Bayesian Mixture Modeling for Spatial Poisson Process Intensities, with Applications to Extreme Value Analysis

Athanasios Kottas [1] and Bruno Sansó

Department of Applied Mathematics and Statistics, Baskin School of Engineering, University of California, 1156 High Street, MS: SOE2, Santa Cruz, CA 95064, USA

(REVISED VERSION: March 16, 2006)

Abstract: We propose a method for the analysis of a spatial point pattern, which is assumed to arise as a set of observations from a spatial non-homogeneous Poisson process. The spatial point pattern is observed in a bounded region, which, for most applications, is taken to be a rectangle in the space where the process is defined. The method is based on modeling a density function, defined on this bounded region, that is directly related with the intensity function of the Poisson process. We develop a flexible nonparametric mixture model for this density using a bivariate Beta distribution for the mixture kernel and a Dirichlet process prior for the mixing distribution. Using posterior simulation methods, we obtain full inference for the intensity function and any other functional of the process that might be of interest. We discuss applications to problems where inference for clustering in the spatial point pattern is of interest. Moreover, we consider applications of the methodology to extreme value analysis problems. We illustrate the modeling approach with three previously published data sets. Two of the data sets are from forestry and consist of locations of trees. The third data set consists of extremes from the Dow Jones index over a period of 1303 days.

Keywords: Bayesian nonparametrics; Bivariate Beta density; Dirichlet process mixture model; Random intensity function; Spatial point patterns

[1]Corresponding author.Tel.: 831-459-5536. Fax: 831-459-4829.

E-mail addresses: thanos@ams.ucsc.edu (A. Kottas), bruno@ams.ucsc.edu (B. Sansó).

1 Introduction

We propose a Bayesian nonparametric model for the intensity function of a spatial nonhomogeneous Poisson process (NHPP). The intensity function characterizes the distribution of a spatial NHPP and, thus, its estimation is an important problem in the analysis of spatial point patterns. Moreover, under a model-based framework, inference for the intensity function yields inference for other important functionals of the spatial NHPP.

We refer to Diggle (2003) for review of parametric likelihood and classical nonparametric inference approaches for spatial Poisson processes (as well as other point processes that extend the Poisson process structure). Other useful references include Møller and Waagepetersen (2004), where, in addition to discussion of likelihood methods, there is a review of more recent work on simulation-based inference for spatial point processes. After the proposed methodology is presented, we devote a section (Section 2.4 below) to discussion of existing methods for spatial NHPPs, including comparison of certain Bayesian nonparametric approaches with our model.

Our modeling approach is based on a mixture formulation for the spatial NHPP intensity function. Hence, as we discuss in Section 2.4, it has analogies with the work of Wolpert and Ickstadt (1998a) and Ishwaran and James (2004) as well as with the non-Bayesian work of Brix (1999). However, our mixture model is directly for densities offering practical advantages in terms of model formulation (including the choice of mixture kernel and the extend of mixing over its parameters), prior specification, and implementation of computational methods for posterior inference. We utilize an equivalent representation of the intensity function, over the bounded region where the spatial point pattern is observed, through a positive scalar parameter and a density function. Extending the work in Kottas (2005) for NHPPs over time, we develop a nonparametric mixture model for this density. Particular emphasis is placed on the choice of the mixture kernel, for which we use a flexible family of bivariate Beta densities. By using a kernel with a bounded support we avoid edge effect problems. A Dirichlet process (DP) prior (Ferguson, 1973, 1974) is used for the mixing distribution. The resulting mixture model enables highly flexible data-driven intensity shapes. Using well-established techniques for DP mixture

models, we develop a posterior simulation method, which yields full and exact inference for the intensity function.

In addition to illustrations with standard problems involving forestry data sets, we consider applications of the methodology to extreme value analysis problems. Spatial NHPP models can be utilized to study the distribution of extreme values. Pickands (1971) considers the process given by the levels and times at which a process crosses a given threshold. Following asymptotic arguments, the process is Poisson and its intensity has a specific parametric form. The traditional parametric approach requires that trends and clusters are removed from the data. This is usually done in an ad-hoc empirical fashion, and it is heavily dependent on the actual value of the threshold that is used. More sophisticated parametric clustering of extreme values is developed in Walshaw (1999) and, more generally, in Bottolo et al. (2003). A nonparametric model for the intensity function provides a more flexible description of the trends and clusters that may be present in the process of extremes. The utility of our nonparametric mixture model in this setting is demonstrated with a data set consisting of extremes from the Dow Jones index.

The plan of the paper is as follows. In Section 2 we describe the model including details on prior specification, the computational approach to posterior inference, and the connection of the proposed method with the existing literature. Section 3 discusses applications of the methodology to the analysis of extreme values. Section 4 provides data illustrations. We conclude with a summary and discussion of possible extensions in Section 5.

2 The modeling approach

2.1 The nonparametric mixture model

Assuming the usual Borel σ-field of R^2, a spatial Poisson point process, defined on a (measurable) space $S \subseteq R^2$, is a random countable subset Π of S. It is governed by a stochastic mechanism that induces two properties for random variables $N(A) = |\Pi \cap A|$, i.e., the number of points of Π lying in measurable subsets A of S. Specifically, for any finite collection A_1,...,A_k of pairwise disjoint measurable subsets of S, the random variables $N(A_1)$,...,$N(A_k)$ are independent; and for any measurable subset A of S, $N(A)$ follows

a Poisson distribution with mean $\int_A \lambda(\boldsymbol{z})\mathrm{d}\boldsymbol{z}$. Here, $\lambda(\cdot)$ is the intensity function of the spatial NHPP, a non-negative measurable function defined on S such that $\int_A \lambda(\boldsymbol{z})\mathrm{d}\boldsymbol{z} < \infty$ for all bounded subsets A of S. Thus the distributional properties of a spatial NHPP are determined by its intensity function $\lambda(\cdot)$, or, equivalently, by the mean measure of the process, $\Lambda(A) = \int_A \lambda(\boldsymbol{z})\mathrm{d}\boldsymbol{z}$ defined for all measurable subsets A of S. For theoretical background on spatial Poisson processes, see, for instance, Cressie (1993), Kingman (1993), and Daley and Vere-Jones (2003).

We develop a prior probability model for the intensity function $\lambda(\cdot)$ of a spatial NHPP. Inference for λ is based on the spatial point pattern $\{\boldsymbol{y}_1, ..., \boldsymbol{y}_n\}$ observed in a bounded region $D \subset S$, where $\boldsymbol{y}_i = (y_{i1}, y_{i2})$ includes the coordinates for the location of the i-th point. In the applications we are interested in, as well as for most approaches discussed in the literature, D is a bounded rectangle, and, thus, without loss of generality, we assume that $D = (0,1) \times (0,1)$. (Inference for general $D = (a_1, b_1) \times (a_2, b_2)$, as in Section 4.2, can be obtained through a linear transformation.) The likelihood for λ based on data $= \{\boldsymbol{y}_1, ..., \boldsymbol{y}_n\}$ can be expressed as

$$\mathcal{L}(\lambda; \text{data}) \propto \exp\{-\Lambda(D)\} \prod_{i=1}^{n} \lambda(\boldsymbol{y}_i). \tag{1}$$

Letting $\gamma \equiv \Lambda(D) < \infty$, the function $f(\boldsymbol{y}) = \lambda(\boldsymbol{y})/\gamma$, $\boldsymbol{y} = (y_1, y_2) \in D$, is a density over D. Hence, the intensity function $\lambda(\cdot)$ can be equivalently represented by the density function $f(\cdot)$ and the scalar parameter $\gamma > 0$. Based on (1), the likelihood for (γ, f) is given by

$$\mathcal{L}(\gamma, f; \text{data}) \propto \exp(-\gamma)\gamma^n \prod_{i=1}^{n} f(\boldsymbol{y}_i). \tag{2}$$

Therefore, a prior probability model for $f(\cdot)$, along with a prior for γ, will induce a prior model for $\lambda(\cdot)$. Evidently, γ affects only the scale of $\lambda(\cdot)$, and, thus, a flexible prior model for the density $f(\cdot)$ will enable flexible modeling for the intensity function.

We thus seek to develop a nonparametric prior model for densities, defined on the unit square, which allows general density shapes. We employ a mixture model,

$$f(\boldsymbol{y}) \equiv f(\boldsymbol{y} \mid G) = \int_{\Theta} k(\boldsymbol{y} \mid \boldsymbol{\theta})\mathrm{d}G(\boldsymbol{\theta}), \tag{3}$$

where $k(\boldsymbol{y} \mid \boldsymbol{\theta})$ is a parametric kernel defined on D, with parameter vector $\boldsymbol{\theta} \in \Theta \subseteq R^d$. Therefore, there is no need for the artificial enlargement of D, as for some of the existing

approaches. The mixture is nonparametric because the random mixing distribution G is assigned a nonparametric prior. Specifically, a DP prior is taken for G, denoted by $\text{DP}(G \mid \alpha, G_0)$. Here α is the total mass (precision) parameter of the DP, and G_0 is the base (centering) distribution. The use of the DP prior in the generic mixture setting of the form in (3) is standard when the mixture kernel is supported by R, R^+ or R^d. Related work dates back to Ferguson (1983), Lo (1984), Kuo (1986), and Brunner and Lo (1989). All of these earlier approaches build on several key theoretical results given by Antoniak (1974) who studied mixtures of DPs, i.e., DPs with random α and/or random G_0 hyperparameters. We refer to MacEachern and Müller (2000) and Müller and Quintana (2004) for reviews of more recent work with DP mixture models. However, we note that less standard is DP mixture modeling (and, more generally, Bayesian nonparametric modeling) for densities defined on bounded regions of R^2. The well-studied DP mixture of bivariate normals model is not optimal in this context, since inference for the density and, thus, also for the intensity function would be subject to edge effects, especially for point patterns with observations close to the boundaries of D.

Given the flexibility of the Beta distribution, a bivariate distribution with Beta marginals emerges as a natural choice for the mixture kernel $k(\boldsymbol{y} \mid \boldsymbol{\theta})$. We work with the following bivariate Beta density

$$k(y_1, y_2 \mid \mu_1, \mu_2, \tau_1, \tau_2, \psi) = \text{be}(y_1 \mid \mu_1, \tau_1)\text{be}(y_2 \mid \mu_2, \tau_2)\{1 + \psi(y_1 - \mu_1)(y_2 - \mu_2)\}, \quad (4)$$

whence $\boldsymbol{\theta} = (\mu_1, \mu_2, \tau_1, \tau_2, \psi)$, and the parameter space $\Theta \subseteq R^5$. Here, $y_j \in (0,1)$, $j = 1, 2$, and $\text{be}(y \mid \mu, \tau)$ denotes the Beta density parameterized in terms of the mean $\mu \in (0,1)$ and a scale parameter $\tau > 0$, i.e.,

$$\text{be}(y \mid \mu, \tau) = \frac{y^{\mu\tau - 1}(1-y)^{\tau(1-\mu)-1}}{\text{Be}(\mu\tau, \tau(1-\mu))},$$

where $\text{Be}(a, b) = \int_0^1 u^{a-1}(1-u)^{b-1}du$, $a > 0$, $b > 0$. For (4) to be a valid density, the range of values for parameter ψ must satisfy $1 + \psi(y_1 - \mu_1)(y_2 - \mu_2) \geq 0$, for all $(y_1, y_2) \in D$. This restriction yields $\underline{C}(\mu_1, \mu_2) \leq \psi \leq \overline{C}(\mu_1, \mu_2)$, where $\underline{C}(\mu_1, \mu_2) = -\left[\max\{\mu_1\mu_2, (1-\mu_1)(1-\mu_2)\}\right]^{-1} < 0$, and $\overline{C}(\mu_1, \mu_2) = -\left[\min\{\mu_1(\mu_2 - 1), \mu_2(\mu_1 - 1)\}\right]^{-1} > 0$. Straightforwardly, the marginals associated with (4) are $\text{be}(\cdot \mid \mu_j, \tau_j)$ densities,

$j = 1, 2$. Moreover, the correlation resulting from (4) is given by

$$\psi \left[\mu_1(1-\mu_1)\mu_2(1-\mu_2)(1+\tau_1)^{-1}(1+\tau_2)^{-1}\right]^{1/2}.$$

The parameter ψ can be viewed as a dependence parameter; $\psi = 0$ corresponds to the case of independence, and, having specified the other four parameters, ψ controls the range of correlation values.

The family of densities in (4) belongs to the Sarmanov class of densities (see Kotz et al., 2000, Chapter 44, for more details and further references). It yields several different shapes, including unimodal densities as well as non-increasing densities with most of their mass near the boundaries of D. As an illustration, Figure 1 shows density contour plots for six combinations of parameter values: the upper left panel corresponds to $(\mu_1, \mu_2, \tau_1, \tau_2, \psi) = (0.5, 0.5, 10, 10, 0)$; the upper middle panel to $(0.5, 0.5, 1, 1, 0.9)$; the upper right panel to $(0.5, 0.1, 1, 10, 0)$; the lower left panel to $(0.9, 0.1, 1, 10, 0.7)$; the lower middle panel to $(0.5, 0.9, 10, 10, 0.7)$; and the lower right panel to $(0.5, 0.9, 10, 0.1, 0.7)$.

2.2 Prior specification

To build the 5-variate distribution $G_0 \equiv G_0(\mu_1, \mu_2, \tau_1, \tau_2, \psi)$, we work with independent components for μ_j and τ_j, $j = 1, 2$, and a conditional distribution for ψ given μ_1 and μ_2. Specification of the distribution for μ_1, μ_2, and ψ is facilitated by the bounded support for these parameters of the mixture kernel defined in (4). In particular, we take a uniform distribution on $(0, 1)$ for μ_1 and μ_2, and a uniform distribution over $(\underline{C}(\mu_1, \mu_2), \overline{C}(\mu_1, \mu_2))$ for ψ. We use inverse gamma distributions for the τ_j with fixed shape parameters $a_j > 1$ and random scale parameters β_j, so that the mean, given β_j, is $\beta_j/(a_j - 1)$, $j = 1, 2$. From a practical point of view, this choice of G_0 is attractive as it leads to a DP mixture model whose implementation requires elicitation of only three hyperparameters, β_1, β_2, and α. Moreover, empirical results based on the analysis of data sets with varying sample sizes (including the ones discussed in Section 4) indicate that this specification for G_0 is sufficiently flexible in uncovering clusters and non-standard patterns suggested by the data.

We assign exponential priors $p(\beta_j)$, with means c_j, to β_j, $j = 1, 2$, and a gamma prior

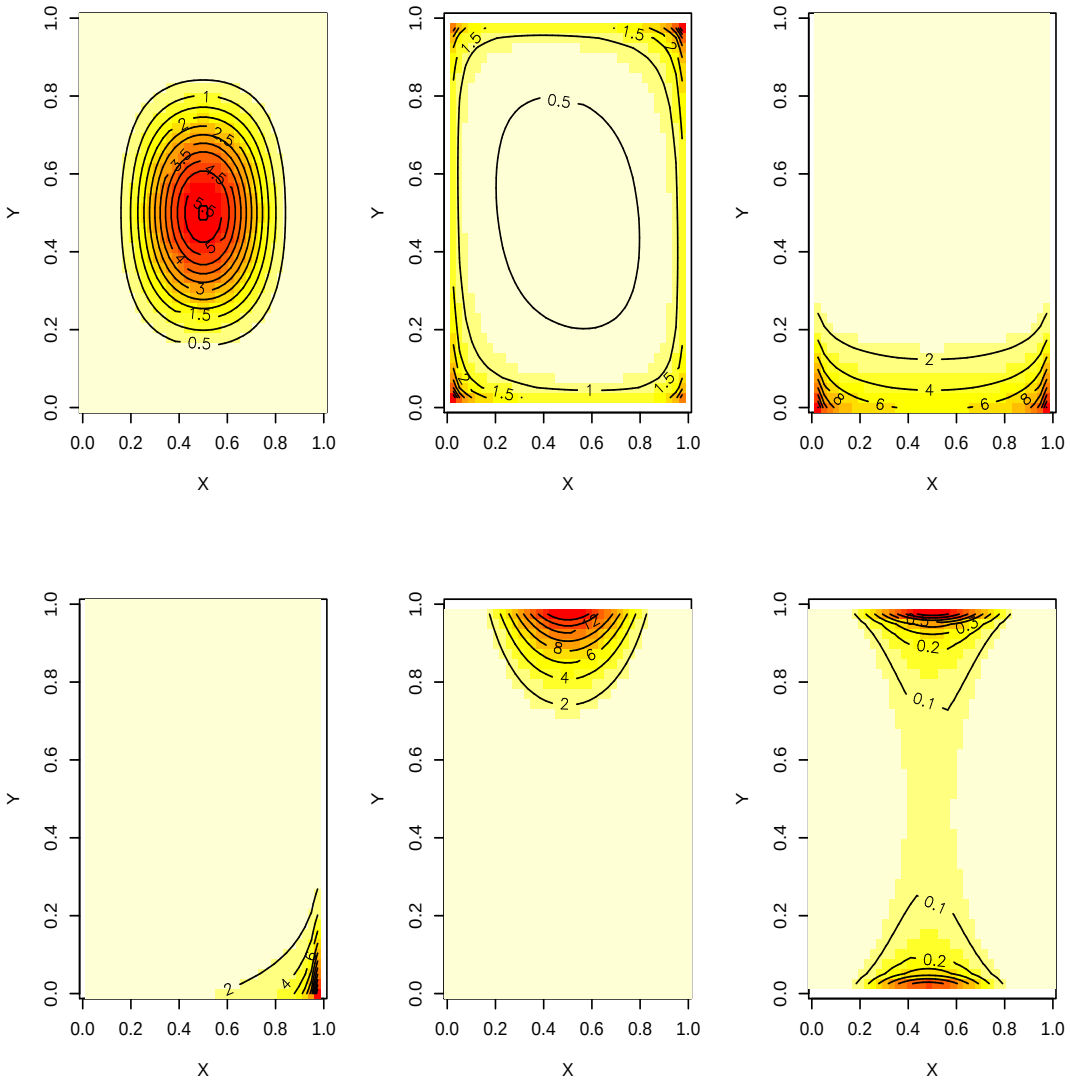

Figure 1: Contour plots of the bivariate Beta density in (4), with (X,Y) corresponding to (y_1, y_2). See Section 2.1 for the parameter values associated with each panel.

$p(\alpha)$ to α, with mean and rate parameters a_α and b_α, respectively. In the examples of Section 4 we have used $a_j = 2$, $j = 1, 2$, which implies infinite variances for the respective inverse gamma distributions. To specify the c_j, we note that the τ_j control the variance, $\mu_j(1-\mu_j)/(1+\tau_j)$, for the marginals of the mixture kernel (4). Using marginal prior means for μ_j and τ_j, a proxy for this variance is $0.25(1+c_j)^{-1}$. Therefore, c_j can be specified based on, say, $0.25(1+c_j)^{-1} = (R/4)^2$, where R is a guess at the range that will be needed so that the mixture components can capture clusters in the spatial point pattern.

Prior elicitation for the precision parameter α of the DP prior is facilitated by the role it plays in the DP mixture model. In particular, α controls the number, n^*, of distinct mixture components, i.e., the number of distinct $\boldsymbol{\theta}_i$, where $\boldsymbol{\theta}_i = (\mu_{1i}, \mu_{2i}, \tau_{1i}, \tau_{2i}, \psi_i)$ is the mixing parameter vector corresponding to $\boldsymbol{y}_i$. Larger values of α yield higher prior probabilities for larger n^*. For example, a useful approximation (for moderately large n) to the marginal prior mean for n^* is given by $\mathrm{E}(n^*) \approx (a_\alpha/b_\alpha)\log\{1+(nb_\alpha/a_\alpha)\}$ (see, e.g., Antoniak, 1974; Escobar and West, 1995; Liu, 1996).

Finally, regarding the prior $p(\gamma)$ for γ, one could use a gamma distribution, or, in the interest of obtaining a more automatic approach to prior specification, a non-informative prior can be used. Following the methodology used to obtain reference priors (see, e.g., Bernardo, 2005), we can use the prior structure specified above to obtain a marginal likelihood for γ, say, $\mathcal{L}^*(\gamma; \text{data})$, by integrating out all other parameters over their (proper) priors. In fact, $\log \mathcal{L}^*(\gamma; \text{data}) \propto -\gamma + n\log\gamma$. The Fisher's information based on this marginalized likelihood yields the reference prior $p(\gamma) \propto \gamma^{-1}$, which is used for all three data examples in Section 4.

2.3 Posterior inference

Combining the likelihood in (2) with the mixture model for the density $f(\cdot)$ given by (3), and with the prior structure discussed in Section 2.2, the full Bayesian model can be expressed as

$$p(\alpha)p(\beta_1)p(\beta_2)\mathrm{DP}(G \mid \alpha, G_0(\beta_1,\beta_2))p(\gamma)\exp(-\gamma)\gamma^n \prod_{i=1}^{n} k(\boldsymbol{y}_i \mid \boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i \mid G), \qquad (5)$$

where $p(\boldsymbol{\theta}_i \mid G)$ denotes the *density* for the $\boldsymbol{\theta}_i$ under their distribution G. Note that, under the mixture model structure, the $\boldsymbol{\theta}_i$, given G, are i.i.d. from G. Thus, if we marginalize over the $\boldsymbol{\theta}_i$, the product term in (5) is replaced with $\prod_{i=1}^{n} \int_{\Theta} k(\boldsymbol{y}_i \mid \boldsymbol{\theta}_i) \mathrm{d}G(\boldsymbol{\theta}_i)$.

Under either of the prior choices for γ discussed in Section 2.2, it can be seen from (5), that the marginal posterior for γ is a gamma distribution. In particular, under the $p(\gamma) \propto \gamma^{-1}$ prior, $p(\gamma|\text{data})$ is a gamma$(n, 1)$ distribution.

Moreover, the full posterior corresponding to model (5) can be expressed as

$$p(\gamma, G, \boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2|\text{data}) = p(\gamma|\text{data}) p(G, \boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2|\text{data}),$$

where $\boldsymbol{\vartheta} = \{\boldsymbol{\theta}_i : i = 1, ..., n\}$ collects all the latent mixing parameter vectors. Hence, to explore $p(\gamma, G, \boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2 \mid \text{data})$, it suffices to sample from $p(G, \boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2 \mid \text{data})$, i.e., the posterior for the DP mixture part of model (5). We use a combination of posterior simulation techniques for DP mixture models to obtain posterior draws $\{G_\ell, \boldsymbol{\vartheta}_\ell, \alpha_\ell, \beta_{1\ell}, \beta_{2\ell} : \ell = 1, ..., L\}$ from $p(G, \boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2|\text{data})$.

First, based on results from Antoniak (1974), we have the following decomposition,

$$p(G, \boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2 \mid \text{data}) = p(G \mid \boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2) p(\boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2 \mid \text{data}). \tag{6}$$

Here, $p(G \mid \boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2) = \mathrm{DP}(G \mid \alpha', G_0')$, where $\alpha' = \alpha + n$, and

$$G_0'(\boldsymbol{\theta} \mid \boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2) = \frac{\alpha}{\alpha + n} G_0(\boldsymbol{\theta} \mid \beta_1, \beta_2) + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{\boldsymbol{\theta}_i}(\boldsymbol{\theta}), \tag{7}$$

where $\boldsymbol{\theta} = (\mu_1, \mu_2, \tau_1, \tau_2, \psi)$, and $\delta_z(\cdot)$ denotes a point mass at z. Moreover, $p(\boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2 \mid \text{data})$ is the posterior that results after marginalizing G in (5) over its DP prior,

$$p(\boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2 \mid \text{data}) \propto p(\alpha) p(\beta_1) p(\beta_2) p(\boldsymbol{\vartheta} \mid \alpha, \beta_1, \beta_2) \prod_{i=1}^{n} k(\boldsymbol{y}_i \mid \boldsymbol{\theta}_i).$$

The marginal joint prior for $\boldsymbol{\vartheta}$ arises from a Pólya urn scheme, described in Blackwell and MacQueen (1973), which is implicit in the DP structure. Specifically,

$$p(\boldsymbol{\vartheta} \mid \alpha, \beta_1, \beta_2) = g_0(\boldsymbol{\theta}_1) \prod_{i=2}^{n} \left\{ \frac{\alpha}{\alpha + i - 1} g_0(\boldsymbol{\theta}_i \mid \beta_1, \beta_2) + \frac{1}{\alpha + i - 1} \sum_{m=1}^{i-1} \delta_{\boldsymbol{\theta}_m}(\boldsymbol{\theta}_i) \right\}, \tag{8}$$

where g_0 is the density of G_0.

Several posterior simulation methods have been suggested in the literature for sampling from the marginal posterior that results after integrating out the random mixing distribution in a DP mixture model (see, e.g., Escobar and West, 1998; MacEachern, 1998; Neal, 2000). To obtain posterior samples from $p(\boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2 \mid \text{data})$, we have found that a version of algorithm 5 from Neal (2000) strikes a good balance between ease of implementation and efficiency. Briefly, we use a Gibbs sampler updating each of the $\boldsymbol{\theta}_i$ with a Metropolis step, which uses the prior full conditional for $\boldsymbol{\theta}_i$ as the proposal distribution. Based on the Pólya urn structure of $p(\boldsymbol{\vartheta} \mid \alpha, \beta_1, \beta_2)$ in (8), the prior full conditional for $\boldsymbol{\theta}_i$ is a mixed distribution with point masses $(\alpha + n - 1)^{-1}$ at $\boldsymbol{\theta}_m$, $m \neq i$, and continuous mass $\alpha(\alpha + n - 1)^{-1}$ on G_0. After all the $\boldsymbol{\theta}_i$ are updated in a Gibbs sampler cycle, we obtain a number, n^*, of distinct values, $\boldsymbol{\theta}^*_\ell = (\mu^*_{1\ell}, \mu^*_{2\ell}, \tau^*_{1\ell}, \tau^*_{2\ell}, \psi^*_\ell)$, $\ell = 1, ..., n^*$, in the vector $\boldsymbol{\vartheta}$. The full conditional for β_j, $j = 1, 2$, is a gamma distribution with shape parameter $n^* a_j + 1$ and rate parameter $c_j^{-1} + \sum_{\ell=1}^{n^*} (\tau^*_{j\ell})^{-1}$. Finally, we update α using the augmentation method discussed in Escobar and West (1995).

Next, we use the approach described in Gelfand and Kottas (2002) and Kottas (2006) to sample from the marginal posterior of the mixing distribution G. Based on (6), this can be accomplished by combining the posterior draws $\boldsymbol{\vartheta}_\ell = (\boldsymbol{\theta}_{1\ell}, .., \boldsymbol{\theta}_{n\ell})$, α_ℓ, $\beta_{1\ell}$, $\beta_{2\ell}$ from $p(\boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2 \mid \text{data})$ with additional draws G_ℓ from $p(G \mid \boldsymbol{\vartheta}_\ell, \alpha_\ell, \beta_{1\ell}, \beta_{2\ell})$. As discussed above, the distribution $p(G \mid \boldsymbol{\vartheta}, \alpha, \beta_1, \beta_2)$ is given by a DP, which can be sampled using its constructive definition (Sethuraman and Tiwari, 1982; Sethuraman, 1994). Based on this definition, a realization from $\text{DP}(G \mid \alpha, G_0)$ is (almost surely) of the form $G = \sum_{m=1}^{\infty} \omega_m \delta_{\phi_m}$, where $\omega_1 = z_1$, $\omega_m = z_m \prod_{r=1}^{m-1}(1 - z_r)$, $m = 2,3,...$, with z_r i.i.d. from a $\text{Beta}(1, \alpha)$ distribution, and, independently, ϕ_m i.i.d. from G_0. We obtain the draws G_ℓ using a partial sum approximation to the countable sum representation above. Specifically, we take $G_\ell = \sum_{m=1}^{M} w_{m\ell} \delta_{\boldsymbol{\phi}_{m\ell}}$, where $w_{1\ell} = z_{1\ell}$, $w_{m\ell} = z_{m\ell} \prod_{r=1}^{m-1}(1 - z_{r\ell})$, $m = 2, ..., M-1$, $w_{M\ell} = 1 - \sum_{m=1}^{M-1} w_{m\ell} = \prod_{r=1}^{M-1}(1 - z_{r\ell})$, with $z_{r\ell}$ i.i.d. $\text{Beta}(1, \alpha_\ell + n)$, and, independently, $\boldsymbol{\phi}_{m\ell}$ i.i.d. from $G'_0(\cdot \mid \boldsymbol{\vartheta}_\ell, \alpha_\ell, \beta_{1\ell}, \beta_{2\ell})$ given in (7). The partial sum approximation can be made arbitrarily accurate. For example, a simple approach to specifying M, uses the result $\text{E}(\sum_{m=1}^{M-1} w_{m\ell} \mid \alpha_\ell) = 1 - \{(\alpha_\ell + n)/(\alpha_\ell + n + 1)\}^{M-1}$, to choose M that makes, say, $\{(n + \max_\ell \alpha_\ell)/(n + 1 + \max_\ell \alpha_\ell)\}^{M-1}$ arbitrarily small.

The samples $\{G_\ell : \ell = 1, ..., L\}$ from the posterior of the mixing distribution provide full inference for the intensity function at any spatial location $\boldsymbol{y} \in D$. Note that $f_\ell(\boldsymbol{y}) = \sum_{m=1}^{M} w_{m\ell} k(\boldsymbol{y} \mid \boldsymbol{\phi}_{m\ell})$, $\ell = 1, ..., L$, are draws from the posterior of the density function $f(\boldsymbol{y} \mid G)$ at $\boldsymbol{y}$. Therefore, using posterior draws γ_ℓ for γ, we obtain samples $\gamma_\ell f_\ell(\boldsymbol{y})$, $\ell = 1, ..., L$, from the posterior of the intensity function $\lambda(\boldsymbol{y} \mid \gamma, G) = \gamma f(\boldsymbol{y} \mid G)$ at $\boldsymbol{y}$. Working with a grid $\{\boldsymbol{y}_c : c = 1, ..., C\}$ of spatial locations over D, we can provide inference for the NHPP intensity surface. We work with means and medians based on the posterior samples $\{\gamma_\ell f_\ell(\boldsymbol{y}_c) : \ell = 1, ..., L\}$ at each $\boldsymbol{y}_c$ to produce point estimates, referred to in Section 4 as posterior mean and posterior median intensity estimates, respectively. We use interquartile ranges from the $\{\gamma_\ell f_\ell(\boldsymbol{y}_c) : \ell = 1, ..., L\}$ posterior samples at each $\boldsymbol{y}_c$ to provide a measure of the uncertainty associated with the point estimates; in Section 4, we refer to this estimate as the posterior interquartile range intensity estimate.

2.4 Discussion and literature review

Here, we review approaches to inference for spatial NHPPs, including Bayesian nonparametric methods, in an effort to indicate our contribution within this research field. Note that, instead of a NHPP, the classical literature uses the term *Cox process* (or doubly stochastic Poisson process), i.e., a NHPP with random intensity function. Under the Bayesian nonparametric framework, the intensity function is treated as a random realization from a (prior) process, and, thus, the data could also be viewed as arising from a doubly stochastic Poisson process.

One of the first approaches to Bayesian nonparametric inference for spatial NHPP intensities can be found in Heikkinen and Arjas (1998), where piecewise constant functions, driven by Voronoi tesselations and Markov random field priors, were used to model the intensity function. See also Heikkinen and Arjas (1999) for an extension to a model for spatial point patterns influenced by concomitant variables.

A different line of research involves log-Gaussian Cox process models, i.e., spatial NHPPs directed by a random intensity function, which is modeled, on the logarithmic scale, with a stationary real-valued Gaussian process. Møller et al. (1998) study properties of log-Gaussian Cox processes and discuss empirical Bayesian inference for the intensity

surface. Extensions to spatio-temporal settings are considered in Brix and Diggle (2001) and Brix and Møller (2001).

Both of the approaches discussed above require, for full Bayesian inference, difficult to tune computational schemes; the former uses reversible jump Markov chain Monte Carlo methods; the latter a Metropolis-adjusted Langevin algorithm. Simulation-based model fitting (as discussed in Section 2.3) for the proposed mixture model of Section 2.1 is, arguably, easier to implement.

Our modeling approach is closer, in spirit, to the Bayesian nonparametric approaches developed by Wolpert and Ickstadt (1998a) and Ishwaran and James (2004). Both of these approaches utilize a mixture representation for the intensity function, $\lambda(\boldsymbol{y}) \equiv \lambda(\boldsymbol{y}; H, \boldsymbol{\varphi}) = \int_\Theta m(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\varphi}) H(d\boldsymbol{\theta})$, $\boldsymbol{y} \in D$, where $m(\cdot; \boldsymbol{\theta}, \boldsymbol{\varphi})$ is a non-negative kernel (that is not a density) specified up to a parameter vector $(\boldsymbol{\theta}, \boldsymbol{\varphi})$, with $\boldsymbol{\theta} \in \Theta \subseteq R^d$, and $H(d\boldsymbol{\theta})$ is the mixing measure. A parametric prior is assigned to $\boldsymbol{\varphi}$. The prior distribution for $H(d\boldsymbol{\theta})$ is an inhomogeneous gamma process with shape parameter $a(d\boldsymbol{\theta})$, a finite measure on Θ, and scale parameter $b(\boldsymbol{\theta})$, a positive integrable function on Θ. This prior structure implies that for each Borel-measurable $C \subset \Theta$, $H(C) = \int_C b(\boldsymbol{\theta}) \Gamma(d\boldsymbol{\theta})$, where $\Gamma(d\boldsymbol{\theta})$ follows a gamma process distribution over Θ with shape parameter $a(d\boldsymbol{\theta})$. A gamma process is an independent increment process such that $\Gamma(C)$ has a gamma$(a(C), 1)$ distribution, for each Borel-measurable $C \subset \Theta$. Note that, under the framework of Wolpert & Ickstadt (1998a), the inhomogeneous gamma random field for the mixing measure can be replaced by another Lévy random field. Details on the corresponding posterior simulation methods are given in Wolpert and Ickstadt (1998b); applications to regression settings are discussed by Ickstadt and Wolpert (1999) and Best et al. (2000).

Note that the illustrations given in Wolpert and Ickstadt (1998a), Ickstadt and Wolpert (1999) and Best et al. (2000) are based on the Poisson-gamma random field model, i.e., the special case of the mixture model described above with a constant scale function $b(\boldsymbol{\theta}) \equiv b$. A connection of this model with DP mixture models is possible, recalling the connection between the DP and the gamma process, i.e., the fact that $\Gamma(\cdot)/\Gamma(\Theta)$ follows a DP distribution with parameter $a(\cdot)$ (Ferguson, 1973, 1974). (In our notation, $a(\cdot) \equiv \alpha P_0(\cdot)$, where $\alpha = a(\Theta)$, and P_0 is the probability measure over Θ corresponding to distribution G_0.)

For example, as discussed in Ickstadt and Wolpert (1999), if $m(\cdot;\boldsymbol{\theta},\boldsymbol{\varphi})$ is the Gaussian density on R^2 with mean vector $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\varphi}$ (which might be diagonal), and $b(\boldsymbol{\theta}) \equiv 1$, then the normalized intensity function, $\lambda(\cdot)/\Lambda(D) \equiv \int_\Theta m(\cdot;\boldsymbol{\theta},\boldsymbol{\varphi})\Gamma(d\boldsymbol{\theta})/\Gamma(\Theta)$, is defined by a location DP mixture of bivariate normals. Our approach differs in that the mixture representation in (3) is used directly for the density $f(\cdot) \equiv f(\cdot \mid G)$, which, along with parameter $\gamma = \Lambda(D)$, defines the intensity function over the bounded region D. Potential advantages of the model formulation developed in Section 2.1 include the following. The mixture $f(\cdot \mid G)$ is fully nonparametric and, hence, it allows clustering for the location parameters as well as the scale parameters and the dependence parameter of the mixture kernel in (4). (Note that $\lambda(\cdot; H, \boldsymbol{\varphi})$ is a semiparametric mixture model. In particular, the choice for the kernel function $m(\cdot;\boldsymbol{\theta},\boldsymbol{\varphi})$ is, typically, a scaled Gaussian density, with mean vector $\boldsymbol{\theta}$ and uncorrelated components, whence the mixing is with respect to the mean vector only.) Moreover, the fact that the bivariate Beta density kernel (4) is supported by D leads to inference for the intensity function, which avoids edge effects problems. Finally, the methods for prior specification (Section 2.2) and posterior simulation for full inference (Section 2.3) build on a familiar Bayesian density estimation framework, and are, thus, at least as easy to implement as the respective methods of the related Bayesian work discussed above.

Finally, we note the work of Brix (1999) who studied generalized gamma measures (G-measures) and defined shot-noise G Cox processes, which are Cox processes driven by kernel smoothed G-measures. Because G-measures include gamma processes, shot-noise G Cox processes include as a special case the Poisson-gamma random field model. Spatio-temporal modeling with shot-noise G Cox processes is considered by Brix and Chadœuf (2002). However, this work does not involve a Bayesian model formulation; in particular, estimation is based on a minimum contrast method (Møller et al., 1998) rather than the likelihood function. Related and, in fact, more general probability models for Cox processes are studied by Møller (2003) and Møller and Torrisi (2005), though the focus of this work is on probabilistic aspects of the spatial processes rather than on statistical modeling and applications.

3 Applications to extreme value analysis problems

Methods for the analysis of the distribution of extreme values have been widely developed in the literature. Coles (2001) provides a very readable account of the most popular models and inferential methods. Three approaches are commonly used in the univariate case: the Generalized Extreme Value Distribution (GEV), the Generalized Pareto Distribution (GPD), and the point process approach. Suppose we have an i.i.d. sequence of random variables, $X_1, X_2, \ldots$, and let $M_N = \max\{X_1, ..., X_N\}$. It can be shown that, if there exist sequences $a_N > 0$ and b_N such that $\Pr((M_N - b_N)/a_N \leq z) \to H(z)$ for a non-degenerate distribution H, then H is a GEV. The GEV is a three parameter family of distributions given by $H(x) = \exp\left\{-\left(1 + (x-\mu)\xi\psi^{-1}\right)_+^{-\xi^{-1}}\right\}$, where $\psi > 0$ and μ, ξ are real-valued parameters. (Here, $z_+ = \max\{0, z\}$ denotes the positive part of z.) For a sequence of observations in time, inference on the parameters of the GEV is obtained by assuming that the maxima over a given time unit are distributed as $H(x)$.

An alternative approach arises by considering the exceedances over a given threshold. Inference is based on the fact that, if the GEV limit holds for a random variable X, then the distribution of $X - u$ conditionally on $X > u$ is, for large enough u, approximately equal to the GPD, i.e., $G(x) = 1 - \left(1 + x\xi\sigma^{-1}\right)_+^{-\xi^{-1}}$ where $\sigma = \psi + \xi(u - \mu)$. Here u is intended as the threshold. Inference based on exceedences can be performed by postulating the following model: (a) The number n of exceedences over the threshold u in any unit of time has a Poisson distribution. (b) Conditionally on $n \geq 1$, the values of the excesses are i.i.d. with a GPD.

A full point process approach consists of considering the times of exceedances, say t, and the excess values, say s, as a bivariate point process. The point process approach for the analysis of the distribution of extreme values was originally introduced in Pickands (1971). Smith (1989; 2003) uses the fact that, under suitable normalization, the process of exceedances behaves as a bivariate NHPP. Asymptotic theory (as $N \to \infty$) is used to justify that the intensity of this NHPP process is given by $\lambda(t, s) = \psi^{-1}\left(1 + \xi\psi^{-1}(s - \mu)\right)_+^{-\xi^{-1}-1}$, over the domain $(0, T] \times (u, \infty)$, where T is the largest time point in the original time series. The former implies that, for a set $A = [t_1, t_2] \times [y, \infty)$,

$\Lambda(A) = \int_A \lambda(t,s)\mathrm{d}s\mathrm{d}t = (t_2 - t_1)\left(1 + \xi\psi^{-1}(y-\mu)\right)^{-\xi^{-1}}$, when $y > u$ and $1+\xi\psi^{-1}(y-\mu) > 0$. Time-varying changes in the intensity of the processes can be modeled by assuming that $\lambda(t,s) = \psi_t^{-1}(1 + \xi_t\psi_t^{-1}(s-\mu_t))^{-\xi_t^{-1}-1}$, so that the three parameters that define the intensity are allowed to vary with time.

More generally, and in order to avoid asymptotic arguments that might not be appropriate for small or moderate sample sizes N of the original time series, we can study the distribution of the extremes using the nonparametric model developed in Section 2.1 for the intensity λ. Inference proceeds by fixing a threshold u and observing the corresponding data $= \{(t_1, s_1), \ldots, (t_n, s_n)\}$. We assume that the bivariate NHPP of exceedances is defined on $(0,\infty)\times(u,V]$ and observed in $D = (0,T]\times(u,V]$, where V is a specified upper bound for the exceedances values. This is not a restrictive assumption, since the value of V can be taken to be large enough to practically imply zero intensity above V. Using the methodology described in Section 2.1, we can obtain full inference for the intensity function $\lambda(t,s)$. Such inference provides quantitative information about the times when large values are more likely to occur. It also provides information about which ranges those extreme values are likely to attain.

Of interest here is also the process of exceedances over time. According to the mapping theorem for the Poisson process, the point process, defined by the projection of (t,s) to t, is a NHPP over time with intensity function $\lambda^*(t) = \int_u^V \lambda(t,s)\mathrm{d}s$. Hence, the likelihood for the point pattern $\{t_1, ..., t_n\}$ of the times of exceedances is proportional to $\exp(-\gamma^*)\prod_{i=1}^n \lambda^*(t_i)$. Since $\gamma^* = \int_0^T \lambda^*(t)\mathrm{d}t = \int_0^T\int_u^V \lambda(t,s)\mathrm{d}s\mathrm{d}t = \gamma$, under the mixture model of Section 2.1 for the bivariate intensity $\lambda(t,s)$, we obtain the following induced mixture model for $\lambda^*(t) \equiv \lambda^*(t \mid \gamma, G)$,

$$\lambda^*(t \mid \gamma, G) = \gamma\int_u^V f(t,s \mid G)\mathrm{d}s = \int_u^V\int_\Theta k(t,s \mid \boldsymbol{\theta})\mathrm{d}G(\boldsymbol{\theta})\mathrm{d}s = \int_\Theta \mathrm{be}(t \mid \mu_1, \tau_1)\mathrm{d}G(\boldsymbol{\theta}).$$

Therefore, posterior inference for the marginal intensity function $\lambda^*(t)$, at any time point $t \in (0,T)$, is readily available using, again, the samples G_ℓ from the posterior of the mixing distribution G in the bivariate nonparametric mixture model (obtained with the computational technique described in Section 2.3). We illustrate both types of inferences discussed above in Section 4.2.

4 Data illustrations

In this section we illustrate our methods using three previously published data sets. The first example corresponds to the locations of 62 redwood seedlings, in a square of 23 meters. The data are presented in Diggle (2003) with a reference to Ripley (1977). The second data set includes the locations of 514 maple trees in a 19.6 acre square plot in Lansing Woods, Clinton County, Michigan, USA. These data are also presented in Diggle (2003). Both data sets are available from `www.maths.lancs.ac.uk/~diggle/`. The third example involves an application to the study of extremes. The data consists of a transformation of the daily closing prices of the Dow Jones Index. These data are obtained from Coles (2001) and are available from `www.maths.bris.ac.uk/~masgc/ismev/summary.html`.

Following the prior specification approach discussed in Section 2.2, for all three data sets, we used exponential priors for β_j with means $c_j = 50$, $j = 1, 2$, corresponding to $R \approx 1/4$. This choice resulted in posterior learning for β_1 and β_2, which increased with the sample size. We have also experimented with other values for c_j in the range from 20 (based on $R \approx 1/2$) to 200 (based on $R \approx 1/7$), noticing little sensitivity in the posterior inference for the intensity function. We also studied sensitivity to the prior choice for α; results are discussed in Sections 4.1 and 4.2 below.

Regarding the posterior simulation algorithm described in Section 2.3, we observed good mixing and fast convergence for all three data examples. In all cases, the reported inferences are based on $L = 15,000$ posterior samples, obtained after a (conservative) burn-in period of $30,000$ iterations.

4.1 Forestry data sets

Figure 2 presents the estimated intensity for the data on redwood seedlings. It has been noted in the literature that the data have a clustering pattern. This can be attributable to the clustering of seedlings around stumps. The positions of the stumps are unknown. Diggle (2003) discusses modeling approaches using Poisson cluster processes. We notice that our model is able to clearly identify the different clusters in the distribution of the redwood. To illustrate the robustness of posterior inference for the intensity function, Figure

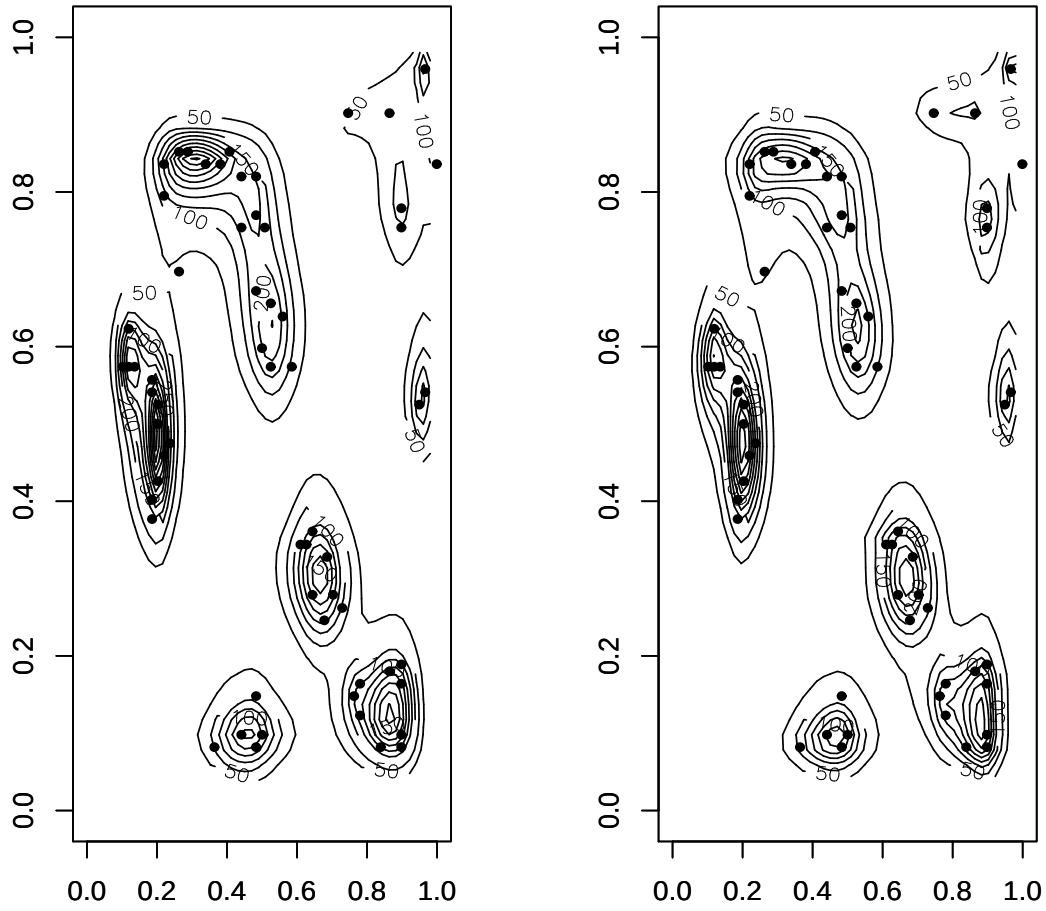

Figure 2: Redwood seedlings data. Contour plots of posterior mean intensity estimates under gamma$(2, 0.7)$ and gamma$(2, 0.15)$ priors for α (left and right panels, respectively). In both panels, the dots indicate the locations of the redwood seedlings.

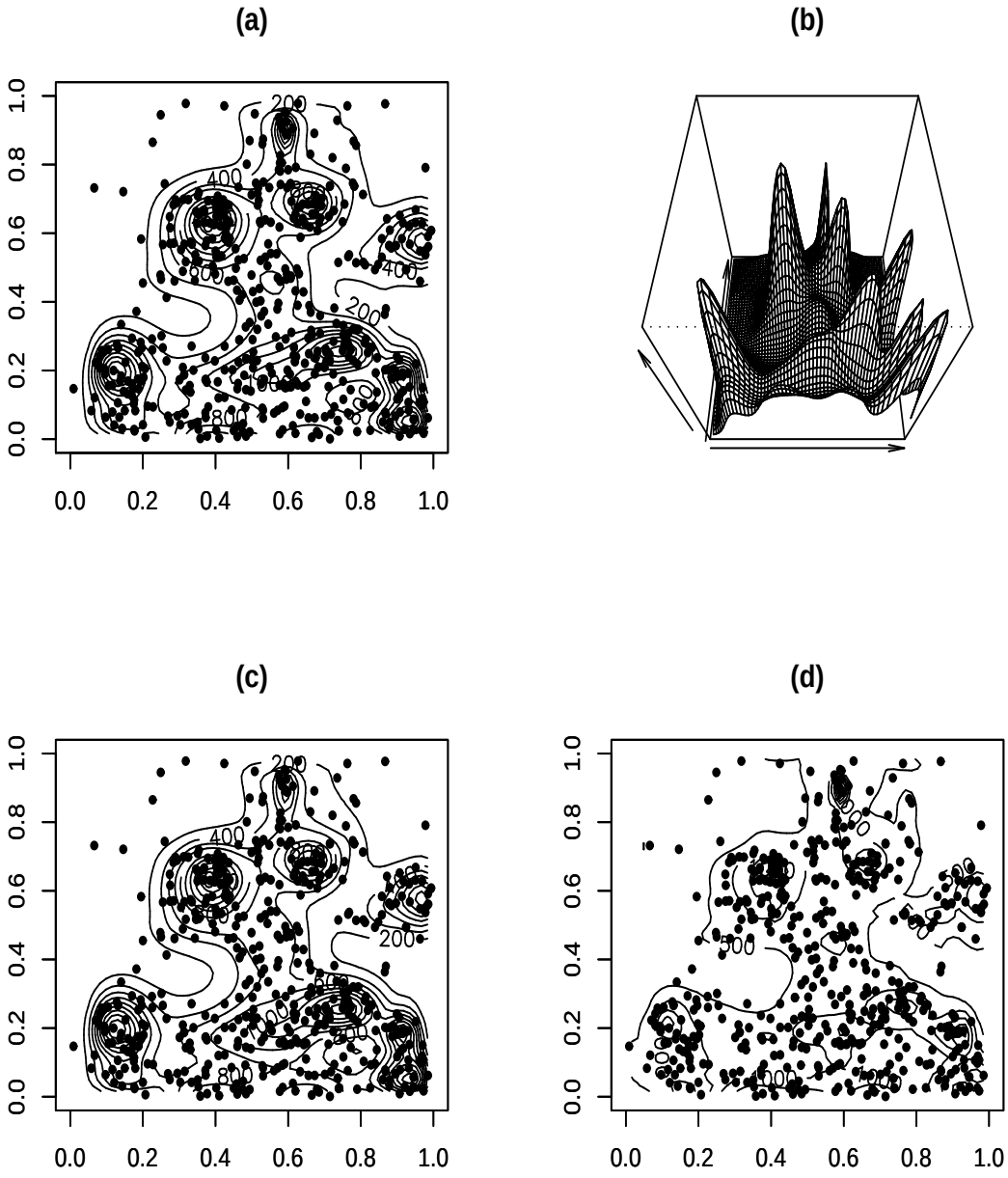

Figure 3: Maples data. Panels (a) and (b) include the posterior mean intensity estimate (contour plot and perspective plot, respectively). Panels (c) and (d) plot contour plots for the posterior median and posterior interquartile range intensity estimates, respectively. The dots superimposed on the contour plots denote the locations of the maple trees.

2 includes posterior mean intensity estimates under two prior choices for α, gamma$(2, 0.7)$ and gamma$(2, 0.15)$ priors, corresponding to $\mathrm{E}(n^*) \approx 9$ and $\mathrm{E}(n^*) \approx 23$, respectively. We observed similar robustness in the posterior median and posterior interquartile range intensity estimates (not shown). Regarding inference for n^*, the posterior 25% percentile, posterior median, and posterior 75% percentile are given by 14, 16, and 18, under the gamma$(2, 0.7)$ prior for α, and by 18, 20, and 23, under the gamma$(2, 0.15)$ prior for α.

Figure 3 shows the results for the maple trees data based on a gamma$(2, 0.197)$ prior for α, which implies $\mathrm{E}(n^*) \approx 40$. (The estimates were practically indistinguishable under two different prior choices for α corresponding to $\mathrm{E}(n^*) \approx 20$ and $\mathrm{E}(n^*) \approx 30$.) The posterior median for n^* is 44; posterior 25% and 75% percentiles are given by 39 and 50, respectively. The hypothesis that there is substantial spatial variation in the intensity of maples in Lansing Woods is confirmed. We notice that panels (a) and (c) are comparable to the results presented by Diggle (2003) in the lower panel of Figure 8.3, p 120. Our method has the advantage of providing a full probabilistic assessment of the uncertainties involved.

4.2 An example from extreme value analysis

We work with 1,303 returns for the daily closing prices of the Dow Jones index. These cover a period of about five years. The returns are calculated as the logarithm of ratios of successive observations. In order to study the behavior of the extremes, Coles (2001) suggests a parametric model based on the generalized Pareto distribution. An ad hoc de-clustering of the data is done and a threshold of $u = 2$ is used, resulting in $n = 37$ exceedances. Following this suggestion, we initially considered as the domain of the Poisson process a region defined by the rectangle $(0, 1303] \times (2, 5]$. The resulting posterior median and posterior interquartile range intensity estimates are included in Figure 4, which indicates three fairly well defined clusters.

A better understanding of how the intensity varies with time is given by the marginal intensities. Figure 5 shows posterior point estimates (based on posterior means) and 95% pointwise interval estimates for the marginal intensity over time. Panels (a), (b), and (c) correspond to the use of different thresholds, $u = 1.5$, $u = 1.75$, and $u = 2$, respectively.

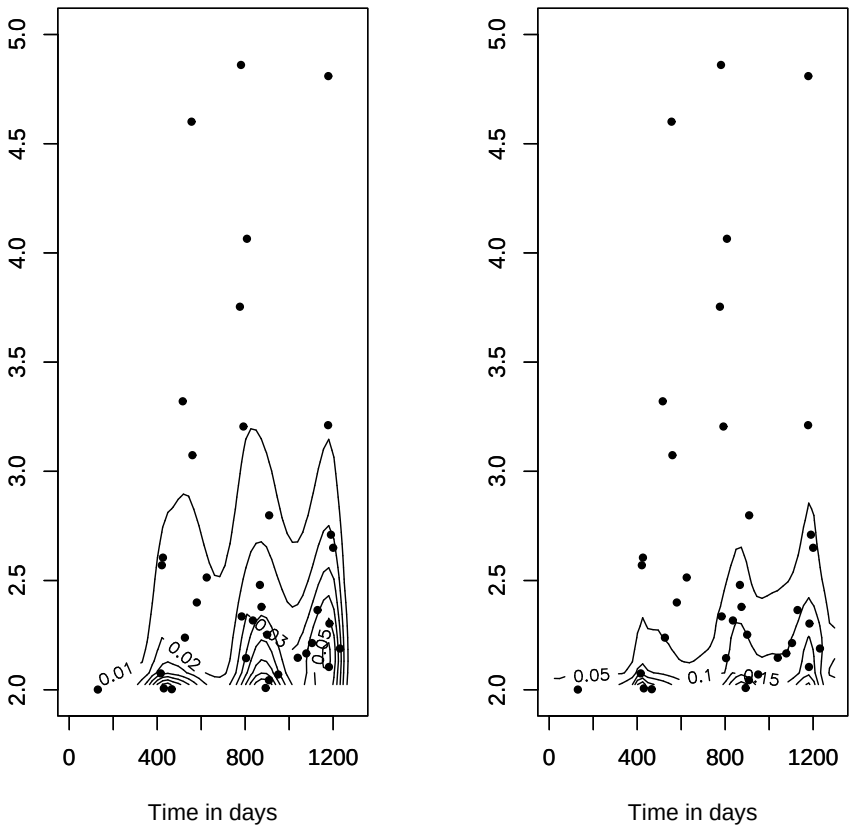

Figure 4: Dow Jones index data. Contour plots of posterior median (left panel) and posterior interquartile range (right panel) intensity estimates under $u = 2$. In both panels, the dots denote the observed times of exceedances with the corresponding excess values.

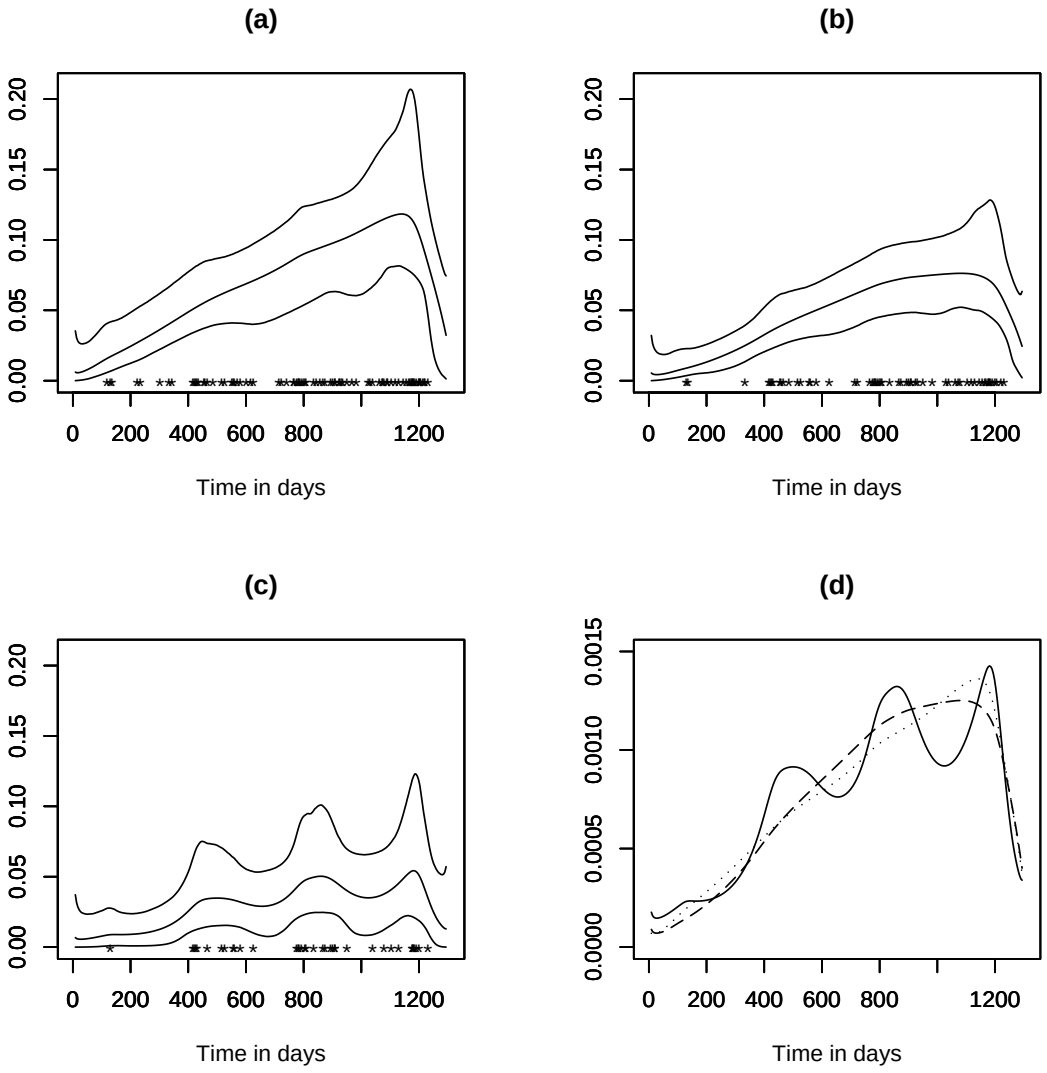

Figure 5: Dow Jones index data. Panels (a), (b), and (c) show posterior mean and 95% interval estimates for the marginal intensity function over time, under $u = 1.5$, $u = 1.75$, and $u = 2$, respectively. In each case, the observed times of exceedances are plotted on the horizontal axis. Panel (d) compares posterior mean estimates for the marginal density function, with the dotted, dashed, and solid line corresponding to $u = 1.5$, $u = 1.75$, and $u = 2$ respectively.

The associated number of exceedances are $n = 86$, $n = 60$, and $n = 37$. In all three cases, we used fairly dispersed priors for α yielding $\mathrm{E}(n^*) \approx 14$. (Again, posterior intensity estimates were not significantly affected by less dispersed priors for α.) We observe that, in all cases, there is a roughly increasing trend. Decreasing the threshold increases the number of points and has the effect of blurring the clustering. This is also clear from the comparison presented in panel (d) of Figure 5.

5 Summary and future work

We have proposed a Bayesian nonparametric approach to the modeling of spatial non-homogeneous Poisson processes. The problem of estimating the intensity of the process in a bounded domain is equivalent to that of density estimation. We have developed a non-parametric mixture model where the kernel is bounded in the unit square and the mixing distribution has a DP prior. The fact that the kernel is of bounded support is important in order to avoid spurious edge effects. We have used a 5-parameter mixture kernel given by a family of bivariate Beta densities that yields a wide range of density shapes. Moreover, we have discussed applications of the methodology to extreme value analysis problems, in particular, to the problem of describing the distribution of exceedances over a threshold. We have argued that the proposed nonparametric mixture modeling approach might be a useful alternative/addition to the related existing Bayesian nonparametric methods, some of which were reviewed in detail. We have illustrated the methodology with three data sets. The results show that the model is able to accurately describe clusters and trends in the data.

Although we have focused on estimation for Poisson process intensity functions, our approach is also applicable to bivariate density estimation problems where the support is bounded. The proposed model can be extended in several directions. A practically important one is to semiparametric regression settings, which allow for the inclusion of location-specific and/or individual-specific covariates. Another extension is that of embedding the present model in a hierarchical structure, so that two or more interacting spatial point processes can be considered jointly. A natural example of the importance

of such a model is given by forestry applications, where it would be desirable to consider the distribution of two or more species jointly. Finally, it would be of interest to develop spatio-temporal model formulations for the analysis of replicated spatial point patterns. We will report on some of these extensions in a future manuscript.

Acknowledgements

The authors thank two anonymous reviewers as well as Antonio Pievatolo (joint guest editor of the Special Issue on "Bayesian Inference in Stochastic Processes") for useful comments. The work of the first author was supported in part by NSF grant DMS-0505085.

References

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

Bernardo, J.M., 2005. Reference analysis. Handbook of Statistics 25, eds. D.K. Dey and C.R. Rao, Amsterdam: Elsevier, pp. 17–90.

Best, N.G., Ickstadt, K., Wolpert, R.L., 2000. Spatial Poisson regression for health and exposure data measured at disparate resolutions. J. Amer. Statist. Assoc. 95, 1076–1088.

Blackwell, D., MacQueen, J.B., 1973. Ferguson distributions via Pólya urn schemes. Ann. Statist. 1, 353–355.

Bottolo, L., Consonni, G., Dellaportas, P., Lijoi, A., 2003. Bayesian analysis of extremes values by mixture modeling. Extremes 6, 25–47.

Brix, A., 1999. Generalized gamma measures and shot-noise Cox processes. Adv. Appl. Prob. 31, 929–953.

Brix, A., Chadœuf, J., 2002. Spatio-temporal modelling of weeds by shot-noise G Cox processes. Biometrical J. 44, 83–99.

Brix, A., Diggle, P.J., 2001. Spatiotemporal prediction for log-Gaussian Cox processes. J. Roy. Statist. Soc. Ser. B 63, 823–841.

Brix, A., Møller, J., 2001. Space-time multi type log Gaussian Cox processes with a view to modelling weeds. Scand. J. Statist. 28, 471–488.

Brunner, L.J., Lo, A.Y., 1989. Bayes methods for a symmetric unimodal density and its mode. Ann. Statist. 17, 1550–1566.

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. New York: Springer-Verlag.

Cressie, N.A.C., 1993. Statistics for Spatial Data. (Revised Edition). New York: Wiley.

Daley, D.J., Vere-Jones, D., 2003. An Introduction to the Theory of Point Processes. (Second Edition). New York: Springer.

Diggle, P.J., 2003. Statistical Analysis of Spatial Point Patterns, Second Edition. London, Arnold.

Escobar, M., West, M., 1995. Bayesian density estimation and inference using mixtures. J. Amer. Statist. Assoc. 90, 577–588.

Escobar, M.D., West, M., 1998. Computing nonparametric hierarchical models. In Practical Nonparametric and Semiparametric Bayesian Statistics, eds. D. Dey, P. Müller and D. Sinha, New York: Springer, pp. 1–22.

Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. Ann. Statist. 1, 209-230.

Ferguson, T.S., 1974. Prior distributions on spaces of probability measures. Ann. Statist. 2, 615-629.

Ferguson, T.S., 1983. Bayesian density estimation by mixtures of normal distributions. In Recent advances in statistics, eds. M.H. Rizvi, J.S. Rustagi and D. Siegmund, New York: Academic Press, pp. 287–302.

Gelfand, A.E., Kottas, A., 2002. A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. J. Comput. Graph. Statist. 11, 289–305.

Heikkinen, J., Arjas, E., 1998. Non-parametric Bayesian estimation of a spatial Poisson intensity. Scand. J. Statist. 25, 435–450.

Heikkinen, J., Arjas, E., 1999. Modeling a Poisson forest in variable elevations: A nonparametric Bayesian approach. Biometrics 55, 738–745.

Ickstadt, K., Wolpert, R.L., 1999. Spatial regression for marked point processes. In Bayesian Statistics 6, eds. J.M. Bernardo, J.O. Berger, P. Dawid, A.F.M. Smith. Oxford University Press, 323–341.

Ishwaran, H., James, L.F., 2004. Computational methods for multiplicative intensity models using weighted gamma processes: Proportional hazards, marked point processes, and panel count data. J. Amer. Statist. Assoc. 99, 175–190.

Kingman, J.F.C., 1993. Poisson Processes. Clarendon Press, Oxford.

Kottas, A., 2005. Bayesian nonparametric mixture modeling for the intensity function of non-homogeneous Poisson processes. Technical Report AMS 2005-02, Department of Applied Mathematics and Statistics, University of California, Santa Cruz. (Available at: `http://www.ams.ucsc.edu/reports/trview.php`)

Kottas, A., 2006. Nonparametric Bayesian survival analysis using mixtures of Weibull distributions. Journal of Statistical Planning and Inference. 136, 578–596.

Kotz, S., Balakrishnan, N., Johnson, N.L., 2000. Continuous Multivariate Distributions, Volume 1: Models and Applications. (Second Edition). New York: Wiley.

Kuo, L., 1986. Computations of mixtures of Dirichlet processes. SIAM J. Scientific Statist. Comput. 7, 60–71.

Liu, J.S., 1996. Nonparametric hierarchical Bayes via sequential imputations. Ann. Statist. 24, 911–930.

Lo, A. Y., 1984. On a class of Bayesian nonparametric estimates: I. Density estimates. Ann. Statist. 12, 351–357.

MacEachern, S.N., 1998. Computational methods for mixture of Dirichlet process models. In Practical Nonparametric and Semiparametric Bayesian Statistics, eds. D. Dey, P. Müller and D. Sinha, New York: Springer, pp. 23–43.

MacEachern, S.N., Müller, P., 2000. Efficient MCMC Schemes for Robust Model Extensions Using Encompassing Dirichlet Process Mixture Models. In Robust Bayesian Analysis, eds. F. Ruggeri and D. Rios-Insua, New York: Springer, pp. 295–316.

Møller, J., 2003. Shot noise Cox processes. Adv. Appl. Prob. 35, 614–640.

Møller, J., Torrisi, G.L., 2005. Generalised shot noise Cox processes. Adv. Appl. Prob. 37, 48–74.

Møller, J., Waagepetersen, R.P., 2004. Statistical Inference and Simulation for Spatial Point Processes. Chapman & Hall, Boca Raton.

Møller, J., Syversveen, A.R., Waagepetersen, R.P., 1998. Log Gaussian Cox processes. Scand. J. Statist. 25, 451–482.

Müller, P., Quintana, F.A., 2004. Nonparametric Bayesian data analysis. Statist. Science 19, 95–110.

Neal, R.M., 2000. Markov chain sampling methods for Dirichlet process mixture models. J. Comput. Graph. Statist. 9, 249–265.

Pickands, J., 1971. The two-dimensional Poisson process and extremal processes. J. Appl. Prob. 8, 745–756.

Ripley, B.D., 1977. Modelling spatial patterns (with discussion). J. Roy. Statist. Soc. Ser. B 39, 172–212.

Sethuraman, J., 1994. A constructive definition of Dirichlet priors. Statist. Sinica 4, 639–650.

Sethuraman, J., Tiwari, R.C., 1982. Convergence of Dirichlet measures and the interpretation of their parameter. In Statistical Decision Theory and Related Topics III, eds. S. Gupta. and J.O. Berger, New York: Springer-Verlag, pp. 305–315.

Smith, R. L., 1989. Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone (with discussion). Statist. Science 4, 367–377.

Smith, R. L., 2003. Statistics of extremes, with applications in environment, insurance and finance. Chapter 1 of, Extreme Values in Finance, Telecommunications and the Environment, edited by B. Finkenstadt and H. Rootzen, Chapman and Hall/CRC Press, London, pp. 1–78.

Walshaw, D., 1999. Extremes of mixed environmental processes. In Bayesian Statistics 6, eds. J.M. Bernardo, J.O. Berger, P. Dawid, A.F.M. Smith. Oxford University Press, 849–858.

Wolpert, R.L., Ickstadt, K., 1998a. Poisson/Gamma random field models for spatial statistics. Biometrika 85, 251–267.

Wolpert, R.L., Ickstadt, K., 1998b. Simulation of Lévy random fields. In Practical Nonparametric and Semiparametric Bayesian Statistics, eds. D. Dey, P. Müller and D. Sinha, New York: Springer, pp. 227–242.