Bayesian Modeling of Continuously Marked Spatial Point Patterns

Matthew A. Bognar[1]

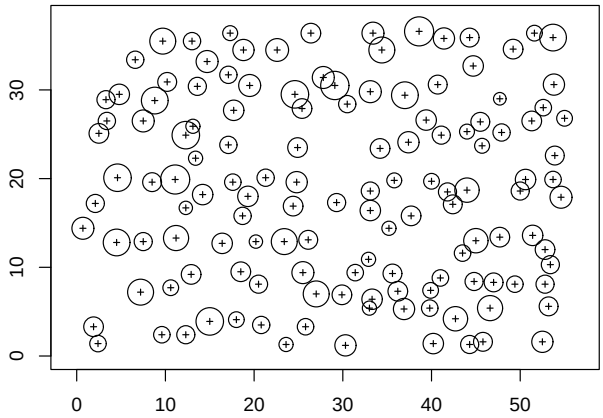
[1]University of Iowa, Iowa City, Iowa, USA

Summary

Many analyses of continuously marked spatial point patterns assume that the density of points, with differing marks, is identical. However, as noted in the originative paper of Goulard et al. (1996), such an assumption is not realistic in many situations. For example, a stand of forest may have many more small trees than large, hence the model should allow for a higher density of points with small marks. In addition, as suggested by Ogata & Tanemura (1985), the interaction between points should be a function of their mark, allowing, for example, the range of interaction for large trees to exceed that of smaller trees. The aforementioned articles use frequentist inferential techniques, but interval estimation presents difficulties due to the complex distributional properties of the estimates. We suggest the use of Bayesian inferential techniques. Although a Bayesian approach requires a complex, computational implementation of (reversible jump) MCMC methodology, it enables a wide variety of inferences (including interval estimation). We demonstrate our approach by analyzing the well known Norway spruce dataset.

Keywords: Markov chain Monte Carlo (MCMC), reversible jump MCMC, pairwise interacting point process, mark chemical activity function

Figure 1: Location of $n = 134$ Norway spruce trees in a 56×38 meter field. Character size is proportional to the trunk diameter.



1 Introduction

A *spatial point pattern* describes the spatial location of events in a region. For example, a spatial point pattern may describe the location of trees in a forest or the location of glacial deposits (drumlin) in a field. Many times a continuous mark is associated with each observed point; for example, Figure 1 depicts a *marked spatial point pattern* of $n = 134$ Norway spruce trees in a 56×38 meter field located in Germany (Fiksel 1984). The location of each tree is plotted with a + while the trunk diameter of each tree is plotted as a circle proportional to the trunk diameter.

A salient feature of the spruce dataset is that there is a higher density of small trees growing in the region. Moreover, it can be seen, upon closer inspection, that the tree locations exhibit *inhibition* in their spacing. In other words, only a few trees are close together (say within 2 meters); a completely random (i.e. binomial) process with the same number of points would exhibit many more close points. Quite simply, there is *interaction* between the trees, most likely due to biological competition. This inhibitive spacing is sometimes called *spatial regularity*, and constitutes the type of patterns considered in the remainder of this paper. Less obvious, perhaps, is that the spatial regularity depends upon the marks; large trees (with large trunk diameters) tend to grow quite removed from one another, while smaller trees are better capable

of growing in closer proximity. Practitioners commonly ignore the marks because of the heightened modeling complexity. However, such information is crucial to fully understanding the underlying process, and hence should not be ignored.

In modeling *unmarked* spatial point patterns, Poisson processes (with known normalization constant) are commonly used, although such models are unable to account for interaction between points. *Gibbsian point processes* can account for such interaction, but involve an *intractable* normalizing constant which complicates inference (Gibbsian models were originally used in statistical physics to study the interaction of particles in fluids and gases). Because Ripley (1977) believed that the interaction between individuals may depend upon the distance between them, he suggested using a *pairwise interacting point process* (a special case of a Gibbs process) which describes the interaction between *pairs* of points by a function (called a *pair potential function*) of the interpoint Euclidean distance. A pairwise interacting point process is the most common form of a Gibbs point process. For more information about spatial point processes, see for example Baddeley & Turner (2000), Bognar (2005), Bognar & Cowles (2004), Diggle (2003), Diggle et al. (1987, 1994), Harkness & Isham (1983), Heikkinen & Penttinen (1999), Mateu & Montes (2001), Møller & Waagepetersen (2004), Ogata & Tanemura (1981, 1984, 1986, 1989), Penttinen (1984), Stoyan & Stoyan (1998) and van Lieshout (2000).

Inference for *marked* spatial point processes has received much less attention in the literature. Fiksel (1984) generalized the inferential method of Takacs (1986) to accommodate marked Gibbsian processes, while Ogata & Tanemura (1985) extended their approximate maximum method using virial expansions to the marked case. Jensen & Møller (1991) presented a theoretical justification for using maximum pseudolikelihood techniques in Markov type processes (including marked processes), while Baddeley & Turner (2000) described a technique for obtaining maximum pseudolikelihood estimates for (marked) spatial point processes using the Berman-Turner device (Berman & Turner 1992). Goulard et al. (1996) suggested that the chemical activity be a function of the mark, extending the previous models seen in the literature. These frequentist inferential approaches provide for point estimation, but because of the complex distributional properties of the estimates, interval estimation (in general) remains difficult. This paper focuses on continuous mark spaces, the simpler discrete case having been thoroughly studied. See Baddeley & Møller (1989), Degenhardt (1999), and Stoyan & Penttinen (2000) for more information on marked spatial point processes.

Bayesian inference for spatial point processes has been notably underrepresented in the literature. In the *unmarked* case, Heikkinen & Penttinen (1999) described a non-parametric estimator of the pair potential function based on a Bayesian smoothing technique. Unfortunately, interval estimation appears

more challenging. Bognar & Cowles (2004) suggest an efficient MCMC algorithm (utilizing importance sampling) which allows sampling from the full posterior distribution; the posterior realizations can be used to perform point, interval, among other inferences. Bognar (2005) described a Gibbsian model that allowed for spatial inhomogeneity in the density of points. Unfortunately the latter two papers do *not* address the incorporation of marks. See Møller & Waagepetersen (2004) (§9.2) for more information on Bayesian inference for (unmarked) spatial point patterns.

The aim of this paper is to describe the underlying spatial structure in continuously marked Gibbsian point processes by 1) modeling the interaction between points, 2) allowing the interaction between points to depend on the attached marks, and 3) allow for inhomogeneity in the density of points with differing marks (the seminal paper by Goulard et al. (1996) addressed the above issues in a *frequentist* setting). Unlike the frequentist inferential paradigm, a Bayesian approach (the focus of this paper) provides the framework for a plethora of inferences, including interval estimation, not readily obtainable via classical methodology.

The remainder of this paper begins with an introduction to marked Gibbs point processes in Section 2. A Bayesian model is suggested in Section 3, including a description of the likelihood and prior specification. Bayesian inference will be based upon MCMC simulations from the full posterior distribution; the algorithm is outlined in Section 4. Because the model contains an intractable normalization constant, the complexity of the MCMC algorithm is intensified; the technical and computational challenges are discussed in Section 5. The spruce dataset is analyzed in Section 6, and concluding remarks constitute Section 7.

2 Marked Gibbsian point processes

Let V be an arbitrary set; typically V is a bounded subset of $\mathbb{R}^2$ or $\mathbb{R}^3$. Let $(V, \mathcal{B}, \lambda)$ be a measure space where the σ-field $\mathcal{B}$ contains all singletons, λ denotes Lebesgue measure, and $\lambda(V) < \infty$. Let Ω_n denote the set of all configurations of n points in V. Letting $\Omega_0 = \{\emptyset\}$, the space of finite point patterns in V is thus $\Omega \stackrel{\text{def}}{=} \cup_{n=0}^{\infty} \Omega_n$. Let $(\Omega, \mathcal{F}, \mu)$ be the exponential space over V (Carter & Prenter 1972) where $\mathcal{F}$ is a σ-field on Ω, and the probability measure μ is assumed to be the distribution of a homogeneous Poisson process on V with intensity measure λ (i.e. the number of points in $B \in \mathcal{B}$ has a Poisson distribution with mean $\lambda(B)$). Hence for $F \in \mathcal{F}$,

$$\mu(F) \quad = \quad \exp[-\lambda(V)] \Bigg[\mathrm{I}(\emptyset \in F) +$$

$$\sum_{n=1}^{\infty}\left(\frac{1}{n!}\int_V\cdots\int_V \mathrm{I}(\{x_1,\ldots,x_n\}\in F)\,d\lambda(x_1)\ldots d\lambda(x_n)\right)\Bigg].$$

A finite (unmarked) point process X on V is a random variable on $(\Omega,\mathcal{F},\mu)$. The density of X with respect to μ is denoted by p.

Let $(M,\mathcal{M},\nu)$ be a measure space where M is the mark space, $\nu(M)=1$, and the σ-field $\mathcal{M}$ on M contains all singletons. To construct a finite marked spatial point process, consider the Cartesian product $(V\times M,\mathcal{B}\otimes\mathcal{M},\lambda\otimes\nu)$. Elements of V, M, and $V\times M$ are respectively called *points*, *marks*, and *marked points* and are denoted by x_i, m_i, and (x_i,m_i). A finite marked point process on V with marks in M is a finite point process on $V\times M \stackrel{\text{def}}{=} \{(\omega_1,\omega_2):\omega_1\in V,\omega_2\in M\}$; i.e. each realization is a member of $\Omega_{V\times M}$ (which, loosely speaking, is the set of all possible finite point patterns in V with marks contained in M). The marked Poisson point process can be considered as a Poisson process (with intensity λ) on V with marks being i.i.d. with distribution ν. In what follows, let $\Gamma\stackrel{\text{def}}{=}\Omega_{V\times M}$, $\mathcal{G}\stackrel{\text{def}}{=}\mathcal{B}\otimes\mathcal{M}=\{\omega_1\times\omega_2:\omega_1\in\mathcal{B},\omega_2\in\mathcal{M}\}$, and define η to be the unique probability on $(\Gamma,\mathcal{G})$ such that $\eta(A_1\times A_2)=\lambda(A_1)\nu(A_2)$ for all $A_1\in\mathcal{B}$ and $A_2\in\mathcal{M}$.

Let $\mathbf{x}=(x_1,\ldots,x_n)$ denote the point locations and $\mathbf{m}=(m_1,\ldots,m_n)$ the respective marks. According to Goulard et al. (1996), the distribution of the marked Gibbs process (with respect to the marked Poisson process with intensity η) is

$$p(\mathbf{x},\mathbf{m})=Z^{-1}\exp[-U(\mathbf{x},\mathbf{m})]$$

where $U(\mathbf{x},\mathbf{m})$ is an *energy function* and

$$Z=\int_\Gamma \exp[-U(\mathbf{x},\mathbf{m})]\,d\eta(\mathbf{x},\mathbf{m})$$

is a *partition function* which normalizes p.

A *marked pairwise interacting point process* is a special case of a marked Gibbs point process where the energy function is completely determined by the *mark chemical activity function* and the *mark pair potential function*. Assume that the energy function takes the form

$$U(\mathbf{x},\mathbf{m})=\sum_{i=1}^{n}\alpha(m_i)+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\phi(x_i,x_j,m_i,m_j).$$

The mark chemical activity function α describes the ability of the system to absorb a point with mark m, while the mark pair potential function ϕ characterizes the interaction between pairs of points by a function of the inter-point Euclidean distance $\|x_i-x_j\|$ and respective marks m_i and m_j. Allowing for inhomogeneity in the mark chemical activity function (e.g. allowing $\alpha(m_i)\neq\alpha(m_j)$ if $m_i\neq m_j$) enables the density of points to depend upon the marks.

If the number of points n is not conditioned upon, then allowing $\phi < 0$ may produce an infinite normalizing constant Z and cause simulation difficulties (Kelly & Ripley 1976, Gates & Westcott 1986, Møller 1999). Thus, in what follows it is assumed that ϕ is non-negative, gauranteeing $Z < \infty$.

3 Model Details

In modeling the mark pair potential function ϕ, consider a family of mark pair potential functions ϕ_θ, indexed by a parameter vector $\theta = (\theta_1, \ldots, \theta_m)$, where $\phi_\theta : V \times V \times M \times M \to [0, \infty)$. It is assumed without loss of generality that θ if of fixed, known dimension. As a tangible example, consider the mark pair potential function (which is utilized in Section 6)

$$\phi_\theta(x_i, x_j, m_i, m_j) = \begin{cases} \infty & \text{if } e_{ij} \leq b_{hc} \\ h & \text{if } b_{hc} < e_{ij} \leq b \\ 0 & \text{if } e_{ij} > b \end{cases} \tag{1}$$

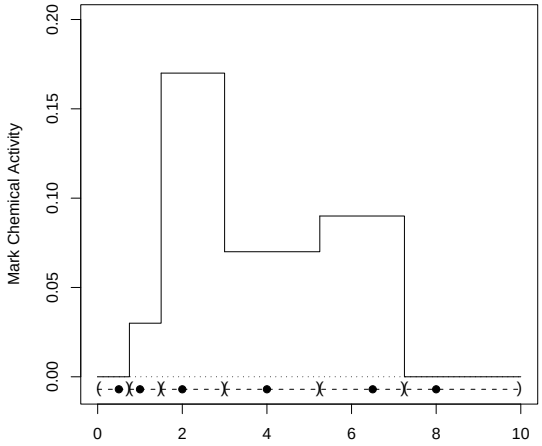
where

$$e_{ij} = \|x_i - x_j\| \left(\frac{m_i + m_j}{2\bar{m}} \right)^{-d}, \tag{2}$$

$\bar{m} = n^{-1} \sum_{i=1}^{n} m_i$, and $\theta = (b_{hc}, b, h, d)$ (Strauss 1975, was first to suggest a similar potential function for unmarked point patterns). For two points with "average" marks (i.e. $(m_i + m_j)/2\bar{m} = 1$), the *hard-core interaction distance* b_{hc} describes the minimum inter-point distance, the *interaction distance* b characterizes the distance at which two points cease to interact, and the *Straussian parameter* h describes the strength of interaction between pairs of points ($h > 0$ indicates spatial regularity, while $h = 0$ implies no spatial interaction). The *scaling parameter* d (see Ogata & Tanemura 1985) in effect scales b_{hc} and b according to the marks m_i and m_j (see Section 6 for a detailed interpretation of d).

The mark chemical activity function α is assumed to be a member of the parametric family of mark chemical activity functions α_ψ, indexed by a parameter vector $\psi = (\psi_1, \ldots, \psi_l)$, where $\alpha_\psi : M \to [0, \infty)$. More specifically, assume that α_ψ describes a one-dimensional *partition model* where the mark space M is partitioned via *Voronoi tesselations* (Voronoi 1908, Green & Sibson 1978). A Voronoi tessellation of M is defined by, say, $k + 2$ *generating points* $C_0 < C_1 < \cdots < C_k < C_{k+1}$ where each generating point $C_i \in M$ defines a region which consists of all points in M closer in Euclidean distance to C_i than to any other generating point. By associating a (constant) height, say $H_i \geq 0$, with the region defined by C_i ($i = 0, 1, \ldots, k, k + 1$), a non-parametric approximation of the mark chemical activity function α is obtained (it is assumed that $H_0 = H_{k+1} = 0$). An example of α_ψ on $M = (0, 10)$ where $k = 4$, $C_0 = 0.5$, $C_1 = 1$, $C_2 = 2$, $C_3 = 4$, $C_4 = 6.5$,

Figure 2: Example of a mark chemical activity function α_ψ when $k = 4$. See text for details.



$C_5 = 8$ and $H_1 = 0.03$, $H_2 = 0.17$, $H_3 = 0.07$, $H_4 = 0.09$ is depicted in Figure 2. The large dots represent the six generating points $C_0, C_1, \ldots, C_4, C_5$ (each respective region is depicted with parentheses). Partition models have been explored in the Statistics literature by Bognar (2005), Heikkinen & Arjas (1998), among others.

Using the above parameterization, the energy function becomes

$$U_{\theta,\psi}(\mathbf{x}, \mathbf{m}) = \sum_{i=1}^{n} \alpha_\psi(m_i) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \phi_\theta(x_i, x_j, m_i, m_j)$$

where $\alpha_\psi(m_i)$ is the height of the region that contains mark m_i (i.e. if m_i is closest to C_j, then $\alpha_\psi(m_i) = H_j$), and ϕ_θ is the mark pair potential function (perhaps of the form (1)). Hence, the marked Gibbs distribution of $(\mathbf{x}, \mathbf{m})$ conditional on (θ, ψ) is

$$p(\mathbf{x}, \mathbf{m}|\theta, \psi) = Z^{-1}(\theta, \psi) \exp[-U_{\theta,\psi}(\mathbf{x}, \mathbf{m})]$$

where $U_{\theta,\psi}$ describes the *total energy* and

$$Z(\theta, \psi) = \int_\Gamma \exp[-U_{\theta,\psi}(\mathbf{x}, \mathbf{m})]\, d\eta(\mathbf{x}, \mathbf{m})$$

is an *intractable* normalizing constant depending on θ and ψ.

If there are k regions in the tessellation α_ψ, then the likelihood becomes

$$L(\theta, \psi) = Z^{-1}(\theta, \psi) \exp\left[-\sum_{i=1}^{n} \alpha_\psi(m_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \phi_\theta(x_i, x_j, m_i, m_j) \right] \quad (3)$$

where $\theta = (\theta_1, \ldots, \theta_m)$, $\psi = (C_0, C_1, \ldots, C_k, C_{k+1}, H_1, \ldots, H_k) \stackrel{\text{def}}{=} (C, H)$, and $Z(\theta, \psi)$ is an intractable function of the parameters (making exact likelihood based inference impossible). It is assumed without loss of generality that θ has fixed, known dimension m. Because k is (typically) unknown, a heirarchical model is implemented with a prior distribution on k, say $p(k)$. With k regions, the external hidden variates (C, H) are assumed to be a priori independent of θ. A priori, C is assumed to follow some point process on M; typically C is assumed to follow a binomial process on M (i.e. a Poisson process conditional on k). The region heights $H_1, \ldots, H_k$ are i.i.d. and are independent of C, a priori; for example $H_1, \ldots, H_k \stackrel{\text{iid}}{\sim} \text{Unif}(0, H_u)$ for some $H_u > 0$, say. Because the process (C, H) is a priori independent of θ, the joint prior is $p(\theta, \psi) = p(\theta, C, H) = p(\theta)p(C, H|k)p(k)$. It may be possible to allow for non-independent heights H, or for a more complex process for the generating points C, if desired.

In contrast to the non-parametric model set forth, it may be possible, a priori, to model the mark chemical activity function with a smooth function, possibly reducing the number of parameters. Although the complexity of the model may be reduced, such an approach may provide for less modeling flexibility.

4 Posterior Simulation

Bayesian inference will be based upon MCMC simulations from the full posterior distribution $p(\theta, \psi|\mathbf{x}, \mathbf{m}) \propto L(\theta, \psi)p(\theta, \psi)$. Since the number of regions in the tessellation k will be dictated by the data, a reversible jump MCMC update (Green 1995) will be used when a region is either added (increasing the dimension of the parameter space by two) or removed (reducing the dimensionality by two). For moves in which the dimension of the parameter space remains unchanged, a standard Metropolis-Hastings update (Metropolis et al. 1953, Hastings 1970) it utilized.

The algorithm begins by arbitrarily choosing the number of regions $k^{(0)}$ for the Voronoi tessellation of M. Then, given $k^{(0)}$, an arbitrary starting value for the sampler can be chosen, say $(\theta^{(t)}, \psi^{(t)}) = (\theta^{(t)}, C^{(t)}, H^{(t)})$ where $t = 0$, $\theta^{(t)} = (\theta_1^{(t)}, \ldots, \theta_m^{(t)})$, $C^{(t)} = (C_0^{(t)}, \ldots, C_{k^{(t)}+1}^{(t)})$, and $H^{(t)} = (H_1^{(t)} \ldots, H_{k^{(t)}}^{(t)})$. The algorithm proceeds by randomly choosing a move type; with probability $a_p + r_p$ the sampler attempts to change the number of regions in the tessel-

lation, and with probability $1 - a_p - r_p$ a non-dimension changing move is chosen.

If the number of regions in the tessellation is to remain unchanged, denote the current parameter vector $(\theta^{(t)}, \psi^{(t)})$ by $\zeta^{(t)} = (\zeta_1^{(t)}, \ldots, \zeta_{m+2k^{(t)}+2}^{(t)})$. Randomly choose (independent of $\zeta^{(t)}$) an element, say $\zeta_i^{(t)}$, to update (each element need not have equal probability of being chosen). Draw a candidate value, say ζ_i^*, from some proposal distribution $q_i(\zeta_i^*|\zeta^{(t)})$ and define $\zeta^* = (\zeta_1^{(t)}, \ldots, \zeta_{i-1}^{(t)}, \zeta_i^*, \zeta_{i+1}^{(t)}, \ldots, \zeta_{m+2k^{(t)}+2}^{(t)})$. Accept the proposal, that is let $\zeta^{(t+1)} = \zeta^*$, with probability

$$\begin{aligned}
\alpha &= \min\left[1, \frac{p(\zeta^*|\mathbf{x},\mathbf{m})}{p(\zeta^{(t)}|\mathbf{x},\mathbf{m})}\frac{q_i(\zeta_i^{(t)}|\zeta^*)}{q_i(\zeta_i^*|\zeta^{(t)})}\right] \\
z &= \min\left[1, \frac{L(\zeta^*)}{L(\zeta^{(t)})}\frac{p(\zeta^*)}{p(\zeta^{(t)})}\frac{q_i(\zeta_i^{(t)}|\zeta^*)}{q_i(\zeta_i^*|\zeta^{(t)})}\right] \\
&= \min\left[1, \frac{\exp[-U_{\zeta^*}(\mathbf{x},\mathbf{m})]Z(\zeta^{(t)})}{\exp[-U_{\zeta^{(t)}}(\mathbf{x},\mathbf{m})]Z(\zeta^*)}\frac{p(\zeta^*)}{p(\zeta^{(t)})}\frac{q_i(\zeta_i^{(t)}|\zeta^*)}{q_i(\zeta_i^*|\zeta^{(t)})}\right], \qquad (4)
\end{aligned}$$

otherwise reject ζ^* and let $\zeta^{(t+1)} = \zeta^{(t)}$. Increment t and randomly choose another move type.

Regions in the tessellation are added and removed with probability a_p and r_p, respectively. Suppose there are $k^{(t)}$ regions currently in the tessellation. To add a region, choose u_1, say, uniformly in $(C_0^{(t)}, C_{k^{(t)}+1}^{(t)})$ and let $C' = u_1$. Then, choose u_2, say, from some proposal density $q(u_2|\zeta^{(t)})$ and let $H' = u_2$. Letting $\zeta^* = (\zeta^{(t)}, C', H')$, accept the new region, that is let $\zeta^{(t+1)} = \zeta^*$, with probability (see Green 1995, for details)

$$\alpha = \min\left[1, \frac{\exp[-U_{\zeta^*}(\mathbf{x},\mathbf{m})]Z(\zeta^{(t)})}{\exp[-U_{\zeta^{(t)}}(\mathbf{x},\mathbf{m})]Z(\zeta^*)}\frac{p(\zeta^*)}{p(\zeta^{(t)})}\frac{r_p(C_{k^{(t)}+1}^{(t)} - C_0^{(t)})}{a_p(k^{(t)}+1)q(H'|\zeta^{(t)})}\ \mathcal{J}_a\right], \qquad (5)$$

where the Jacobian $\mathcal{J}_a$ is

$$\mathcal{J}_a = \left|\frac{\partial\zeta^*}{\partial(\zeta^{(t)}, u_1, u_2)}\right| = \left|\frac{\partial(\zeta^{(t)}, C', H')}{\partial(\zeta^{(t)}, u_1, u_2)}\right| = 1.$$

The removal of a region involves randomly choosing (with equal probability) one of the $k^{(t)}$ regions $C_1^{(t)}, \ldots, C_{k^{(t)}}^{(t)}$. If $C_i^{(t)}$ was chosen, let $\zeta^* = (\theta^{(t)}, C^{(t)} \backslash C_i^{(t)}, H^{(t)} \backslash H_i^{(t)})$. Accept the removal, that is let $\zeta^{(t+1)} = \zeta^*$, with probability

$$\alpha = \min\left[1, \frac{\exp[-U_{\zeta^*}(\mathbf{x},\mathbf{m})]Z(\zeta^{(t)})}{\exp[-U_{\zeta^{(t)}}(\mathbf{x},\mathbf{m})]Z(\zeta^*)}\frac{p(\zeta^*)}{p(\zeta^{(t)})}\frac{a_p k^{(t)} q(H_i^{(t)}|\zeta^*)}{r_p(C_{k^{(t)}+1}^{(t)} - C_0^{(t)})}\ \mathcal{J}_r\right], \qquad (6)$$

where the Jacobian $\mathcal{J}_r$ is the *inverse* of the Jacobian $\mathcal{J}_a$ had we been attempting to add the generating point $C_i^{(t)}$ and respective height $H_i^{(t)}$ to ζ^*. Hence, $\mathcal{J}_r = 1$. For notational and computational convenience, reorder and relabel the generating points in ascending order after executing an $a-$move or $r-$move. For example, suppose the region C' (and mark H') is successfully added in an $a-$move, yielding $\zeta^{(t+1)} = (\zeta^{(t)}, C', H')$. If $C_{i-1}^{(t)} < C' < C_i^{(t)}$, then let $\zeta^{(t+1)} = (\theta^{(t+1)}, \psi^{(t+1)}) = (\theta^{(t+1)}, C^{(t+1)}, H^{(t+1)})$ where $\theta^{(t+1)} = \theta^{(t)}$, $C^{(t+1)} = (C_0^{(t)}, \ldots, C_{i-1}^{(t)}, C', C_i^{(t)}, \ldots, C_{k^{(t)}+1}^{(t)})$, and $H^{(t+1)} = (H_1^{(t)}, \ldots, H_{i-1}^{(t)}, H', H_i^{(t)}, \ldots, H_{k^{(t)}}^{(t)})$.

Because the intractable functions $Z(\zeta^{(t)})$ and $Z(\zeta^*)$ do *not* cancel in the acceptance probabilities (4), (5), and (6), the intractable ratio

$$r \stackrel{\text{def}}{=} \frac{Z(\zeta^{(t)})}{Z(\zeta^*)} \tag{7}$$

must be estimated within every iteration of the sampler.

5 Estimation of the intractable ratio

As motivated by Bognar (2005) and Bognar & Cowles (2004), the intractable ratio r in (7) can be estimated by importance sampling (Smith & Gelfand 1992). Crucial to the implementation of importance sampling is the MCMC algorithm of Geyer & Møller (1994) for simulating (marked) spatial point patterns (*importance samples*) from, say, $p(\mathbf{x}, \mathbf{m}|\zeta') = Z^{-1}(\zeta') \exp[-U_{\zeta'}(\mathbf{x}, \mathbf{m})]$ for any $\zeta' = (\theta', \psi') = (\theta'_1, \ldots, \theta'_m, C'_0, \ldots, C'_{k'+1}, H'_1, \ldots, H'_{k'})$ (where $Z(\zeta')$ is finite). In the literature, $Z^{-1}(\zeta') \exp[-U_{\zeta'}(\mathbf{x}, \mathbf{m})]$ is known as the *importance sampling density*. The algorithm is outlined later in this Section; fortunately, knowledge of $Z(\zeta')$ is not needed.

Suppose importance samples $(\mathbf{x}^{(l)}, \mathbf{m}^{(l)}) = (x_1^{(l)}, \ldots, x_{n^{(l)}}^{(l)}, m_1^{(l)}, \ldots, m_{n^{(l)}}^{(l)})$ $(l = 1, \ldots, L)$ are simulated (after burn-in) from the importance sampling density $p(\mathbf{x}, \mathbf{m}|\zeta')$. Then $r = Z(\zeta^{(t)})/Z(\zeta^*)$ can be estimated by

$$\hat{r} \stackrel{\text{def}}{=} \sum_{l=1}^{L} \frac{\exp[-U_{\zeta^{(t)}}(\mathbf{x}^{(l)}, \mathbf{m}^{(l)})]}{\exp[-U_{\zeta'}(\mathbf{x}^{(l)}, \mathbf{m}^{(l)})]} \left(\sum_{l=1}^{L} \frac{\exp[-U_{\zeta^*}(\mathbf{x}^{(l)}, \mathbf{m}^{(l)})]}{\exp[-U_{\zeta'}(\mathbf{x}^{(l)}, \mathbf{m}^{(l)})]} \right)^{-1}. \tag{8}$$

If the chain is ergodic, then

$$\begin{aligned} \frac{1}{L} \sum_{l=1}^{L} \frac{\exp[-U_{\zeta^{(t)}}(\mathbf{x}^{(l)}, \mathbf{m}^{(l)})]}{\exp[-U_{\zeta'}(\mathbf{x}^{(l)}, \mathbf{m}^{(l)})]} &\stackrel{\text{a.s.}}{\to} \int_\Gamma \frac{\exp[-U_{\zeta^{(t)}}(\mathbf{x}, \mathbf{m})]}{\exp[-U_{\zeta'}(\mathbf{x}, \mathbf{m})]} p(\mathbf{x}, \mathbf{m}|\zeta') \, d\eta(\mathbf{x}, \mathbf{m}) \\ &= \frac{1}{Z(\zeta')} \int_\Gamma \exp[-U_{\zeta^{(t)}}(\mathbf{x}, \mathbf{m})] \, d\eta(\mathbf{x}, \mathbf{m}) \end{aligned}$$

$$= \quad \frac{Z(\zeta^{(t)})}{Z(\zeta')}.$$

The denominator of (8) converges almost surely to $Z(\zeta^*)/Z(\zeta')$, hence $\hat{r} \stackrel{\text{a.s.}}{\to} r$.

A poor choice of importance sampling density $p(\mathbf{x}, \mathbf{m}|\zeta')$ will require many more importance samples to approximate r for any given degree of accuracy. In other words, to accurately estimate r with a computationally *feasible* number of importance samples requires that $p(\mathbf{x}, \mathbf{m}|\zeta')$ is concentrated near $p(\mathbf{x}, \mathbf{m}|\zeta^*)$ *and* $p(\mathbf{x}, \mathbf{m}|\zeta^{(t)})$. Informally speaking, this is accomplished by choosing a ζ' close to both $\zeta^{(t)}$ and ζ^*. Hence, the proposal density should be concentrated near $\zeta^{(t)}$, ensuring that ζ^* is relatively close to $\zeta^{(t)}$, allowing ζ' to be chosen close to both $\zeta^{(t)}$ and ζ^*. A proposal density not concentrated near $\zeta^{(t)}$ can cause poor estimates of r, which can induce poor mixing (as is well known, such a proposal density can lead to low acceptance rates, causing poor mixing). On the other hand, if the proposal density is too highly concentrated about $\zeta^{(t)}$, the sampler will traverse the sample space very slowly and have poor mixing properties. The proposal density, therefore, should be concentrated, but not too concentrated, near $\zeta^{(t)}$ to enable optimal sampler mixing. Clearly, choosing proposal densities is important; see Bognar & Cowles (2004) for more guidance.

Although Geyer & Møller (1994) did not explicitly describe an algorithm for simulating continuously marked spatial point patterns, the generality of their algorithm does allow for such simulations (as noted by the authors). A simple implementation of Geyer and Møller's algorithm for simulating from $p(\mathbf{x}, \mathbf{m}|\zeta')$ is now described. The sampler is initialized by letting $l = 0$ and simulating an unmarked spatial point pattern $\mathbf{x}^{(l)} = (x_1^{(l)}, \ldots, x_{n^{(l)}}^{(l)})$ with, say, $n^{(l)} = n$ points distributed uniformly in V. Then, generate marks $\mathbf{m}^{(l)} = (m_1^{(l)}, \ldots, m_{n^{(l)}}^{(l)})$ (independently of $\mathbf{x}^{(l)}$) uniformly in the interval (C'_{min}, C'_{max}) where $C'_{min} = (C'_0 + C'_1)/2$ and $C'_{max} = (C'_{k'} + C'_{k'+1})/2)$. Within each iteration the sampler randomly chooses between four move types: 1) add a (marked) point to the pattern, 2) remove a point, 3) move a point, and 4) change a mark. Denote the move probabilities by a_p^{imp}, r_p^{imp}, x_p^{imp}, and m_p^{imp} respectively, where $a_p^{imp} + r_p^{imp} + x_p^{imp} + m_p^{imp} = 1$ and (for simplicity) $a_p^{imp} = r_p^{imp}$. To move a point, randomly choose a point (all points are chosen with equal probability), say $x_i^{(l)}$, from the current pattern $\mathbf{x}^{(l)}$, propose a new location x_i^* uniformly in V, and let $\mathbf{x}^* = (x_1^{(l)}, \ldots, x_{i-1}^{(l)}, x_i^*, x_{i+1}^{(l)}, \ldots, x_{n^{(l)}}^{(l)})$. Accept the move, that is let $(\mathbf{x}^{(l+1)}, \mathbf{m}^{(l+1)}) = (\mathbf{x}^*, \mathbf{m}^{(l)})$, with probability $\min[1, \exp[-U_{\zeta'}(\mathbf{x}^*, \mathbf{m}^{(l)})]/\exp[-U_{\zeta'}(\mathbf{x}^{(l)}, \mathbf{m}^{(l)})]]$. To change a mark, randomly choose (with equal probability) a mark, say $m_i^{(l)}$, generate a candidate mark m_i^* uniformly in $(m_i^{(l)} - \epsilon, m_i^{(l)} + \epsilon)$ for some $\epsilon > 0$, and let $\mathbf{m}^* = (m_1^{(l)}, \ldots, m_{i-1}^{(l)}, m_i^*, m_{i+1}^{(l)}, \ldots, m_{n^{(l)}}^{(l)})$. Accept the move with probability $\min[1, \exp[-U_{\zeta'}(\mathbf{x}^{(l)}, \mathbf{m}^*)]/\exp[-U_{\zeta'}(\mathbf{x}^{(l)}, \mathbf{m}^{(l)})]]$. To add a marked

point to the pattern, generate a new point x^* uniformly in V, choose a mark m^* uniformly in (C'_{min}, C'_{max}), and let $(\mathbf{x}^{(l+1)}, \mathbf{m}^{(l+1)}) = (\mathbf{x}^*, \mathbf{m}^*) = (\mathbf{x}^{(l)}, x^*, \mathbf{m}^{(l)}, m^*)$ with probability $\min[1, (\exp[-U_{\zeta'}(\mathbf{x}^*, \mathbf{m}^*)]\lambda(V))/(\exp[-U_{\zeta'}(\mathbf{x}^{(l)}, \mathbf{m}^{(l)})](n^{(l)}+1))]$. To remove a point, randomly choose a point (with equal probability), say $x_i^{(l)}$. Accept the removal of the marked point, that is let $(\mathbf{x}^{(l+1)}, \mathbf{m}^{(l+1)}) = (\mathbf{x}^*, \mathbf{m}^*) = (\mathbf{x}^{(l)} \backslash x_i^{(l)}, \mathbf{m}^{(l)} \backslash m_i^{(l)})$, with probability $\min[1, (\exp[-U_{\zeta'}(\mathbf{x}^*, \mathbf{m}^*)]n^{(l)})/(\exp[-U_{\zeta'}(\mathbf{x}^{(l)}, \mathbf{m}^{(l)})]\lambda(V))]$. After each update, randomly choose another move type and repeat. Note that it is possible to informally observe convergence (and determine an appropriate burn-in period) of the MCMC sampler by tracking number of points $n^{(l)}$ and the total energy $U_{\zeta'}(\mathbf{x}^{(l)}, \mathbf{m}^{(l)})$ over time l.

6 Analysis of the Norway spruce dataset

The Norway spruce marked spatial point pattern $(\mathbf{x}, \mathbf{m}) = (x_1, \ldots, x_{134}, m_1, \ldots, m_{134})$, depicted in Figure 1, was observed in a 56×38 meter plot of forest V. Since spatial regularity appears to be present, and since there is likely a "minimum" inter-tree distance needed to sustain life, the mark pair potential function (1) was utilized to model the spatial interaction. For simplicity, the hard-core interaction distance b_{hc} was fixed at 1.044031 meters $(= \min_{i \neq j} \|x_i - x_j\|)$ yielding $\theta = (b, h, d)$. The interaction distance b describes the distance at which pairs of trees cease to interact, the Straussian parameter h (assume $h > 0$ to ensure that Z is finite) describes the strength of interaction (inhibition), and the scaling parameter d describes how b and b_{hc} scale according to the attached marks. As before, a one-dimensional Voronoi tessellation α_ψ (described by the external process (C, H)) is used to model the mark chemical activity function α where it is assumed that the mark space $M = (0, 50)$.

6.1 Prior specification

Given k, there are $2k+5$ model parameters; the mark chemical activity function parameters $\psi = (C, H) = (C_0, C_1, \ldots, C_k, C_{k+1}, H_1, \ldots, H_k)$ and the mark pair potential function parameters $\theta = (b, h, d)$. The priors were specified as follows: $b \sim \text{Unif}(1.044031 = b_{min}, 10 = b_{max})$, $h \sim \text{N}(\mu_h = 1, \sigma_h = 100)$ (truncated to $\mathcal{R}^+$), $d \sim \text{Unif}(-1 = d_{min}, 5 = d_{max})$, $k \sim \text{Pois}(\lambda_k = 5)$ (truncated to $1, 2, \ldots$), $C_0, C_1, \ldots, C_k, C_{k+1} | k \overset{\text{iid}}{\sim} \text{Unif}(0 = C_{min}, 50 = C_{max})$, and $H_1, \ldots, H_k | k \overset{\text{iid}}{\sim} \text{Unif}(0 = H_{min}, 2 = H_{max})$.

The prior specification on the heirarchical part of the model effects the amount of smoothing of the mark chemical activity function. Suppressing the number of regions k, or decreasing the variance of the H_i's, or both, will

produce stronger smoothing. As described by Bognar (2005) (in a different setting), placing very diffuse priors on the H_i's can *also* produce strong smoothing (the same phenomenon applies here). Hence, the *least* amount of smoothing occurs with a *moderately* diffuse prior on the H_i's and with a λ_k that encourages multiple tessellation regions. Interestingly, Green (1995) noted a similar type of phenomenon in his analysis of a one-dimensional multiple change point problem, as did Bognar (2005) in an analysis of a spatially inhomogeneous Gibbsian point process.

6.2 MCMC details

To perform inference, we seek to simulate from the full posterior distribution $p(\theta, \psi|\mathbf{x}, \mathbf{m})$. To simplify the construction of proposal densities, only one component of the parameter vector was updated at a time. At the beginning of each iteration, a move type was randomly chosen: update the interaction distance ($b-$move), update the Straussian parameter ($h-$move), update the scaling parameter ($d-$move), move a randomly chosen generating point ($C-$move), update the height of a region ($H-$move), add a region to the tessellation ($a-$move), and remove a region from the tessellation ($r-$move). The move probabilities were, respectively, $b_p = h_p = d_p = 0.10$, $C_p = H_p = 0.15$, and $a_p = r_p = 0.20$.

At the beginning of iteration t, suppose the current parameter vector is $\zeta^{(t)} = (\theta^{(t)}, \psi^{(t)})$. If a $b-$move is selected, obtain a candidate interaction distance b^* from a Unif$(b^{(t)} - 0.5, b^{(t)} + 0.5)$ distribution and let $\zeta^* = (b^*, h^{(t)}, d^{(t)}, \psi^{(t)})$. The intractable ratio $r = Z(\zeta^*)/Z(\zeta^{(t)})$ in the Metropolis-Hastings acceptance probability (4) must be estimated before accepting/rejecting the candidate. To this end, $L = 5{,}000$ point patterns (following a 2,000 iteration burn-in) were generated from $p(\mathbf{x}, \mathbf{m}|\zeta')$, where $\zeta' = (0.5[b^* + b^{(t)}], h^{(t)}, d^{(t)}, \psi^{(t)}) = 0.5[\zeta^* + \zeta^{(t)}]$, using the algorithm of Geyer & Møller (1994). The authors noted that the algorithm mixes most rapidly for large a_p^{imp} and r_p^{imp}; hence $x_p^{imp} = m_p^{imp} = 0.15$ and $a_p^{imp} = r_p^{imp} = 0.35$ were used. The simulated marked spatial point patterns $(\mathbf{x}^{(l)}, \mathbf{m}^{(l)})$ $(l = 1, \ldots, L)$ from $p(\mathbf{x}, \mathbf{m}|\zeta')$ were used to estimate r via (8), the estimate was plugged into the acceptance probability (4), and the candidate b^* was accepted/rejected accordingly. After incrementing t, another move type was randomly chosen and executed. If $b^* \notin (b_{min} = 1.044031, b_{max} = 10)$, then the candidate b^* was rejected via the prior.

The last simulated post burn-in pattern, $(\mathbf{x}^{(L)}, \mathbf{m}^{(L)})$, was used to initialize Geyer and Møller's algorithm in the subsequent iteration, allowing a shorter (2,000 iteration) burn-in period. Since the values of ζ' experience relatively little change from iteration to iteration, initializing the algorithm with $(\mathbf{x}^{(L)}, \mathbf{m}^{(L)})$ allows for a shorter burn-in period than starting from, say,

a binomial process with uniform marks in (C_{min}, C_{max}). Trace plots of n and the total energy $U_{\zeta'}$ over time l confirm the feasibility of this computational shortcut.

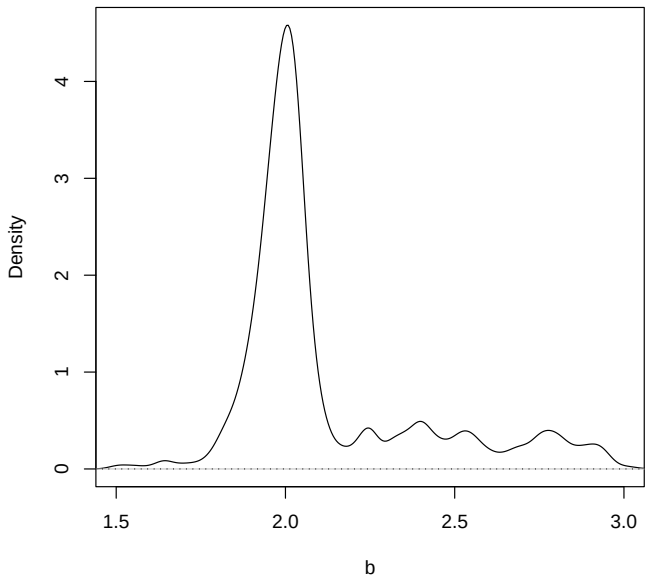
The $h-$move, $d-$move, $C-$move, and $H-$move proceeded similarly. For an $h-$move, h^* was drawn from a Unif$(h^{(t)} - 0.75, h^{(t)} + 0.75)$ distribution, while d^* was drawn from a Unif$(d^{(t)} - 0.75, d^{(t)} + 0.75)$ distribution in a $d-$move. For a $C-$move, a generating point, say $C_i^{(t)}$, was randomly chosen, and a candidate location C_i^* was drawn from a Unif$(C_i^{(t)} - 2.5, C_i^{(t)} + 2.5)$ distribution conditional on $C_i^* \in (C_{i-1}^{(t)}, C_{i+1}^{(t)})$ (i.e. if $C_i^* \notin (C_{i-1}^{(t)}, C_{i+1}^{(t)})$ then the candidate was rejected via the proposal). For an $H-$move, one of the current heights, say $H_i^{(t)}$, was randomly chosen, and a candidate height H_i^* was drawn from a Unif$(H_i^{(t)} - 0.075, H_i^{(t)} + 0.075)$ distribution. In each case, the importance sampling locale was $\zeta' = 0.5[\zeta^{(t)} + \zeta^*]$. Note that if $h^* < 0$, $d^* \notin (-1, 5)$, or $H_i^* \notin (0, 2)$, then the candidate was rejected by the prior. The Metropolis-Hastings acceptance rates were 20.3%, 50.9%, 40.0%, 58.9%, and 51.4% for a $b-$move, $h-$move, $d-$move, $C-$move, and $H-$move respectively.

An $a-$move proceeded by choosing a new generating point C' uniformly in (C_{min}, C_{max}), drawing a candidate height H' from $q(H'|\zeta^{(t)}) \sim$ Unif$(\alpha_{\psi^{(t)}}(C') - 0.05, \alpha_{\psi^{(t)}}(C') + 0.05)$ (i.e. propose a height uniformly within 0.05 of the current height at C'), generating $L =$ 5,000 point patterns from $p(\mathbf{x}, \mathbf{m}|\zeta')$ where $\zeta' = (\theta^{(t)}, C^{(t)}, H^{(t)})$, estimating r via (8), plugging the estimate into (5), and accepting the new region with probability (5). If $H' \notin (0, 2)$, then the candidate was rejected via the prior. An $r-$move proceeded as in Section 4 where r was estimated via (8) using $\zeta' = (\theta^{(t)}, C^{(t)} \backslash C_i^{(t)}, H^{(t)} \backslash H_i^{(t)})$. If an $r-$move was attempted when $k^{(t)} = 1$, the move was rejected by the prior and $\zeta^{(t+1)} = \zeta^{(t)}$.

Six separate samplers were run in parallel on six Intel Xeon 2.4GHz workstations running Linux. With the coding was in C++, each machine executed 25,000 posterior iterations, including a 5,000 iteration burn-in (the post burn-in iterations were combined). It took approximately 31 hours to perform the 150,000 updates (approximately 186 total computing hours). The Bayesian Output Analysis (BOA) software (Smith 2001) was used to analyze the 120,000 post burn-in iterations. Trace plots of b, h, and d show good mixing behavior: the Gelman and Rubin (Gelman & Rubin 1992, Brooks & Gelman 1998) corrected scale reduction factors for b, h, and d were 1.010, 1.001, and 1.015 respectively. Because of the changing dimensionality, evaluating mixing and convergence of the external variates C and H is more difficult (see Brooks & Giudici 1999, for current work in this area).

Because the intractable ratio r in the acceptance probability is *estimated*, this implies that the established theory for MCMC samplers does not ap-

Figure 3: Marginal posterior distribution of the interaction distance b.



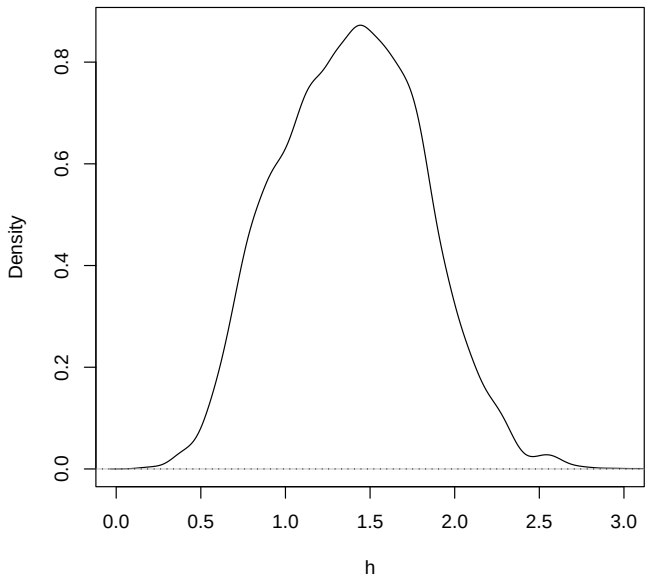
ply. Nevertheless, based upon anecdotal evidence from simulation studies (not described herein), the aforementioned approximation does not appear to prevent the sampler from converging to a distribution that closely *resembles* the true posterior. From a practical point of view, this is probably of little consequence.

6.3 Results

The estimated marginal posterior distribution of the interaction distance b is displayed in Figure 3. The posterior mean of b is approximately 2.123 meters (Monte Carlo error 0.008), with median and mode of 2.017 and 2.008 meters respectively. The 95% equal tail interval is $(1.810, 2.874)$ meters. The jagged nature of the marginal posterior is due to ϕ_θ, and hence the marginal posterior distribution of b, $p(b|\mathbf{x}, \mathbf{m})$, being discontinuous (Bognar & Cowles 2004). In short, the spruce trees cease to interact with one another at a distance of (approximately) two meters.

Figure 4 displays the estimated marginal posterior distribution of the Straussian parameter h. The estimated posterior mean, median, and mode of

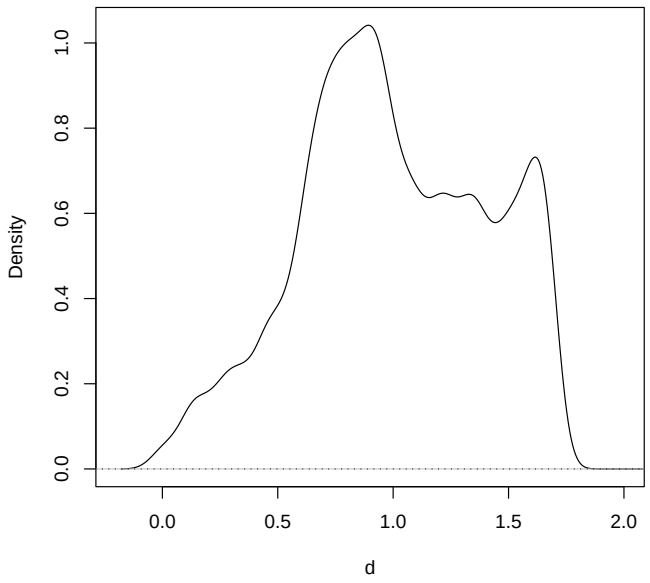
Figure 4: Marginal posterior distribution of the Straussian parameter h.



h are 1.386 (Monte Carlo error 0.010), 1.393, and 1.445 respectively. The 95% equal tail interval is $(0.629, 2.206)$. The amount of uncertainty in h is surprising; however, the ability to reliably obtain interval estimates (and make such an observation) demonstrates the favorability of the current approach.

Figure 5 displays the estimated marginal posterior distribution of the scaling parameter d. The estimated posterior mean, median, and mode of d are 1.002 (Monte Carlo error 0.010), 0.960, and 0.893 respectively. The 95% equal tail interval is $(0.187, 1.675)$. The high posterior probability that d is positive, i.e. $P(d > 0|\mathbf{x}, \mathbf{m}) \simeq 0.997$, indicates that large trees (with large diameter trunks) have larger b_{hc} and b than smaller trees. For instance, recall that the posterior mean of b is 2.123. Now, for trees with a larger than average (i.e. larger than $\bar{m} \simeq 25$) trunk diameter, say $(m_i + m_j)/2 = 30$, then from (2), $e_{ij} = \|x_i - x_j\|[(m_i + m_j)/(2\bar{m})]^{-d} = \|x_i - x_j\|(30/25)^{-1.002} = \|x_i - x_j\|0.833$. This indicates, from (1), inflated interaction and hard-core distances; namely $\phi_\theta(x_i, x_j, m_i, m_j) = h$ when $b_{hc} < e_{ij} \leq b \iff b_{hc}/0.833 < \|x_i - x_j\| \leq b/0.833 \iff 1.200 b_{hc} < \|x_i - x_j\| \leq 1.200 b$. In other words, a pair of trees with an average trunk diameter of 30 cm have a hard-core interaction distance of $1.200 b_{hc} = 1.200(1.044031) = 1.253$ meters and interaction distance of $1.200 b = 1.200(2.123) = 2.548$ meters. Similarly, pairs of trees with a smaller

Figure 5: Marginal posterior distribution of the scaling parameter d.

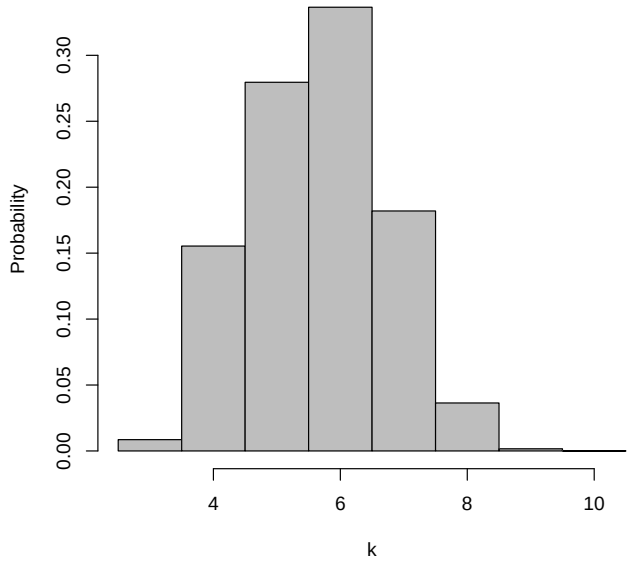


than average trunk diameter, say $(m_i + m_j)/2 = 18$, have hard-core and interaction distances of $0.720(1.044031) = 0.752$ and $0.720(2.123) = 1.529$ meters respectively.

The estimated marginal posterior distribution of the number of regions in the tessellation k is depicted in Figure 6. As alluded to earlier, the prior on the H_i's was not overly diffuse (which suppresses k) *or* informative (which allows little variability in the H_i's), producing a desirable amount of smoothing of the mark chemical activity function (displayed in Figure 7 below).

To estimate the mark chemical activity function, the height of $\alpha_{\psi^{(t)}}$ was recorded, for each t, at 200 equally spaced grid points in $M = (0, 50)$. At each grid point, the (pointwise) posterior mean and 95% equal tail credible set were found from the 120,000 evaluations. Figure 7 displays the empirical mark chemical activity function (as a histogram), the posterior mean (solid line), and 95% credible set (dashed lines) of α_ψ. Usefully, the posterior mean of α_ψ looks like a smoothed version of the empirical mark chemical activity function. In addition, the ability to obtain Bayesian pointwise credible sets allows one to witness, unlike the frequentist approaches, the (large) uncertainty in the point estimates, once again highlighting the additional inferential depth available in the current setting.

Figure 6: Marginal posterior distribution of the number of regions in the Voronoi tessellation k.



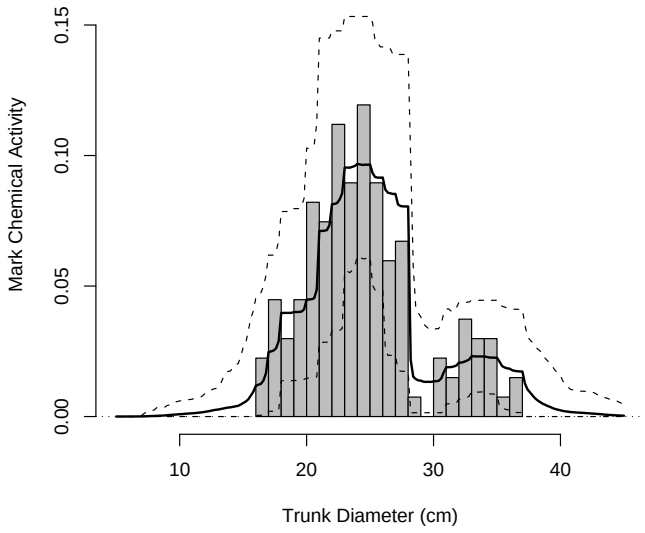
7 Discussion

Although computationally intensive, Bayesian inference for continuously marked Gibbsian point processes is now feasible with careful, efficient coding of the requisite MCMC sampler. A Bayesian approach enables point estimation, but unlike the frequentist techniques where the sampling distributions of the estimates are not well understood (complicating interval estimation), the Bayesian paradigm provides a convenient framework for interval estimation as well as a variety of other inferences.

As suggested by Ogata & Tanemura (1985) (and Goulard et al. 1996), adding the scaling parameter d is a convenient means of incorporating marks into the model. A Bayesian approach allows one to approximate the posterior probability that $d > 0$ (e.g. the posterior probability that large trees have a larger distance of interaction than small trees). In the frequentist paradigm, testing for such scaling remains unclear due to the complex distributional properties of the estimates.

Goulard et al. (1996) suggested that the chemical activity function (which

Figure 7: Estimated posterior mean and 95% pointwise credible set of the mark chemical activity function α_ψ.



regulates the density of points in the pattern) should be a function of the marks. While the authors were able to estimate the mark chemical activity function, the complex distributional properties of the estimates impeded the construction of (pointwise) confidence intervals. A Bayesian approach, on the other hand, allows the construction of (pointwise) credible sets for the mark chemical activity function, enabling an assessment of variability.

The prior on the tessellation region heights H_i $(i = 1, \ldots, k)$ (and on the number of regions k) influences the amount of smoothing of the mark chemical activity function. The prior $p(\theta, \psi)$ on the model parameters, therefore, should not only reflect ones prior knowledge, but the desired amount of smoothing.

References

Aykroyd, R. G. (2002), 'Approximations for Gibbs distribution normalising constants', *Statistics and Computing* **12**, 391–397.

Baddeley, A. & Møller, J. (1989), 'Nearest-neighbor Markov point processes

and random sets', *International Statistical Review* **57**, 57–89.

Baddeley, A. & Turner, R. (2000), 'Practical maximum pseudolikelihood for spatial point patterns', *Australian and New Zealand Journal of Statistics* **42**, 283–322.

Berman, M. & Turner, R. (1992), 'Approximating point process likelihoods with glim', *Applied Statistics* **41**, 31–38.

Besag, J. (1975), 'Statistical analysis of non-lattice data', *The Statistician* **24**, 179–195.

Besag, J., Milne, R. & Zachary, S. (1982), 'Point process limits of lattice processes', *J. Appl. Prob.* **19**, 210–216.

Bognar, M. A. (2005), 'Bayesian inference for spatially inhomogeneous pairwise interacting point processes', *Computational Statistics and Data Analysis* **49**(1), 1–18.

Bognar, M. A. & Cowles, M. K. (2004), 'Bayesian inference for pairwise interacting point processes', *Statistics and Computing* **14**, 109–117.

Brooks, S. & Gelman, A. (1998), 'General methods for monitoring convergence of iterative simulations', *Journal of Computational and Graphical Statistics* **7**, 434–455.

Brooks, S. & Giudici, P. (1999), Convergence assessment for reversible jump MCMC simulations, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds, 'Bayesian Statistics 6', Oxford University Press, pp. 733–742.

Carter, D. S. & Prenter, P. M. (1972), 'Exponential spaces and counting processes', *Z. Wahr. verw. Geb.* **21**, 1–19.

Chen, M.-H. & Shao, Q.-M. (1997), 'On Monte Carlo methods for estimating ratios of normalizing constants', *The Annals of Statistics* **25**, 1563–1594.

Degenhardt, A. (1999), 'Description of tree distribution patterns and their development through marked Gibbs processes', *Biometrika* **74**, 763–770.

Diggle, P. J. (1983), *Statistical Analysis of Spatial Point Patterns*, Academic Press, London.

Diggle, P. J. (2003), *Statistical Analysis of Spatial Point Patterns*, second edn, Arnold.

Diggle, P. J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. & Tanemura, M. (1994), 'On parameter estimation for pairwise interaction point processes', *International Statistical Review* **62**, 99–117.

Diggle, P. J., Gates, D. J. & Stibbard, A. (1987), 'A nonparametric estimator for pairwise interaction point processes', *Biometrika* **74**, 763–770.

Fiksel, T. (1984), 'Estimation of parameterized pair potentials of marked and unmarked Gibbsian point processes', *Elektron. Inform. Kybernet.* **20**, 270–278.

Fiksel, T. (1988), 'Estimation of interaction potentials of Gibbsian point processes'.

Gates, D. J. & Westcott, M. (1986), 'Clustering estimates for spatial point distributions with unstable potentials', *Annals of the Institute of Statistical Mathematics* **38**, 123–135.

Gelman, A. & Meng, X.-L. (1998), 'Simulating normalizing constants: From importance sampling to bridge sampling to path sampling', *Statistical Science* **13**, 163–185.

Gelman, A., Roberts, G. O. & Gilks, W. R. (1996), Efficient Metropolis jumping rules, *in* J. M. Bernardo, A. P. Berger, A. P. Dawid & A. F. M. Smith, eds, 'Bayesian Statistics 5', Oxford University Press, pp. 599–607.

Gelman, A. & Rubin, D. B. (1992), 'Inference from iterative simulation using multiple sequences (with discussion)', *Statistical Science* **7**, 457–511.

Geyer, C. J. (1999), Likelihood inference for spatial point processes, *in* O. E. Barndorff-Nielsen, W. S. Kendall & M. N. M. van Lieshout, eds, 'Stochastic Geometry: Likelihood and Computation', Chapman and Hall/CRC, London, pp. 79–140.

Geyer, C. J. & Møller, J. (1994), 'Simulation procedures and likelihood inference for spatial point processes', *Scandinavian Journal of Statistics* **21**, 359–373.

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996), *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Goulard, M., Särkkä, A. & Grabarnik, P. (1996), 'Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method', *Scandinavian Journal of Statistics* **23**, 365–379.

Green, P. J. (1995), 'Reversible jump Markov chain Monte Carlo computation and Bayesian model determination', *Biometrika* **82**, 711–732.

Green, P. J. & Sibson, R. (1978), 'Computing dirichlet tesselations in the plane', *The Computer Journal* **21**, 168–173.

Harkness, R. D. & Isham, V. (1983), 'A bivariate spatial point pattern of ants' nests', *Applied Statistics* **32**, 293–303.

Hastings, W. K. (1970), 'Monte Carlo sampling methods using Markov chains and their applications', *Biometrika* **57**, 97–109.

Heikkinen, J. & Arjas, E. (1998), 'Non-parametric Bayesian estimation of a spatial Poisson intensity', *Scandinavian Journal of Statistics* **25**, 435–450.

Heikkinen, J. & Penttinen, A. (1999), 'Bayesian smoothing in the estimation of the pair potential function of Gibbs point patterns', *Bernoulli* **5**, 1119–1136.

Jensen, E. B. V. & Nielsen, L. S. (2001), A review on inhomogeneous spatial point processes, *in* I. V. Basawa & C. C. Heyde, eds, 'Selected Proceedings of the Symposium on Inference for Stochastic Processes', Vol. 37, IMS Lecture notes, pp. 297–318.

Jensen, J. L. & Møller, J. (1991), 'Pseudolikelihood for exponential family models of spatial point processes', *Annals of Applied Probability* **3**, 445–461.

Kelly, F. P. & Ripley, B. D. (1976), 'On Strauss' model for clustering', *Biometrika* **63**, 357–360.

Kendall, W. S. & Møller, J. (2000), 'Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes', *Advances in Applied Probability* **32**, 844–865.

Mateu, J. & Montes, F. (2001), 'Likelihood inference for Gibbs processes in the analysis of spatial point patterns', *International Statistical Review* **69**, 81–104.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953), 'Equations of state calculations by fast computing machines', *Journal of Chemical Physics* **21**, 1087–1091.

Møller, J. (1999), Markov chain Monte Carlo and spatial point processes, *in* O. E. Barndorff-Nielsen, W. S. Kendall & M. N. M. van Lieshout, eds, 'Stochastic Geometry: Likelihood and Computation', Chapman and Hall/CRC, London, pp. 141–172.

Møller, J. & Nicholls, G. K. (2004), 'Perfect simulation for sample-based inference'. Statistics and Computing, to appear.

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

Ogata, Y. & Tanemura, M. (1981), 'Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure', *Annals of the Institute of Statistical Mathematics* **33**, 315–338.

Ogata, Y. & Tanemura, M. (1984), 'Likelihood analysis of spatial point patterns', *Journal of the Royal Statistical Society B* **46**, 496–518.

Ogata, Y. & Tanemura, M. (1985), 'Estimation of interaction potentials of marked spatial point patterns through the maximum likelihood method', *Biometrics* **41**, 421–433.

Ogata, Y. & Tanemura, M. (1986), Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns, *in* I. S. Francis, B. F. J. Manly & F. C. Lam, eds, 'Proceedings of the Pacific Statistical Congress', Elsevier, Amsterdam, pp. 150–154.

Ogata, Y. & Tanemura, M. (1989), 'Likelihood estimation of soft-core interaction potentials for Gibbsian point patterns', *Annals of the Institute of Statistical Mathematics* **41**, 583–600.

O'Sullivan, D. & Unwin, D. J. (2003), *Geographic Information Analysis*, Wiley, Hoboken, New Jersey.

Penttinen, A. K. (1984), *Modeling interaction in spatial point patterns: Parameter estimation by the maximum likelihood method*, Jyväskylä Studies in Computer Science, Economics and Statistics 7.

Propp, J. G. & Wilson, D. B. (1996), 'Exact sampling with coupled Markov chains and applications to statistical mechanics', *Random Structures and Algorithms* **9**, 223–252.

Ripley, B. D. (1977), 'Modeling spatial patterns (with discussion)', *Journal of the Royal Statistical Society B* **39**, 172–212.

Ripley, B. D. (1981), *Spatial Statistics*, John Wiley and Sons, New York.

Ripley, B. D. (1988), *Statistical Inference for Spatial Point Processes*, Cambridge University Press.

Ripley, B. D. & Kelly, F. P. (1977), 'Markov point processes', *J. Lond. Math. Soc.* **15**, 188–192.

Smith, A. F. M. & Gelfand, A. E. (1992), 'Bayesian statistics without tears: A sampling resampling perspective', *The American Statistician* **46**, 84–88.

Smith, B. J. (2001), 'Bayesian Output Analysis (BOA) software'. Copyright (c) 2001 Brian J. Smith, http://www.public-health.uiowa.edu/boa.

Stoyan, D. & Penttinen, A. (2000), 'Recent applications of point process methods in forestry statistics', *Statistical Science* **15**, 61–78.

Stoyan, D. & Stoyan, H. (1998), 'Non-homogeneous Gibbs process models for forestry − a case study', *Biometrical Journal* **40**, 521–531.

Strand, L. (1972), A model for stand growth, *in* 'IUFRO Third Conference Advisory Group of Forest Statisticians', Institut National de la Recherche Agronomique, Paris, pp. 207–216.

Strauss, D. (1975), 'A model for clustering', *Biometrika* **62**, 467–475.

Takacs, R. (1986), 'Estimator for the pair-potential of a Gibbsian point process', *Math. Operationsf. Statist. Ser. Statist.* **17**, 429–433.

van Lieshout, M. N. M. (2000), *Markov Point Processes and Their Applications*, Imperial College Press, London.

Venables, W. N. & Ripley, B. D. (1994), *Modern Applied Statistics with S-Plus*, Springer-Verlag, New York.

Voronoi, M. G. (1908), 'Nouvelles applications des paramèters continus à la théorie des formes quadratiques', *J. Reine Angew. Math.* **134**, 198–287.

Author Affiliation

Matthew A. Bognar
241 Schaeffer Hall
University of Iowa
Iowa City, IA 52242

matthew-bognar@uiowa.edu

Phone: (319)335-0712
Phone: (319)337-3474
Fax: (319)335-3017