\left(s\right) &\sim \text{GP}\left(0, C_r\left(s - s'; \phi_r\right)\right); s, s' \in D \\
\log K\left(s\right) &= \mu_K\left(s; \beta_K\right) + \theta_K\left(s\right) \\
\theta_K\left(s\right) &\sim \text{GP}\left(0, C_K\left(s - s'; \phi_K\right)\right); s, s' \in D \\
\begin{matrix}\beta_\Lambda, \beta_r, \beta_r \\ \phi_\Lambda, \phi_r, \phi_K\end{matrix} &\sim \text{priors}
\end{aligned}
\tag{4.21}
$$

where $\beta_{(\cdot)}$ is the parameter vector in the mean surface function and $C_{(\cdot)}\left(s - s'; \phi_{(\cdot)}\right)$ is the stationary covariance function.

In this model, the intensity surfaces are deterministic functions of $r\left(s\right)$, $K\left(s\right)$, and $\Lambda_0\left(s\right)$, $s \in D$. The joint likelihood for the $J + 1$ conditionally independent spatial point patterns is

$$
\prod_{j=1}^{J}\left\{\exp\left(-\int_D \Delta\Lambda_j\left(s\right)ds\right)\prod_{i=1}^{n_j}\Delta\Lambda_j\left(x_{ji}\right)\right\}\cdot\exp\left(-\int_D \Lambda_0\left(s\right)ds\right)\prod_{i=1}^{n_0}\Lambda_0\left(x_{0i}\right). \tag{4.22}
$$

The stochastic integrals in (4.22) cannot be worked with directly. In order to fit

this model, we have to approximate them by Riemann sums. We divide the geographical region D into M small cells and assume the point processes have homogenous intensity within each cell. Let $\Delta\Lambda_j(m)$ and $\Lambda_0(m)$ be the average intensity in cell m. Let the area of cell m be $A(m)$. The joint likelihood after this discretization is

$$\begin{aligned}&\prod_{j=1}^{J}\left[\exp\left(-\sum_{m=1}^{M}\Delta\Lambda_j(m)A(m)\right)\prod_{m=1}^{M}\Delta\Lambda_j(m)^{n_{jm}}\right]\\&\cdot\exp\left(-\sum_{m=1}^{M}\Lambda_0(m)A(m)\right)\prod_{m=1}^{M}\Lambda_0(m)^{n_{0m}},\end{aligned}\tag{4.23}$$

where n_{jm} is the number of point in cell m in period j.

If we assume the growth rate $r(s)$ and the capacity $K(s); s\in D$ are homogeneous in each cell m, and denote them as $r(m)$ and $K(m)$, we derive a version of dynamics in discrete time and space for the intensity,

$$\Delta\Lambda_j(m)=r(m)\Lambda_{j-1}(m)\left[1-\frac{\Lambda_{j-1}(m)}{K(m)}\right]\Delta t.\tag{4.24}$$

Based on the assumptions in model (4.21) that $\theta_r(s)$, $\theta_K(s)$ and $\theta_\Lambda(s)$ are Gaussian processes, we have

$$\begin{aligned}\log r(m)&=\mu_r(m,\beta_r)+\theta_r(m),\\\text{where }\theta_r&=(\theta_r(1),\ldots,\theta_r(M))\sim N_M(0,R_M(\phi_r));\\\log K(m)&=\mu_K(m,\beta_K)+\theta_K(m),\\\text{where }\theta_K&=(\theta_K(1),\ldots,\theta_K(M))\sim N_M(0,R_M(\phi_K));\\\log\Lambda_0(m)&=\mu_\Lambda(m,\beta_\Lambda)+\theta_\Lambda(m);\\\text{where }\theta_\Lambda&=(\theta_\Lambda(1),\ldots,\theta_\Lambda(M))\sim N_M(0,R_M(\phi_\Lambda))\end{aligned}\tag{4.25}$$

where $R_M\left(\phi_{(\cdot)}\right)$ is the $M\times M$ covariance matrix derived from the correlation function $C\left(s-s';\phi_{(\cdot)}\right)$.

A criticism of models (4.21) and (4.25) which we noted above is that they are deterministic conditional on $r(s)$, $K(s)$ and $\Lambda_0(s)$. We can relax this restriction according to the mean-reverting Ornstein-Uhlenbeck process model (4.14). To formulate this model in discrete time and space, we can discretize equation (4.16) into a difference equation for period $j = 1, \ldots, J$:

$$\Delta Z_j = \left[\alpha^M - \zeta I_M Z_j\right] \Delta t + \sigma_Z \left(W_j - W_{j-1}\right), \tag{4.26}$$

$$\text{and } \log r_j = Z_j = Z_0 + \sum_{l=1}^{j} \Delta Z_l,$$

where $\Delta Z_j = \{Z_j(m); m = 1, \ldots, M\}$, $W_j = \{W_j(m); m = 1, \ldots, M\}$, and for any cell m,

$$\alpha_m^M = \log r(m) = \mu_r(m, \beta_r) + \theta_r(m),$$

$$\text{and } Z_0(m) = \mu_{z_0}(m, \beta_z) + \theta_{z_0}(m).$$

In this more flexible model, the log growth rate has a state space equation (4.26), where $W_j - W_{j-1}, j = 1, \ldots, J$ are independent Gaussian random fields since W_t in (4.14) is a spatial Brownian motion.

4.2.2 Approximation Using Kernel Convolution

The computational difficulty in fitting this model is the large number of cells (e.g. 2500 in the examples in Section 4.4.2) and therefore the high dimension of the correlation matrix R_M. There are numerous strategies for handling this "large M" problem (Vecchia 1998; Furrer et al. 2006; etc.). To reduce the dimensionality of this model, we use the kernel convolution approximation technique proposed by Xia and Gelfand (2006) to approximate the distributions of θ_r, θ_K and θ_Λ. A brief account of this method is as follow.

We define a region D_r that covers the region D under study. We divide the region D_r into L blocks, each of which has centroid s_l^* and area $A(l)$, $l = 1, \ldots, L$. We

define $V_l, l = 1, \ldots, L$ to be independent standard-normal random variables, where V_l is associated with block l and let $\varphi(s - s_l^*)$ be the kernel function. The kernel convolution approximation to a Gaussian random field $\theta = \{\theta(s), s \in D\}$ is defined as

$$\tilde{\theta}(s) = \sum_{l=1}^{L} \sqrt{A(l)}\varphi(s - s_l^*; \phi) V_l; s \in D \tag{4.27}$$

The original Gaussian random field θ is assumed to be mean-zero and have a stationary covariance function $C(s - s'; \phi)$, whose spectral density is $\hat{C}(\omega; \phi)$. For $\tilde{\theta}$ to approximate θ, the kernel in (4.27) is shown to be the inverse Fourier transform of $\sqrt{\hat{C}(\omega; \phi)}$:

$$\varphi(s - s'; \phi) = (2\pi)^{-2} \int_{\mathcal{R}^2} e^{-i\omega^T(s-s')} \sqrt{\hat{C}(\omega; \phi)} d\omega \tag{4.28}$$

For example, the Matérn class covariance function in $\mathcal{R}^2$ is

$$C_{\psi,\xi,\nu}(u) = \frac{\pi\psi}{2^{\nu-1}\Gamma(\nu+1)\xi^{2\nu}} (\xi ||u||)^{\nu} \kappa_{\nu}(\xi ||u||) \tag{4.29}$$

where the variance σ^2 is proportional to $\psi/\nu\xi^{2\nu}$. The kernel mixture approximation to a random field with Matérn class covariance has the following kernel:

$$\begin{aligned} \varphi(u; \psi, \xi, \nu) &= (2\pi)^{-1} C_{\psi^{1/2},\xi,(\nu-1)/2}(u) \\ &= \frac{\pi\psi^{1/2}}{2^{(\nu-3)/2}\Gamma\left(\frac{\nu+1}{2}\right)\xi^{\nu-1}} (\xi ||u||)^{\frac{\nu-1}{2}} \kappa_{(\nu-1)/2}(\xi ||u||). \end{aligned} \tag{4.30}$$

Note that (4.30) requires that the smoothness parameter $\nu > 1$. Xia and Gelfand (2006) also discuss how to select the covering region D_r and the number of blocks in D_r.

If we use Matérn covariance in our Model (4.25), we can approximate θ_r, θ_K and θ_Λ with the kernel in (4.30) and $3L$ independent standard normal random variables:

$$V^r \sim N(0, I_L), V^K \sim N(0, I_L), V^\Lambda \sim N(0, I_L).$$

Define a $M \times L$ matrix $H(\phi)$

$$H(\phi)_{ml} = \sqrt{A(l)}\varphi(s_m - s_l^*; \phi); m = 1, \ldots, M; l = 1, \ldots, L$$

where s_m is the centroid of cell m and $\hat{s}_l$ the centroid of block l. The kernel mixture approximation to $\theta_r(m)$, $\theta_K(m)$ and $\theta_\Lambda(m), m = 1, \ldots, M$ is

$$\tilde{\theta}_r = H(\phi_r) V^r; \tilde{\theta}_K = H(\phi_K) V^K; \tilde{\theta}_\Lambda = H(\phi_\Lambda) V^\Lambda. \tag{4.31}$$

Because $\tilde{\theta}_r$, $\tilde{\theta}_r$, $\tilde{\theta}_\Lambda$ are linear combinations of a much lower dimensional independent set of V^r, V^K, V^Λ ($L \ll M$), computation is expedited.

4.3 Bayesian Inference and Prediction

We provide brief details on both inference and prediction associated with the models in Section 4.2.1.

4.3.1 Bayesian Inference

With regard to inference for model (4.21) and (4.25), there are three latent surfaces discretized over the grid M: $r(m)$, $K(m)$ and $\Lambda_0(m)$. The parameters and latent variables in this model include the β_r, β_K and β_Λ in the parametric trend surfaces, the discretized spatial random effects $\theta_r(m)$, $\theta_K(m)$, $\theta_\Lambda(m)$, $m = 1, \ldots, M$ and the parameters ϕ_r, ϕ_K, ϕ_Λ in the covariance functions. As in the previous section, we use the kernel convolution processes $\tilde{\theta}_r$, $\tilde{\theta}_K$, $\tilde{\theta}_\Lambda$ to approximate θ_r, θ_K, θ_Λ:

$$\log r \approx \mu_r(\beta_r) + \tilde{\theta}_r; \;\; \log K \approx \mu_K(\beta_K) + \tilde{\theta}_K; \;\; \log \Lambda_0 \approx \mu_\Lambda(\beta_\Lambda) + \tilde{\theta}_\Lambda. \tag{4.32}$$

The priors and hyper-priors for the model parameters and latent variables are assumed to take the form

$$\begin{aligned} \beta_r, \beta_K, \beta_\Lambda &\sim \pi(\beta_r) \cdot \pi(\beta_K) \cdot \pi(\beta_\Lambda) \\ \phi_r, \phi_K, \phi_\Lambda &\sim \pi(\phi_r) \cdot \pi(\phi_K) \cdot \pi(\phi_\Lambda) \\ V^r, V^K, V^\Lambda &\sim N(0, I_L) \cdot N(0, I_L) \cdot N(0, I_L) \end{aligned} \tag{4.33}$$

where specification of the hyper-priors for $\beta_{(\cdot)}$ and $\phi_{(\cdot)}$ depends on the particular application. For example, the trend surface for the log growth rate may take the linear regression form: $\mu_r(\beta_r) = X(s)\beta_r$, then we can specify a disperse normal prior $N\left(0, \sigma_\beta^2 = 10^8\right)$for β_r. The parameter in the correlation function ϕ_r may represent a value proportional to the variance, hence we can specify a Inverse-Gamma prior for it. Note that $\Delta\Lambda_j$ and Λ_0 in the discretized likelihood (4.23) are deterministic functions of the parameters and latent variables defined by (4.24), (4.31) and (4.32). Therefore the joint posterior is proportional to

$$\begin{aligned}
&\prod_{j=1}^{J}\left[\exp\left(-\sum_{m=1}^{M}\Delta\Lambda_j(m)A(m)\right)\prod_{m=1}^{M}\Delta\Lambda_j(m)^{n_{jm}}\right] \\
&\cdot\exp\left(-\sum_{m=1}^{M}\Lambda_0(m)A(m)\right)\prod_{m=1}^{M}\Lambda_0(m)^{n_{0m}} \\
&\cdot\pi(\beta_r)\pi(\beta_K)\pi(\beta_\Lambda)N(0,I_L)N(0,I_L)N(0,I_L)\pi(\phi_r)\pi(\phi_K)\pi(\phi_\Lambda).
\end{aligned} \tag{4.34}$$

Furthermore, each of β_r, β_K, β_Λ, ϕ_r, ϕ_K, ϕ_Λ may represent multiple parameters; for example, we have $\phi_r = \{\psi_r, \xi_r, \nu_r\}$ if we use Matérn class covariance and kernel for the spatial random effect in the log growth rate.

We simulate the posterior distributions of the model parameters and latent variables in (4.34) using a Markov Chain Monte Carlo algorithm. Because the intensities in the likelihood function are very irregular nonlinear functions of the model parameters and latent variables, it is very difficult to obtain derivatives, and hence a *directed* MCMC (e.g., Langevin diffusions in Benes et al. 2002 and Robert and Casella 1999), we use the random-walk Metropolis-Hastings algorithm in the posterior simulation. Each parameter is updated in turn in every iteration of the simulation and the latent variables $\left\{V^r, V^K, V^\Lambda\right\}$ are sampled in three blocks.

4.3.2 Prediction

The prediction problem in the context of our spatio-temporal model is about forecasting the future intensity and point pattern for the spatial point process. Indeed, we can hold out the observed point pattern in a future time period. Then, if we obtain expected growth for that period, we can compare with observed growth to *validate* our model. The predictive intensity relies highly on the model specification. For the logistic growth function, conditioning on the posterior samples of $\{\beta_r, \beta_K, \beta_\Lambda, \phi_r, \phi_K, \phi_\Lambda\}$ and $\left\{V^r, V^K, V^\Lambda\right\}$, we can extrapolate the intensity $\Delta\Lambda_{J+1}(s)$ in period $J+1$ at any location $s \in D$ by calculating

$$\mu_r(s, \beta_r);\ \mu_K(s, \beta_K);\ \mu_\Lambda(s, \beta_\Lambda)$$

$$\tilde{\theta}_r(s) = \sum_{l=1}^{L} H(s - s_l^*; \phi_r) V_l^r;$$

$$\tilde{\theta}_K(s) = \sum_{l=1}^{L} H(s - s_l^*; \phi_K) V_l^K;$$

$$\tilde{\theta}_\Lambda(s) = \sum_{l=1}^{L} H(s - s_l^*; \phi_\Lambda) V_l^\Lambda.$$

and then using (4.18) and (4.19) recursively. Because we can obtain a predictive sample for $\Delta\Lambda_{J+1}(s)$ from the posterior samples, we can easily compute any quantity of interest pertaining to the predictive distribution of $\Delta\Lambda_{J+1}(s)$, such as $E[\Delta\Lambda_{J+1}(s)]$, $Median[\Delta\Lambda_{J+1}(s)]$ and a predictive interval for $\Delta\Lambda_{J+1}(s)$.

For any subregion $D_1 \subset D$ of interest, we can aggregate $\Delta\Lambda_{J+1}(s)$ over $s \in D_1$ to obtain the predictive distribution of the Poisson intensity $\Delta\Lambda_{J+1}(D_1)$:

$$\Delta\Lambda_{J+1}(D_1) = \int_{D_1} \Delta\Lambda_{J+1}(s)\, ds \approx \sum_{m \cap D_1 \neq \phi} \Delta\Lambda_{J+1}(m).$$

We can then use the predictive distribution of $\Delta\Lambda_{J+1}(D_1)$ to find the predictive

distribution for $N_{J+1} \sim Po\left(\Delta\Lambda_{J+1}\left(D_1\right)\right)$. And, as above, with holdout data, we can compare observed growth with predicted.

4.4 Examples and Results

4.4.1 Simulation Example

In order to see how well we can learn about the true process, we first illustrate the fitting of models (4.21) and (4.25) with a simulated data set. In a study region D of 24×24 square miles shown as the central block in Figure 4.2, we simulate an initial point pattern representing the locations of the existent houses in a city, and 16 consecutive years of growth. Figure 4.3 displays the growth of the simulated urban area, where the first plot shows the existent houses at the time we start our observation, and the other five show the locations of new houses in the next 5 successive years.

The data are simulated as follows. As in the real house construction data set, the observed spatio-temporal point process is discrete in time (we use annual data here). The point patterns in our simulation comprise locations of new buildings in each year. The intensity for the development of these point patterns has the transition model defined by (4.25), where we need three spatial processes $r\left(s\right)$, $K\left(s\right)$ and $\Lambda_0\left(s\right)$, $s \in D$. The initial intensity $\Lambda_0\left(s\right)$, $s \in D$ also generates the initial point pattern, e.g. the locations of the existent houses when we start our observation.

In practice, we can only simulate the latent spatial processes at a finite number of grid points. We therefore divide the region into 1600 (40×40) equally spaced grid cells. In order to use the kernel convolution methods in Section 4.2.2, we put our study region in a larger area D_r of 40×40 square miles (see Xia and Gelfand 2006 for justifications for the selection of the area and grid). We overlay the region D_r with a 10×10 grid at spacing of 4 miles and thus obtain 100 blocks shown in Figure 4.2, each of which has an area of 16 square miles. The latent variables V^r, V^K and V^Λ

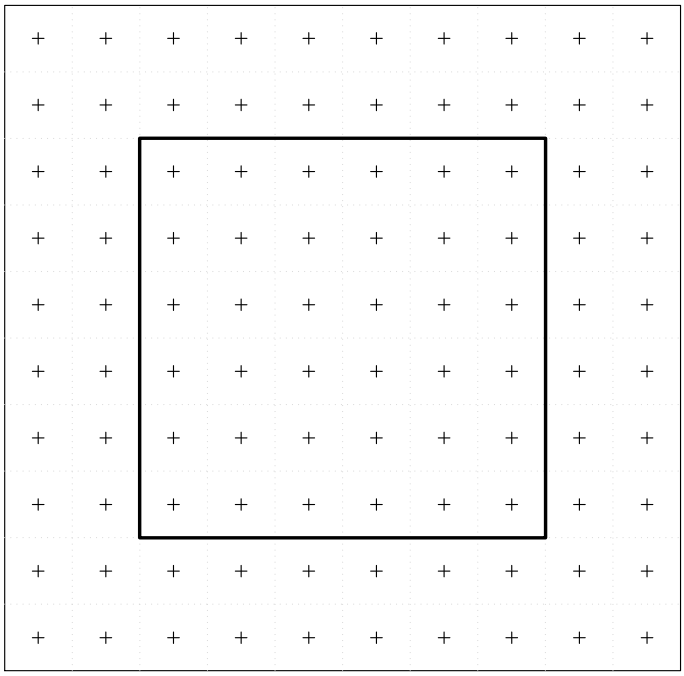

Figure 4.2: The region under study and the bigger area and grid for the kernel convolution approximation (see Section 4.4.1).

are defined at the centroid (see Figure 4.2) of each square.

The latent variables $r(m)$, $K(m)$ and $\Lambda_0(m)$ of each cell in D are defined in (4.25). The log initial intensity $\log \Lambda_0(m)$ has a mean surface μ_Λ of the following structure: if the centroid of m is less than 4 miles from the center of the entire region, we let $\mu_\Lambda(m) = \beta_1$, otherwise we let $\mu_\Lambda(m) = \beta_0$, where $\beta_1 > \beta_0$. By doing so, we generate a densely populated "downtown" area with the diameter of 4 miles, and less populated "suburbs". The spatial random effect θ_Λ is assumed to be a isotropic Gaussian process with Matérn class covariance function $C(\psi_\Lambda, \xi_\Lambda, \nu)$ defined by (4.29). The vector of $\{\theta_\Lambda(m), m = 1, \ldots, 1600\}$ is then simulated using the kernel

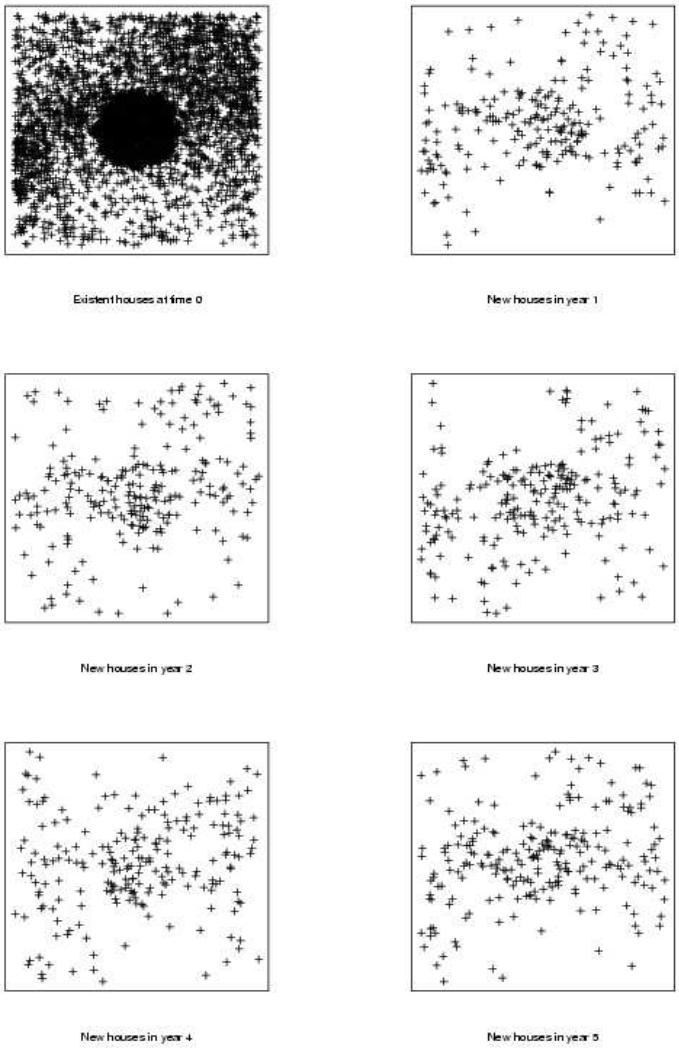

Figure 4.3: Existing houses at time 0 and new houses constructed in 5 successive years for the simulated example.

convolution method with the corresponding Matérn kernel. We assume $\log r(m)$ has a homogeneous mean μ_r and θ_r has a Matérn class covariance $C(\psi_r, \xi_r, \nu)$. Finally $\log K(m)$ has the homogeneous mean μ_K and θ_K has a Matérn class covariance $C(\psi_K, \xi_K, \nu)$. $\{\theta_r(m)$ and $\theta_K(m)$, $m = 1, \ldots, 1600\}$ are also simulated using the kernel convolution method. The actual values of these parameters are presented in

Table 4.2. We fix the smoothness parameter ν to be $\frac{3}{2}$.

At the centroid of each cell m, we use the simulated $r(m)$, $K(m)$ and $\Lambda_0(m)$ and the transition equation (4.24) recursively to obtain the intensity $\Delta\Lambda_j(m)$ for each of the following years. Within each cell, we simulate the points using a Poisson process with homogeneous intensity equal to $\Delta\Lambda_j(m)$. We simulate the initial point pattern with intensity $\Lambda_0(m)$ and then 16 point patterns with $\Delta\Lambda_j(m)$, $j = 1, \ldots, 16$. We use the initial point pattern and the next 15 to fit our model and leave the 16th period out for prediction and model validation. Table 4.1 summarizes the number of points in each of the following years.

Year	Initial	One	Two	Three	Four	Five	Six
# of houses	4914	224	236	227	214	226	237
Year	Seven	Eight	Nine	Ten	Eleven	Twelve	Thirteen
# of houses	262	267	261	286	293	300	309
Year	Fourteen	Fifteen	Sixteen				
# of houses	311	341	347				

Table 4.1: Number of new houses in 16 years.

We fit the discretized model with the same 40×40 grid to the data obtained above. We use very vague priors for the parameters in the mean function. Because ψ and ξ are weakly identified (Zhang 2004), we only use vagues prior for $\psi_{(\cdot)}$'s and very informative priors for $\xi_{(\cdot)}$'s:

$$\pi(\beta_0), \pi(\beta_1) \overset{ind}{\sim} \mathrm{N}\left(0, 10^{-8}\right); \ \pi(\mu_r) \sim \mathrm{N}\left(0, 10^{8}\right); \ \pi(\mu_K) \sim \mathrm{N}\left(0, 10^{8}\right);$$

$$\pi(\psi_\Lambda) \sim \text{log-N}\left(0, 10^{8}\right); \ \pi(\psi_K) \sim \text{log-N}\left(0, 10^{8}\right); \ \pi(\psi_r) \sim \text{log-N}\left(0, 10^{8}\right);$$

$$\pi(\xi_\Lambda) \sim \text{log-N}(-2.5, 2); \ \pi(\xi_K) \sim \text{log-N}(-2.5, 2); \ \pi(\xi_K) \sim \text{log-N}(-2.5, 2);$$

where log-norm$(-2.5, 1)$ has the mean=0.223 and variance=0.318.

Our posterior is proportional to (4.34) by the kernel convolution approximation. We use the random-walk Metropolis-Hastings mentioned in Section 4.3.1 to simulate

posterior samples. The algorithm is tuned to obtain acceptance rates close to the theoretically optimal values in Robert and Casella (1999). We obtain 50,000 samples from the algorithm and discard the first 20,000 as burn-in. For the posterior inference, we use 3,000 subsamples from the remaining 30,000 samples, with a thinning equal to 10. The posterior median and 95% equal-tail quantile for the model parameters are presented in Table 4.2. Evidently we are recovering the true parameter values very well. Figure 4.4 displays the contour plot of the posterior median surfaces for the initial intensity, growth rate and carrying capacity, compared with the actual surfaces. Up to the uncertainty in the model we seem to approximate the actual intensity surface quite well.

Model Parameters	True Value	Posterior Median	95% Equal-tail Interval
β_0	3.0	2.998	$(2.815, 3.211)$
β_1	1.0	0.897	$(0.741, 1.091)$
μ_r	-3.0	-2.991	$(-3.135, -2.855)$
μ_K	5.0	5.011	$(4.844, 5.188)$
ψ_Λ	2.0×10^{-3}	2.37×10^{-3}	$(1.62 \times 10^{-3}, 3.23 \times 10^{-3})$
ψ_r	1.0×10^{-3}	1.35×10^{-3}	$(9.07 \times 10^{-3}, 1.95 \times 10^{-3})$
ψ_K	1.0×10^{-3}	7.46×10^{-4}	$(7.91 \times 10^{-5}, 2.18 \times 10^{-3})$
ξ_Λ	0.2	0.204	$(0.171, 0.251)$
ξ_r	0.2	0.288	$(0.21, 0.376)$
ξ_K	0.6	0.241	$(0.148, 0.505)$

Table 4.2: Simulation example parameters and their posterior inference.

We use the Bayesian prediction in Section 4.3.2 to obtain the predictive distributions for the intensity $\Delta\Lambda_{J+1}(m)$ in each cell. In Figure 4.5 we display the true intensity surface at period $J+1=16$ and the predictive intensity surface, using the medians of the predictive samples for $\Delta\Lambda_{J+1}(m)$, $m = 1, \ldots, M$. We can see the prediction captures the major spatial variation of the actual intensity surface.

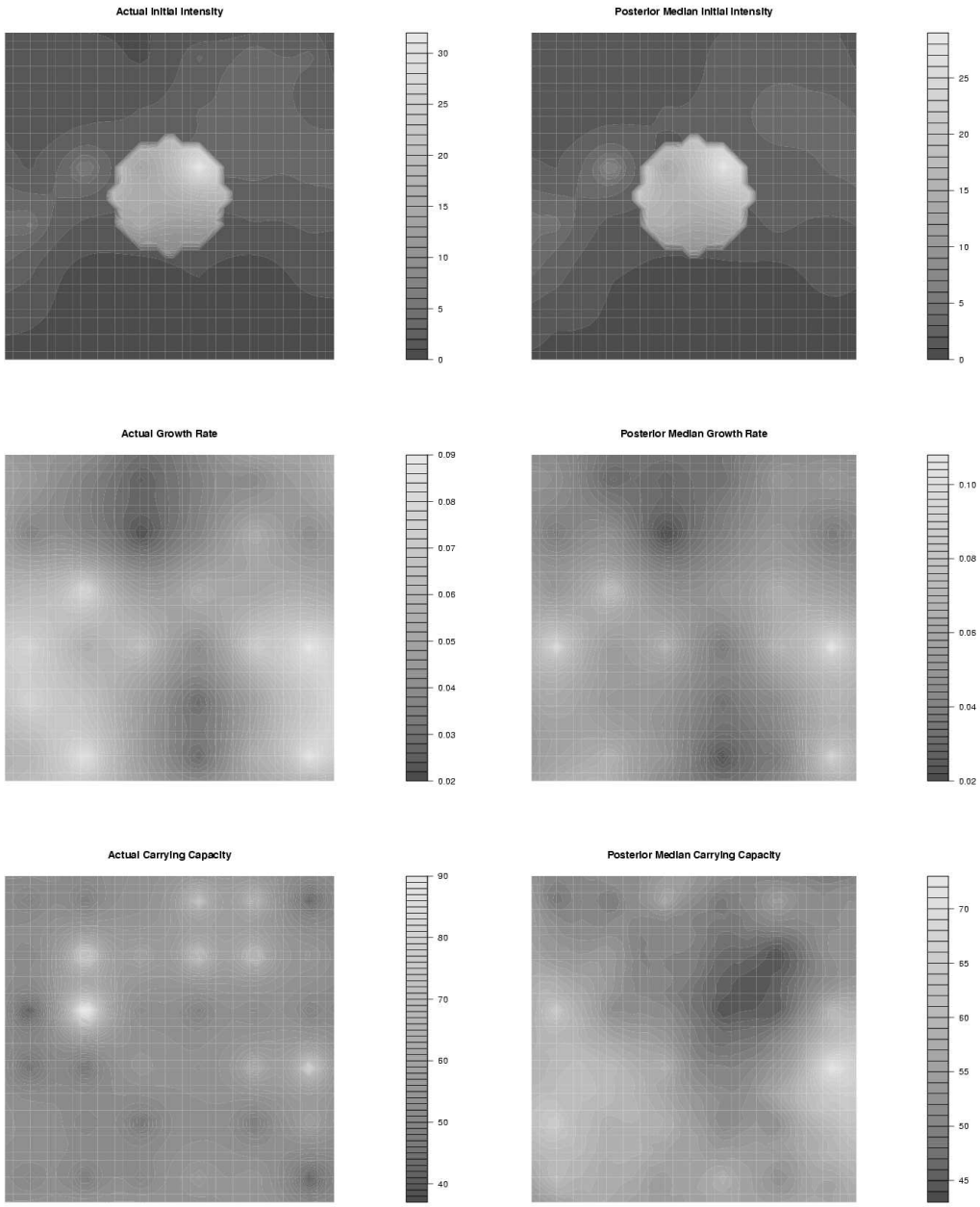

Figure 4.4: Actual and posterior-median initial intensity, growth rate and carrying capacity in the simulated example.

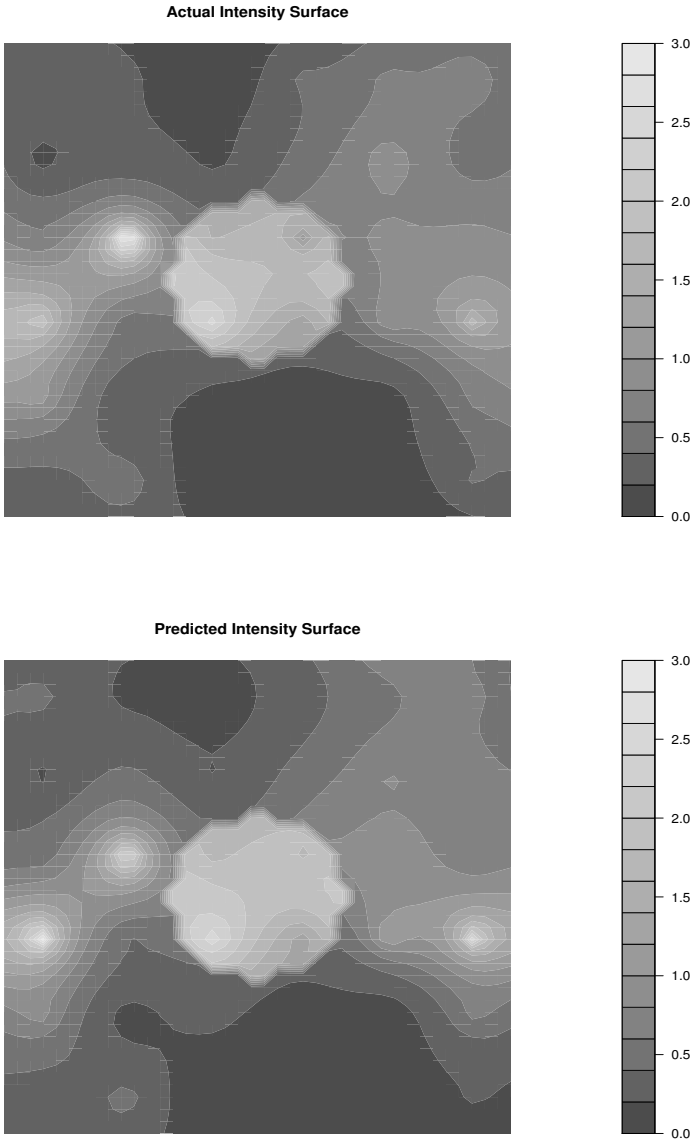

Figure 4.5: Actual and predicted intensity surfaces for period J+1 in the simulated example

4.4.2 Residential House Construction Data for Irving, TX

Our real house construction data consist of the geo-coded locations and years of the newly constructed residential houses in Irving, TX from 1901 to 2002. Figure 4.1 demonstrates how the city develops from early 1950's to late 1960's. It seems that

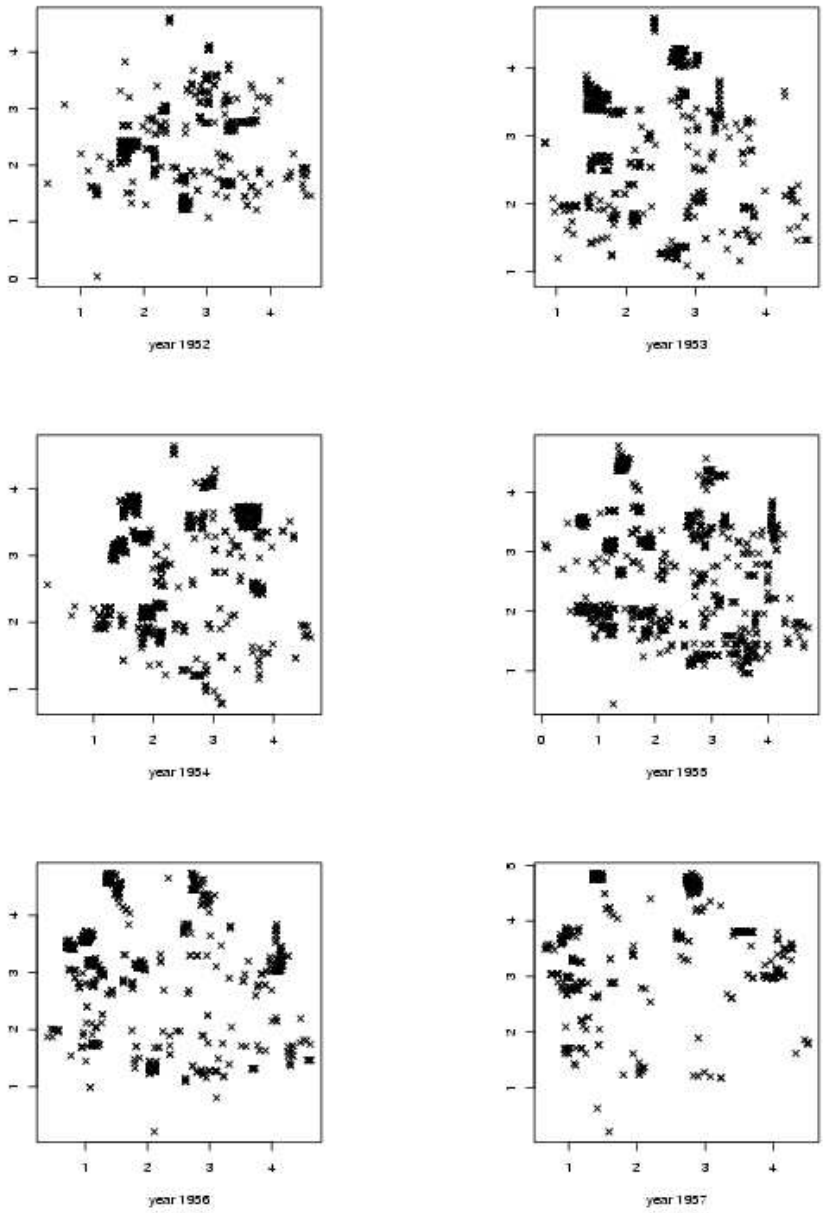

Figure 4.6: New residential houses constructed in Irving, TX from 1952 to 1957.

Irving started to develop after WWII. The current outline of the city is not much different from that of late 1960's because city had been almost fully developed by the early 1970's. For our data analysis, we select the period from 1951 through 1968

when the urban development was substantial. The number of new residential houses for each of the 18 year is recorded in Table 4.3. Figure 4.6 shows the geographical distribution and locations of new houses in six consecutive years (1952–1957) during that period. In our analysis, we use the data for year 1951–1966 to fit our model and leave year 1967 and 1968 out for prediction and model validation.

Year	Before 1951	1952	1953	1954	1955	1956	1957
# of houses	1957	747	1006	1288	1332	807	629
Year	1958	1959	1960	1961	1962	1963	1964
# of houses	910	759	950	918	860	954	1034
Year	1965	1966	1967	1968			
# of houses	851	531	654	583			

Table 4.3: Number of new houses from 1952 to 1968.

As shown in the central block of Figure 4.7, our study region D in this example is a square of 5.6×5.6 square miles with Irving, TX in the middle. This region is selected to exclude other urban areas of Dallas county. We divide the region into 2500 (50×50) equally spaced grid cells. In order to use the kernel convolution approximation, we put our study region in the middle of a larger area D_r of 10×10 square miles. We again overlay the region by a 10×10 grid at spacing of 1 mile and define the latent variables V^r, V^K and V^Λ at the centroids (centroids shown as small circles in Figure 4.7) of the resulting 100 blocks.

The log initial intensity, $\log \Lambda_0(m)$, in this case is assumed to have a constant mean surface μ_Λ. We also assume a constant mean, μ_r, for the log growth rate and μ_K for the log carrying capacity. θ_Λ, θ_r and θ_K have Matérn class covariance $C(\psi_\Lambda, \xi_\Lambda, \nu)$, $C(\psi_r, \xi_r, \nu)$ and $C(\psi_K, \xi_K, \nu)$ respectively, with ν equal to $\frac{3}{2}$.

We again use very vague priors for the parameters in the mean function. For the same reason as in the simulated example, we only use vague priors for $\psi_{(\cdot)}$'s and

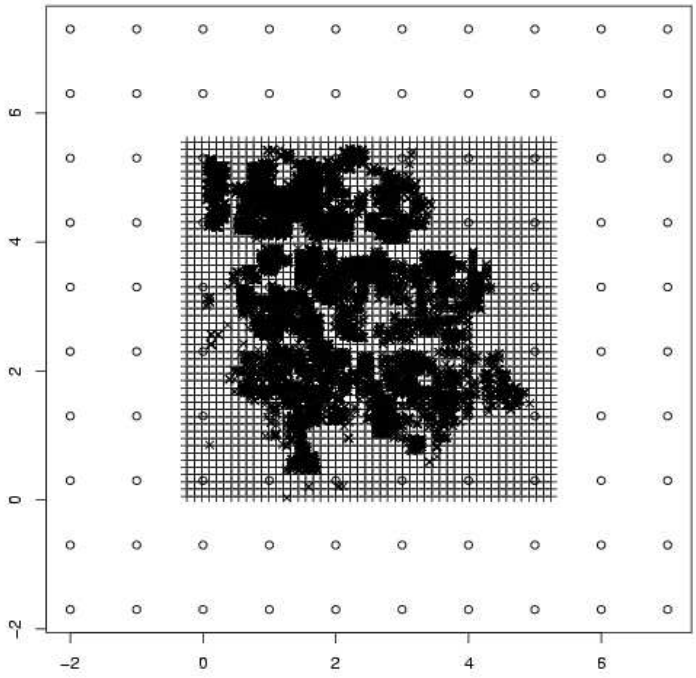

Figure 4.7: The gridded study region encompassing Irving, TX and the bigger area for the kernel convolution approximation.

informative priors for $\xi_{(\cdot)}$'s:

$$\pi(\mu_\Lambda) \sim \text{N}\left(0, 10^{-8}\right);\ \pi(\mu_r) \sim \text{N}\left(0, 10^{8}\right);\ \pi(\mu_K) \sim \text{N}\left(0, 10^{8}\right);$$

$$\pi(\psi_\Lambda) \sim \text{log-N}\left(0, 10^{8}\right);\ \pi(\psi_K) \sim \text{log-N}\left(0, 10^{8}\right);\ \pi(\psi_r) \sim \text{log-N}\left(0, 10^{8}\right);$$

$$\pi(\xi_\Lambda) \sim \text{log-N}(0.5, 1);\ \pi(\xi_K) \sim \text{log-N}(0.5, 1);\ \pi(\xi_K) \sim \text{log-N}(0.5, 1);$$

We use the same random-walk Metropolis-Hastings algorithm as in the simulation example to simulate posterior samples with the same tuning of acceptance rates. We obtain 250,000 samples from the algorithm and discard the first 100,000 as burn-in. For the posterior inference, we use 5,000 subsamples from the remaining 150,000 samples, with a thinning equal to 30. The posterior mean, median and 95% equal-tail

quantile for the model parameters are presented in Table 4.4. Figure 4.8 displays the contour plot of the posterior median surfaces for the initial intensity, growth rate and carrying capacity.

Model Parameters	Posterior Mean	Posterior Median	95% Equal-tail Interval
μ_Λ	-0.6464	-0.5764	$(-1.2540, -0.2013)$
μ_r	-2.6467	-2.6458	$(-2.8476, -2.4582)$
μ_K	2.7747	2.6049	$(1.7955, 4.5789)$
ψ_Λ	1.3456	1.3578	$(1.1964, 1.5210)$
ψ_r	3.2675	3.2839	$(2.9858, 3.4921)$
ψ_K	35.4332	38.0502	$(0.7739, 67.4488)$
ξ_Λ	0.04231	0.04399	$(0.03765, 0.05109)$
ξ_r	0.4765	0.4808	$(0.3119, 0.6320)$
ξ_K	0.6429	0.6578	$(0.4197, 0.8953)$

Table 4.4: Posterior inference for Irving, TX data

In Figure 4.9 we display the predictive intensity surface in year 1967 and 1968, using the medians of the predictive samples for $\Delta\Lambda_{J+1}(m)$, $m = 1, \ldots, M$. We also overlay the actual point patterns in 1967 and 1968 on the predictive intensity surface. Figure 4.9 shows our model can forecast the major areas of high intensity, hence high growth very well.

4.5 Discussion

In our data analysis, we apply the parsimonious model (4.21) without the time-varying growth rate to a simulated example and the Irving, TX data. In the short term, when the fundamental elements in the structured model are stable, this is advantageous as we demonstrate its functionality through the real house construction data for Irving, TX. In the long term, the growth rate $r(t, s)$ is expected to change, and therefore we will have to employ model (4.14) for the growth rate. But in this case, the more complicated model can be fitted since many more periods of data

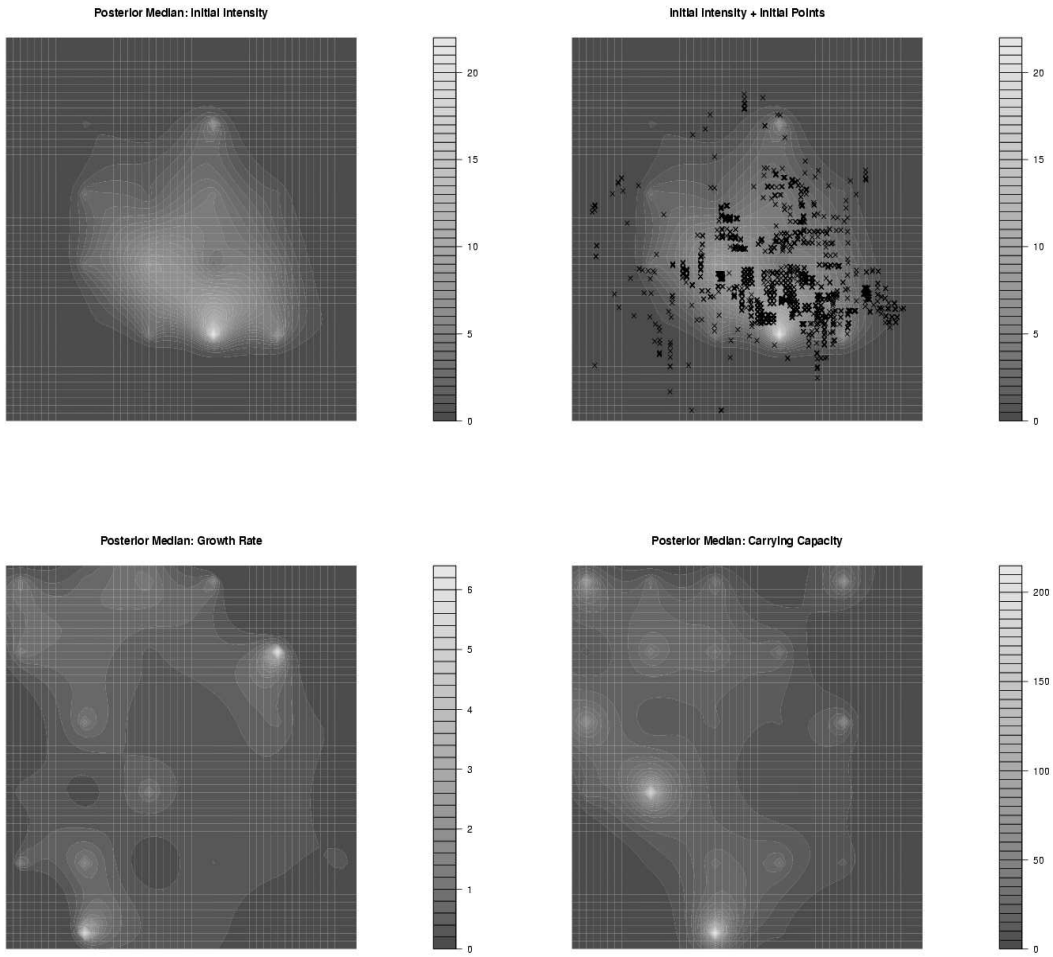

Figure 4.8: Posterior-median initial intensity, growth rate and carrying capacity in the Irving, TX example.

will be available. Of course, (4.14) is only an idealized model for the growth rate, presuming a stable mean and stationary process in both time and space. However, if the damping effect of growth is controlled by the logistic model, it is not unreasonable to assume the growth rate is mean-reverting. Additionally the logistic model, in a fixed functional form, may be too restrictive as well as the time-independent carrying capacity. These issues will be considered in future research.

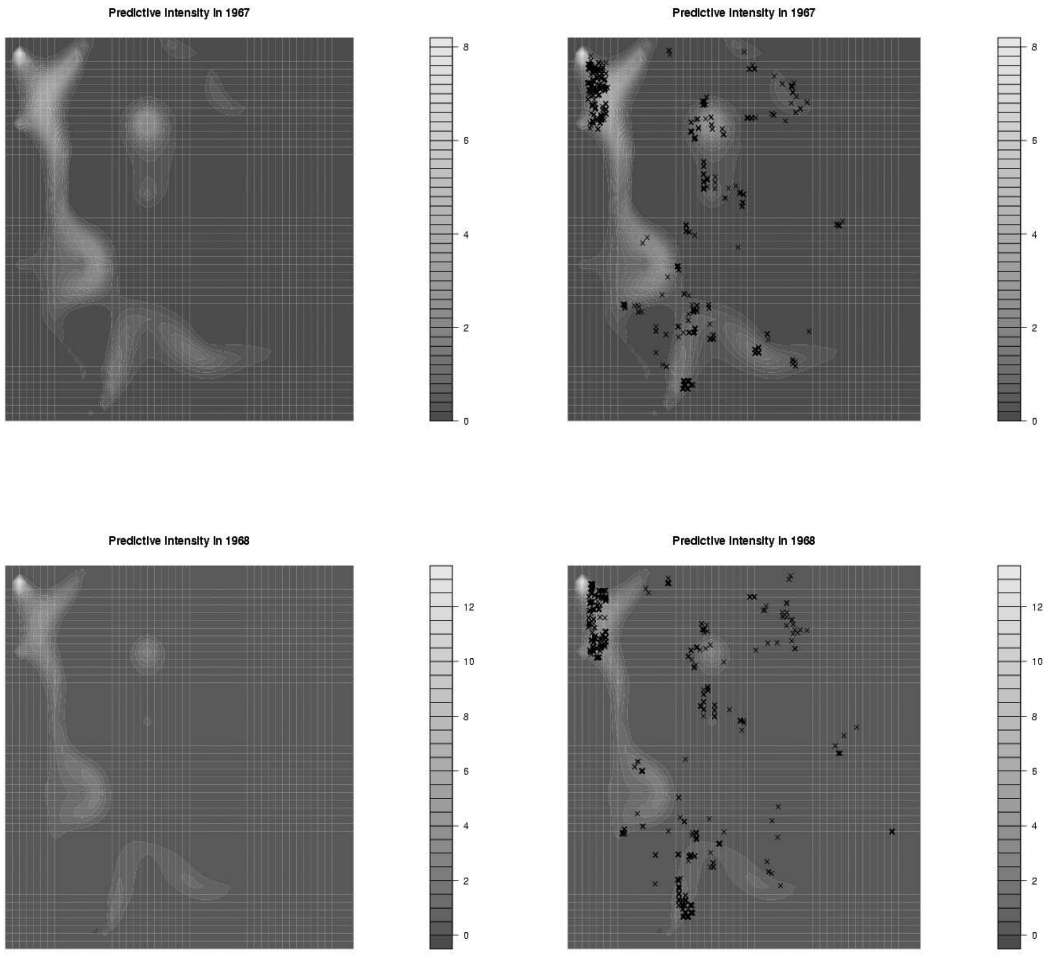

Figure 4.9: Predicted intensity surfaces and actual point patterns for year 1967 and 1968 in the Irving, TX example.

We demonstrate in our data analysis that the Bayesian hierarchical model can preserve the complicated model structure and achieve good estimation and prediction. The major challenges in fitting our proposed model are: (i) the handling of a large data set that has thousands of spatially correlated observations; (ii) the evaluation of a likelihood that involves stochastic integrals to be approximated with discretization; and (iii) a likelihood that does not allow an easy formulation of an

efficient Metropolis-Hastings algorithm. In dealing with first two challenges, we use the process-convolution approximation in Xia and Gelfand (2006) and the discretization method in Benes et al. (2002). Though the simulation results are encouraging, further investigation of these approximations or alternatives would be helpful. For (iii), we apply the random-walk Metropolis algorithm to the posterior simulation, which is liable to create large auto-correlation in the sampling chain. The nonlinear and recursive structure of our likelihood makes most of the current Metropolis methods inapplicable, encouraging future research for a more efficient Metropolis-Hastings algorithm for this class of problems.

Despite all these concerns, the current model is the first attempt to incorporate a structured growth model into a spatial-temporal point process. The structured model and its statistical inference affords valuable insights into the mechanism of the urban development problem. We hope our approach will spark more research related to the integration of scientific mechanisms and statistical models, and the development of novel statistical methodologies in this field.

Appendix A

Gibbs Sampler for the GSDP Models

1. Full conditionals for the Z's.

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

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

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

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

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

in which $h_i(\eta)$ is the i-th column vector of $H_n(\eta)$, $H_{(-i)}(\eta)$ the $(n-1)\times(n-1)$ sub-matrix obtained from $H_n(\eta)$ by deleting the i-th row and column, and $Z^{(-i)}_{t,l}$ is the $n-1$ dimensional vector obtained from $Z_{t,l}$ by deleting the i-th element. Notice that both $\tilde{\mu}^i_{t,l}$ and $\tilde{H}_i(\eta)$ are scalars. Let us indicate with $\psi = \{X_t, \beta, \theta^*, \tau^2, \sigma^2, \phi, \mu_l, l > 1, \eta\}$ the vector of parameters of the model other than the $Z_{t,l}$'s. Then, the full conditional of $Z_{t,l}(s_i)$ is given by

$$[Z_{t,l}(s_i)|Y_t, Z_{t,l}(s_j), Z_{t,m}(s_i), m \neq l, j \neq i, \psi] \propto [Z_{t,l}(s_i)|Z_{t,l}(s_j), j \neq i, \psi] \times$$

$$\sum_{m=1}^{K} \exp\left[-\frac{1}{2\tau^2}\left\{y_t(s_i) - X_t(s_i)^T \beta - \theta^*_m(s_i)\right\}^2\right] I_{\mathcal{Z}_{t,m}(s_i)},$$

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

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

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

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

and

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

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

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

2. Full conditional for the θ^*'s.

We can update all the θ^*'s at once for all locations. Let us consider at each point

$s \in D$ the partition induced on the space of the Z 's by the allocation process, that is, for $t = 1, \ldots, T$ and $l = 1, \ldots, K-1$, consider the sets $\mathcal{Z}_{t,l}(s) = \{s \in D : Z_{t,1}(s) < 0, \ldots, Z_{t,l-1}(s) < 0, Z_{t,l}(s) \geq 0\}$, and $\mathcal{Z}_{t,K}(s) = \{s \in D : Z_{t,1}(s) < 0, \ldots, Z_{t,K-1}(s) < 0\}$. Then, $I(\mathcal{Z}_{t,l}) = diag\{I_{\mathcal{Z}_{t,l}(s_1)}, \ldots, I_{\mathcal{Z}_{t,l}(s_n)}\}$ is the diagonal matrix whose i-th entry is equal to one when the component l is chosen at location s_i. Immediately, the full conditional for $\theta_l^* = \{\theta_l^*(s_1), \ldots, \theta_l^*(s_n)\}$ is given by

$[\theta_l^* | Y_t, Z_t, t = 1, \ldots, T, \beta, \tau^2, \sigma^2, \phi] \propto$

$$\exp\left\{-\frac{1}{2\tau^2}\sum_{t=1}^{T}(y_t - X_t^T\beta - \theta_l^*)^T I(\mathcal{Z}_{t,l})(y_t - X_t^T\beta - \theta_l^*)\right\} \exp\left\{-\frac{1}{2\sigma^2}\,\theta_l^{*T} R_n^{-1}(\phi)\,\theta_l^*\right\}$$

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

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

Once we know θ_l^* and Z_t for all $l = 1, \ldots, K$ and $t = 1, \ldots, T$, we can compute each θ_t as a function of (θ_l^*, Z_t). The full conditionals for θ_l^*'s in the spatio-temporal dynamical model are far more complicated. We are still able to update θ_l^* at all locations, but it has to be conditioned on all the other value θ_j^*'s with $j \neq l$.

For the dynamic linear model (2.17), we expand accumulated spatial random effect

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

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

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

Hence, the likelihood × prior can be written as proportional to

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

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

$$\times\exp\left\{-\frac{1}{2\sigma^2}\theta_l^{*T}R^{-1}(\phi)\theta_l^*\right\}.$$

Let us define

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

The expression above becomes

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

$$\times\exp\left\{-\frac{1}{2\sigma^2}\theta_l^{*T}R^{-1}(\phi)\theta_l^*\right\},$$

from which we can deduce

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

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

3. Full conditionals for $\beta,\tau^2,\sigma^2,\phi,\mu$ and η.

Assume $\beta\sim N_p(\beta_0,\Sigma_0)$. Then, $[\beta|X_t,Y_t,Z_t,\theta_t,\tau^2]\sim N(\hat{\beta},\hat{\Sigma}_\beta)$, where

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

.

4. Full conditional for τ^2.

Assume $\tau^2 \sim IGamma(\alpha_\tau, \beta_\tau)$. Then $[\tau^2|X_t, Y_t, \theta_t, \beta] \sim IG(\tilde{\alpha}_\tau, \tilde{\beta}_\tau)$ where

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

5. Full conditional for σ^2.

Assume $\sigma^2 \sim IG(\alpha_\sigma, \beta_\sigma)$. Then, $[\sigma^2|\theta_l^*, \phi] \sim IGamma(\tilde{\alpha}_\sigma, \tilde{\beta}_\sigma)$ where

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

6. Full conditional for ϕ.

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

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

7. Full conditional for μ.

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

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

8. Full conditional for η.

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

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

Appendix B

Posterior Simulation for the SDP Disease Mapping Models

Here, we provide the details for MCMC posterior simulation for the spatial and spatio-temporal models discussed in Sections 2.1 and 2.4, respectively. In both cases, the posterior of the model can be explored using a Gibbs sampler that combines standard MCMC techniques for DP mixtures (West et al. 1994; Bush and MacEachern 1996) with updates for the latent z_{it}.

1. Spatial model

Under model (3.2), the full conditional for each z_{it} can be expressed as

$$p(z_{it} \mid ..., \text{data}) \propto \exp(-n_{it} \exp(z_{it}))N(z_{it} \mid \mu + \theta_{it} + \tau^2 y_{it}, \tau^2).$$

We can sample from this full conditional introducing an auxiliary variable u_{it}, with positive values, such that

$$p(z_{it}, u_{it} \mid ..., \text{data}) \propto N(z_{it} \mid \mu + \theta_{it} + \tau^2 y_{it}, \tau^2) 1_{(0<u_{it}<\exp(-n_{it}\exp(z_{it})))}.$$

The Gibbs sampler is extended to draw from $p(u_{it} \mid z_{it}, \text{data})$ and $p(z_{it} \mid u_{it}, ..., \text{data})$. The former is a uniform distribution over $(0, \exp(-n_{it} \exp(z_{it})))$. The latter is a

$\mathrm{N}(\mu+\theta_{it}+\tau^2 y_{it}, \tau^2)$ distribution truncated over the interval $(-\infty, \log(-n_{it}^{-1}\log u_{it}))$. Alternatively, adaptive rejection sampling can be used to draw from the full conditional for z_{it} noting that its density is log-concave.

Having updated all the z_{it}, the mixing parameters θ_t, $t = 1, ..., T$, and hyperparameters μ, τ^2, σ^2, ϕ, can be updated as in the spatial DP mixture model, with z_t playing the role of the data vector y_t. (We refer to the Appendix in Gelfand et al. 2005 for details.) All these updates require computations involving the matrix $R_n(\phi)$. To approximate the entries of this matrix, we use Monte Carlo integrations based on sets of locations distributed independently and uniformly over each region B_i, $i = 1, ..., n$. Note that, with the discrete uniform prior for ϕ, these calculations need only be performed once at the beginning of the MCMC algorithm.

2. Spatio-temporal model

The posterior for model (3.5) is given by expression (3.8). The form of the full conditionals for the z_{it} is similar to the one for the spatial model, and, thus, either auxiliary variables or adaptive rejection sampling can be used to update these parameters.

For each $t = 1, ..., T$, the full conditional for η_t,

$$p(\eta_t|..., \text{data}) \propto p(\eta_t|\{\eta_j : j \neq t\}, \sigma^2, \phi) \prod_{\ell=t}^{T} \mathrm{N}_n(z_\ell|d_\ell + \nu^{\ell-t}\eta_t, \tau^2 I_n)$$

where $d_\ell = (\beta_0 + \beta_1 \ell)1_n + \sum_{m=1, m\neq t}^{\ell} \nu^{\ell-m}\eta_m$, $\ell = t, ..., T$. The product term above is proportional to a $\mathrm{N}_n(\eta_t|\mu_t, \Sigma_t)$ density, with $\mu_t = (\sum_{\ell=t}^{T} \nu^{2(\ell-t)})^{-1} \sum_{\ell=t}^{T} \nu^{\ell-t}(z_\ell - d_\ell)$ and $\Sigma_t = \tau^2(\sum_{\ell=t}^{T} \nu^{2(\ell-t)})^{-1} I_n$. Let T^{*-} be the number of distinct η_j in $\{\eta_j : j \neq t\}$, η_j^{*-}, $j = 1, ..., T^{*-}$, be the distinct values, and T_j^- be the size of the cluster corresponding to η_j^{*-}. The prior full conditional $p(\eta_t|\{\eta_j : j \neq t\}, \sigma^2, \phi)$ is a mixed distribution with point masses $T_j^-(\alpha + T - 1)^{-1}$ at the η_j^{*-} and continuous mass

$\alpha(\alpha + T - 1)^{-1}$ on the $\mathrm{N}_n(0, \sigma^2 R_n(\phi))$ distribution. Hence, $p(\eta_t|..., \text{data})$ is also a mixed distribution with point masses, proportional to $T_j^- q_j$, at the η_j^{*-} and continuous mass, proportional to αq_0, on an n-variate normal distribution with covariance matrix $H_t = (\Sigma_t^{-1} + \sigma^{-2} R_n^{-1}(\phi))^{-1}$ and mean vector $H_t \Sigma_t^{-1} \mu_t$. Here, q_j is the value of the $\mathrm{N}_n(\mu_t, \Sigma_t)$ density at η_j^{*-}, and

$$q_0 = \int N_n(u|0, \sigma^2 R_n(\phi)) N_n(u|\mu_t, \Sigma_t) du,$$

an integral that is available analytically.

Updating σ^2 and ϕ proceeds as in the spatial model. The full conditional for τ^2 is an inverse gamma distribution, and β_0 and β_1 have normal full conditionals. Finally, working with a discrete uniform prior for ν, we sample directly from its discretized full conditional.

Bibliography

Albert, J. and Chib, S. (1993), "Bayesian analysis of binary and polychotomous reponse data," *Journal of the American Statistical Association*, 88, 669–679.

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

Banerjee, S., Wall, M., and Carlin, B. (2003b), "Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota," *Biostatistics*, 4, 123–142.

Benes, V., Bodlak, K., Møller, J., and Waagepetersen, R. (2002), "Bayesian Analysis of Log Gaussian Cox Processes for Disease Mapping," Working paper, Department of Mathematical Sciences, Aalborg University.

Bernardinelli, L., Clayton, D., and Montomoli, C. (1995), "Bayesian estimates of disease maps: How important are priors?" *Statistics in Medicine*, 14, 2411–2432.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995), "Bayesian computation and stochastic systems (with discussion)," *Statistical Science*, 10, 3–66.

Besag, J., York, J., and Mollie, A. (1991), "Bayesian image restoration with two applications in spatial statistics," *Annals of the Institute of Statistical Mathematics*, 43, 1–59.

Best, N., Ickstadt, K., and Wolpert, R. (2000), "Spatial Poisson regression for health and exposure data measured at disparate resolutions," *Journal of the American Statistical Association*, 95, 1076–1088.

Blackwell, D. and MacQueen, J. (1973), "Ferguson distributions via Pólya urn scheme." *Annals of Statistics*, 1, 353–355.

Böhning, D., Dietz, E., and Schlattmann, P. (2000), "Space-time mixture modelling of public health data," *Statistics in Medicine*, 19, 2333–2344.

Brix, A. and Diggle, P. (2001), "Spatiotemporal prediction for log-Gaussian Cox processes," *Journal of the Royal Statistical Society: Series B*, 63, 823–841.

Brix, A. and Møller, J. (2001), "Space-time multi type log Gaussian Cox Processes with a View to Modelling Weeds." *Scandinavian Journal of Statistics*, 28, 471–488.

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

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

Christensen, O. and Waagepetersen, R. (2002), "Bayesian prediction of spatial count data using generalized linear mixed models," *Biometrics*, 58, 280–286.

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.

Clayton, D. and Kaldor, J. (1987), "Empirical Bayes estimates of age-standardized relative risks for use in disease mapping," *Biometrics*, 43, 671–681.

Cox, D. (1955), "Some statistical methods connected with series of events." *Journal of the Royal Statistical Society: Series B*, 17, 129–164.

Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley, 2nd ed.

Cressie, N. and Chan, N. (1989), "Spatial modeling of regional variables." *Journal of the American Statistical Association*, 84, 393–401.

Daley, D. and Vere-Jones, D. (1988), *Introduction to the Theory of Point Processes*, New York: Springer Verlag.

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.

Diggle, P. (2005a), "A partial likelihood for spatio-temporal point processes," Working Paper, Johns Hopkins University.

— (2005b), "Spatio-temporal point processes: methods and applications," Working Paper, Johns Hopkins University.

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

Elliott, P., Wakefield, J., Best, N., and Briggs, D. (2000), *Spatial Epidemiology: Methods and Applications*, Oxford: University Press.

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

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

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

— (1974), "Prior distributions on spaces of probability measures," *Annals of Statistics*, 2, 615–629.

Fernandez, C. and Green, P. (2002), "Modelling spatially correlated data via mixtures: a Bayesian approach," *Journal of the Royal Statistical Society: Series B*, 64, 805–805.

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

Furrer, R., Genton, M., and Nychka, D. (2006), "Covariance tapering for interpolation of large spatial dataset," To appear in *Journal of Computational and Graphical Statistics*.

Gard, T. (1988), *Introduction to Stochastic Differential Equations*, New York: Marcel Dekker.

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

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

Giudici, P., Knorr-Held, L., and Rasser, G. (2000), "Modelling categorical covariates in Bayesian disease mapping by partition structures," *Statistics in Medicine*, 19, 2579–2593.

Glaeser, E. (2003), "Sprawl and urban growth," Handbook of Regional and Urban Economics, Vol. 4, forthcoming.

Green, P. and Richardson, S. (2002), "Hidden Markov models and disease mapping." *Journal of the American Statistical Association*, 97, 1055–1070.

Griffin, J. and Steel, M. (2004), "A class of dependent Dirichlet Processes," Tech. rep., University of Kent at Canterbury.

Heagerty, P. and Lele, S. (1998), "A composite likelihood approach to binary spatial data," *Journal of the American Statistical Association*, 93, 1099–1111.

Higdon, D., Swall, J., and Kern, J. (1999), "Nonstationary spatial modelling," in *Bayesian Statistics 6*, eds. Bernardo, J., Berger, J., Dawid, A., and Smith, A., Oxford University Press, pp. 761–768.

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," *Canadian Journal of Statistics*, 30, 269–283.

Itô, K. (1984), *Foundations of Stochastic Differential Equations in Infinite Dimensional Spaces*, Philadelphia: Society for Industrial and Applied Mathematics.

Karatzas, I. and Shreve, S. (1991), *Brownian Motion and Stochastic Calculus*, New York: Springer Verlag, 2nd ed.

Karr, A. (1991), *Point Processes and Their Statistical Inference*, New York: Marcel Dekker, 2nd ed.

Kelsall, J. and Wakefield, J. (2002), "Modeling spatial variation in disease risk: a geostatistical approach," *Journal of the American Statistical Association*, 97, 692–701.

Kent, J. (1989), "Continuity properties of random fields," *Annals of Probability*, 17, 1432–1440.

Knorr-Held, L. and Rasser, G. (2000), "Bayesian detection of clusters and discontinuities in disease maps," *Biometrics*, 56, 13–21.

Kot, M. (2001), *Elements of Mathematical Ecology*, Cambridge Press.

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

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

Militino, A., Ugarte, M., and Dean, C. (2001), "The use of mixture models for identifying high risks in disease mapping," *Statistics in Medicine*, 20, 2035–2049.

Møller, J., Syversveen, A., and Waagepetersen, R. (1998), "Log Gaussian Cox processes," *Scandanavian Journal of Statistics*, 25, 451–482.

Møller, J. and Waagepetersen, R. (2004), *Statistical Inference and Simulation for Spatial Point Processes*, Chapman and Hall/CRC Press.

Neal, R. (2000), "Markov Chain sampling methods for Dirichlet process mixture models," *Journal of Computational and Graphical Statistics*, 9, 249–265.

Ogata, Y. (1998), "Space-time point-process models for earthquake occurrences," *Annals of the Institute for Statistical Mathematics*, 50, 379–402.

Oksendal, B. (2002), *Stochastic Differential Equations*, New York: Springer Verlag, 5th ed.

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

Pascutto, C., Wakefield, J., Best, N., Richardson, S., Bernardinelli, L., Staines, A., and Elliott, P. (2000), "Statistical issues in the analysis of disease mapping data," *Statistics in Medicine*, 19, 2493–2519.

Robert, C. and Casella, G. (1999), *Monte Carlo Statistical Methods*, New York: Springer Verlag, 1st ed.

Rossi-Hansberg, E. and Wright, M. (2005), "Urban structure and growth," NBER Working Papers 11262, National Bureau of Economic Research, Inc.

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

Schach, S. (1971), "Weak convergence results for a class of multivariate Markov processes," *Annals of Mathematical Statistics*, 42, 451–465.

Schlattmann, P. and Böhning, D. (1993), "Mixture models and disease mapping," *Statistics in Medicine*, 12, 1943–1950.

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.

Short, M., Carlin, B., and Gelfand, A. (2005), "Covariate-adjusted spatial CDF's for air pollutant data." *Journal of Agricultural, Biological, and Environmental Statistics*, 10, 259–275.

Vecchia, A. (1998), "Estimation and identification for continuous spatial process," *Journal of the Royal Statistical Society: Series B*, 50, 297–312.

Wakefield, J. and Morris, S. (1999), "Spatial dependence and errors-in-variables in environmental epidemiology," in *Bayesian Statistics 6*, eds. Bernardo, J., Berger, J., Dawid, A., and Smith, A., Oxford University Press, pp. 657–684.

Wakefield, J. and Shaddick, G. (2006), "Health-exposure modelling and the ecological fallacy," *Biostatistics*, 7.

Waller, L., Carlin, B., Xia, H., and Gelfand, A. (1997), "Hierarchical spatio-temporal mapping of disease rates," *Journal of the American Statistical Association*, 92, 607–617.

West, M., Müller, P., and Escobar, M. (1994), "Hierarchical priors and mixture models, with application in regression and density estimation," in *Aspects of Uncertainty: A Tribute to D.V. Lindley*, eds. Smith, A. and Freeman, P., New York: Wiley, pp. 363–386.

Xia, G. and Gelfand, A. (2006), "Stationary process approximation for the analysis of large spatial datasets," Working Paper, Duke University.

Zhang, H. (2004), "Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics," *Journal of the American Statistical Association*, 99, 250–261.

Biography

Jun Duan was born in Nanjing, China on April 8, 1978. He was conferred the Bachelor of Science in Chemistry by Nanjing University in June, 1999. He received his Master of Arts in Economics from Duke University in December, 2005. Jun Duan is a member of the American Economic Association, American Statistical Association and Institute of Mathematical Statistics.