Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Computational Statistics & Data Analysis 49 (2005) 1–18
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Bayesian inference for spatially inhomogeneous pairwise interacting point processes

Matthew A. Bognar*,[1]

Department of Statistics and Actuarial Science, University of Iowa, 241 Schaeffer Hall, Iowa City, Iowa 52242, USA

Received 31 December 2002; received in revised form 12 April 2004; accepted 14 April 2004

Abstract

Spatial point patterns are frequently modeled with pairwise interacting point processes. Unfortunately, inference in these models is complicated by the presence of an intractable function of the parameters in the likelihood. Because of the relative computational simplicity, *frequentist* inference in pairwise interacting point processes has dominated the literature. However, a Bayesian approach has not been computationally feasible until recently. Since the Metropolis–Hastings acceptance probability contains a ratio of two likelihoods evaluated at differing parameter values, the resulting intractable *ratio* complicates the required application of MCMC. In this article, we describe how to obtain *Bayesian* inferences without conditioning on the number of points in the pattern, allowing the modeling of spatial inhomogeneity in the density of points. After describing our importance sampling within MCMC algorithm, we analyze the well-known Irish drumlin data set using a hard-core Straussian model.

Keywords: Bayesian inference; Spatial point pattern; Gibbs point process; Inhomogeneous pairwise interacting point process; Metropolis–Hastings algorithm; Reversible jump MCMC; Importance sampling

1. Introduction

The spatial location of events in a region constitutes a *spatial point pattern*. A spatial point pattern may describe, for example, the location of trees in a forest, the location of ant nests in a field, or the location of amacrine cells in the eye. A well-known example of a spatial point pattern is depicted in Fig. 1; the data (O'Sullivan and

* Tel.: +319-335-0799; fax: +319-335-3017.
E-mail address: mbognar@stat.uiowa.edu (M.A. Bognar).
[1] **In memory of Alison Huisenga**.

0167-9473/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2004.04.008

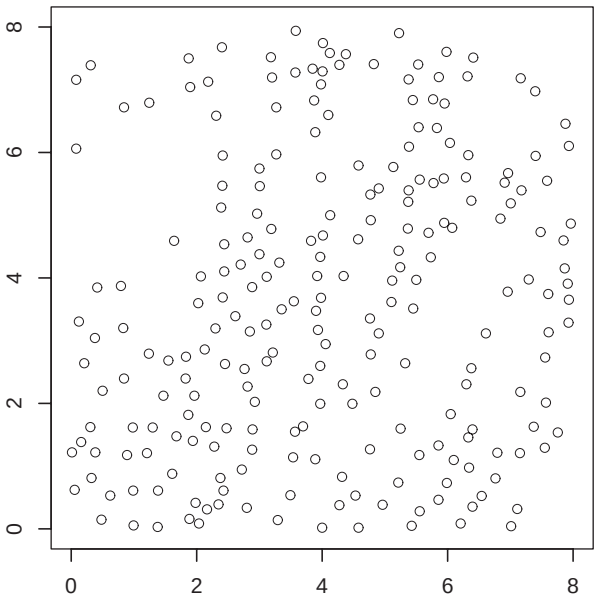

Fig. 1. Location of $n = 232$ drumlin in an 8×8 km^2 field.

Unwin, 2003) describe the location of 232 drumlin (glacial drift) in an 8×8 km^2 field. The data set suggests that the spatial locations of drumlin can exhibit regularity (inhibition) in their spacing. In other words, there is *interaction* between individuals, and our goal is to describe the underlying spatial structure. Unfortunately, the Poisson process (with known normalizing constant) is unable to account for such interaction. *Gibbs point processes*, which *can* account for interaction between points, involve an *intractable* normalizing constant which complicates inference (such is the case in the sequel). For more information, see Diggle (2003), Møller and Waagepetersen (2003), and van Lieshout (2000).

Gibbs point processes, originally used in statistical physics to study the interaction of particles in fluids and gases, are commonly used to model spatial point patterns in bounded regions. The simplest and most common form of a Gibbs point process is a *pairwise interacting point process* (Ripley, 1977). Motivated by the observation that the interaction between individuals may depend upon the distance between them, a pairwise interacting point process describes the interaction between *pairs* of points by a function of (typically) the interpoint Euclidean distance called a *pair potential function*. Some examples of pairwise interacting point processes can be found in Baddeley and Turner (2000), Diggle (2003), Diggle et al. (1987, 1994), Harkness and Isham (1983), Heikkinen and Penttinen (1999), Mateu and Montes (2001), Ogata and Tanemura (1981, 1984, 1986, 1989), Penttinen (1984), and Stoyan and Stoyan (1998).

Numerous frequentist techniques have been developed to allow inference in pairwise interacting point processes: approximate maximum likelihood (Ogata and Tanemura, 1981, 1984, 1986), maximum pseudo-likelihood (Besag et al., 1982), Monte Carlo likelihood (Geyer, 1999), among other methods (e.g. Diggle et al., 1987; Fiksel, 1984; Takacs, 1986). While point estimation is routine, the distributional properties of the estimates are not well understood; thus, parametric bootstrap methods are commonly used for interval estimation (although it is unclear if this is a sound practice). Also, some techniques yield unbiased estimates for patterns with weak to moderate interaction, but yield-biased estimates when there is strong inhibition (or vice versa). Researchers must take care when choosing a technique for analysis.

A Bayesian model has been comparatively underrepresented in the literature. Heikkinen and Penttinen (1999) described a non-parametric estimator of the pair potential function based on a Bayesian smoothing technique. Modeling the pair potential function as a step function, and utilizing the Marquardt algorithm, the authors developed a Monte Carlo algorithm which provides for estimation of the posterior mode. It appears as though interval estimation is more challenging, however. Bognar and Cowles (2004) demonstrated 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. Additionally, they demonstrated in a simulation study that the estimates appear to be approximately unbiased, in a frequentist sense, for any amount of interaction (the coverage appeared correct as well).

In this article, we will (1) not condition on the number of points in the pattern, and (2) allow for spatial inhomogeneity in the density of points. The latter constitutes the main theme and contribution of this article. While it is often claimed that number of points in the pattern provides little information about the interaction structure (Gates and Westcott, 1986; Ripley, 1988), we have found that the differences in a Bayesian context can be noteworthy.

This paper begins with an introduction to Gibbs and pairwise interacting point processes in Section 2. In Section 3, a Bayesian model is proposed; specifically, the likelihood (for inhomogeneous pairwise interacting point processes) and prior are described. The goal is to perform inference by simulating, using the MCMC algorithm outlined in Section 4, from the posterior distribution. As alluded to earlier, because the model contains an intractable normalization constant, the MCMC algorithm for posterior simulation is of heightened complexity; Section 5 addresses this ramification. Finally, the Irish drumlin data set is analyzed in Section 6, and concluding remarks are provided in Section 7.

2. Pairwise interacting point processes

Consider a point pattern $\mathbf{x} = \{x_1, \ldots, x_n\}$ with n points observed in a bounded region $V \subset \mathscr{R}^d$. Let $(V, \mathscr{B}, \lambda)$ be a measure space where $\mathscr{B}$ is the Borel σ-field on V containing all singletons, λ is Lebesgue measure where $0 < \lambda(V) < \infty$, and the number of points in $B \in \mathscr{B}$ has a Poisson distribution with mean $\lambda(B)$. For $n > 0$, 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 $\Omega = \bigcup_{n=0}^{\infty} \Omega_n$. A point process on V is a random variable on the exponential space $(\Omega, \mathscr{F}, \mu)$ (Carter and Prenter, 1972) where $\mathscr{F}$ is a σ-field on Ω, and μ is the homogeneous Poisson process with intensity measure λ. Thus, for $F \in \mathscr{F}$,

$$\mu(F) = \exp[-\lambda(V)] \left[I(\emptyset \in F) + \sum_{n=1}^{\infty} \left(\frac{1}{n!} \int_{V^n} I(\{x_1, \ldots, x_n\} \in F) \prod_{i=1}^{n} \mathrm{d}\lambda(x_i) \right) \right].$$

A Gibbs point process is any random variable on $(\Omega, \mathscr{F}, \mu)$ having a density p with respect to μ. The density is usually specified as $p(\mathbf{x}) = \exp[-U(\mathbf{x})]/Z$ where $U(\mathbf{x})$ is an *energy function* and $Z = \int_{\Omega} \exp[-U(\mathbf{x})]\, \mathrm{d}\mu(\mathbf{x})$ is a *partition function* which normalizes p.

We consider a family of pair potential functions $\{\phi_\theta(s) : \theta \in \Theta\}$, of Euclidean distance s, indexed by a parameter vector $\theta = (\theta_1, \ldots, \theta_m)$. The non-negative *chemical activity* function ξ will allow the modeling of trends in the density of points ($\xi : x \to [0, \infty)$ for all $x \in V$). A pairwise interacting point process is a special case of a Gibbs point process where the energy function U is completely determined by the chemical activity and pair potential. Letting $\|\cdot\|$ denote Euclidean distance, the probability density function (with respect to the homogeneous spatial Poisson process of rate 1) is

$$p(\mathbf{x}|\theta, \xi) = Z^{-1}(\theta, \xi) \exp \left[-\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \phi_\theta(\|x_i - x_j\|) \right] \prod_{i=1}^{n} \xi(x_i)$$

$$\stackrel{\text{def}}{=} Z^{-1}(\theta, \xi) g(\mathbf{x}|\theta) \prod_{i=1}^{n} \xi(x_i),$$

where

$$Z(\theta, \xi) = \int_{\Omega} g(\mathbf{x}|\theta) \prod_{i=1}^{n} \xi(x_i)\, \mathrm{d}\mu(\mathbf{x})$$

is an *intractable* normalizing constant which depends on θ and ξ. The model, which employs type I inhomogeneity (Jensen and Nielsen, 2001), has been studied by Baddeley and Turner (2000), Ogata and Tanemura (1986), and Stoyan and Stoyan (1998).

Kelly and Ripley (1976) observe that $\phi_\theta < 0$ may produce an infinite normalizing constant Z and cause simulation difficulties (Gates and Westcott, 1986; Møller, 1999). If $\phi_\theta(s) \geqslant 0$ for all $s > 0$, or if $\phi_\theta(s) = \infty$ for all s less than some positive *hard-core distance*, then $Z(\cdot) < \infty$ (Diggle et al., 1987). In this article, we focus on inference in patterns with spatial inhibition (where $\phi_\theta(s) \geqslant 0$ for all $s > 0$).

3. Bayesian model

We employ a *partition model* for modeling the chemical activity function ξ. A partition model partitions the space V into k disjoint regions via, for example, *Voronoi*

tesselations (Voronoi, 1908; Green and Sibson, 1978). A Voronoi tessellation of V is given by k *generating points* $C_1,\ldots,C_k \in V$, where the ith region (sometimes called a tile) consists of all points in V closer in Euclidean distance to C_i than to any other generating point. By associating a height, say H_i ($\geqslant 0$), with the ith tile ($i=1,\ldots,k$), we can approximate the chemical activity surface ξ (note that H_i is *constant* over its respective tile). As noted by the Referee, we can easily avoid negative tile heights by letting $\xi(x)=\exp[H(x)]$ where $H : x \to \mathscr{R}$ for all $x \in V$. We use the former parameterization in what follows, however.

It may be possible, a priori, to model the chemical activity function via some smooth parametric function, if desired. Such an approach could provide for less complicated models, but may yield less flexibility than our aforementioned non-parametric technique.

Assume, without loss of generality, that $\theta=(\theta_1,\ldots,\theta_m)$ has known dimension. Because the number of tiles k is (typically) unknown, we construct a hierarchical model with a prior distribution on k, say $p(k)$. Given k, the external hidden variates $(C,H) \stackrel{\text{def}}{=} (C_1,\ldots,C_k,H_1,\ldots,H_k)$ are a priori independent of θ, and are defined by the researcher, in practice. A priori, C is assumed to follow some point process; typically, given k, C will follow a binomial process in V. A priori, the tile heights $H_1,\ldots,H_k$ are i.i.d. and are independent of C; for example $H_1,\ldots,H_k$ may be a priori i.i.d. uniform on $(0,H_u)$ for some $H_u>0$. Because the process (C_i,H_i) $(i=1,\ldots,k)$ is independent of statistical model for the point pattern $\mathbf{x}$, the joint prior can be written as $p(\theta,k,C,H)=p(\theta)p(C,H|k)p(k)$ where, for simplicity, we let $p(\theta,C,H) \stackrel{\text{def}}{=} p(\theta,k,C,H)$, notationally ignoring k. Heikkinen and Arjas (1998) applied this hierarchical type of construction in modeling the intensity of an inhomogeneous spatial Poisson process. Although their model assumed no interaction between points, it did allow for non-independent tile heights (this may also be possible for the current setting).

If there are k tiles, we write the likelihood as

$$L(\theta,C,H) \stackrel{\text{def}}{=} Z^{-1}(\theta,C,H)g(\mathbf{x}|\theta)\prod_{i=1}^{n} H(x_i), \tag{1}$$

where $H(x_i)$ denotes the height of the tile that contains the point x_i (i.e. if x_i is closest to C_j, then $H(x_i)=H_j$). Unfortunately, because $Z(\theta,C,H)$ is an intractable function of the parameters, inference based on the exact likelihood is impossible.

4. Posterior simulation

We now describe how to simulate from the full posterior distribution

$$p(\theta,C,H|\mathbf{x}) \propto L(\theta,C,H)p(\theta,C,H).$$

Since the sampler will dictate the number of tiles k, a reversible jump MCMC update (Green, 1995) is needed to allow the sampler to traverse models with differing numbers of tiles. A standard Metropolis–Hastings update (Metropolis et al., 1953; Hastings, 1970) will suffice for moves in which the dimension of the parameter space remains unchanged.

The algorithm begins by choosing an arbitrary number of tiles $k^{(t)}$ followed by an arbitrary starting value

$$(\theta^{(t)}, C^{(t)}, H^{(t)}) = (\theta_1^{(t)}, \ldots, \theta_m^{(t)}, C_1^{(t)}, \ldots, C_{k^{(t)}}^{(t)}, H_1^{(t)}, \ldots, H_{k^{(t)}}^{(t)}),$$

where $t = 0$. A move type is chosen at random; with probability $a_p + r_p$ the sampler attempts to change the number of tiles, and with probability $1 - a_p - r_p$ the sampler chooses a move in which the number of tiles remains unchanged. If the latter is chosen, we can let $\zeta^{(t)} = (\zeta_1^{(t)}, \ldots, \zeta_{m+2k^{(t)}}^{(t)}) \stackrel{\text{def}}{=} (\theta^{(t)}, C^{(t)}, H^{(t)})$ denote the current parameter vector, and randomly choose (independent of $\zeta^{(t)}$) a component to be updated, say $\zeta_i^{(t)}$ (all components are chosen with equal probability). A candidate value ζ_i^* is drawn from some proposal distribution $q_i(\zeta_i^*|\zeta^{(t)})$ yielding the proposed parameter vector $\zeta^* = (\zeta_1^{(t)}, \ldots, \zeta_{i-1}^{(t)}, \zeta_i^*, \zeta_{i+1}^{(t)}, \ldots, \zeta_{m+2k^{(t)}}^{(t)}) \stackrel{\text{def}}{=} (\theta^*, C^*, H^*)$. Accept the candidate value, that is let $\zeta^{(t+1)} = \zeta^*$, with probability

$$\begin{aligned}
\alpha &= \min\left[1, \frac{p(\zeta^*|\mathbf{x})}{p(\zeta^{(t)}|\mathbf{x})} \frac{q_i(\zeta_i^{(t)}|\zeta^*)}{q_i(\zeta_i^*|\zeta^{(t)})}\right] \\
&= \min\left[1, \frac{L(\zeta^*)}{L(\zeta^{(t)})} \frac{p(\zeta^*)}{p(\zeta^{(t)})} \frac{q_i(\zeta_i^{(t)}|\zeta^*)}{q_i(\zeta_i^*|\zeta^{(t)})}\right] \\
&= \min\left[1, \frac{g(\mathbf{x}|\theta^*) \prod_{i=1}^n H^*(x_i) Z(\zeta^{(t)})}{g(\mathbf{x}|\theta^{(t)}) \prod_{i=1}^n H^{(t)}(x_i) Z(\zeta^*)} \frac{p(\zeta^*)}{p(\zeta^{(t)})} \frac{q_i(\zeta_i^{(t)}|\zeta^*)}{q_i(\zeta_i^*|\zeta^{(t)})}\right], \qquad (2)
\end{aligned}$$

otherwise reject the candidate value and set $\zeta^{(t+1)} = \zeta^{(t)}$. Increment t and choose another move type at random.

The sampler attempts to add and remove tiles with probabilities a_p and r_p, respectively. Suppose the current state $\zeta^{(t)} = (\theta^{(t)}, C^{(t)}, H^{(t)})$ has $k^{(t)}$ tiles. To add a single tile, the reversible jump MCMC update proceeds by choosing, say, u_1 uniformly in V and letting the candidate generating point $C_{k^{(t)}+1}^* = u_1$. Then choose u_2 from some proposal density $q(u_2|\zeta^{(t)})$ and let the respective tile height $H_{k^{(t)}+1}^* = u_2$. The candidate vector, $\zeta^* = (\theta^{(t)}, C^{(t)}, C_{k^{(t)}+1}^*, H^{(t)}, H_{k^{(t)}+1}^*) \stackrel{\text{def}}{=} (\theta^*, C^*, H^*)$, is accepted via Green's acceptance probability

$$\alpha = \min\left[1, \frac{\prod_{i=1}^n H^*(x_i) Z(\zeta^{(t)})}{\prod_{i=1}^n H^{(t)}(x_i) Z(\zeta^*)} \frac{p(\zeta^*)}{p(\zeta^{(t)})} \frac{r_p \lambda(V)}{a_p (k^{(t)} + 1) q(H_{k^{(t)}+1}^*|\zeta^{(t)})} \mathscr{J}_a\right]. \qquad (3)$$

Note that Green's acceptance probability (3) contains a Jacobian $\mathscr{J}_a$. Here, the Jacobian of the transformation (see Green, 1995) is

$$\mathscr{J}_a = \left|\frac{\partial \zeta^*}{\partial(\zeta^{(t)}, u_1, u_2)}\right| = \left|\frac{\partial(\theta^{(t)}, C^{(t)}, C_{k^{(t)}+1}^*, H^{(t)}, H_{k^{(t)}+1}^*)}{\partial(\theta^{(t)}, C^{(t)}, u_1, H^{(t)}, u_2)}\right| = 1.$$

To remove a tile, randomly choose one of the $k^{(t)}$ current tiles, say the ith, for removal (all tiles are chosen with equal probability). Denoting the current and proposed parameter vectors as $\zeta^{(t)}=(\theta^{(t)}, C^{(t)}, H^{(t)})$ and $\zeta^{*}=(\theta^{(t)}, C^{(t)}\setminus C_i^{(t)}, H^{(t)}\setminus H_i^{(t)})=(\theta^{*}, C^{*}, H^{*})$, respectively, accept the tile removal with probability

$$\alpha=\min\left[1, \frac{\prod_{i=1}^{n} H^{*}(x_i)Z(\zeta^{(t)})}{\prod_{i=1}^{n} H^{(t)}(x_i)Z(\zeta^{*})}\frac{p(\zeta^{*})}{p(\zeta^{(t)})}\frac{a_p k^{(t)} q(H_i^{(t)}|\zeta^{*})}{r_p \lambda(V)}\mathscr{J}_r\right]. \tag{4}$$

The Jacobian $\mathscr{J}_r$ is the *inverse* of the Jacobian $\mathscr{J}_a$ had we been attempting to add the tile $C_i^{(t)}$ and respective height $H_i^{(t)}$ to $\zeta^{*}=(\theta^{(t)}, C^{(t)}\setminus C_i^{(t)}, H^{(t)}\setminus H_i^{(t)})=(\theta^{*}, C^{*}, H^{*})$. Hence, of course, $\mathscr{J}_r=1$.

Notice that the intractable functions $Z(\zeta^{(t)})$ and $Z(\zeta^{*})$ do *not* cancel in the acceptance probabilities (2)–(4). Hence, the intractable ratio

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

must be estimated within every iteration of the sampler.

5. Estimation of the intractable ratio

Importance sampling (Smith and Gelfand, 1992) has proven to be a useful tool for estimating r. Via the MCMC algorithm described by Geyer and Møller (1994), it is possible to generate spatial point patterns (importance samples) from

$$p(\mathbf{x}|\theta', C', H')\overset{\text{def}}{=}p(\mathbf{x}|\zeta')=Z^{-1}(\zeta')g(\mathbf{x}|\theta')\prod_{i=1}^{n}H'(x_i)$$

for any ζ' where $\phi_{\theta'}\geqslant 0$ *without* knowledge of $Z(\zeta')$ (the algorithm is described below). The importance samples, say $\mathbf{x}^{(l)}=(x_1^{(l)},\ldots,x_{n^{(l)}}^{(l)})$, $l=1,\ldots,L$, will be used to estimate r. Specifically, suppose $\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(L)}$ are realized (after burn-in) from $p(\mathbf{x}|\zeta')$ for some ζ'. We can estimate $r=Z(\zeta^{(t)})/Z(\zeta^{*})$ in (2)–(4) by

$$\hat{r}\overset{\text{def}}{=}\sum_{l=1}^{L}\frac{g(\mathbf{x}^{(l)}|\theta^{(t)})\prod_{i=1}^{n^{(l)}}H^{(t)}(x_i^{(l)})}{g(\mathbf{x}^{(l)}|\theta')\prod_{i=1}^{n^{(l)}}H'(x_i^{(l)})}\left(\sum_{l=1}^{L}\frac{g(\mathbf{x}^{(l)}|\theta^{*})\prod_{i=1}^{n^{(l)}}H^{*}(x_i^{(l)})}{g(\mathbf{x}^{(l)}|\theta')\prod_{i=1}^{n^{(l)}}H'(x_i^{(l)})}\right)^{-1}. \tag{5}$$

If the chain is ergodic, then

$$\begin{aligned}\frac{1}{L}\sum_{l=1}^{L}\frac{g(\mathbf{x}^{(l)}|\theta^{(t)})\prod_{i=1}^{n^{(l)}}H^{(t)}(x_i^{(l)})}{g(\mathbf{x}^{(l)}|\theta')\prod_{i=1}^{n^{(l)}}H'(x_i^{(l)})}&\overset{\text{a.s.}}{\to}\int_{\Omega}\frac{g(\mathbf{x}|\theta^{(t)})\prod_{i=1}^{n}H^{(t)}(x_i)}{g(\mathbf{x}|\theta')\prod_{i=1}^{n}H'(x_i)}p(\mathbf{x}|\zeta')\,\mathrm{d}\mu(\mathbf{x})\\&=\frac{1}{Z(\zeta')}\int_{\Omega}g(\mathbf{x}|\theta^{(t)})\prod_{i=1}^{n}H^{(t)}(x_i)\,\mathrm{d}\mu(\mathbf{x})\\&=\frac{Z(\zeta^{(t)})}{Z(\zeta')}.\end{aligned}$$

Similarly,

$$\frac{1}{L}\sum_{l=1}^{L}\frac{g(\mathbf{x}^{(l)}|\theta^*)\prod_{i=1}^{n^{(l)}}H^*(x_i^{(l)})}{g(\mathbf{x}^{(l)}|\theta')\prod_{i=1}^{n^{(l)}}H'(x_i^{(l)})}\xrightarrow{\text{a.s.}}\frac{Z(\zeta^*)}{Z(\zeta')}$$

and therefore $\hat{r}\xrightarrow{\text{a.s.}}r$.

We stress that two MCMC algorithms are being used; the main MCMC sampler in Section 4, and the following MCMC algorithm for generating importance samples. Remember, because the Metropolis–Hastings acceptance probability in the main sampler contains the intractable ratio r, then r must be estimated within *each* iteration. In response, we execute, within each iteration of the main sampler, the following MCMC sampler to obtain the spatial point patterns needed to estimate r. Clearly, running an MCMC sampler within *every* iteration of the main sampler is computationally expensive. However, with current computing technology, such an implementation is feasible.

For completeness, the algorithm of Geyer and Møller (1994) for generating spatial point patterns is now outlined. To generate spatial point patterns from $p(\mathbf{x}|\zeta')$, let $l=0$ and generate an initial point pattern $\mathbf{x}^{(l)}=(x_1^{(l)},\ldots,x_{n^{(l)}}^{(l)})$ with, say, $n^{(l)}=n$ points distributed uniformly in V. Within each iteration, the sampler randomly chooses between three move types: an a^x-move (add a point to the pattern), an r^x-move (remove a point), and an m^x-move (move a point). The probability of choosing each move type is a_p^x, r_p^x, and m_p^x, respectively, where $a_p^x+r_p^x+m_p^x=1$ and (for simplicity) $a_p^x=r_p^x$. For an m^x-move, randomly choose a point (all points are chosen with equal probability), say $x_i^{(l)}$, from the current pattern $\mathbf{x}^{(l)}$ and propose a new location x_i^* uniformly in V. Letting $\mathbf{x}^*=(x_1^{(l)},\ldots,x_{i-1}^{(l)},x_i^*,x_{i+1}^{(l)},\ldots,x_{n^{(l)}}^{(l)})$ and $n^{(l+1)}=n^{(l)}$, accept the move, that is let $\mathbf{x}^{(l+1)}=\mathbf{x}^*$, with probability $\min[1,\{g(\mathbf{x}^*|\theta')H'(x_i^*)\}/\{g(\mathbf{x}^{(l)}|\theta')H'(x_i^{(l)})\}]$, otherwise set $\mathbf{x}^{(l+1)}=\mathbf{x}^{(l)}$. For an a^x-move, choose a point, say x^*, uniformly in V. Letting $\mathbf{x}^*=(\mathbf{x}^{(l)},x^*)$, accept the new point x^*, that is let $\mathbf{x}^{(l+1)}=\mathbf{x}^*$ and $n^{(l+1)}=n^{(l)}+1$, with probability $\min[1,\exp\{-\sum_{i=1}^{n^{(l)}}\phi_{\theta'}(\|x_i^{(l)}-x^*\|)\}H'(x^*)\lambda(V)/(n^{(l)}+1)]$, else set $\mathbf{x}^{(l+1)}=\mathbf{x}^{(l)}$ and $n^{(l+1)}=n^{(l)}$. For an r^x-move, choose a point at random (all points are chosen with equal probability), say $x_i^{(l)}$, from the current pattern $\mathbf{x}^{(l)}$. Accept the removal of the point, that is let $\mathbf{x}^{(l+1)}=\mathbf{x}^{(l)}\setminus x_i^{(l)}$ and $n^{(l+1)}=n^{(l)}-1$, with probability $\min[1,n^{(l)}/\{\exp\{-\sum_{j\neq i}\phi_{\theta'}(\|x_j^{(l)}-x_i^{(l)}\|)\}H'(x_i^{(l)})\lambda(V)\}]$, else let $\mathbf{x}^{(l+1)}=\mathbf{x}^{(l)}$ and $n^{(l+1)}=n^{(l)}$. After an m^x, a^x, or r^x move is executed, increment l, choose another move type at random, and repeat. By monitoring the number of points $n^{(l)}$ and the *total potential energy* (TPE), $\sum_{i=1}^{n^{(l)}-1}\sum_{j=i+1}^{n^{(l)}}\phi_{\theta'}(\|x_i^{(l)}-x_j^{(l)}\|)$, over time l, it is possible to witness the convergence of $n^{(l)}$ and the total potential energy, and hence, informally, the sampler.

To avoid having to obtain an unduly large importance sample $\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(L)}$, the accurate estimation of r requires that the importance sampling density $p(\mathbf{x}|\zeta')$ be not unlike *both* target densities $p(\mathbf{x}|\zeta^*)$ and $p(\mathbf{x}|\zeta^{(t)})$ (i.e. ζ' should be close to both ζ^* and $\zeta^{(t)}$). A poor choice of importance sampling density will necessitate the generation of many more importance samples to approximate r for any given degree of accuracy. Hence, the proposal density should be concentrated near $\zeta^{(t)}$ ensuring that ζ^* is relatively close to $\zeta^{(t)}$ which in turn ensures that ζ' can be chosen close to both $\zeta^{(t)}$ and ζ^*.

A proposal density not concentrated near $\zeta^{(t)}$ can cause r (and hence the acceptance probability) to be poorly estimated, which can induce poor mixing behavior (Bognar and Cowles, 2004) (notwithstanding the issue of importance sampling, such a proposal density can also create low acceptance rates, causing poor mixing). Furthermore, if the proposal density is *highly* concentrated near $\zeta^{(t)}$, mixing will be poor since the sampler will move very slowly. Clearly, the proposal densities should be chosen with care.

Chen and Shao (1997) described, in a non-spatial setting, a similar estimator to (5). Their technique, called *ratio importance sampling*, was shown, in its optimal implementation, to have a smaller asymptotic relative mean squared error than bridge and path sampling (Gelman and Meng, 1998). Bognar and Cowles (2004) described a similar but more efficient method for estimating the intractable ratio r called *grid-based importance sampling*. Through clever setup and implementation, grid-based importance sampling allows the reusing of previously generated importance samples, increasing computational efficiency. When conditioning on n, they also found that optimal ratio importance sampling and (5), when used to estimate r in the acceptance probability, produced comparable inferences. Unlike our importance sampling approach described in Section 5, grid-based importance sampling is only feasible (due to storage considerations) when the parameter space is of low dimension.

6. Analysis of the Irish drumlin data

Recall the drumlin data $\mathbf{x} = \{x_1, \ldots, x_{232}\}$ observed in V, an 8×8 km^2 field, plotted in Fig. 1. Because the drumlin locations exhibit regularity (i.e. there are very few pairs of points that are close), because there appears to be inhomogeneity in the density of points, and because there are *no* pairs of points that are *very* close, a natural choice is an inhomogeneous hard-core *Straussian model* (Strauss, 1975) where $\theta = (b_{\mathrm{hc}}, b, h)$ and

$$\phi_\theta(s) = \begin{cases} \infty & s \leqslant b_{\mathrm{hc}}, \\ h & b_{\mathrm{hc}} < s \leqslant b, \\ 0 & s > b. \end{cases} \tag{6}$$

The *hard-core* parameter b_{hc} is the minimum distance between two drumlin locations, the *Straussian* parameter h (we assume $h > 0$) describes the strength of inhibition, and the *interaction distance* b ($> b_{\mathrm{hc}}$) describes the distance at which pairs of points cease to interact.

A uniform prior on $(0, 0.14455 = \min_{i \neq j} \|x_i - x_j\|)$ was placed on b_{hc}, and given b_{hc}, a uniform prior on $(b_{\mathrm{hc}}, 8)$ was placed on b (a flat prior on $(b_{\mathrm{hc}}, \infty)$ could be utilized as well). An $\mathrm{N}(\mu_h = 1, \sigma_h = 10)$ prior (truncated to $\mathscr{R}^+$) was placed on h. A Poisson prior with mean $\lambda_k = 5$ was placed on the number of tiles k, subject to $k \geqslant 1$. Given the number of tiles k, the generating points C_i were assumed to be i.i.d. uniform in V, $i = 1, \ldots, k$. The prior on the corresponding tile heights H_i $(i = 1, \ldots, k)$ were i.i.d. $\mathrm{Unif}(0, 20)$.

Note that, in general, decreasing the number of tiles k, or decreasing the variance of H_i, or both, produces stronger smoothing. Hence, the priors on k and H_i $(i=1,\ldots,k)$ should be chosen to reflect the desired amount of smoothing of the chemical activity. The degree of smoothing is further considered in Section 6.2.

6.1. Sampler details

Six move types were used to obtain draws from the posterior $p(b_{\mathrm{hc}}, b, h, C, H|\mathbf{x})$: b_{hc}-move (update the hard-core interaction distance), b-move (update the interaction distance), h-move (update the Straussian parameter), C-move (move a generating point), H-move (update the height of a tile), a-move (add a tile), and an r-move (remove a tile). A move type was randomly chosen within each iteration of the main sampler. The move probabilities were $b_{\mathrm{hc},p}=b_p=h_p=0.1$ and $C_p=H_p=a_p=r_p=0.175$, respectively.

At the beginning of iteration t, suppose the current parameter vector was $\zeta^{(t)}=(b_{\mathrm{hc}}^{(t)}, b^{(t)}, h^{(t)}, C^{(t)}, H^{(t)})$. Drawing from Section 4, if, say, a non-dimension changing b-move was chosen, a candidate interaction distance b^* was obtained from a Unif $(b^{(t)}\pm 0.05)$ distribution. Using the estimation procedure in Section 5, $L=15,000$ point patterns (following a 5000 iteration burn-in, discussed below) were generated from $p(\mathbf{x}|\zeta')$ where $\zeta'=(b_{\mathrm{hc}}^{(t)}, 0.5[b^*+b^{(t)}], h^{(t)}, C^{(t)}, H^{(t)})$. Geyer and Møller (1994) observe that their MCMC algorithm (described in Section 5) mixes most rapidly when a_p^x and r_p^x are large; hence we let $m_p^x=0.1$ and $a_p^x=r_p^x=0.45$. After realizing $\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(L)}$ from $p(\mathbf{x}|\zeta')$, the intractable ratio $r=Z(\zeta^{(t)})/Z(\zeta^*)$ in the acceptance probability (2) was estimated via (5). Plugging estimate (5) into (2), the candidate was accepted with probability (2). Then, t was incremented, and another move type was randomly chosen and executed. If $b^*\notin(b_{\mathrm{hc}}^{(t)}, 8)$, then the prior ratio equals 0 and the candidate b^* was rejected.

Because Geyer and Møller's algorithm for generating patterns from $p(\mathbf{x}|\zeta')$ was initialized with a binomial process (i.e. a Poisson process conditional on the number of points n), convergence is slowest when ζ' describes spatial regularity. Trace plots of n and the total potential energy were used to determine the burn-in period. Trace plots were generated under various values of ζ'; in all cases, convergence occurs quickly (certainly by the 5000th iteration, our burn-in period), and mixing is quite rapid.

The dominated coupling from the past technique of Kendall and Møller (2000) can be used to allow *exact* sampling from Strauss (among other) spatial point processes, eliminating the need to determine the length of burn-in in Geyer and Møller's algorithm in Section 5. However, because Møller and Nicholls (1999) noted that such simulations can be too slow to be useful in practice, we did not attempt to use this technique in our analysis.

The b_{hc}-move, h-move, C-move, and H-move proceeded similarly to a b-move. For a b_{hc}-move, the candidate b_{hc}^* was drawn from a Unif$(b_{\mathrm{hc}}^{(t)}\pm 0.025)$ distribution and the importance sampling locale was $\zeta'=(0.5[b_{\mathrm{hc}}^*+b_{\mathrm{hc}}^{(t)}], b^{(t)}, h^{(t)}, C^{(t)}, H^{(t)})$. For an h-move, h^* was drawn from a Unif$(h^{(t)}\pm 0.5)$ distribution and $\zeta'=(b_{\mathrm{hc}}^{(t)}, b^{(t)}, 0.5[h^*+h^{(t)}], C^{(t)}, H^{(t)})$. For a C-move, a generating point, say $C_i^{(t)}$, was randomly chosen, a candidate location

C_i^* was drawn uniformly from a square of area 4 centered about $C_i^{(t)}$, and $\zeta' = (b_{\text{hc}}^{(t)}, b^{(t)}, h^{(t)}, C_1^{(t)}, \ldots, C_{i-1}^{(t)}, C_i^{\text{mid}}, C_{i+1}^{(t)}, \ldots, C_{k^{(t)}}^{(t)}, H^{(t)})$ where C_i^{mid} was the mid-point between $C_i^{(t)}$ and C_i^*. For an H-move, one of the current heights, say $H_i^{(t)}$, was randomly chosen, a candidate height H_i^* was drawn from a $\text{Unif}(H_i^{(t)} \pm 1.5)$ distribution, and $\zeta' = (b_{\text{hc}}^{(t)}, b^{(t)}, h^{(t)}, C^{(t)}, H_1^{(t)}, \ldots, H_{i-1}^{(t)}, 0.5[H_i^* + H_i^{(t)}], H_{i+1}^{(t)}, \ldots, H_{k^{(t)}}^{(t)})$. Note that if $b_{\text{hc}}^* \notin (0, b^{(t)})$, $h^* < 0$, $H_i^* \notin (0, 20)$, or $C_i^* \notin V$, then the candidate was automatically rejected by the prior.

As was alluded to in Section 5, since the sampler must move in moderate steps, the Metropolis–Hastings acceptance rates (and autocorrelation) can become inflated (depending upon implementation, Gelman et al. (1996) suggest acceptance rates between approximately 15% and 50%). Consequently, the acceptance rates were 20.0%, 43.0%, 49.1%, 27.7%, and 64.9% for a b-move, b_{hc}-move, h-move, C-move, and H-move, respectively. See Bognar and Cowles (2004) for more guidance on choosing proposal densities.

As in Section 4, an a-move proceeds by generating $C_{k^{(t)}+1}^*$ uniformly in V, drawing a candidate height $H_{k^{(t)}+1}^*$ from $q(H_{k^{(t)}+1}^* | \zeta^{(t)}) = \text{Unif}(H^{(t)}(C_*^{(t)}) \pm 1.5)$ where $C_*^{(t)}$ is the closest generating point to $C_{k^{(t)}+1}^*$ (i.e. propose a height uniformly within 1.5 of the current height at $C_{k^{(t)}+1}^*$), generating $L = 15,000$ point patterns from $p(\mathbf{x}|\zeta')$ where $\zeta' = (b_{\text{hc}}^{(t)}, b^{(t)}, h^{(t)}, C^{(t)}, C_{k^{(t)}+1}^*, H^{(t)}, 0.5[H_{k^{(t)}+1}^* + H^{(t)}(C_*^{(t)})])$, estimating r via (5), plugging in the estimate into (3), and accepting the new tile with probability (3). If $H_{k^{(t)}+1}^* \notin (0, 20)$, then the candidate was automatically rejected via the prior. An r-move proceeded as in Section 4 where r was estimated via (5) using $\zeta' = (b_{\text{hc}}^{(t)}, b^{(t)}, h^{(t)}, C^{(t)}, H_1^{(t)}, \ldots, H_{i-1}^{(t)}, 0.5[H_i^{(t)} + H^*(C_i^{(t)})], H_{i+1}^{(t)}, \ldots, H_{k^{(t)}}^{(t)})$. If an r-move is attempted when $k^{(t)} = 1$, the move is automatically rejected by the prior and $\zeta^{(t+1)} = \zeta^{(t)}$.

Ten separate chains were run in parallel on 10 Intel Xeon 2.4 GHz workstations running Linux. Each computer completed 17,500 posterior iterations, including a 5000 iteration burn-in (the post-burn-in iterations were combined). With the coding in C++, the time to obtain the 175,000 updates was approximately 14 hours (140 total computing hours). Analysis of the 125,000 post-burn-in iterations was performed in part using the Bayesian Output Analysis (BOA) software (Smith, 2001). Trace plots of b_{hc}, b, and h show good mixing behavior: the Gelman and Rubin (Gelman and Rubin, 1992; Brooks and Gelman, 1998) corrected scale reduction factors for b_{hc}, b, and h were 1.049, 1.003, and 1.017, respectively, while the multivariate potential scale reduction factor was 1.055. Because of the changing dimensionality, evaluating mixing and convergence of C and H is more difficult (see Brooks and Giudici (1999) for current work in this area). We informally address this issue below, however.

Probably of little practical concern, in theory the sampler could jump from $\zeta^{(t)}$ to ζ' and back again. Because the respective ratios $Z(\zeta^{(t)})/Z(\zeta')$ and $Z(\zeta')/Z(\zeta^{(t)})$ are *approximated* in the acceptance probability, their product will not equal one, as it should. Hence, the likelihood function $L(b_{\text{hc}}, b, h, C, H)$ is not constant, which implies that the established theory for MCMC samplers does not apply. Nevertheless, based upon anecdotal evidence from simulation studies (not described herein), the aforementioned approximation does not prevent the sampler from converging to a distribution that closely *resembles* the true posterior.

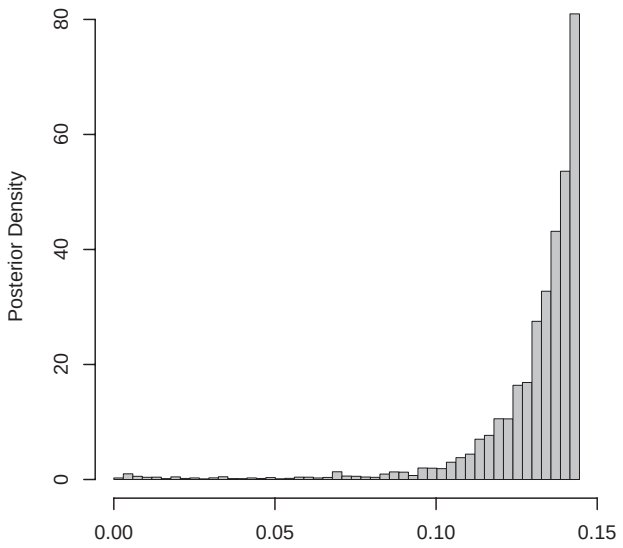

Fig. 2. Estimated marginal posterior distribution of the hard-core interaction distance b_{hc}.

6.2. Results

The estimated marginal posterior distribution of the hard-core distance b_{hc} is displayed as histogram in Fig. 2. The posterior mean of b_{hc} is approximately 0.130 km (Monte Carlo error 0.0005), with median of 0.136 km. The posterior mode is just under $0.14455 = \min_{i \neq j} \|x_i - x_j\|$ km. The 95% equal tail interval is (0.071, 0.144) km.

The estimated marginal posterior distribution of the interaction distance b is displayed in Fig. 3. The posterior mean of b is approximately 0.262 km (Monte Carlo error 0.0002), with median and mode of 0.263 and 0.256 km, respectively. The 95% equal tail interval is $(0.243, 0.278)$ km. The jagged nature of the marginal posterior is due to ϕ_θ, and hence the marginal posterior distribution of b, $p(b|\mathbf{x})$, being discontinuous (Bognar and Cowles, 2004).

Fig. 4 displays the estimated marginal posterior distribution of the Straussian parameter h. The estimated posterior mean, median, and mode of h are 1.073 (Monte Carlo error 0.0051), 1.067, and 1.063, respectively. The 95% equal tail interval is $(0.655, 1.540)$. The amount of uncertainty in the point estimate is surprising, especially considering the size of the data set. Unlike the frequentist approaches, the ability to reliably obtain interval estimates (and make such an observation) demonstrates the favorability of the current approach.

For each post-burn-in iteration, the pair potential function $\phi_\theta(s)$ was evaluated at 200 values of s ranging from 0 to 0.3 km. The posterior mean of $\phi_\theta(s)$ was approximated at each value of s by taking the mean of the 125,000 evaluations. The approximated

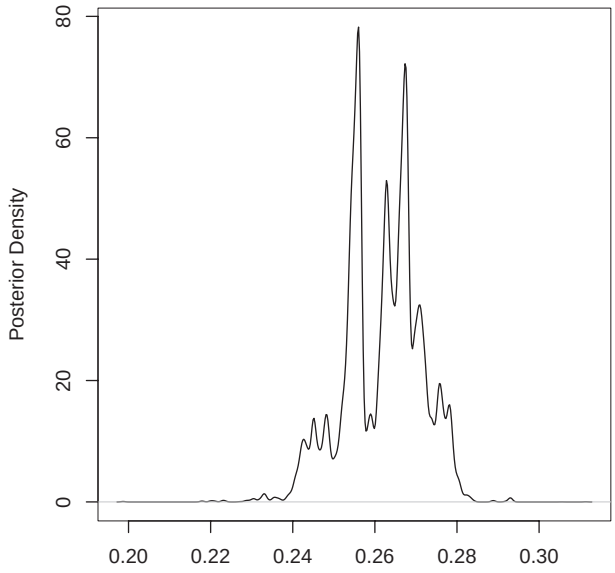

Fig. 3. Estimated marginal posterior distribution of the interaction distance b.

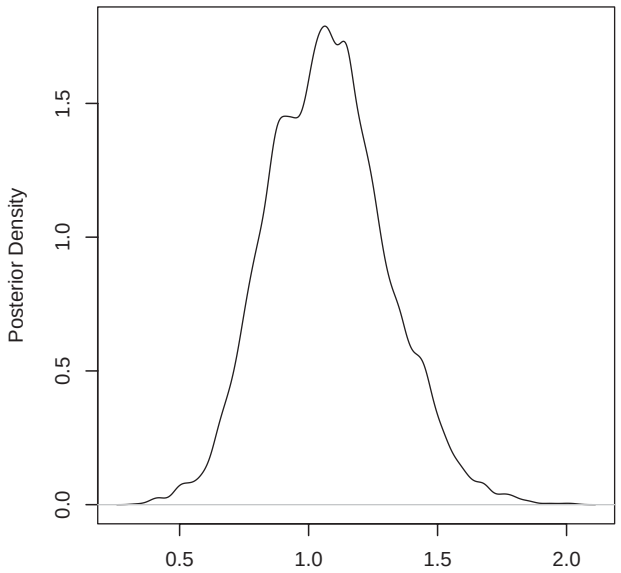

Fig. 4. Estimated marginal posterior distribution of the Straussian parameter h.

pair potential function is depicted in Fig. 5. By finding the quantiles of the 125,000 evaluations at each value of s, we were able to compute 95% pointwise credible intervals.

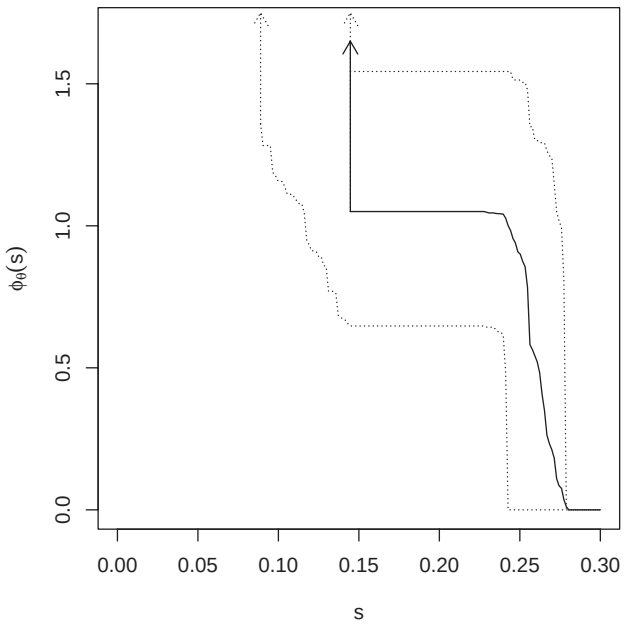

Fig. 5. Estimated posterior mean (with 95% pointwise credible intervals, dotted lines) of the pair potential function $\phi_\theta(s)$ for s ranging from 0 to 0.3 km.

To estimate the chemical activity surface, a 51×51 grid was constructed over V. For each post-burn-in iteration, the height of the chemical activity function was recorded at each grid location. The mean of the 125,000 evaluations was computed at the $51^2 = 2601$ grid points, and is graphically depicted in Fig. 6 as a contour plot.

It is also possible to estimate the marginal posterior distribution of the chemical activity at any point in V. Fig. 7 depicts the marginal posterior distribution of the chemical activity (per square kilometer) at locations $(1,6)$, $(4,1)$, and $(6,4)$ (which are depicted in Fig. 6 by A, B, and C, respectively). The posterior mean, median, and 95% credible interval at $A = (1,6)$ are 2.910 (Monte Carlo error $= 0.0319$), 2.736, and $(1.030, 5.797)$, respectively. Similarly, at $B = (4,1)$ the respective estimates are 9.428 (0.0474), 9.169, and $(6.519, 13.869)$; while at $C = (6,4)$, the estimates are 8.482 (0.0367), 8.387, and $(5.732, 11.768)$.

We informally addressed the mixing behavior of C and H by monitoring the height of the chemical activity surface (from a single chain) at locations A, B, and C. There appears to be good mixing behavior based upon the trace plots.

The estimated marginal posterior distribution of the number of tiles k is depicted in Fig. 8. In Q-move, the product of the prior and proposal ratios, under a $\text{Unif}(0, H_u)$ prior on the H's, is $\lambda_k r_p[(1-\exp(-\lambda_k))(k^{(t)}+1)^2 a_p H_u]^{-1}$. Placing a very diffuse prior on tile heights H (i.e. H_u large) deflates the aforementioned ratio and in turn deflates the acceptance probability (3), causing the sampler to always choose the simplest model with $k = 1$ tiles. Recall that strong smoothing of the chemical activity surface is attained when using strong, informative priors on the H's since little variability in the H's will

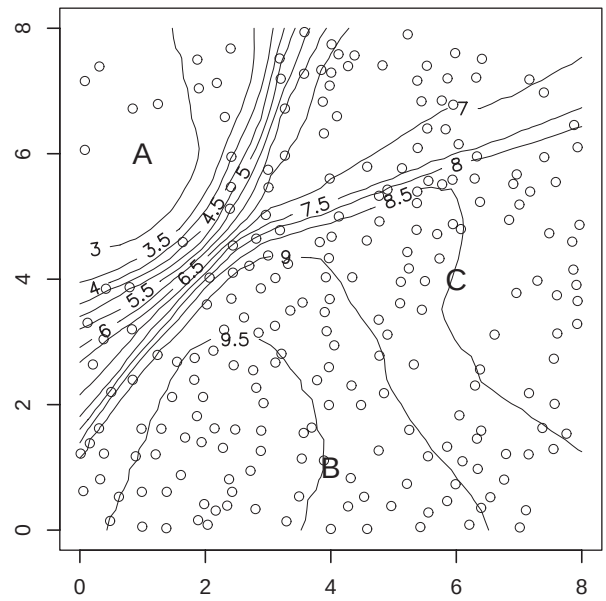

Fig. 6. Estimated posterior mean of the chemical activity surface (per square kilometer).

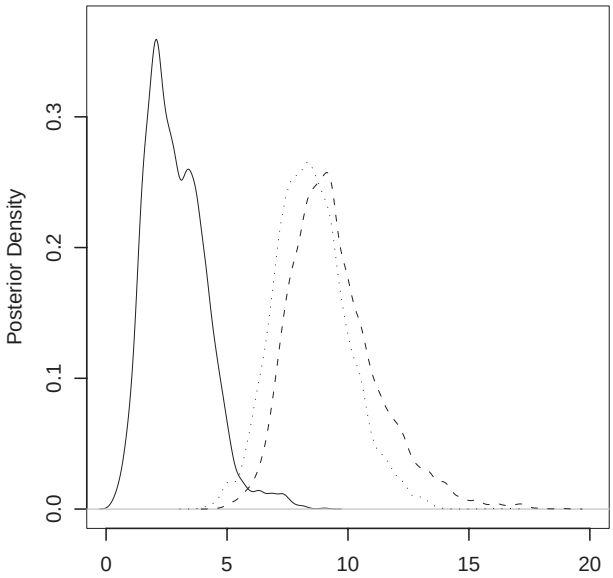

Fig. 7. Estimated marginal posterior distribution of the chemical activity (per square kilometer) at locations $A = (1, 6)$ (solid line), $B = (4, 1)$ (dashed line), and $C = (6, 4)$ (dotted line).

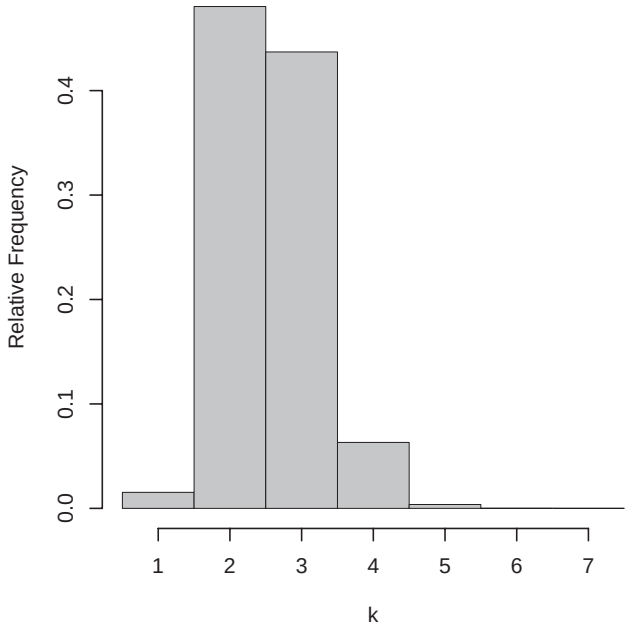

Fig. 8. Estimated marginal posterior distribution of the number of tiles k.

be encouraged by the prior. The same occurs with a very diffuse prior on the H's since such a prior suppresses the number of tiles k, increasing smoothing. Thus, the *least* amount of smoothing occurs for a slightly diffuse, but not overly diffuse, prior on the tile heights. Interestingly, Green (1995) noted a similar phenomenon in his analysis of a one-dimensional multiple change point problem.

7. Discussion

Our algorithm to obtain Bayesian inference in spatially inhomogeneous pairwise interacting point processes, although computationally intensive, is feasible with current computing technology. Like the frequentist approaches, our Bayesian technique provides for routine point estimation. However, Mateu and Montes (2001) and Diggle et al. (1994) show that some frequentist approaches can produce biased estimates in certain circumstances; but, as illustrated by Bognar and Cowles (2004), a Bayesian model does not appear to exhibit a similar characteristic.

Because the sampling distributions of the estimates are unclear, frequentists typically use parametric bootstrap procedures to obtain interval estimates. However, the behavior of such estimates is unclear in the current setting. In contrast, the Bayesian paradigm provides a convenient framework not only for interval estimation, but for many other posterior inferences.

Because there is no off-the-shelf software that can perform the analysis, practitioners must code their own MCMC samplers to fit Bayesian models. This, in and of itself,

is not a drawback since frequentists must code their analyses as well. In the end, however, we feel that a Bayesian model provides more detailed inferential insights. Additionally, our Bayesian technique can be used to model marked pairwise interacting point processes, models with multi-scale pair potential functions (Penttinen, 1984), and models in which θ is of unknown dimension.

Acknowledgements

I wish to thank the Editor, Associate Editor, and Referees for their very insightful comments and feedback, and Drs. Mary Kathryn Cowles and Osnat Stramer for their helpful input.

References

Baddeley, A., Turner, R., 2000. Practical maximum pseudolikelihood for spatial point patterns. Austral. New Zealand J. Statist. 42, 283–322.

Besag, J., Milne, R., Zachary, S., 1982. Point process limits of lattice processes. J. Appl. Probab. 19, 210–216.

Bognar, M.A., Cowles, M.K., 2004. Bayesian inference for pairwise interacting point processes. Statist. Comput. 14, 109–117.

Brooks, S., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graphical Statist. 7, 434–455.

Brooks, S., Giudici, P., 1999. Convergence assessment for reversible jump MCMC simulations. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, Vol. 6. Oxford University Press, Oxford, 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. Ann. Statist. 25, 1563–1594.

Diggle, P.J., 2003. Statistical Analysis of Spatial Point Patterns, 2nd Edition. Arnold, Paris.

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

Diggle, P.J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D., Tanemura, M., 1994. On parameter estimation for pairwise interaction point processes. Internat. Statist. Rev. 62, 99–117.

Fiksel, T., 1984. Estimation of parameterized pair potentials of marked and unmarked Gibbsian point processes. Elektron. Inform. Kybernet. 20, 270–278.

Gates, D.J., Westcott, M., 1986. Clustering estimates for spatial point distributions with unstable potentials. Ann. Inst. Statist. Math. 38, 123–135.

Gelman, A., Meng, X.-L., 1998. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. Statist. Sci. 13, 163–185.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences (with discussion). Statist. Sci. 7, 457–511.

Gelman, A., Roberts, G.O., Gilks, W.R., 1996. Efficient Metropolis jumping rules. In: Bernardo, J.M., Berger, A.P., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, Vol. 5. Oxford University Press, Oxford, pp. 599–607.

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

Geyer, C.J., Møller, J., 1994. Simulation procedures and likelihood inference for spatial point processes. Scand. J. Statist. 21, 359–373.

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. Comput. J. 21, 168–173.
Harkness, R.D., Isham, V., 1983. A bivariate spatial point pattern of ants' nests. Appl. Statist. 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. Scand. J. Statist. 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: Basawa, I.V., Heyde, C.C. (Eds.), Selected Proceedings of the Symposium on Inference for Stochastic Processes, Vol. 37. IMS Lecture Notes, pp. 297–318.
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. Adv. Appl. Probab. 32, 844–865.
Mateu, J., Montes, F., 2001. Likelihood inference for Gibbs processes in the analysis of spatial point patterns. Internat. Statist. Rev. 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. J. Chem. Phys. 21, 1087–1091.
Møller, J., 1999. Markov chain Monte Carlo and spatial point processes. In: Barndorff-Nielsen, O.E., Kendall, W.S., van Lieshout, M.N.M. (Eds.), Stochastic Geometry: Likelihood and Computation. Chapman & Hall/CRC, London, pp. 141–172.
Møller, J., Nicholls, G.K., 1999. Perfect simulation for sample-based inference. Research Report R-99-2011, http://www.math.auc.dk/%7Ejm/perfect.sim.temp.ps.gz, Department of Mathematical Sciences, Aalborg University. To appear in: Statistics and Computing, (conditionally accepted). Statist. Comput., to appear.
Møller, J., Waagepetersen, R.P., 2003. 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. Ann. Inst. Statist. Math. 33, 315–338.
Ogata, Y., Tanemura, M., 1984. Likelihood analysis of spatial point patterns. J. Roy. Statist. Soc. B 46, 496–518.
Ogata, Y., Tanemura, M., 1986. Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In: Francis, I.S., Manly, B.F.J., Lam, F.C. (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. Ann. Inst. Statist. Math. 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 Stud. Comput. Sci. Econom. Statist. 7.
Ripley, B.D., 1977. Modeling spatial patterns (with discussion). J. Roy. Statist. Soc. B 39, 172–212.
Ripley, B.D., 1988. Statistical Inference for Spatial Point Processes. Cambridge University Press, Cambridge.
Smith, B.J., 2001. Bayesian Output Analysis (BOA) software. Copyright (c) 2001 Brian J. Smith, http://www.public-health.uiowa.edu/boa.
Smith, A.F.M., Gelfand, A.E., 1992. Bayesian statistics without tears: a sampling resampling perspective. Amer. Statist. 46, 84–88.
Stoyan, D., Stoyan, H., 1998. Non-homogeneous Gibbs process models for forestry—a case study. Biometrical J. 40, 521–531.
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. Oper. Statist. Ser. Statist. 17, 429–433.
van Lieshout, M.N.M., 2000. Markov Point Processes and Their Applications. Imperial College Press, London.
Voronoi, M.G., 1908. Nouvelles applications des paramèters continus à la théorie des formes quadratiques. J. Reine Angew. Math. 134, 198–287.