Bayesian Analysis (2017) **12**, Number 1, pp. 1–30

Bayesian Inference and Model Assessment for Spatial Point Patterns Using Posterior Predictive Samples

Thomas J. Leininger[*] and Alan E. Gelfand[†]

Abstract. Spatial point pattern data describes locations of events observed over a given domain, with the number of and locations of these events being random. Historically, data analysis for spatial point patterns has focused on rejecting complete spatial randomness and then on fitting a richer model specification. From a Bayesian standpoint, the literature is growing but primarily considers versions of Poisson processes, focusing on specifications for the intensity. However, the Bayesian literature on, e.g., clustering or inhibition processes is limited, primarily attending to model fitting. There is little attention given to full inference and scant with regard to model adequacy or model comparison.

The contribution here is full Bayesian analysis, implemented through generation of posterior point patterns using composition. Model features, hence broad inference, can be explored through functions of these samples. The approach is general, applicable to any generative model for spatial point patterns.

The approach is also useful in considering model criticism and model selection both in-sample and, when possible, out-of-sample. Here, we adapt or extend familiar tools. In particular, for model criticism, we consider Bayesian residuals, realized and predictive, along with empirical coverage and prior predictive checks through Monte Carlo tests. For model choice, we propose strategies using predictive mean square error, empirical coverage, and ranked probability scores. For simplicity, we illustrate these methods with standard models such as Poisson processes, log-Gaussian Cox processes, and Gibbs processes. The utility of our approach is demonstrated using a simulation study and two real datasets.

Keywords: Cox process, cross-validation, Gibbs process, Markov chain Monte Carlo, nonhomogeneous Poisson process, predictive residuals, ranked probability scores, realized residuals, Strauss process.

1 Introduction

Spatial point pattern data refers to the field of spatial analysis that examines spatial locations of events observed over a given domain, with the number and locations of these events being random. Analysis of such data involves understanding the underlying process generating these events. This includes learning whether locations of points can be explained say by associated covariate-driven intensity surfaces or perhaps through a clustering or inhibition mechanism. As examples, a point pattern may consist of the

[*] Duke University, Durham, NC, USA, tjl13@stat.duke.edu
[†] Duke University, Durham, NC, USA, alan@stat.duke.edu

DOI: 10.1214/15-BA985

locations of trees in a forest or the locations of crimes in a city. Potential covariate information might include climate variables or socio-economic variables, respectively. There may be extra information attached to the event (marks), such as the species of the tree or the type of crime.

Point pattern analysis often begins by exploring whether such a point pattern exhibits complete spatial randomness (CSR), i.e., whether locations occur independently and uniformly over the domain. Anticipating rejection of CSR, a more complex model will be specified for the data, often on mechanistic or behavioral grounds. Within the Bayesian framework, the literature is growing but primarily considers versions of Poisson processes, focusing on specifications for the intensity (see below). The Bayesian literature on other useful processes, such as processes where the intensity for the model need not be available as in clustering or inhibition models, is limited and primarily focuses on model fitting. There is little attention given to full inference with uncertainty and scant with regard to model criticism or model comparison.

Our contribution is to provide a fully model-based Bayesian approach to posterior inference, model validation, and model selection for spatial point process models. Implementation is through generation of posterior point patterns using composition. Model features, hence broad inference, can be explored through functions of these samples. The approach is general, applicable to any generative model for spatial point patterns.

More precisely, the approach we propose is that which has emerged as the dominant strategy for Bayesian data analysis these days, simulation from the posterior distribution for the model to provide full inference with uncertainty estimates. The novelty here is that we focus on simulating posterior predictive point patterns. In particular, using bracket notation for densities, we write our model in the general form $[\mathcal{S}|\boldsymbol{\theta}][\boldsymbol{\theta}]$, where $\mathcal{S}$ denotes a point pattern realization and $\boldsymbol{\theta}$ denotes model parameters. We observe $\mathcal{S}_{obs}$, and after we fit the model we obtain posterior samples $\boldsymbol{\theta}_l^*$ from $[\boldsymbol{\theta}|\mathcal{S}_{obs}]$. We then use composition to create posterior predictive samples $\mathcal{S}_l^*$ from $[\mathcal{S}|\mathcal{S}_{obs}]$ by drawing $\mathcal{S}_l^*$ from $[\mathcal{S}|\boldsymbol{\theta}_l^*]$. Inference follows by creating posterior samples of any function or feature say h of $\mathcal{S}$ as $\{h(\mathcal{S}_l^*), l = 1, 2, \dots, L\}$ from $[h(\mathcal{S})|\mathcal{S}_{obs}]$. Of course, the $\boldsymbol{\theta}_l$'s enable learning directly about the posterior distribution of the function $b(\boldsymbol{\theta})$ if b is available explicitly. If not, we show how to use Campbell's Theorem (Illian et al., 2008) with the $\mathcal{S}_l^*$'s to learn about $b(\boldsymbol{\theta})$.

The story from a Bayesian perspective is: if we can fit the model and if we can sample point patterns under the model, we can implement arbitrary inference. As an aside, if we can sample, we can also develop prior–posterior comparison to assess Bayesian learning.

Moreover, if posterior point patterns can be generated, they can help in assessing model adequacy and performing model selection. We adapt and extend familiar tools. In particular, for model adequacy, we propose examination of Bayesian residuals (drawing on the work of Baddeley et al., 2005), both realized and predictive, as well as empirical coverage and prior predictive checks through Monte Carlo tests. For model selection, we use predictive mean square error, empirical coverage and ranked probability scores. We consider both in-sample and, when possible, out-of-sample approaches. In this regard, the essence of our contribution is developed in Sections 4 and 5 below, after we review some basic theory for point patterns and several classes of point pattern models.

The point pattern literature is very large, covering theoretical development and computational tools for analyzing many types of point patterns, including simple point patterns, multivariate point patterns, marked point patterns, and spatiotemporal point patterns (see, e.g., Møller and Waagepetersen, 2003; Illian et al., 2008; Gelfand et al., 2010). Furthermore, useful software such as the R package `spatstat` (Baddeley and Turner, 2005) allows extensive point pattern analysis.

Important Bayesian contributions have been made by Møller and colleagues (Møller et al., 1998, 2006; Møller and Waagepetersen, 2007; Berthelsen and Møller, 2008). In addition, there has been a recent strand which considers Poisson process models, focusing on a rich range of specifications for the intensity. See, e.g., Kottas and Sansó (2007) and Taddy and Kottas (2012). Some recent Bayesian work (e.g., Illian et al., 2012; King et al., 2012) has employed integrated nested Laplace approximation (INLA) (Rue et al., 2009) for inference. Work that follows our inference paradigm in the context of sequential evolution of point patterns is considered in Møller and Rasmussen (2012). See also suggestions in a conference address by Møller (2012). With regard to model criticism, they suggest using *posterior* predictive model checking. We demonstrate that in many cases *prior* predictive model checking is needed for effective assessment of model adequacy. Finally, we find some quite recent Bayesian model checking work, primarily validating intensities, in Taddy (2010), Zhou et al. (2015), and Xiao et al. (2015).

For simplicity, in elaborating our inference and model assessment approach, we focus on a subset of common models: namely, the homogeneous Poisson process (HPP); the nonhomogeneous Poisson process (NHPP); the log-Gaussian Cox process (LGCP); and a Gibbs process for inhibition, specifically the Strauss process. We use a simulation study as well as two real data examples to illuminate our ideas. In the interest of space, we do not consider other attractive models such as the Neyman–Scott processes and shot noise processes (Banerjee et al., 2014) or the recent determinantal point processes (Lavancier et al., 2015).

The remainder of the paper begins in Section 2 by briefly reviewing basic point process models and some useful results for point patterns. Section 3 reviews methods for Bayesian fitting of these models and generating point patterns under these models. Section 4 elaborates our model-based approach to point pattern inference. Section 5 discusses model criticism and model comparison strategies. In Section 6, we present a brief simulation study and then, in Section 7, we offer two real data examples to further demonstrate our approach. Section 8 closes with a summary of our contribution and provides several ideas for future work.

2 Some basic modeling and theory

Again, we denote a point pattern realization over $\mathbb{R}^2$ by $\mathcal{S}$. The point pattern will be observed over a domain of interest, denoted by D, a bounded region in $\mathbb{R}^2$. As a result, the observed point pattern will be finite but with a random number of points. It will be denoted by $\mathcal{S} \cap D$ and the finite set of points comprising this pattern will be $\{\mathbf{s}_i\}_{i=1}^n$, where $n \equiv N(D)$ is the number of points observed in D. The number of points in any set $A \subseteq D$ will be denoted by $N(A)$. The distribution for the locations of the points must have a valid density $f_n(\cdot\,;\boldsymbol{\theta})$ for any n and parameters values $\boldsymbol{\theta}$. Since the points are unordered and labeled arbitrarily, this *location* density $f_n(\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n; \boldsymbol{\theta})$ must be symmetric in its arguments.

2.1 Moment measures and Campbell's Theorem

Moment measures are characteristics of a point process. The first-order moment measure, called the intensity, is denoted by $\lambda(\mathbf{s})$ and is used to define $\lambda(A) \equiv \mathbb{E}[N(A)] = \int_A \lambda(\mathbf{s})d\mathbf{s}$. The second-order moment measure, called the second-order intensity, is denoted by $\gamma(\mathbf{s}, \mathbf{s}')$, and addresses the covariance structure. If, for bounded sets A and B, $\gamma(A \times B) \equiv \mathbb{E}_{\mathcal{S}} \sum_{\mathbf{s},\mathbf{s}' \in \mathcal{S}} 1(\mathbf{s} \in A, \mathbf{s}' \in B)$, then $\gamma(\mathbf{s}, \mathbf{s}')$ is defined as the function satisfying $\gamma(A \times B) = \int 1((\mathbf{s}, \mathbf{s}') \in A \times B)\gamma(\mathbf{s}, \mathbf{s}')d\mathbf{s}'d\mathbf{s}$. If $A \cap B = \emptyset$, $\gamma(A \times B) = \mathbb{E}[N(A)N(B)]$. The pair correlation function (PCF), also called the reweighted second-order intensity, is defined as $\tilde{g}(\mathbf{s}, \mathbf{s}') = \gamma(\mathbf{s}, \mathbf{s}')/\lambda(\mathbf{s})\lambda(\mathbf{s}')$ and provides a standardized version of the second-order measure (see, e.g., Illian et al., 2008, p. 220). A process is said to be second-order reweighted stationary if $\tilde{g}(\mathbf{s}, \mathbf{s}')$ can be simplified to be a function of $d = ||\mathbf{s} - \mathbf{s}'||$, in which case the PCF is written as $\tilde{g}(d)$.

The Papangelou conditional intensity (Illian et al., 2008) inherits the parameters of the model and is defined as

$$\lambda(\mathbf{s}|\mathcal{S};\boldsymbol{\theta}) = \begin{cases} \dfrac{f(\mathcal{S} \cup \{s\};\boldsymbol{\theta})}{f(\mathcal{S};\boldsymbol{\theta})} & \text{if } \mathbf{s} \notin \mathcal{S}, \text{ and} \\[2ex] \dfrac{f(\mathcal{S};\boldsymbol{\theta})}{f(\mathcal{S} \setminus \{\mathbf{s}\};\boldsymbol{\theta})} & \text{if } \mathbf{s} \in \mathcal{S}, \end{cases} \tag{1}$$

where $f(\cdot\,;\boldsymbol{\theta})$ is the finite point process density (usually defined with respect to a homogeneous Poisson process with intensity 1; see, e.g., Møller and Waagepetersen, 2007) and $\mathcal{S} \setminus \{\mathbf{s}\}$ denotes $\mathcal{S}$ with $\{\mathbf{s}\}$ removed.

The main theoretical tool we employ here is Campbell's Theorem (see, e.g., Illian et al., 2008), which gives the expectation of the summation over $\mathcal{S} \cap D$ of a function $h(\mathbf{s})$ (restriction to D ensures that expectations exist). It states that

$$\mathbb{E}_{\mathcal{S}\cap D}\Big[\sum_{\mathbf{s}_i\, \in \mathcal{S}\cap D} h(\mathbf{s}_i)\Big] = \int_D h(\mathbf{s})\lambda(\mathbf{s})\, d\mathbf{s}. \tag{2}$$

For example, letting $g(\mathbf{s}) = \mathbf{1}(s \in A)$ for some set $A \subset D$, Campbell's Theorem says that $\sum_{\mathbf{s}_i \in \mathcal{S}} \mathbf{1}(\mathbf{s}_i \in A)$ is an unbiased estimator for $\int_D \mathbf{1}(s \in A)\lambda(\mathbf{s})\, d\mathbf{s} = \int_A \lambda(\mathbf{s})\, d\mathbf{s} = \lambda(A)$. Anticipating the discussion of inference in Section 4, (2) suggests how, for a given h, we can directly create a Monte Carlo integration for the left side with posterior samples, hence a Monte Carlo integration for the posterior mean of the integral on the right side.

Similarly, Campbell's Theorem has a bivariate form for h, a function of two points in $\mathcal{S}$:

$$\mathbb{E}_{\mathcal{S}\cap D}\Big[\sum_{\substack{\mathbf{s}_i,\, \mathbf{s}_j\, \in \mathcal{S}\cap D \\ i \neq j}} h(\mathbf{s}_i, \mathbf{s}_j)\Big] = \int_D \int_D h(\mathbf{s}, \mathbf{s}')\gamma(\mathbf{s}, \mathbf{s}')\, d\mathbf{s}\, d\mathbf{s}'. \tag{3}$$

(3) is useful for exploring second-order properties of a point process and enables similar Monte Carlo integration for the posterior mean of the right side.

A more general result is the Georgii–Nguyen–Zessin (GNZ) formula (Georgii, 1976; Nguyen and Zessin, 1979), which applies to h of the form $h(\mathbf{s};\mathcal{S}\backslash\{\mathbf{s}\})$ and gives the equality

$$\mathbb{E}_{\mathcal{S}\cap D}\Big[\sum_{\mathbf{s}_i\in\mathcal{S}} h(\mathbf{s}_i,\mathcal{S}\backslash\{\mathbf{s}_i\})\Big] = \mathbb{E}_{\mathcal{S}\cap D}\left[\int_D h(\mathbf{s},\mathcal{S})\lambda(\mathbf{s}|\mathcal{S})d\mathbf{s}\right], \tag{4}$$

where $\lambda(\mathbf{s}|\mathcal{S})$ is the Papangelou conditional intensity. Again, Monte Carlo integration enables the posterior mean for the right side.

2.2 Some standard models

A Poisson process (Illian et al., 2008) with a spatially varying intensity $\lambda(\mathbf{s})$ is referred to as a nonhomogeneous Poisson process (NHPP). Here, $N(A)$ is distributed as Poisson$(\lambda(A))$ and, if A and B are disjoint, then $N(A)$ and $N(B)$ are independent conditional on $\lambda(\mathbf{s})$. The spatially varying intensity may include a regression component and is often specified in the form $\lambda(\mathbf{s}) = \lambda_0 \exp\{x^T(\mathbf{s})\beta\}$ where λ_0 is baseline intensity and $x(\mathbf{s})$ is a vector of covariates at location $\mathbf{s}$. In general notation, we will write $\lambda(\mathbf{s};\boldsymbol{\theta})$. For a realization $\mathcal{S}$, the NHPP likelihood is

$$\begin{aligned} f_{\mathcal{S}}(\mathcal{S};\boldsymbol{\theta}) &= \frac{\exp\{-\lambda(D;\boldsymbol{\theta})\}\big(\lambda(D;\boldsymbol{\theta})\big)^n}{n!} \times n! \prod_{\mathbf{s}_i\in\mathcal{S}} \frac{\lambda(\mathbf{s}_i;\boldsymbol{\theta})}{\lambda(D;\boldsymbol{\theta})} \\ &= \exp\{-\lambda(D;\boldsymbol{\theta})\} \prod_{\mathbf{s}_i\in\mathcal{S}} \lambda(\mathbf{s}_i;\boldsymbol{\theta}). \end{aligned} \tag{5}$$

The case with λ a constant is referred to as a homogeneous Poisson process (HPP).

When $\lambda(\mathbf{s})$ is a realization of a non-negative stochastic process, then we have a Cox process. The log-Gaussian Cox process (LGCP) is characterized by the log of the intensity surface arising from a Gaussian process (GP) realization (Møller et al., 1998). If $Z(\mathbf{s})$ is a GP with mean $m(\mathbf{s})$ and covariance function $c(\mathbf{s},\mathbf{s}')$ and the intensity is written as $\lambda(\mathbf{s}) = \lambda_0 \exp\{x^T(\mathbf{s})\beta + Z(\mathbf{s})\}$, the LGCP likelihood takes the form

$$f_{\mathcal{S}}(\mathcal{S};\boldsymbol{\theta}) = \exp\left\{-\lambda_0 \textstyle\int_D \exp\{x^T(\mathbf{s})\beta+Z(\mathbf{s})\}d\mathbf{s}\right\}(\lambda_0)^n \exp\left\{\textstyle\sum_{\mathbf{s}_i\in\mathcal{S}}(x^T(\mathbf{s}_i)\beta+Z(\mathbf{s}_i))\right\}. \tag{6}$$

The integral in (6) is stochastic and is never available explicitly.

A point process is a Gibbs process with pairwise interactions if its finite point process density can be written as $f(\mathcal{S};\boldsymbol{\theta}) = \exp\{-Q(\mathcal{S};\boldsymbol{\theta})\}$ where

$$Q(\mathcal{S};\boldsymbol{\theta}) = c_0(\boldsymbol{\theta}) + \sum_{\mathbf{s}_i\in\mathcal{S}} h_1(\mathbf{s}_i;\boldsymbol{\theta}) + \sum_{\mathbf{s}_i,\mathbf{s}_j\in\mathcal{S},i\neq j} h_2(\mathbf{s}_i,\mathbf{s}_j;\boldsymbol{\theta}). \tag{7}$$

Here, $c_0(\boldsymbol{\theta})$ is an unknown (usually intractable) constant making the density integrate to 1 and h_k denotes a potential of order k, with h_2 usually being a function of interpoint distance, $d_{ij} = ||\mathbf{s}_i - \mathbf{s}_j||$.

The Strauss process (Strauss, 1975) is a Gibbs process which sets $h_2(d) = -\log\gamma$ if $d \leq R$ and 0 otherwise. $h_2 \geq 0$ is required for integrability which implies that $0 \leq \gamma \leq 1$. Specifying $h_1(\mathbf{s}) = \beta$ provides a constant first-order intensity, resulting in a homogeneous Strauss process. The finite point process density for the homogeneous Strauss process is then

$$f(\mathcal{S};\boldsymbol{\theta}) = e^{-c_0(\beta,\gamma)}\,\beta^{N(D)}\,\gamma^{\mathbf{s}_R(\mathcal{S})}, \tag{8}$$

where $\mathbf{s}_R(\mathcal{S})$ counts the number of pairs of points $(\mathbf{s}_i, \mathbf{s}_j) \subset \mathcal{S} \cap D$ with $||\mathbf{s}_i - \mathbf{s}_j|| \leq R$. We see that, given R, $\mathbf{s}_R(\mathcal{S})$ is a *sufficient* statistic. Viewed as a function of R, it will be useful below for model checking. Working with Gibbs processes in general, and the Strauss process in particular, is facilitated by the Papangelou conditional intensity (Illian et al., 2008) which takes the form

$$\lambda(\mathbf{s}|\mathcal{S};\boldsymbol{\theta}) = \beta\,\gamma^{s_R(\mathcal{S}\cup\{s\}) - s_R(\mathcal{S}\setminus\{s\})}. \tag{9}$$

Conveniently, the unknown normalizing constant cancels out.

3 Bayesian fitting and sampling

In Section 3.1, we provide a brief review of Bayesian model fitting for the models introduced in Section 2.2. Some require advanced Markov chain Monte Carlo (MCMC) algorithms to obtain posterior samples of model parameters. Section 3.2 discusses sampling point patterns under the various models.

3.1 Bayesian fitting for standard models

Nonhomogeneous Poisson processes

For the NHPP model, again we specify the intensity as $\lambda(\mathbf{s}) = \lambda_0 \exp\{x^T(\mathbf{s})\beta\}$ and plug into (5). A gamma prior distribution for λ_0 provides a conjugate prior distribution, but no conjugate prior specifications exist for the regression coefficients $\{\beta_j\}$ due to the integral in the likelihood; a normal distribution is usually employed. Fitting the model now requires MCMC with a Gibbs step for λ_0 and a Metropolis–Hastings step for the β_j. We find that a random walk Metropolis–Hastings step for each β_j is usually adequate. The integral in the exponent has no explicit form. Typically, numerical integration is used by discretizing the domain D and evaluating the function $\exp\{x^T(\mathbf{s})\beta\}$ at the centroids of the grid cells.

Log-Gaussian Cox processes

Prior specification for the LGCP requires the mean function $m(\mathbf{s})$ and the correlation function $c(\mathbf{s}, \mathbf{s}')$ for the Gaussian process. With regard to specifying $Z(\mathbf{s})$, we suggest using $m(\mathbf{s}) \equiv -c(\mathbf{s}, \mathbf{s})/2$ to provide $\mathbb{E}[e^{Z(\mathbf{s})}] = 1$, which, along with a zero-mean specification for any regression coefficients, roughly sets $\mathbb{E}[\lambda(\mathbf{s})] = \lambda_0$ *a priori*. Møller et al. (1998) provide some discussion about the choice of covariance function; some care is needed in specifying the priors for the hyperparameters. With the Matérn covariance function in the form $\sigma^2\rho(||\mathbf{s}-\mathbf{s}'||;\phi)$, σ^2 and ϕ are not identifiable (Zhang (2004)), suggesting that informative priors for one of these parameters will be needed for well-behaved model fitting.

In the absence of prior knowledge, we suggest estimating ϕ at its minimum contrast estimate using the pair correlation function (Møller et al., 1998), which we denote by $\tilde{\phi}$. In our experience (based on extensive simulation not presented here), this estimate seems to be more robust than the K-function minimum contrast estimate. With ϕ fixed, σ^2 will now be well identified. We use either a log-normal or gamma prior for σ^2, preferably centered around its minimum contrast estimate $\tilde{\sigma}^2$.

Sampling λ_0 and the β_j can be handled as discussed previously for the NHPP. Sampling the Z's cannot be done efficiently through Gibbs sampling as in the usual geostatistical setting. Simple Metropolis–Hastings samplers get stuck easily in local modes; more advanced MCMC methods are required. A common approach is to use a Metropolis-adjusted Langevin algorithm (MALA), as discussed in Møller et al. (1998) and Christensen et al. (2005). Girolami and Calderhead (2011) provide some extensions, including Hamiltonian Monte Carlo methods, which require less tuning. Murray et al. (2010) and Murray and Adams (2010) develop an elliptical slice sampling (ESS) algorithm for latent Gaussian fields and their hyperparameters.

We employ elliptical slice sampling here; it is easy to implement and requires no matrix inversions or estimation of the Fisher information matrix. We found Algorithm 2 in Murray and Adams (2010) to work well for updating the hyperparameters with elliptical slice sampling for updating Z. Each of the algorithms for fitting LGCPs requires discretizing Z to a finite-dimensional grid over the domain D in order to evaluate the integral in the exponent of the likelihood function (6). After discretizing, Monte Carlo integration is used, evaluating the function $\exp\{x^T(\mathbf{s})\beta + Z(\mathbf{s})\}$ at the centroids of the grid cells, similar to what was done for the NHPP model. Waagepetersen and Schweder (2006) show that the approximation converges to the exact value as the size of the grid cells goes to zero.

Gibbs processes

In the Gibbs process likelihood (7), the normalizing constant, being a function of the model parameters, complicates model fitting. Frequentist estimation generally proceeds by maximizing the pseudolikelihood, i.e., the product of the Papangelou conditional intensities, which removes the normalizing constant. Baddeley and Turner (2000) describe how to use the Berman–Turner device (Berman and Turner, 1992) to obtain maximum pseudolikelihood estimates. King et al. (2012) provide a Bayesian version in which the pseudolikelihood is again used. To avoid using the pseudolikelihood, Møller et al. (2006) discuss an auxiliary variable approach in which the auxiliary variable comes from the same state space as the point pattern. In their approach, the normalizing constant cancels in the Metropolis–Hastings ratio. Berthelsen and Møller (2006) further study this approach and demonstrate its use for Strauss processes and we employ it below.

3.2 Sampling methods for standard models

Our proposed approach relies on simulating posterior point patterns given the observed point pattern, i.e., generating $\mathcal{S}^*$ from $[\mathcal{S}|\mathcal{S}_{\text{obs}}]$. This will be done through composition using a posterior parameter draw. Hence, we need to be able to generate a point pattern under a specified model, given the values of the parameters for that model.

Generating an NHPP realization, given an intensity $\lambda(\mathbf{s})$, is done using the Lewis–Shedler thinning approach (Lewis and Shedler, 1979). We draw a point pattern from an HPP with intensity $\lambda_{max} \equiv \sup_{\mathbf{s}\in D} \lambda(\mathbf{s})$ and then thin the sampled points using rejection sampling. Generating an LGCP realization employs a similar approach for a given realization $\lambda(\mathbf{s})$. Since the Gaussian process involves an infinite number of random variables, a discretization is made and a Gaussian process realization is generated on the associated tiled surface. Then, λ_{max} is calculated and the Lewis–Shedler approach is applied to produce a sample.

Generating Gibbs process realizations can be done using an MCMC chain with a birth–death algorithms, as in Illian et al. (2008), Section 3.6.3. Summaries such as n or $\sum_i h_1(\mathbf{s}_i) + \sum_{i\neq j} h_2(\mathbf{s}_i, \mathbf{s}_j)$ are monitored until convergence seems to be achieved. Alternatively, Berthelsen and Møller (2002, 2003) develop a perfect simulation algorithm to simulate from spatial point processes such as Strauss processes. Their method, using dominated coupling from the past, provides a simulation from the exact desired distribution, whereas the birth–death algorithms only provide an approximation.

4 Inference

4.1 The general inference approach

As noted in the Introduction, we write our model in the general form $[\mathcal{S}|\boldsymbol{\theta}][\boldsymbol{\theta}]$. We observe $\mathcal{S}_{obs}$ and, after we fit the model, we obtain posterior samples $\boldsymbol{\theta}_l^*$ from $[\boldsymbol{\theta}|\mathcal{S}_{obs}]$. Then, using composition, i.e., by drawing $\mathcal{S}_l^*$ from $[\mathcal{S}|\boldsymbol{\theta}_l^*]$, we obtain posterior predictive samples $\{\mathcal{S}_l^*, l = 1, 2, \ldots, L\}$.

Returning to Campbell's theorem, it was noted that summing over the indicator function $\mathbf{1}(\mathbf{s}_i \in A)$ provides an unbiased estimator for $E[N(A)] = \lambda(A; \boldsymbol{\theta})$ whose usual Bayes estimate is $\mathbb{E}[\lambda(A; \boldsymbol{\theta})|\mathcal{S}_{obs}]$. If $\lambda(A; \boldsymbol{\theta})$ is available explicitly, a Monte Carlo integration for $\mathbb{E}[\lambda(A; \boldsymbol{\theta})|\mathcal{S}_{obs}]$ is $\frac{1}{L}\sum_l \lambda(A; \boldsymbol{\theta}_l^*)$. When we cannot calculate $\lambda(A; \boldsymbol{\theta})$, we note that $\mathbb{E}[\lambda(A; \boldsymbol{\theta})|\mathcal{S}_{obs}] = \mathbb{E}[N(A)|\mathcal{S}_{obs}] \approx \frac{1}{L}\sum_{l=1}^L \sum_{\mathbf{s}_{li}^* \in \mathcal{S}_l^*} \mathbf{1}(\mathbf{s}_{li}^* \in A)$, providing the desired Monte Carlo integration. Of course, the members of the set $\{\sum_{\mathbf{s}_{li}^* \in \mathcal{S}_l^*} \mathbf{1}(\mathbf{s}_{li}^* \in A), l = 1, 2, \ldots, L\}$ provide posterior predictive samples of $N(A)$.

More generally, we may be interested in inference on $b(\boldsymbol{\theta})$, some characteristic of the point process (examples below), based upon the posterior $[b(\boldsymbol{\theta})|\mathcal{S}_{obs}]$. With posterior samples, $\{\boldsymbol{\theta}_l^*\}$ and an explicit $b(\cdot)$, we obtain $\{b(\boldsymbol{\theta}_l^*)\}$ for such inference, as usual. If interest is in the predictive distribution$[h(\mathcal{S})|\mathcal{S}_{obs}]$ where $h(\mathcal{S})$ is a feature of the point pattern (examples below), then the set $\{\mathcal{S}_l^*\}$ provides the set $\{h(\mathcal{S}_l^*)\}$ for inference. For a function $v(\mathcal{S}, \boldsymbol{\theta})$ of both the point pattern and the parameters, if v is available explicitly we can use $\{\boldsymbol{\theta}_l^*, \mathcal{S}_l^*\}$ to generate samples from $[v(\mathcal{S}, \boldsymbol{\theta})|\mathcal{S}_{obs}]$.

A challenge is that often, $b(\boldsymbol{\theta})$ is not available explicitly. Then, the strategy is to find $h(\mathcal{S})$ such that $E(h(\mathcal{S})|\boldsymbol{\theta}) = b(\boldsymbol{\theta})$. Now, to obtain $b(\boldsymbol{\theta}_l^*)$, for each $\boldsymbol{\theta}_l^*$, we need to generate samples $\mathcal{S}_{lb}^*$, yielding a Monte Carlo integration for $b(\boldsymbol{\theta}_l^*)$, that is, $\frac{1}{B}\sum_b h(\mathcal{S}_{lb}^*)$. A rich class of such $b(\boldsymbol{\theta})$'s arises through Campbell's Theorem. From (2), the right side provides $b_g(\boldsymbol{\theta}) = \int_D g(\mathbf{s})\lambda(\mathbf{s};\boldsymbol{\theta})ds$. For a given $\boldsymbol{\theta}_l^*$, the foregoing posterior samples provide a Monte Carlo integration for the left side. Similar opportunities are available for the bivariate version of Campbell's Theorem in (3).

Apart from $\lambda(A;\boldsymbol{\theta})$, examples of $b(\boldsymbol{\theta})$'s from Section 2 include $\lambda(\mathbf{s};\boldsymbol{\theta}), \gamma(d;\boldsymbol{\theta})$, and $\tilde{g}(d;\boldsymbol{\theta})$. The usual distance-based measures such as the G- and K-functions (Illian et al., 2008) are defined through functions $v(\mathcal{S},\boldsymbol{\theta})$ such that $\mathbb{E}[v(\mathcal{S},\boldsymbol{\theta})|\boldsymbol{\theta}] = G(d;\boldsymbol{\theta})$ or $\mathbb{E}[v(\mathcal{S},\boldsymbol{\theta})|\boldsymbol{\theta}] = K(d;\boldsymbol{\theta})$ (Banerjee et al., 2014). The inhomogeneous K-function (Baddeley et al., 2000) is another example. In this way, we obtain *model-based* estimates of these quantities rather than the customary empirical estimates. The former provide inference under a model; the latter may be viewed as more exploratory. Another example of $v(\mathcal{S},\boldsymbol{\theta})$ arises through the *realized* residuals, motivated by frequentist residual analysis as discussed in Baddeley et al. (2005, 2008) and below. A simple version would consider the posterior distribution, $[N(A) - \lambda(A;\boldsymbol{\theta})|\mathcal{S}_{obs}]$.

A further example is the Papangelou conditional intensity in (1) where $v(\mathcal{S},\boldsymbol{\theta})$ takes the form $\lambda(\mathbf{s}|\mathcal{S};\boldsymbol{\theta})$. Then, the GNZ result (4) provides further $b(\boldsymbol{\theta})$'s of interest, i.e., here, the right side is $b_g(\boldsymbol{\theta}) = \mathbb{E}_{\mathcal{S}\cap D}[\int_D g(\mathbf{s},\mathcal{S})\lambda(\mathbf{s}|\mathcal{S};\boldsymbol{\theta})d\mathbf{s}]$ with Monte Carlo integration for the left side.

Examples of $h(\mathcal{S})$ of interest include $N(A)$, $[N(A), N(B)]$, $N(A)/N(D)$, along with the posterior distribution of conditional events, e.g., $[N(A)|N(B) = m; \mathcal{S}_{obs}]$. A further example is the *predictive* residual with posterior distribution, $[N_{obs}(A) - N(A)|\mathcal{S}_{obs}]$. Altogether, we see a strategy for implementation of rich posterior inference for general spatial point pattern models.

Bayesian residual analysis

Residuals are a common tool for model assessment. In particular, Baddeley et al. (2005, 2008) develop various notions of residuals for point patterns. For example, they define a *raw* residual, analogous to the standard residual from a regression model, as

$$R_{\hat{\boldsymbol{\theta}}}(B) \equiv N(B) - \int_B \hat{\lambda}(\mathbf{s}|\mathcal{S})d\mathbf{s}, \tag{10}$$

for $B \subseteq D$ where $\hat{\lambda}(\mathbf{s}|\mathcal{S}) \equiv \lambda(\mathbf{s}|\mathcal{S};\hat{\boldsymbol{\theta}})$ is the estimated Papangelou conditional intensity function. In the Bayesian setting, we would work with the *realized* residual, which removes the hat in (10), and consider its posterior.

More generally, Baddeley et al. (2005) define the h-weighted innovation measure as

$$I(B,h,\lambda) \equiv \sum_{\mathbf{s}_i \,\in\, \mathcal{S}\cap B} h(\mathbf{s}_i, \mathcal{S}\backslash\{\mathbf{s}_i\}) - \int_B h(\mathbf{s},\mathcal{S})\lambda(\mathbf{s}|\mathcal{S})d\mathbf{s}. \tag{11}$$

The innovations have mean 0 under the true model, as can be seen using (4). Choices of h include $h(\mathbf{s}, \mathcal{S}) = 1/\lambda(\mathbf{s}|\mathcal{S})$ which defines the inverse λ residuals, in the spirit of Stoyan and Grabarnik (1991). With $h(\mathbf{s}, \mathcal{S}) = 1/\sqrt{\lambda(\mathbf{s}|\mathcal{S})}$, an analogue of the Pearson residual from Poisson regression arises. Estimators are obtained by inserting an estimator of $\lambda(\mathbf{s}|\mathcal{S})$. Again, in the Bayesian setting, we work with realized residuals.

From a Bayesian perspective, the posterior distribution of $\int_B h(\mathbf{s}, \mathcal{S})\lambda(\mathbf{s}|\mathcal{S})d\mathbf{s}$ and, in fact, $I(B, h, \lambda)$ would be studied. In particular, these innovations are of the form $v(\mathcal{S}, \boldsymbol{\theta})$ and so their posteriors can be obtained as described in Section 4.1. We can use the posterior mean, $\mathbb{E}[\int_B h(\mathbf{s}, \mathcal{S})\lambda(\mathbf{s}|\mathcal{S})d\mathbf{s} \,|\, \mathcal{S}\,]$, to obtain a point estimate and can also examine whether 0 falls in a given credible interval.

With regard to validation, under a given model, should credible intervals created from these innovation distributions over many sets be expected to achieve empirical coverage of 0 at roughly the nominal level? For the raw/realized innovations, the answer is no. The raw innovations compare an observed count with the posterior distribution for the *expectation* of that count. Though we hope the expectations are close to the raw innovations, the credible intervals provide coverage for the expected counts rather than for the counts themselves. Thinking of the regression analogue, the raw innovations are akin to employing the distribution $[y - \mu_y|\text{Data}]$ when we should employ the distribution for the *predictive* innovations, $[y - y_{\text{pred}}|\text{Data}]$.

Instead, we adopt *predictive residuals*,

$$R_{\text{pred}}(B) = N_{\text{obs}}(B) - N_{\text{pred}}(B), \tag{12}$$

where, as above, posterior samples $\mathcal{S}_l^*$ supply the draws $N_{(l)}^*(B)$, hence the posterior predictive distribution of N_{pred} and, thus, of $R_{\text{pred}}(B)$.

Finally, for an h-scaled innovation as in (11), Baddeley et al. (2005) define the smoothed innovation field $r(\mathbf{u}; \boldsymbol{\theta})$ at location $\mathbf{u} \in D$ as

$$\begin{aligned} r(\mathbf{u}; \boldsymbol{\theta}) &= e(\mathbf{u}) \int_D k(\mathbf{u} - \mathbf{v}) dI(\mathbf{v}, \mathbf{h}, \boldsymbol{\theta}) \\ &= e(\mathbf{u}) \left[\sum_{\mathbf{s}_i \in \mathcal{S}} k(\mathbf{u} - \mathbf{s}_i)\mathbf{h}(\mathbf{s}_i, \mathcal{S}\backslash\{\mathbf{s}_i\}) - \int_D k(\mathbf{u} - \mathbf{v})\mathbf{h}(\mathbf{v}, \mathcal{S})\lambda(\mathbf{v}|\mathcal{S}; \boldsymbol{\theta}) d\mathbf{v} \right], \end{aligned} \tag{13}$$

where $k(\mathbf{s})$ is a probability density on $\mathbb{R}^2$ used as a smoothing kernel and $e(\mathbf{u}) \equiv 1/\int_D k(\mathbf{u} - \mathbf{v})d\mathbf{v}$ is an edge correction. This field puts positive atoms at each $\mathbf{s}_i \in \mathcal{S}$ and a negative value elsewhere and then smoothes using the kernel. So, a comparison is made between the intensity estimate under a model and an empirical estimate of the intensity. Positive values indicate locations where the empirical intensity was higher than the model intensity, conversely for negative values.

Baddeley et al. (2005) estimate $\boldsymbol{\theta}$ to obtain a residual field, $r(\mathbf{u}; \hat{\boldsymbol{\theta}})$. With a posterior distribution for $\lambda(\mathbf{s}; \boldsymbol{\theta})$, illustratively, we can obtain a posterior distribution for $r(\mathbf{u}; \boldsymbol{\theta})$ for the NHPP and LGCP models. Additionally, one can create a plot showing those regions that have a credible interval (for the smoothed innovation) which contains 0, those regions that have a credible interval above 0, and those below 0. Such plots are demonstrated in the data examples of Section 7.

	Model Criticism	**Model Comparison**
Out-of-sample	Predictive residual analysis, Residual field analysis (informal), Empirical coverage	Posterior vs. empirical (informal), Predictive mean square error, Ranked probability score
In-sample	Prior predictive MC tests, Discrepancy measures (DGSV)	Posterior vs. empirical (informal), Predictive mean square error, Ranked probability score

Table 1: Proposed techniques for model criticism and model comparison.

5 Model criticism and model comparison

Model assessment using a fitting/training sample and an independent validation/test sample is now standard practice. With point pattern data, such an approach may not be available. Under a conditionally independent location distribution, as with NHPP's and LGCP's, the answer is yes. However, with an inhibition model, holding out points will alter the nature of the interpoint distances, hence the interaction structure. This will be true in general for a point pattern model, such as a Gibbs process, where there is dependence between the locations of the points.

When cross-validation is permissible, to date there is limited discussion for point processes. Diggle and Marron (1988) adapted leave-one-out cross-validation from Bowman (1984) for bandwidth selection for kernel intensity estimates. For a Bayesian approach where MCMC model fitting is needed, the computational burden required for leave-one-out cross-validation is impractical. However, we can employ holdout, developing training and test datasets. Suppose we decide to administer 20% holdout. We cannot simply remove 20% of the data at random. This will *fix* the size of the point pattern rather than allowing it to be random. Rather, the p-thinning approach, as in Illian et al. (2008), can be applied to create appropriate training and test data. The p-thinning proceeds point-by-point, independently deleting $\mathbf{s}_i \in \mathcal{S}$ with probability $1-p$. This produces a training point pattern $\mathcal{S}^{\text{train}}$ and test point pattern $\mathcal{S}^{\text{test}}$, which are independent, conditional on $\lambda(\mathbf{s})$. In fact, $\mathcal{S}^{\text{train}}$ has intensity $p\lambda(\mathbf{s})$, $\mathcal{S}^{\text{test}}$ has intensity $(1-p)\lambda(\mathbf{s})$, and the revised validation intensity compared with the fitting intensity is $\lambda^{\text{test}}(\mathbf{s}) = (\frac{1-p}{p})\lambda^{\text{train}}(\mathbf{s})$.

As a high level summary of the various criteria which we detail in the remainder of this section, we offer Table 1. We emphasize that all of the proposed techniques are implemented as a post model fitting exercise.

5.1 Model adequacy through empirical coverage

When cross-validation is possible, using a validation sample $\mathcal{S}^{\text{test}}$, posterior predictive point patterns will supply the posterior predictive distribution of, say $N(B)$. The predictive residuals should be centered around zero for an adequate model. If we look at a set of subregions $\{B_k\}$, we expect the empirical coverage to be roughly the nominal level of coverage if the model is adequate. How shall we create a set $\{B_k\}$? Baddeley et al. (2005), Section 11.1 propose to analyze a set of residuals over disjoint partitions B_k of the domain, similar to quadrat counting (see, e.g., Diggle, 2003). With an irregular domain D, division into disjoint subregions of similar size can be time-consuming and is, in fact, unnecessary. We prefer to draw random subregions uniformly over D and then evaluate the residuals or innovations in each subregion. Moreover, there is no reason to require the B_k be disjoint in which case we can draw as many B_k as desired, subject to the requirement that each B_k has the same area. Denote the area of each B_k by $q|D|$ so q represents the size of each B_k relative to D. For various q's we can evaluate the innovation or residual measures on each of the B_k's and obtain the observed empirical coverage of 0.

In the sequel, we take the shape of each B_k to be a square but, depending upon D, there may be some reason to choose the shape more carefully. The use of squares sometimes limits the placement of the B_k when q is large and also access to the edges of D. Work by Sherman and Carlstein (1994), Lahiri (1999), and Lahiri (2003) suggests letting the shape of B_k mimic the shape of D. Furthermore, with randomly placed, overlapping B_k, it can be hard to identify regions where the model fits poorly. Disjoint B_k, as is demonstrated in Illian et al. (2009), alleviate this problem but, with regard to empirical coverage, Bernoulli trials based upon random B_k's will suffice.

In-sample model criticism

When we can not develop a test sample we resort to in-sample model criticism. This leads to familiar work on posterior model checks by Gelman et al. (1996) (henceforth GMS) and work on prior model checks by Dey et al. (1998) (henceforth DGSV). GMS is more common and easier to do. However, it doesn't criticize the model well enough and uses the data twice (once to fit, once to check). DGSV is more computationally demanding but is formally coherent and uses the data only once. Both GMS and DGSV employ Monte Carlo tests in looking at discrepancy measures, $D(\mathcal{S};\boldsymbol{\theta})$ which, for instance, might be $N(A)-\lambda(A;\boldsymbol{\theta})$.

GMS looks at $[D(\mathcal{S};\boldsymbol{\theta})|\mathcal{S}_{obs}]$ and compares it with $[D(\mathcal{S}_{obs};\boldsymbol{\theta})|\mathcal{S}_{obs}]$. The problem is evident. Draws of $\mathcal{S}$ from $[\mathcal{S};\boldsymbol{\theta}|\mathcal{S}_{obs}]$ will look too much like $\mathcal{S}_{obs}$ and discrepancies will look too much like $D(\mathcal{S}_{obs};\boldsymbol{\theta})$; the model checking will not be critical enough. Given that assessing adequacy for point pattern models is difficult, GMS will not be good enough.

DGSV create $\mathcal{S}_l^*$'s from the marginal distribution of $\mathcal{S}$ by drawing $\boldsymbol{\theta}_l^*$ from the prior distribution $[\boldsymbol{\theta}]$ and then $\mathcal{S}_l^*$ from $[\mathcal{S}|\boldsymbol{\theta}_l^*]$. Then, they obtain $[\mathcal{S},\boldsymbol{\theta}|\mathcal{S}_l^*]$ and compare $[D(\mathcal{S}_l^*;\boldsymbol{\theta})|\mathcal{S}_l^*]$ with $[D(\mathcal{S}_{obs};\boldsymbol{\theta})|\mathcal{S}_{obs}]$. DGSV compare the observed discrepancy with the discrepancies you expect under the model; GMS compare the observed discrepancies with what you expect under the model **and** the observed data. The computational demand required for DGSV is evident; one must fit and sample for every $\mathcal{S}_l^*$.

With regard to model checking, Møller and Rasmussen (2012) and Møller (2012) seem to embrace simulation akin to the GMS approach. Monte Carlo tests are proposed to examine discrepancies of the form $D(\mathcal{S}_{obs},\mathcal{S}_l^*,\boldsymbol{\theta}_l^*)=v(\mathcal{S}_{obs},\boldsymbol{\theta}_l^*)-v(\mathcal{S}_l^*,\boldsymbol{\theta}_l^*)$. In-sample, our empirical coverage model criticism check will also suffer the GMS problem; it will not be critical enough. For a collection of B_k's, we look at the set $\{[N_{obs}(B_k)-N(B_k)|\mathcal{S}_{obs}]\}$ and check empirical coverage relative to nominal coverage. We see that the $\mathcal{S}_l^*$'s will be too similar to $\mathcal{S}_{obs}$ so the $N(B_k)$ that we generate given $\mathcal{S}_{obs}$ will tend to look too much like $N_{obs}(B_k)$, since the latter is a function of $\mathcal{S}_{obs}$.

Consider a simpler checking function approach which can be expected to supply model criticism through the prior predictive framework. Suppose $h(\mathcal{S})$ is a function only of the point pattern. For instance, in assessing the adequacy of an HPP or Strauss process model, given a radius $R = r$, suppose we consider the statistic, $\mathbf{s}_r(\mathcal{S})$ discussed at the end of Section 2.2. We can implement a Monte Carlo test for $\mathbf{s}_r(\mathcal{S}_{obs})$ and the set $\{\mathbf{s}_r(\mathcal{S}_b^*), b = 1, 2, \ldots, B\}$ where the $\mathcal{S}_b^*$'s are generated under the model. If there is interaction between the points in $\mathcal{S}$, then as we run through a set of r's (motivated by the size of the region), these Monte Carlo tests should criticize the HPP model but potentially support Strauss process models in the vicinity of a suitable r.

An alternative $h(\mathcal{S})$, working with the $\{B_k\}$ above, is the sample variance across the $\{N(B_k)\}$. This variance would be expected to be smaller under a stationary Strauss process than under an HPP. So, most directly, we could calculate $h(\mathcal{S}_{obs})$ and compare with the collection of $h(\mathcal{S}_b^*)$'s, again using a Monte Carlo test. To enrich the assessment, we could consider varying cell sizes and varying numbers of cells, each providing a Monte Carlo test. Yet another choice might adopt a checking function in the form of a χ^2 statistic, i.e., $v(\mathcal{S};\boldsymbol{\theta}) = \sum_{\{B_k\}}(N(B_k) - \lambda(B_k;\boldsymbol{\theta}))^2/\lambda(B_k;\boldsymbol{\theta})$ to employ as a discrepancy measure above.

5.2 Model comparison

A typical attempt at model selection uses ad-hoc tests of the homogeneity and independence assumptions of CSR but, having decided which assumption to relax, there is no clear procedure for comparing models. Often model comparison is not even considered; a model is adopted on mechanistic or behavioral grounds. Lack of fit using the methods described above can eliminate some models but will not help when choosing among adequately fitting models. Also, informal model comparison is frequently employed. For instance, when appropriate, we might develop posterior intensities to compare with the observed point pattern (or a kernel intensity estimate).

The first discussions of formal Bayesian model selection for point processes appear in Akman and Raftery (1986) and Raftery and Akman (1986), who discuss computing Bayes factors for NHPPs and change point Poisson processes, respectively. Guttorp and Thorarinsdottir (2012) perform model choice via a reversible jump algorithm that allows movement between two nested models. They can then use the work of Akman and Raftery (1986) to compute a Bayes factor.

Model comparison should be done in predictive space since parameters have no meaning across models, raising the question, "What would we be predicting?" Since counts for sets are often of interest, a natural choice would focus on $[N(A)|\mathcal{S}_{obs}]$ for $A \subset D$. In particular, we would compare $N_{obs}(A)$ with $[N(A)|\mathcal{S}_{obs}; M_j]$ for each model, $j = 1, 2, \ldots, J$. Here, for model j with parameters $\boldsymbol{\theta}_j$, we obtain posterior samples, $\boldsymbol{\theta}_{j,l}^*$ and then $\mathcal{S}_{j,l}^*$. Again, we would want to do this out-of-sample through p-thinning, as with NHPP's, LGCP's, and for cluster processes which are superpositions of NHPP's. As for criteria, we can look at predictive mean square error (PMSE), perhaps standardized by the expected number (the usual loss function for Poisson counts) and ranked probability scores (RPS) (Gneiting and Raftery, 2007).

We remind the reader that the RPS arises from a proper scoring rule and offers an informative metric for assessing the performance of a predictive distribution. For count data, the ranked probability score (RPS) is appropriate (Epstein, 1969). For us, the RPS compares the posterior predictive distribution for a cell count with the degenerate distribution associated with the observed cell count using a sum of squares over the set of support values $\{0, 1, 2, \dots\}$. RPS prefers models yielding predictive distributions that are concentrated around the observed value.

If cross-validation is available, we would employ the RPS with our hold-out data, comparing observed counts in subsets to posterior predictive distributions for these counts. Specifically, returning to $\{B_k\}$, for a given model M_j, we can compute an out-of-sample RPS for each B_k, say $RPS_j(B_k)$. Averaging these over k yields a performance measure for M_j. Model selection would choose the model with the smallest average RPS. If holding out data is not possible, we would examine these metrics in-sample.

6 Simulation study

The Duke Forest data example in the next section provides an effective criticism of a NHPP model in favor of a LGCP model. In the online supplement (Leininger and Gelfand, 2015) we offer a simulation investigation comparing a HPP, a NHPP, and a LGCP. Here, we offer a simulation study focused on criticizing the HPP in favor of a Strauss process when the latter is the true process. More precisely, we examine whether a Strauss model is criticized when fitted to data from an HPP, and whether a Strauss model is preferred over an HPP model when fitted to data from a Strauss process.

The two data-generating processes used in the simulation study are an HPP with $\lambda = 100$ and a Strauss process with $(\beta = 250, \gamma = 0.05, R = 0.05)$. These latter choices were made to both generate roughly 100 points on the unit square. The Strauss process was also chosen to imply a strong amount of inhibition, so the Strauss process is similar to the HPP in its first-order intensity but differs strongly in its second-order characteristics.

Two domains were used in order to provide a comparison between the learning available on a small domain versus on a larger domain. By keeping the parameter values the same for the two domains, we achieve low and high intensity settings. The small domain D_1 is the unit square $[0, 1] \times [0, 1]$ and the larger domain D_2 is the square $[0, \sqrt{10}] \times [0, \sqrt{10}]$, such that the larger domain is ten times larger than the smaller domain (which should facilitate seeing inhibition, hence distinguishing the two process models).

The models fitted to each simulated dataset are an HPP model with λ unknown and a Strauss model with (β, γ) unknown and $R = 0.05$ fixed. The HPP model uses a Gamma prior with $E[\lambda] = 100$ and $Var[\lambda] = 0.1$. The Strauss model uses the priors $\beta \sim \text{Uniform}(75, 400)$ and $\gamma \sim \text{Beta}(1, 6)$. The prior for γ implies moderate to strong inhibition since it has a mode around 0.05 and most of its mass is below 0.4. Ten

r	HPP, D_1	Strauss, D_1	HPP, D_2	Strauss, D_2
0.01	0.579	0.915	0.670	0.999
0.02	0.554	0.972	0.650	1.000
0.03	0.544	0.991	0.698	1.000
0.04	0.443	0.991	0.611	1.000
0.05	0.496	0.994	0.583	1.000

Table 2: For each model, the average quantile of $s_r(\mathcal{S}_{obs})$ using Monte Carlo tests, averaged over ten simulations of an HPP(100) on the domains $D_1 = [0,1] \times [0,1]$ and $D_2 = [0, \sqrt{10}] \times [0, \sqrt{10}]$.

r	HPP, D_1	Strauss, D_1	HPP, D_2	Strauss, D_2
0.01	0.289	0.799	0.001	0.583
0.02	0.012	0.585	0.001	0.530
0.03	0.002	0.494	0.001	0.568
0.04	0.001	0.469	0.001	0.534
0.05	0.001	0.454	0.001	0.537

Table 3: For each model, the average quantile of $s_r(\mathcal{S}_{obs})$ using Monte Carlo tests, averaged over ten simulations of a Strauss($\beta = 250, \gamma = 0.05, R = 0.05$) process on D_1 and D_2.

replications of point patterns from each data-generating process were simulated over each domain and both models were fit to each simulated dataset.

First, we compare the coverage and ranked probability scores for the predictive residuals over random subsets of the domain. Again, with a Gibbs process, hold out is not available. Using RPS for model choice, the correct model was chosen only slightly more than 50% of the time, even in the high intensity setting. The coverages of the predictive residuals in all cases were at or above the nominal 90% level. Altogether, first order diagnostics do not distinguish the models.

For a better assessment of model fit we turn to second-order diagnostics. Prior predictive checks were run using the discrepancy function $s_r(\mathcal{S})$. As noted previously, this is a sufficient statistic for the Strauss process and should be able to separate the two models since it focuses on pairwise characteristics of the point pattern.

For each domain, Table 2 obtains the average quantile of $s_r(\mathcal{S}_{obs})$ across the replicates, using Monte Carlo tests, as described in Section 5.1.1, for $r = 0.01$, 0.02, 0.03, 0.04, and 0.05. We see that even on the smaller domain, the Strauss model produces values much smaller than that observed, while the HPP model performs adequately. The Strauss model with associated prior specification is not well-suited for the HPP data, though there can be sensitivity to the prior on γ. We note that there is substantial variation in the quantiles across the replications, arguing for the usefulness of the replications. This variation is mitigated as the domain grows, for example, for D_2 compared with D_1.

Table 3 performs the same comparisons for replications from the Strauss($\beta = 250$, $\gamma = 0.05, R = 0.05$) process. This simulation shows that we can clearly separate the

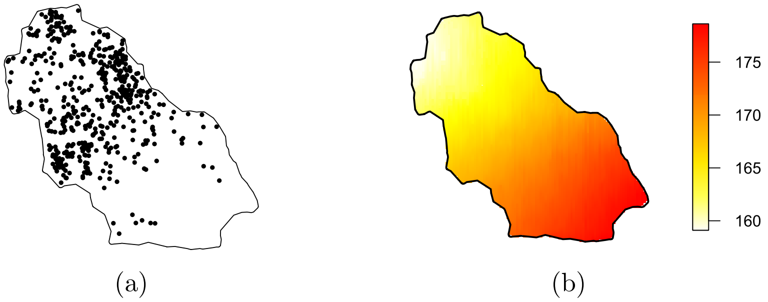


Figure 1: (a) The locations of 530 American sweetgum trees in a tract of Duke Forest and (b) the elevation in meters over the same region.

Strauss processes from the HPP, rejecting the HPP model when a Strauss process is the underlying data-generating process. The HPP model always generates $s_r(\mathcal{S}_b^*)$ values that are too large because it does not include any inhibition. Here, there is less variability in the quantiles across replications than for those in the previous table. This simulation study confirms that $s_r(\mathcal{S})$ is an effective discrepancy function for distinguishing between HPPs and stationary processes with inhibition, such as the Strauss process.

7 Real data examples

We present two real data examples to illustrate the methods proposed in the previous sections. In Section 7.1, we consider an analysis of tree data from Duke Forest, in which we compare a NHPP model and a LGCP model. In Section 7.2, we look at the classic Swedish pines dataset, which exhibits some regularity, and compare an HPP model with several Strauss process models.

7.1 Duke Forest example

We first consider a point pattern consisting of the locations of American sweetgum trees (*Liquidambar styraciflua*) in a subplot of Duke Forest in Durham, North Carolina, USA. Figure 1(a) shows the locations of these trees within the tract of forest. Elevation is also available on a fine grid over the region, as shown in Figure 1(b).

NHPP model

For this data, elevation is expected to be significant in explaining the intensity. Trees may be more likely to grow at certain elevations or elevation may act as a surrogate for other unobserved covariates. In fact, here, elevation may serve as a proxy for soil moisture. A spatial trend surface might also be included. Moreover, since other species are on this tract, some sort of competition covariate could be constructed. However, for now we include only a linear and quadratic trend in elevation, so the regression model

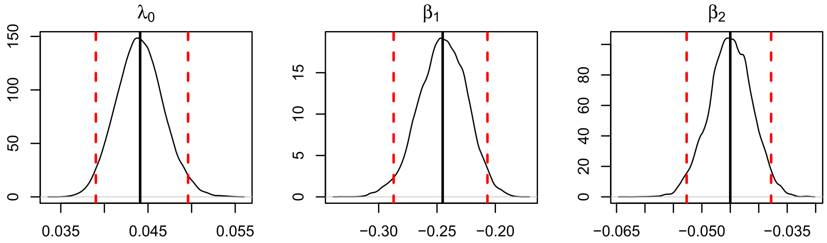


Figure 2: Posterior distributions for the parameters of the NHPP model. The posterior mean is marked by the solid vertical line and the 95% credible intervals are marked by the dashed lines.

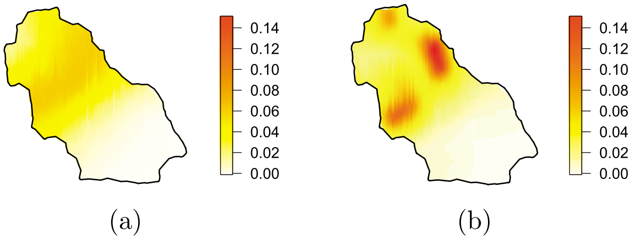


Figure 3: (a) The posterior mean of the intensity surface for the NHPP model and (b) the posterior mean of the intensity surface for the LGCP model.

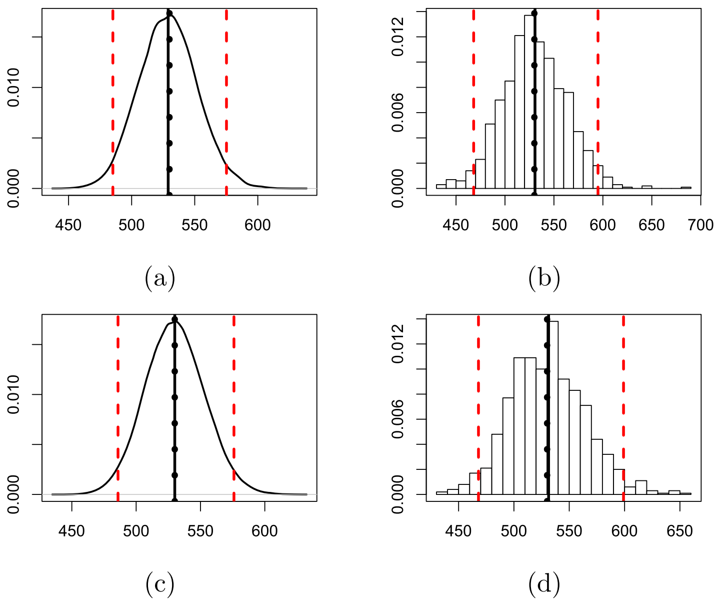
for the intensity of a NHPP is

$$\log\lambda(\mathbf{s}) = \log\lambda_0 + \beta_1\,\texttt{elev}(\mathbf{s}) + \beta_2\,\texttt{elev}^2(\mathbf{s}). \tag{14}$$

We use the prior $\lambda_0 \sim \text{Gamma}(a_\lambda = 1.3, b_\lambda = 50)$, which gives $\mathbb{E}[\lambda_0] = 0.026$. It may be simplest to expect, *a priori*, each $\mathbb{E}[\beta_j] = 0$ and then specify the prior for λ_0 induced by first specifying the prior for $\mathbb{E}[N(D)] = \lambda(D) = \lambda_0|D|$. Our prior for λ_0 implies *a priori* $\mathbb{E}[N(D)] \approx 500$ with a wide variance. For the regression coefficients, we use $\beta_j \overset{\text{ind}}{\sim} \text{Normal}(0, \omega^2)$ for $j = 1, 2$ and a large value ω^2 (e.g., $\omega^2 = 1000$). We ran our MCMC scheme for 10,000 iterations of burn-in and then collected 20,000 posterior samples of the model parameters.

Figure 2 shows the posterior distributions of λ_0, β_1, and β_2; all are significantly different from 0. The X matrix was centered prior to fitting the model so that λ_0 is roughly interpretable as the average intensity across D. At a point $\mathbf{s}^*$ with average elevation, the intensity $\lambda(\mathbf{s}^*)$ is about 0.044. A location that is 5 meters higher in elevation than the average has an intensity that is around $\exp\{5\beta_1 + 5^2\beta_2\} \approx 0.095$ percent of the intensity at the mean elevation. Figure 3(a) provides the posterior mean of the intensity surface under the NHPP model.

We employed the Lewis–Shedler algorithm to generate $L = 1000$ posterior predictive point patterns; $\mathcal{S}_l^*$ arises from a NHPP with intensity $\lambda^{(l)}(\mathbf{s})$, where the lth posterior



(a)

(b)

(c)

(d)

Figure 4: The posterior distributions for (a) $\lambda(D)$ and (b) $N(D)$ under the NHPP model and for (c) $\lambda(D)$ and (d) $N(D)$ under the LGCP model, both using the Duke Forest data.

samples of λ_0, β_1, and β_2 are used to construct $\lambda^{(l)}(\mathbf{s})$. Consider the posterior for $\lambda(D) = \int_D \lambda(\mathbf{s})d\mathbf{s}$. This integral was approximated to evaluate the likelihood in (5) during the model fitting, so these posterior samples have already been computed. Figure 4(a) shows the posterior distribution of $\lambda(D)$; it has a posterior mean of 528.97 with a 95% credible interval of (485.12, 575.18). Figure 4(b) shows the predictive distribution for $N(D)$; it has a posterior mean of 530.70 and a 95% credible interval of (468, 595). The distributions have the same center but the latter has greater spread, as expected.

LGCP model

We fit an LGCP model to compare with the NHPP model. The prior specifications remain the same except for the inclusion of the local Gaussian adjustment to the log-intensity. A Matérn covariance function was used with smoothness $\nu = 3/2$, chosen after discussions with ecologists involved in the project. We fit the model, running 10,000 iterations of burn-in and then taking 100,000 posterior samples. Elevation and squared elevation were again used as covariates. Again, we thin the posterior parameter samples and retain $L = 1000$ posterior point patterns.

Figure 5 shows the posterior distributions for λ_0, β_1, β_2, and σ^2. The minimum contrast estimate $\tilde{\phi}$ was 0.0427. It appears that the linear effect for elevation was again significant here, with a posterior mean similar to that obtained under the NHPP model. The quadratic effect of elevation has a credible interval that does overlap 0, however,

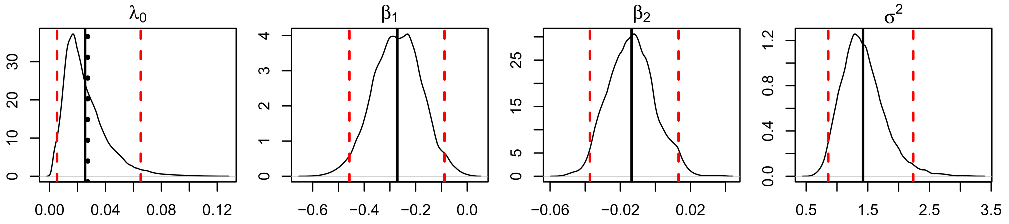


Figure 5: Posterior distributions for the parameters of the LGCP model. The posterior mean is marked by the solid vertical line and the 95% credible intervals are marked by the dashed lines. The HPP MLE $\hat{\lambda}$ is given by the dotted line in the first panel.

Model	p	Raw Innovations	Predictive Residuals
NHPP	0.5	0.29 / 0.18	0.84 / 0.76
LGCP	0.5	0.76 / 0.44	0.98 / 0.90
NHPP	0.8	0.14 / 0.09	0.74 / 0.82
LGCP	0.8	0.76 / 0.29	0.99 / 0.88

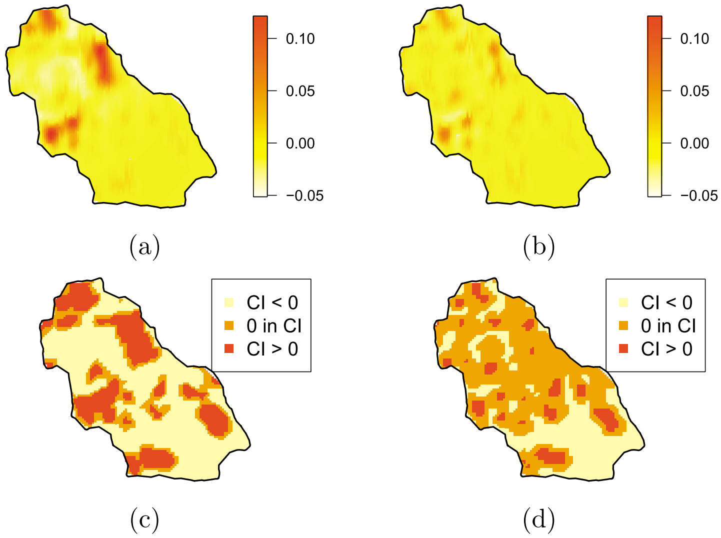
Table 4: Coverage of the 90% credible intervals for the raw innovations and predictive residuals in the Monte Carlo test for thinning levels $p = 0.5, 0.8$ and $q = 0.05$. The coverage on the training dataset is given before the forward slash and the coverage on the test dataset is given after the forward slash.

and the posterior mean is much closer to 0 than for the NHPP. We also note that the posterior mean for λ_0 is essentially the same as the MLE estimate under the HPP model, $\hat{\lambda} = n/|D| = 0.0273$. Figure 3(b) shows the posterior mean intensity surface for $\lambda(\mathbf{s})$ under the LGCP model. It seems to better capture the observed point pattern compared with the posterior mean intensity for the NHPP. Panels (c) and (d) in Figure 4 show the posterior distributions for $\lambda(D)$ and $N(D)$. These posterior distributions are essentially indistinguishable from those obtained using the NHPP model.

Model diagnostics

Table 4 presents the empirical coverage of the raw innovations and the predictive residuals for the Duke Forest data. We first used p-thinning with $p = 0.5$ and $p = 0.8$ to create training and test datasets. We used $K = 200$ squares of size $0.05 \times |D|$ and calculated 90% credible intervals for the raw residuals and 90% prediction intervals for the predictive residuals using both the training data and the test data. We see the severe undercoverage with the raw innovations, particularly for the NHPP. For the predictive residuals, we see that the NHPP exhibits undercoverage even in-sample while the LGCP achieves nominal coverage out of sample and expected elevated coverage in-sample. The results for the inverse λ and Pearson residuals are not shown but are similar to those for the raw residuals. The NHPP seems inadequate while we do not criticize the LGCP.

Finally, in Figure 6 we turn to the smoothed raw innovation fields, as discussed in Section 5.1. A bivariate Gaussian kernel was used with a bandwidth chosen using



(a)

(b)

(c)

(d)

Figure 6: Posterior mean of the smoothed raw innovation fields for the (a) NHPP and (b) LGCP models and posterior coverage plots for the smoothed raw innovation fields for the (c) NHPP and (d) LGCP models. The coverage plots describe whether a pointwise credible interval (CI) contains 0 or whether the interval is completely above or below 0.

cross-validation. We see that the smoothed innovation field for the NHPP model has more extreme negative and positive values than the LGCP model. That is, the NHPP intensity was too low in areas where a lot of data was observed (the high positive values in the smoothed residual field) and too high in areas where data was sparse (the negative values). The LGCP residual field is generally much closer to 0, i.e., it is closer to the empirical intensity.

The bottom row of Figure 6 shows locations which have a pointwise credible interval that contains 0, our proposed companion plot for the smoothed innovation plot. For the NHPP model, about 60% of the locations in D have a raw innovation posterior 95% credible interval which is entirely below 0 and 25% of the locations have a credible interval that is entirely above 0. So, about 15% of the domain is being covered by the residual intervals, which is roughly what we found in Table 4. In contrast, for the LGCP model, 33% of the domain has a raw innovation credible interval entirely below 0, 58% has an interval containing zero, and about 9% has an interval entirely above 0. Again, the NHPP performs inadequately.

Model selection

The foregoing investigation suggests that the NHPP model is inadequate. Nonetheless, to illustrate model comparison, for both the NHPP and LGCP models, we compute the ranked probability scores and predictive mean square error for the training and test data when holding out roughly 50% of the data using p-thinning. We set $K = 200$ and

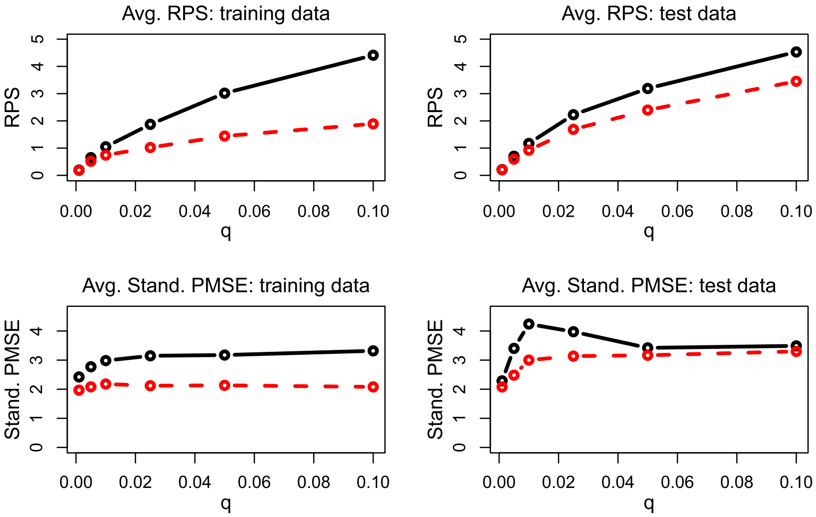


Figure 7: Average RPS and standardized PMSE for the NHPP model (solid black line) and the LGCP model (dashed lines) fitted to the Duke Forest test data for three cross-validation sets with $p = 0.5$.

sample the B_k locations uniformly over D, each B_k being a square of size $q|D|$ with $q \in (0, 0.1)$; $q > 0.1$ limits where the B_k can fall in D. The average RPS and predictive residual coverage were calculated for both the training and test data. We replicated this analysis three times (applying p-thinning to the dataset three separate times) and performed this analysis for each set of training and test data. We then averaged over these replications to provide the shown results.

The top row of Figure 7 compares the average RPS for the two models on both the training and test datasets across different values of q. The same set of B_k was used for both the training and test data. The NHPP model clearly performs worse than the LGCP model. The bottom row of Figure 7 presents the average standardized PMSE for both models in the same format as the top row. The LGCP model again provides an advantage. In terms of RPS, the LGCP provides results that are 15–35% better on the test data for $q \geq 0.05$. The LGCP model provides standardized PMSE on the test data that is 25–40% better for q in the [0.005, 0.025] range and 5–10% better elsewhere.

Using both the RPS and standardized PMSE results, the LGCP model emerges as preferable to the NHPP model (in conjunction with the previous section which showed that the NHPP exhibited some lack of model fit). Section 1 in the online supplementary material contributes extra analyses related to inference and model diagnostics for this example. Section 2 in the online supplementary material uses simulation examples to further investigate the ability to perform model selection between HPPs, NHPPs, and LGCPs.

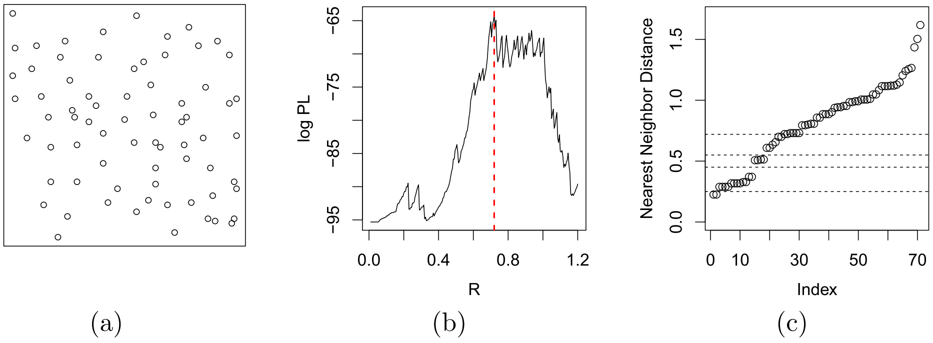


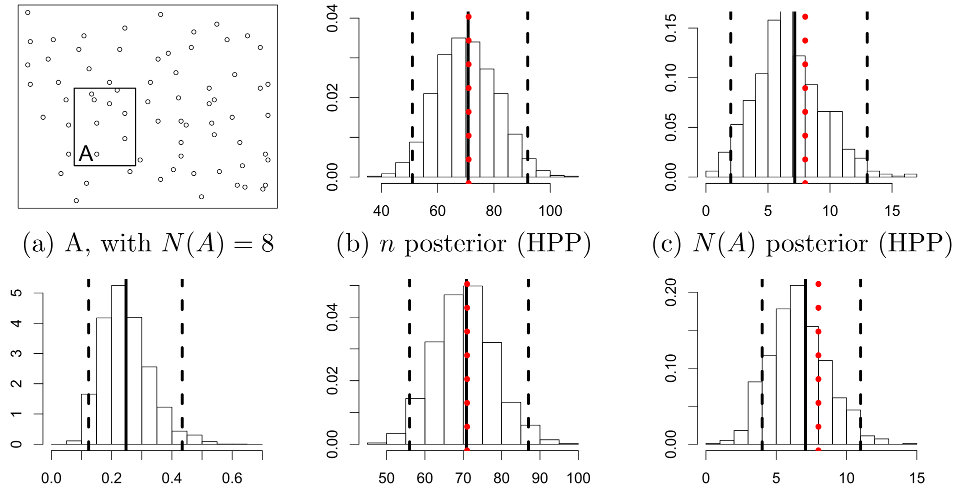
Figure 8: Plots of (a) the Swedish pines data, (b) profile pseudolikelihood for the Strauss model as a function of R, and (c) the (sorted) nearest neighbor distances. The dashed line in (b) indicates the profile maximum pseudolikelihood estimate $\hat{R} = 0.72$. The dashed lines in (c) indicate the candidate R values of 0.25, 0.45, 0.55, and 0.72.

7.2 Swedish pines data

We now fit a HPP and a Strauss model to the Swedish pines data from Ripley (1981) and Baddeley and Turner (2000). The data consists of the locations of 71 pine saplings within a 10 m × 10 m square. Figure 8 shows the data, along with the profile pseudolikelihood of the Strauss model (across radius R), and the ordered nearest neighbor distances for each $\mathbf{s}_i \in S$. We compare the HPP model with four Strauss models, each having a different value for R. The smallest observed interpoint distance is 0.22 with most of the nearest neighbors being greater than 0.5, so the values of R we consider are $R = 0.25$, 0.45, 0.55 and the profile maximum pseudolikelihood estimate, $\hat{R} = 0.72$. These candidate values are shown as dashed horizontal lines in Figure 8(c).

For the HPP model, a Gamma prior was used for λ with $\mathbb{E}[\lambda] = 70/|D|$ and $Var[\lambda] = 0.01$. The HPP model was run for 1,000 iterations of burn-in and then 50,000 posterior samples were collected. The Strauss models all used the uniform prior $\gamma \sim U(0, 0.75)$ and uniform priors for β which varied by model (e.g., the Strauss model with $R = 0.72$ used $\beta \sim U(0.9, 3.3)$). The Strauss models were run for 5,000 burn-in iterations and then 10^6 posterior samples were taken. For the Strauss models, longer runs were required due to a tendency for the chain to not move for long periods of time. Møller et al. (2006) noted this as well; we found it to worsen as γ gets smaller or as R gets larger. For all models, 1,000 posterior predictive point patterns were generated using thinned posterior samples of model parameters.

Figure 9(d) shows the posterior distribution for γ, the interaction potential, under the Strauss model with $R = 0.72$. Because the Strauss model becomes an HPP for $\gamma = 1$, the posterior for γ suggests that interaction is present, since most of the mass is in the range $(0.1, 0.4)$. Figure 9 shows the posterior distributions for n and $N(A)$ under both models, where $A = [2, 4.5] \times [2, 6]$ and $|A| = 0.1|D|$. The posterior summaries for both n and $N(A)$ are similar under both the HPP and Strauss($R = 0.72$) models, though the posteriors under the Strauss model are more concentrated around the observed values.



(a) A, with $N(A) = 8$ (b) n posterior (HPP) (c) $N(A)$ posterior (HPP)

(d) γ posterior (Strauss) (e) n posterior (Strauss) (f) $N(A)$ posterior (Strauss)

Figure 9: Plots of (a) the Swedish pines data with subregion A labeled and the posterior distributions for (b)–(c) n and $N(A)$ under the HPP model, and (d)–(f) γ, n, and $N(A)$ under the Strauss(R=0.72) model. The solid and dashed lines indicate the posterior means and 95% credible intervals, respectively, and the dotted lines indicate the observed values.

Model adequacy and comparison

Table 5 shows the in-sample RPS, PMSE standardized by the expected number, and coverage results for the HPP and Strauss models using 200 random boxes ($\{B_k\}$) placed in D with size $q|D|$. Using predictive residuals, the models all give similar coverage. In fact, the coverage percentages for all of the models are well above the 90% nominal level, which, in-sample, is not unexpected. The RPS results show that the performance of the Strauss and HPP models is very similar, with the largest differences occurring for the largest box size ($q = 0.10$) and large R (more different from the HPP). The standardized PMSE results show improved predictive performance as we go from the HPP to Strauss processes with increasing R's.

As with the simulation example, we turn to second order model comparison using $s_r(\mathcal{S})$. We compare $s_r(\mathcal{S}_{obs})$ with those generated under the prior expectations of each model by simulating 999 point patterns ($\{S_b^*\}$) from the prior predictive distribution of each model. To address sensitivity to the value of r used, we compare the discrepancy function for several values of r: 0.25, 0.45, 0.55, 0.72, and 0.90. The results of the model criticism checks are given in Table 6. The table shows, for each value of r, the Monte Carlo p-value, i.e., the proportion of simulated $s_r(\mathcal{S}_b^*)$'s that are below the observed value $s_r(\mathcal{S}_{obs})$ for each of the five models. In general, for each r, the HPP is nearly significant and always "more" significant than the corresponding Strauss process models. Moreover, for $r = 0.72$, the data formally criticizes the HPP model and the Strauss models with

	HPP	Strauss (R=0.25)	Strauss (R=0.45)	Strauss (R=0.55)	Strauss (R=0.72)
Ranked Probability Score					
$q = 0.005$	0.25	0.25	0.25	0.25	0.24
0.01	0.34	0.34	0.34	0.34	0.33
0.025	0.51	0.52	0.54	0.51	0.49
0.05	0.71	0.74	0.80	0.73	0.69
0.10	1.17	1.04	1.04	0.99	1.07
Standardized PMSE					
$q = 0.005$	1.77	1.62	1.50	1.51	1.54
0.01	1.59	1.46	1.34	1.30	1.27
0.025	1.50	1.38	1.27	1.17	1.09
0.05	1.45	1.38	1.27	1.14	1.02
0.10	1.60	1.39	1.19	1.10	1.07
Empirical Coverage					
$q = 0.005$	1.00	1.00	1.00	0.98	0.98
0.01	1.00	1.00	1.00	1.00	1.00
0.025	1.00	1.00	1.00	1.00	1.00
0.05	1.00	1.00	1.00	1.00	1.00
0.10	1.00	1.00	1.00	1.00	1.00

Table 5: RPS, standardized PMSE, and empirical coverage (nominal 90% intervals) for the HPP and Strauss models on the Swedish pines data.

smaller R. Using a discrepancy function targeting second-order characteristics, within this set of models, the Strauss($R = 0.72$) model is the only one that receives no criticism.

8 Summary and future work

We have presented a general approach for posterior inference, model criticism, and model selection under Bayesian modeling for spatial point pattern data. For models which allow generation of point patterns given parameter values, we assert that rich inference can be straightforwardly done through simulation of posterior predictive point patterns. We propose using p-thinning to perform cross-validation for Poisson and Cox processes, which allows model checking and comparison on an independent test point

Model	$r = 0.25$	0.45	0.55	0.72	0.90
HPP	0.086	0.100	0.053	0.021	0.074
Strauss(R=0.25)	0.435	0.141	0.089	0.052	0.097
Strauss(R=0.45)	0.350	0.483	0.167	0.053	0.086
Strauss(R=0.55)	0.341	0.470	0.416	0.070	0.085
Strauss(R=0.72)	0.344	0.454	0.400	0.359	0.244

Table 6: Monte Carlo p-values of $s_r(\mathcal{S}_{obs})$ for each model fitted to the Swedish pines data.

pattern from the same process (except for a scaling of the intensity function). We offered Bayesian analogues of point pattern residuals and innovations and argue for predictive residuals because they are more suitable for comparing empirical coverage with nominal level of coverage.

For model criticism, we argued for prior predictive model checks in the context of empirical coverage when holding out data is possible. For in-sample criticism, we suggested discrepancy functions tailored to the classes of models under investigation. For model selection, we can first determine whether a candidate model exhibits strong indications of lack of fit. Then, for adequate models, we proposed comparison of predictive inference for set counts with observed counts for these sets. We suggested comparison using PMSE and RPS, averaged over sets, to provide a measure of model performance. We would do this out-of-sample when possible, in-sample otherwise. We caution that the ability to successfully assess model fit and model performance is often hampered by small sample sizes and also by the weak information that a point pattern realization offers about the underlying point process generating it. Often, several models may appear to perform equally well; subject matter insight into the process driving the point pattern would help to make a choice.

Areas for future work include extending our tools to Neyman–Scott processes, shot noise Cox processes, marked point patterns with discrete and continuous marks, inhomogeneous Gibbs processes, and spatio-temporal point patterns.

Supplementary Material

Online Supplementary Material for Bayesian Inference and Model Assessment for Spatial Point Patterns Using Posterior Predictive Samples (DOI: 10.1214/15-BA985SUPP; .pdf). Further analysis of the Duke Forest example is given in the online supplementary material, showing posterior distributions for first- and second-order marginal intensities, the pairwise correlation function, and other features of interest. Then a simulation study is presented, showing the performance of predictive residual coverage and model choice using RPS for various models under several data-generating scenarios. Data is generated under an HPP, an NHPP, and different specifications of LGCPs and then each model is fit to each data scenario. With many observations, the simpler models show signs of lack of fit when the data-generating process is more complex.

References

Akman, V. E. and Raftery, A. E. (1986). "Bayes Factors for Non-homogeneous Poisson Process with Vague Prior Information." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 48(3): 322–329. MR0876843. 13

Baddeley, A., Møller, J., and Pakes, A. G. (2008). "Properties of residuals for spatial point processes." *Annals of the Institute of Statistical Mathematics*, 60(3): 627–649. MR2434415. doi: `http://dx.doi.org/10.1007/s10463-007-0116-6`. 9

Baddeley, A. and Turner, R. (2000). “Practical maximum pseudolikelihood for spatial point patterns.” *Australian & New Zealand Journal of Statistics*, 42(3): 283–322. MR1794056. doi: http://dx.doi.org/10.1111/1467-842X.00128. 7, 22

Baddeley, A. and Turner, R. (2005). “Spatstat: an R package for analyzing spatial point patterns.” *Journal of Statistical Software*, 12(6): 1–42. 3

Baddeley, A., Turner, R., Møller, J., and Hazelton, M. (2005). “Residual analysis for spatial point processes.” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(5): 617–666. MR2210685. doi: http://dx.doi.org/10.1111/j.1467-9868.2005.00519.x. 2, 9, 10, 11

Baddeley, A. J., Moller, J., and Waagepetersen, R. (2000). “Non- and semi-parametric estimation of interaction in inhomogeneous point patterns.” *Statistica Neerlandica*, 54(3): 329–350. MR1804002. doi: http://dx.doi.org/10.1111/1467-9574.00144. 9

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman and Hall/CRC Press, 2 edition. MR3362184. 3, 9

Berman, M. and Turner, T. R. (1992). “Approximating point process likelihoods with GLIM.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41(1): 31–38. 7

Berthelsen, K. K. and Møller, J. (2002). “A primer on perfect simulation for spatial point processes.” *Bulletin of the Brazilian Mathematical Society*, 33(3): 351–367. MR1978833. doi: http://dx.doi.org/10.1007/s005740200019. 8

Berthelsen, K. K. and Møller, J. (2003). “Likelihood and Non-parametric Bayesian MCMC Inference for Spatial Point Processes Based on Perfect Simulation and Path Sampling.” *Scandinavian Journal of Statistics*, 30(3): 549–564. 8

Berthelsen, K. K. and Møller, J. (2006). “Bayesian analysis of Markov point processes.” In Baddeley, A., Gregori, P., Mateu, J., Stoica, R., and Stoyan, D. (eds.), *Case Studies in Spatial Point Processes*, 85–97. New York: Springer-Verlag. MR2232124. doi: http://dx.doi.org/10.1007/0-387-31144-0_4. 7

Berthelsen, K. K. and Møller, J. (2008). “Non-Parametric Bayesian Inference for Inhomogeneous Markov Point Processes.” *Australian & New Zealand Journal of Statistics*, 50(3): 257–272. MR2455432. doi: http://dx.doi.org/10.1111/j.1467-842X.2008.00516.x. 3

Bowman, A. W. (1984). “An Alternative Method of Cross-Validation for the Smoothing of Density Estimates.” *Biometrika*, 71(2): 353. MR0767163. doi: http://dx.doi.org/10.1093/biomet/71.2.353. 11

Christensen, O. F., Roberts, G. O., and Rosenthal, J. S. (2005). “Scaling limits for the transient phase of local Metropolis-Hastings algorithms.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2): 253–268. MR2137324. doi: http://dx.doi.org/10.1111/j.1467-9868.2005.00500.x. 7

Dey, D. K., Gelfand, A. E., Swartz, T. B., and Vlachos, P. K. (1998). "A simulation-intensive approach for checking hierarchical models." *Test*, 7(2): 325–346. 12

Diggle, P. and Marron, J. S. (1988). "Equivalence of Smoothing Parameter Selectors in Density and Intensity Estimation." *Journal of the American Statistical Association*, 83(403): 793–800. MR0963807. 11

Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*. London: Academic Press, 2nd edition. MR0743593. 11

Epstein, E. S. (1969). "A scoring system for probability forecasts of ranked categories." *Journal of Applied Meteorology*, 8: 985–987. 14

Gelfand, A. E., Diggle, P. J., Guttorp, P., and Fuentes, M. (eds.) (2010). *Handbook of Spatial Statistics*. London: Chapman & Hall/CRC Press. MR2761512. doi: `http://dx.doi.org/10.1201/9781420072884`. 3

Gelman, A., Meng, X.-l., and Stern, H. (1996). "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies." *Statistica Sinica*, 6(4): 733–807. MR1422404. 12

Georgii, H.-O. (1976). "Canonical and grand canonical Gibbs states for continuum systems." *Communications in Mathematical Physics*, 48: 31–51. MR0411497. 5

Girolami, M. and Calderhead, B. (2011). "Riemann manifold Langevin and Hamiltonian Monte Carlo methods." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2): 123–214. MR2814492. doi: `http://dx.doi.org/10.1111/j.1467-9868.2010.00765.x`. 7

Gneiting, T. and Raftery, A. E. (2007). "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association*, 102(477): 359–378. MR2345548. doi: `http://dx.doi.org/10.1198/016214506000001437`. 14

Guttorp, P. and Thorarinsdottir, T. L. (2012). "Advances and Challenges in Space-time Modelling of Natural Events." *Advances and Challenges in Space-time Modelling of Natural Events*, 207: 79–102. 13

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Wiley-Interscience. MR2384630. 2, 3, 4, 5, 6, 8, 9, 11

Illian, J. B., Møller, J., and Waagepetersen, R. P. (2009). "Hierarchical spatial point process analysis for a plant community with high biodiversity." *Environmental and Ecological Statistics*, 16: 389–405. MR2749847. doi: `http://dx.doi.org/10.1007/s10651-007-0070-8`. 12

Illian, J. B., Sørbye, S. H., and Rue, H. (2012). "A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA)." *The Annals of Applied Statistics*, 6(4): 1499–1530. MR3058673. doi: `http://dx.doi.org/10.1214/11-AOAS530`. 3

King, R., Illian, J. B., King, S. E., Nightingale, G. F., and Hendrichsen, D. K. (2012). "A Bayesian Approach to Fitting Gibbs Processes with Temporal Random Effects." *Journal of Agricultural, Biological, and Environmental Statistics*. MR3041887. doi: http://dx.doi.org/10.1007/s13253-012-0111-0. 3, 7

Kottas, A. and Sansó, B. (2007). "Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis." *Journal of Statistical Planning and Inference*, 137(10): 3151–3163. MR2365118. doi: http://dx.doi.org/10.1016/j.jspi.2006.05.022. 3

Lahiri, S. (1999). "Asymptotic distribution of the empirical spatial cumulative distribution function predictor and prediction bands based on a subsampling method." *Probability Theory and Related Fields*, 114(1): 55–84. MR1697139. doi: http://dx.doi.org/10.1007/s004400050221. 12

Lahiri, S. N. (2003). "Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs." *Sankhyā: The Indian Journal of Statistics*, 65(2): 356–388. MR2028905. 12

Lavancier, F., Møller, J., and Rubak, E. (2015). "Determinantal point process models and statistical inference." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(4): 853–877. MR3382600. 3

Leininger, T. J. and Gelfand, A. E. (2015). "Online Supplementary Material for Bayesian Inference and Model Assessment for Spatial Point Patterns using Posterior Predictive Samples." *Bayesian Analysis*. doi: http://dx.doi.org/10.1214/15-BA985SUPP. 14

Lewis, P. A. W. and Shedler, G. S. (1979). "Simulation of nonhomogeneous Poisson processes by thinning." *Naval Research Logistics Quarterly*, 26(3): 403–413. MR0546120. doi: http://dx.doi.org/10.1002/nav.3800260304. 8

Møller, J. (2012). "Aspects of spatial point process modelling and Bayesian inference." http://conferences.inf.ed.ac.uk/bayeslectures/moeller.pdf. Presentation at Bayes Lectures 2012, Edinburgh, UK. 3, 12

Møller, J., Pettitt, A. N., Reeves, R., and Berthelsen, K. K. (2006). "An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants." *Biometrika*, 93(2): 451–458. MR2278096. doi: http://dx.doi.org/10.1093/biomet/93.2.451. 3, 7, 22

Møller, J. and Rasmussen, J. G. (2012). "A sequential point process model and Bayesian inference for spatial point patterns with linear structures." *Scandinavian Journal of Statistics*, 39(4): 618–634. MR3000838. doi: http://dx.doi.org/10.1111/j.1467-9469.2011.00769.x. 3, 12

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). "Log-Gaussian Cox Processes." *Scandinavian Journal of Statistics*, 25: 451–482. MR1650019. doi: http://dx.doi.org/10.1111/1467-9469.00115. 3, 5, 6, 7

Møller, J. and Waagepetersen, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton, FL: Chapman and Hall/CRC. 3

Møller, J. and Waagepetersen, R. P. (2007). "Modern Statistics for Spatial Point Processes." *Scandinavian Journal of Statistics*, 34: 643–684. MR2392447. doi: http://dx.doi.org/10.1111/j.1467-9469.2007.00569.x. 3, 4

Murray, I. and Adams, R. P. (2010). "Slice sampling covariance hyperparameters of latent Gaussian models." In: Lafferty, J., Williams, C. K. I., Shawe-Taylor, J., Zemel, R., and Culotta, A. (eds.), *Advances in Neural Information Processing Systems 23*, 1723–1731. 7

Murray, I., Adams, R. P., and Mackay, D. J. C. (2010). "Elliptical slice sampling." *Journal of Machine Learning Research: Workshop and Conference Proceedings*, 9: 541–548. 7

Nguyen, X. X. and Zessin, H. (1979). "Integral and differential characterizations of the Gibbs process." *Mathematische Nachrichten*, 88: 105–115. MR0543396. doi: http://dx.doi.org/10.1002/mana.19790880109. 5

Raftery, A. E. and Akman, V. E. (1986). "Bayesian analysis of a Poisson process with a change-point." *Biometrika*, 73(1): 85–89. MR0836436. doi: http://dx.doi.org/10.1093/biomet/73.1.85. 13

Ripley, B. D. (1981). *Spatial Statistics*. New York: John Wiley & Sons. MR0624436. 22

Rue, H., Martino, S., and Chopin, N. (2009). "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2): 319–392. MR2649602. doi: http://dx.doi.org/10.1111/j.1467-9868.2008.00700.x. 3

Sherman, M. and Carlstein, E. (1994). "Nonparametric Estimation of the Moments of a General Statistics Computed from Spatial Data." *Journal of the American Statistical Association*, 89(426): 496–500. MR1294075. 12

Stoyan, D. and Grabarnik, P. (1991). "Second-order Characteristics for Stochastic Structures Connected with Gibbs Point Processes." *Mathematische Nachrichten*, 151(1): 95–100. MR1121200. doi: http://dx.doi.org/10.1002/mana.19911510108. 10

Strauss, D. J. (1975). "A Model for Clustering." *Biometrika*, 62(2): 467–475. MR0383493. 6

Taddy, M. A. (2010). "Autoregressive Mixture Models for Dynamic Spatial Poisson Processes: Application to Tracking Intensity of Violent Crime." *Journal of the American Statistical Association*, 105(492): 1403–1417. MR2796559. doi: http://dx.doi.org/10.1198/jasa.2010.ap09655. 3

Taddy, M. A. and Kottas, A. (2012). "Mixture Modeling for Marked Poisson Processes." *Bayesian Analysis*, 7(2): 335–362. MR2934954. doi: http://dx.doi.org/10.1214/12-BA711. 3

Waagepetersen, R. and Schweder, T. (2006). "Likelihood-based inference for clustered line transect data." *Journal of Agricultural, Biological, and Environmental Statistics*, 11(3): 264–279. 7

Xiao, S., Kottas, A., Sansó, B., et al. (2015). "Modeling for seasonal marked point processes: An analysis of evolving hurricane occurrences." *The Annals of Applied Statistics*, 9(1): 353–382. MR3341119. doi: `http://dx.doi.org/10.1214/14-AOAS796`. 3

Zhang, H. (2004). "Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics." *Journal of the American Statistical Association*, 99(465): 250–261. MR2054303. doi: `http://dx.doi.org/10.1198/016214504000000241`. 7

Zhou, Z., Matteson, D. S., Woodard, D. B., Henderson, S. G., and Micheas, A. C. (2015). "A spatio-temporal point process model for ambulance demand." *Journal of the American Statistical Association*, 110(509): 6–15. MR3338482. doi: `http://dx.doi.org/10.1080/01621459.2014.941466`. 3

Acknowledgments

The work of both authors was supported in part by NSF CMG 0934595. The authors wish to thank the editor, associate editor, and reviewer for helpful suggestions in improving the paper. The authors also thank Jim Clark and Kai Zhu for providing the Duke Forest data and Robert Wolpert for valuable discussions.