ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2017)
79, *Part* 3, *pp.* 877–897

Bayesian inference for Matérn repulsive processes

Vinayak Rao,

Purdue University, West Lafayette, USA

Ryan P. Adams

Harvard University, Boston, and Twitter, Boston, USA

and David D. Dunson

Duke University, Durham, USA

[Received August 2013. Final revision June 2016]

Summary. In many applications involving point pattern data, the Poisson process assumption is unrealistic, with the data exhibiting a more regular spread. Such repulsion between events is exhibited by trees for example, because of competition for light and nutrients. Other examples include the locations of biological cells and cities, and the times of neuronal spikes. Given the many applications of repulsive point processes, there is a surprisingly limited literature developing flexible, realistic and interpretable models, as well as efficient inferential methods. We address this gap by developing a modelling framework around the Matérn type III repulsive process. We consider some extensions of the original Matérn type III process for both the homogeneous and the inhomogeneous cases. We also derive the probability density of this generalized Matérn process, allowing us to characterize the conditional distribution of the various latent variables, and leading to a novel and efficient Markov chain Monte Carlo algorithm. We apply our ideas to data sets of spatial locations of trees, nerve fibre cells and Greyhound bus stations.

Keywords: Event process; Gaussian process; Gibbs sampling; Matérn process; Point pattern data; Poisson process; Repulsive process; Spatial data

1. Introduction

Point processes find wide use in fields such as astronomy (Peebles, 1974), biology (Waller *et al.*, 2011), ecology (Hill, 1973), epidemiology (Knox, 2004), geography (Kendall, 1939) and neuroscience (Brown *et al.*, 2004). The simplest and most popular model for point pattern data is the Poisson process (see for example Daley and Vere-Jones (2008)); however, implicit in the Poisson process is the assumption of independence of the event locations. This simplification is unsuitable for many real applications in which it is of interest to account for interactions between nearby events.

Point processes on the real line can deviate from Poisson processes either by being more bursty or more refractory. In higher dimensions, these are called clustered and repulsive processes. Our focus is on the latter, characterized by being more regular (underdispersed) than the Poisson process. Reasons for this could be competition for finite resources between trees (Strand, 1972), interaction between rigid objects such as cells (Waller *et al.*, 2011) or the result of planning (e.g.

Address for correspondence: Vinayak Rao, Department of Statistics, Purdue University, Math 236, 150 North University Street, West Lafayette, IN 47907, USA.
E-mail: varao@purdue.edu

 1369–7412/17/79877

the locations of bus stations). Characterizing this repulsion is important towards understanding how disease affects cell locations (Waller *et al*., 2011), how neural spiking history affects stimulus response (Brown *et al*., 2004) or how reliable a network of stations is (Şahin and Süral, 2007).

Developing a flexible and tractable statistical framework to study such repulsion is not straightforward on spaces that are more complicated than the real line. For the latter, the ordering of points leads to a convenient framework based on renewal processes (Daley and Vere-Jones, 2008). This work focuses on a class of repulsive point processes on higher dimensional spaces called the Matérn type III process. First introduced in Matérn (1960, 1986), this involves thinning events of a 'primary' Poisson process that are 'too close to each other'. Starting with the simplest such process (called a hard-core process), we introduce extensions that provide a flexible framework for modelling repulsive processes. We derive the probability density of the resulting process: a characterization that allows us to identify the conditional distribution over the thinned events as a simple Poisson process. This allows us to develop a simple and efficient Markov chain Monte Carlo (MCMC) algorithm for posterior inference. Code implementing different models and algorithms is available from `https://github.com/varao/matern`.

2. Related work

Gibbs processes Daley and Vere-Jones (2008) arose from the statistical physics literature to describe systems of interacting particles. A Gibbs process assigns energy

$$U(S) = \sum_{i=1}^{n} \sum_{1 \leqslant j_1 < \ldots < j_i \leqslant n} \psi_i(s_{j_1}, \ldots, s_{j_i})$$

to any configuration of events $S = (s_1, \ldots, s_n)$, where ψ_i is an ith-order potential term. Usually, interactions are limited to be pairwise and, by choosing these potentials appropriately, one can model different kinds of interactions. The probability density of any configuration is proportional to its exponentiated negative energy and, letting θ parameterize the potential energy, we have

$$p(S|\theta) = \exp\{-U(S;\theta)\}/Z(\theta). \tag{1}$$

Unfortunately, evaluating the normalization constant $Z(\theta)$ is usually intractable, making even sampling from the prior difficult; typically, this requires a coupling from the past approach (Møller and Waagepetersen, 2007). Inference over the parameters usually proceeds by maximum likelihood or pseudolikelihood methods (Møller and Waagepetersen, 2007; Mateu and Montes, 2001) and is slow and expensive.

Determinantal point processes (Hough *et al*., 2006; Scardicchio *et al*., 2009) are another framework for modelling repulsion. Although these processes are mathematically and computationally elegant, they are not intuitive or easy to specify, and with few exceptions (e.g. Affandi *et al*. (2014)) there is little Bayesian work involving such models.

There has also been work using Matérn type I and II processes (see Waller *et al*. (2011)). We mention some limitations of these next. Additionally, our sampling scheme which is outlined later does not extend to these processes, making Bayesian inference via MCMC sampling less natural than with the type III process.

3. Matérn repulsive point processes

Matérn (1960) proposed three schemes, which are now called the Matérn type I, type II and type III hard-core point processes, for constructing repulsive point processes. For the type I process, one samples a *primary* process from a homogeneous Poisson process with intensity λ

and then deletes all points that are separated by less than R. Although the simplicity of this scheme makes it amenable to theoretical analysis, the thinning strategy is often too aggressive. In particular, as λ increases, the thinning effect begins to dominate, and eventually the density of Matérn events begins to decrease with λ.

The Matérn type II process tries to rectify this by assigning each point an 'age'. When there is a conflict between two points, rather than deleting both, the older point survives. Although this construction allows higher event densities, it also allows an event to be thinned by an earlier point that was also thinned. This is slightly unnatural, as we might expect only surviving points to influence future events.

The Matérn type III process addresses these limitations by letting a newer event be thinned only if it falls within radius R of an older event that was not thinned before. The resulting process has several desirable properties. In many applications, this thinning mechanism is more natural than the type I and II processes. In particular, it forms a realistic model for various spatio-temporal phenomena where the latent birth times are not observed and must be inferred. This process also supports higher event densities than the type I and type II processes with the same parameters; in fact, as λ increases, the average number of points in any area increases monotonically to the 'jamming limit', which is the maximum density at which spheres of radius R can be packed in a bounded area (Møller *et al.*, 2010). Monotonicity with respect to the intensity of the primary process is important when we model inhomogeneity by allowing λ to vary with location (see Section 7), with large values of λ implying high densities.

In spite of these properties, the Matérn type III process has not found widespread use in the spatial point process community. Theoretically it is not as well understood as the other two Matérn processes; for instance, there is no closed form expression for the average number of points in any region. Instead, one usually resorts to simulation studies to understand better the modelling assumptions that are implicit to this process.

A more severe impediment to the use of this process (and the other Matérn processes) is that, given a realization, there do not exist efficient techniques for inference over parameters such as λ or R. The few existing inference schemes involve imputing, and then perturbing the thinned events via incremental birth–death steps. This sets up a Markov chain which proceeds by randomly inserting, deleting or shifting the thinned events, with the various event probabilities set up so that the Markov chain converges to the correct posterior over thinned events (Møller *et al.*, 2010; Huber and Wolpert, 2009; Adams, 2009). Given the entire set of thinned events, it is straightforward to obtain samples of the parameters λ and R. However, the incremental nature of these birth–death updates can make the sampler mix quite slowly. The birth–death sampler can be adapted to a coupling from the past scheme to draw perfect samples of the thinned events (Huber and Wolpert, 2009). This can then be used to approximate the likelihood of the Matérn observations or, perhaps, to drive a Markov chain following ideas from Andrieu and Roberts (2009). However, this can be quite inefficient also, with long waiting times until the sampler returns a perfect sample.

Somewhat surprisingly, despite being more complicated than the type I and II processes, we can develop an efficient MCMC sampler for the type III process. Before describing this, we develop more general extensions of the Matérn type III process, providing a flexible and practical framework for modelling repulsive processes.

4. Generalized Matérn type III processes

A Matérn type III hard-core point process on a measurable space $(\mathcal{S}, \Sigma)$ is a repulsive point process parameterized by an intensity λ and an interaction radius R. It is obtained by thinning

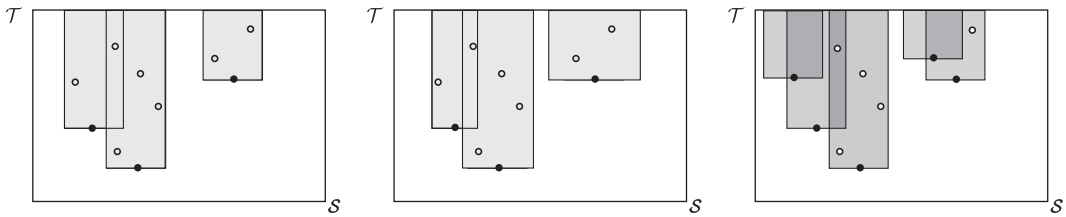

Fig. 1. (a) Matérn type III hard-core point process in one dimension (• (projected onto $\mathcal{S}$), Matérn events; ∘, thinned events; ▫, shadow); (b), (c) Matérn type III processes with varying radii and probabilistic deletion

events of a homogeneous primary Poisson process F with intensity λ. Each event $f \in F$ of the primary process is independently assigned a random mark t, the time of its birth. Without loss of generality, we assume that this takes values in the interval [0, 1], which we call $\mathcal{T}$. The (f_i, t_i) form a Poisson process F^+ on $\mathcal{S} \times \mathcal{T}$ whose intensity is still λ. Call the collection of birth times T^F, and define $F^+ \equiv (F, T^F)$ as the collection of (location, birth time) pairs. T^F induces an ordering on the events in F^+, and a secondary process $G^+ \equiv (G, T^G)$ is obtained by traversing F^+ in this order and deleting all points within a distance R of any earlier undeleted point. We obtain the Matérn process G by projecting G^+ onto $\mathcal{S}$.

Fig. 1(a) shows relevant events for the one-dimensional case. The full dots form the Matérn process G and the empty dots represent thinned events. Both together form the primary process. Define the 'shadow' of a point $(s^*, t^*) \in \mathcal{S} \times \mathcal{T}$ as the indicator function for all locations in $\mathcal{S} \times \mathcal{T}$ that would be thinned by (s^*, t^*). This is the set of all points whose $\mathcal{S}$-co-ordinate is within R of s^*, and whose $\mathcal{T}$-co-ordinate is greater than t^*. Letting I be the indicator function, define the shadow of (s^*, t^*) at (s, t) as

$$\mathscr{H}(s, t; s^*, t^*, R) = I(t > t^*)\, I(\|s - s^*\| < R). \tag{2}$$

The shadow of G^+ is all locations that would be thinned by any element of G^+:

$$\mathscr{H}(s, t; G^+) = 1 - \prod_{(s^*, t^*) \in G^+} \{1 - \mathscr{H}(s, t; s^*, t^*, R)\}. \tag{3}$$

For notational convenience, we drop the dependence of the shadow on R above. The shaded area in Fig. 1(a) shows the shadow of all Matérn events, G^+. All thinned events must lie in the shadow of the Matérn events; otherwise they could not have been thinned. Similarly, Matérn events cannot lie in each other's shadows; however, they *can* fall within the shadow of some thinned event.

The hard-core repulsive process can be generalized in various ways, For instance, instead of requiring all Matérn events to have the same radius, we can assign each an independent radius drawn from some distribution $q(R)$. Such Matérn processes are called soft-core repulsive processes (Huber and Wolpert, 2009). In this case, the primary process can be viewed as a Poisson process on a space whose co-ordinates are location $\mathcal{S}$, birth time $\mathcal{T}$ and radius $\mathcal{R}$. Given a realization $F^+ \equiv (F, T^F, R^F)$, we define a secondary point process $G^+ \equiv (G, T^G, R^G)$ by deleting all points that fall within the radius that is associated with an older, undeleted primary event. The locations G constitute a sample from the soft-core Matérn type III process. Given the triplet (G, T^G, R^G), we can once again calculate the shadow $\mathscr{H}$, which is now defined as

$$\mathscr{H}(s, t; G^+) = 1 - \prod_{(s^*, t^*, r^*) \in G^+} \{1 - \mathscr{H}(s, t; s^*, t^*, r^*)\}. \tag{4}$$

Fig. 1(b) illustrates this; note that the radii of the thinned events are irrelevant.

We propose an approach to soft repulsion by *probabilistically* thinning events of the primary Poisson process. This is a generalization of ideas in the literature, allowing flexible control over the strength of the repulsive effect (the thinning probability) as well as its span (the interaction radius). Our extension further allows more realistic models where the probability of deletion decays smoothly with distance from a previous unthinned point, and we shall deal with this general case below. Write the deletion kernel that is associated with location s^* as $K(\cdot, s^*)$, so that the probability of thinning an event at (s,t) is given by $I(t>t^*)\,K(s,s^*)$. To keep this process efficient, we can use a deletion kernel with a compact support; Fig. 1(c) illustrates the resulting shadow. Where previously the Matérn events defined a black-or-white shadow, now the shadow can have intermediate 'grey' values corresponding to the probability of deletion. Now the shadow at any location $(s,t)\in\mathcal{S}\times\mathcal{T}$ is given by

$$\mathscr{H}(s,t;G^+)=1-\prod_{(s^*,t^*)\in G^+}\{1-I(t>t^*)\,K(s,s^*)\}. \tag{5}$$

For this process, although the thinned events must still lie in the shadow $\mathscr{H}(s,t;G^+)$, Matérn events *can* lie in each other's shadow. We recover the Matérn processes with deterministic thinning by letting K be the indicator function.

Another generalization is to allow the thinning probability to depend on the difference of the birth times of two events. This is useful in applications where the repulsive influence of an event decays as time passes, and the thinning probability is given by $K_1(t,t^*)\,K_2(s,s^*)$. Although we do not study this, we mention it to demonstrate the flexibility of the Matérn framework towards developing realistic repulsive mechanisms.

Finally, we mention that repulsive processes on the real line can be viewed as generalized Matérn type III processes where birth times are observed. Probabilistic thinning recovers a class of self-inhibiting point processes that is commonly used to model neuronal spiking (Brown *et al*., 2004). Similarly, renewal processes can be viewed as a Matérn type III process where the shadow $\mathscr{H}$ has a special Markovian construction.

4.1. Probability density of the Matérn type III point process

Let $(\mathcal{S},\Sigma)$ be a subset of a d-dimensional Euclidean space (Σ is the Borel σ-algebra). For two points $\mathbf{s}_1,\mathbf{s}_2\in\mathcal{S}$, define $\mathbf{s}_1>\mathbf{s}_2$ if $s_{1d}>s_{2d}$. Thus the last co-ordinate defines a partial ordering on $\mathcal{S}$ (and we shall associate this co-ordinate with the birth time of an event). A realization of a Matérn process on $\mathcal{S}$ is obtained by thinning a primary Poisson process and has finite cardinality if the primary process is finite. We restrict ourselves to this case. With probability 1, there is unique ordering of the Poisson elements according to the partial ordering that we have just defined. This will allow us to associate realizations of point processes with unique, increasing sequences of points in $\mathcal{S}$. Below, we describe this space of point process realizations.

For each n, let $\mathcal{S}^n$ be the n-fold product space with the usual product σ-algebra, Σ^n. We refer to elements of $\mathcal{S}^n$ as S^n. Define the union space $\tilde{\mathcal{S}}^{\cup}\equiv\cup_{n=0}^{\infty}\mathcal{S}^n$ (where $\mathcal{S}^0$ is a point corresponding to an empty sequence) and assign it the σ-algebra $\tilde{\Sigma}^{\cup}\equiv\{\cup_{n=0}^{\infty}A^n,\forall\,A^n\in\Sigma^n\}$. $\tilde{\mathcal{S}}^{\cup}$ is the space of all finite sequences in $\mathcal{S}$, and define $(\mathcal{S}^{\cup},\Sigma^{\cup})$ as its restriction to increasing sequences in $\mathcal{S}$. We treat finite point processes as random variables taking values in $\mathcal{S}^{\cup}$ and refer to elements of this space by upper-case letters (e.g. S). For a finite measure μ on $(\mathcal{S},\Sigma)$ (e.g. Lebesgue measure), let μ^n be the n-fold product measure on $(\mathcal{S}^n,\Sigma^n)$. Assign any set $B\in\Sigma^{\cup}$ the measure

$$\mu^{\cup}(B)=\sum_{n=0}^{\infty}\mu^n(B\cap\mathcal{S}^n)=\sum_{n=0}^{\infty}\int_{B\cap\mathcal{S}^n}\mu^n(\mathrm{d}S^n). \tag{6}$$

Now, let S be a realization of a Poisson process with mean measure Λ, and assume that Λ admits a density λ with respect to μ. Writing $|S|$ for the cardinality of S, we have the following theorem.

Theorem 1 (density of a Poisson process). A Poisson process on the space $\mathcal{S}$ with intensity $\lambda(s)$ (where $\Lambda(\mathcal{S}) = \int_{\mathcal{S}} \lambda(s)\mu(\mathrm{d}s) < \infty$) is a random variable taking values in $(\mathcal{S}^{\cup}, \Sigma^{\cup})$ with probability density with respect to the measure $\mu^{\cup}$ given by

$$p(S) = \exp\{-\Lambda(\mathcal{S})\} \prod_{j=1}^{|S|} \lambda(s_j). \tag{7}$$

For a proof, see Daley and Vere-Jones (2008). The density $p(S)$ is called the *Janossy density* and is used in the literature to define the Janossy measure. The Janossy measure is not a probability measure; however, our definition of S as an ordered sequence ensures that

$$\int_{\mathcal{S}^{\cup}} p(S)\, \mu^{\cup}(\mathrm{d}S) = 1. \tag{8}$$

We now return to the Matérn type III process. Recall that events of the augmented primary Poisson process $F^+ = (F, T^F)$ lie in the product space $(\mathcal{S} \times \mathcal{T})$, where $\mathcal{T}$ is the unit interval. Since we order elements by their last co-ordinate, F^+ is a sequence with increasing birth times T^F. Let μ be a measure on $(\mathcal{S} \times \mathcal{T})$; when $\mathcal{S}$ is a subset of $\mathbb{R}^2$, μ is just Lebesgue measure on $\mathbb{R}^3$. By first writing down the density of the augmented primary Poisson process F^+, and then using the thinning construction of the Matérn type III process, we can calculate the probability density of the augmented Matérn type III process $G^+ = (G, T^G)$ with respect to the measure $\mu^{\cup}$.

Theorem 2. Let $G^+ = (G, T^G)$ be a sample from a generalized Matérn type III process, augmented with the birth times. Let λ be its intensity and $\mathscr{H}(s, t; G^+)$ be its shadow following the appropriate thinning scheme. Then, its density with respect to $\mu^{\cup}$ is

$$p(G^+|\lambda) = \exp\Big[-\lambda \int_{\mathcal{S}\times\mathcal{T}} \{1 - \mathscr{H}(s, t; G^+)\}\, \mu(\mathrm{d}s\, \mathrm{d}t)\Big] \lambda^{|G^+|} \times \prod_{g^+ \in G^+} \{1 - \mathscr{H}(g^+; G^+)\}. \tag{9}$$

We include a proof in Appendix A. The product term in expression (9) penalizes Matérn events that fall within the shadow of earlier events; in fact, for deterministic thinning, such an occurrence will have zero probability. The exponentiated integral encourages the shadow to be large, which in turn implies that the events are spread out. Setting the shadow to 0 recovers the Poisson density of equation (7).

A similar result was derived in Huber and Wolpert (2009); they expressed the Matérn type III density with respect to a homogeneous Poisson process with unit intensity. However, their result applied only to the hard-core process. Also, their proof is less direct than ours, proceeding via a coupling from the past construction. It is not clear how such an approach extends to the more complicated extensions that we introduced above.

Theorem 2 shows that, for all parameter settings, the Matérn probability measure is absolutely continuous with respect to $\mu^{\cup}$. Following Schervish (1995), theorem 3.14, we can then apply Bayes's theorem, giving the following corollary.

Corollary 1. Let $G^+ = (G, T^G)$ be a sample from a Matérn type III process augmented with its birth times. Let λ be its intensity, and $\mathscr{H}(s, t; G^+)$ its shadow. Then, given G^+, the conditional distribution of the locations and birth times of the thinned events $\tilde{G}^+ = (\tilde{G}, T^{\tilde{G}})$ is a Poisson process on $\mathcal{S} \times \mathcal{T}$, with intensity $\lambda\, \mathscr{H}(s, t; G^+)$.

Proof. Observe that the primary Poisson process $F^+ = (\tilde{G}^+ \cup G^+)$ (where the union of two ordered sequences is their concatenation followed by reordering). The joint $p(\tilde{G}^+, G^+)$ is the density of F^+ multiplied by the probability that the elements of F^+ are assigned labels 'thinned' or 'not thinned'. The latter depends on the shadow $\mathscr{H}$, and it follows easily from theorem 1 (see equation (24) in Appendix A) that

$$p(\tilde{G}^+, G^+) = \exp\{-\lambda\mu(\mathcal{S}\times\mathcal{T})\}\lambda^n \prod_{(s,t)\in\tilde{G}^+} \mathcal{H}(s,t;G^+) \prod_{(s,t)\in G^+} \{1-\mathscr{H}(s,t;G^+)\}. \tag{10}$$

Plugging this and equation (9) into Bayes's rule, we have

$$p(\tilde{G}^+|G^+) = \frac{p(\tilde{G}^+, G^+)}{p(G^+)} \tag{11}$$

$$= \exp\left\{-\int_{\mathcal{S}\times\mathcal{T}} \lambda\,\mathscr{H}(s,t;G^+)\,\mu(\mathrm{d}s\,\mathrm{d}t)\right\} \prod_{(\tilde{s},\tilde{t})\in\tilde{G}^+} \lambda\,\mathscr{H}(\tilde{s},\tilde{t};G^+). \tag{12}$$

From theorem 1, this is the density of a Poisson process with intensity $\lambda\,\mathscr{H}(s,t;G^+)$. □

This result provides a remarkably simple characterization of the thinned primary events. Rather than having to resort to incremental birth–death schemes that update the thinned Poisson events one at a time (Green, 2003; Adams, 2009), we can simulate all of these jointly from a Poisson process, directly obtaining their number and locations. Such an approach is much simpler and much more efficient, and it is central to the MCMC sampling algorithm that we describe in the next section.

The intuition behind this result is that, for a type III process, a point of the primary Poisson process F can be thinned only by an element of the secondary process. Consequently, given the secondary process, there are no interactions between the thinned events themselves: given the Matérn process, the thinned events are just Poisson distributed. Such a strategy does not extend to Matérn type I and II processes where the fact that thinned events *can* delete each other means that the posterior is no longer Poisson. For instance, for any of these processes, it is not possible for a thinned event to occur by itself within any neighbourhood of radius R (otherwise it could not have been thinned in the first place). However, two or more events *can* occur together. Clearly such a process is not Poisson; rather it has a clustered structure.

5. Bayesian modelling and inference for Matérn type III processes

In what follows, we model an observed sequence of points G as a realization of a Matérn type III process. The parameters governing this are the intensity of the primary process, λ, and the parameters of the thinning kernel, θ. For the hard-core process, θ is just the interaction radius R, whereas, with probabilistic thinning, θ might include an interaction radius R, and a thinning probability p (with $p = 1$ recovering the hard-core model). For the soft-core process, each Matérn event has its own interaction radius which we must infer, and θ would be this collection of radii. In this case we might also assume that the distribution that these radii are drawn from has unknown parameters.

Taking a Bayesian approach, we place priors on the unknown parameters. A natural prior for λ is the conjugate gamma density. The gamma density is also a convenient and flexible prior for the thinning length scale parameter R. For the case of probabilistic thinning where $\theta = (p, R)$, we can place a beta prior on the thinning probability p. For the soft-core model, we model the radii as uniformly distributed over $[r_{\mathrm{L}}, r_{\mathrm{U}}]$ and place conjugate hyperpriors on r_{L} and r_{U}. For

simplicity, we leave out any hyperparameters in what follows and, writing q for the prior on θ, we have

$$\lambda \sim \text{gamma}(a, b), \tag{13}$$

$$\theta \sim q, \tag{14}$$

$$F^+ \equiv (F, T^F) \sim \text{Poisson process}(\lambda), \tag{15}$$

$$G^+ \equiv (G, T^G) \sim \text{Thin}(F^+, \theta). \tag{16}$$

Note that G^+ includes the Matérn events G as well as their birth times T^G; however, we observe only G. Given G, we require the posterior distribution $p(\lambda, \theta|G)$. We will actually work with the augmented posterior $p(\lambda, \theta, F^+, T^G|G)$. In particular, we set up a Markov chain whose state consists of all these variables, and whose transition operator is a sequence of Gibbs steps that conditionally update each of these four groups of variables. We describe the four Gibbs updates below.

5.1. Sampling the thinned events

Given the Matérn events G^+ and the thinning parameters θ, we can calculate the shadow $\mathscr{H}_\theta(s, t; G^+)$ (here we make explicit the dependence of $\mathscr{H}$ on θ). Sampling the thinned events $\tilde{G}^+$ is now a straightforward application of corollary 1: discard the old events, and simulate a new sequence from a Poisson process with intensity $\lambda\, \mathscr{H}_\theta(s, t; G^+)$.

We do this by applying the thinning theorem for Poisson processes (Lewis and Shedler (1979); see also theorem 3 in Section 7). In particular, we first sample a homogeneous Poisson process with intensity λ on $(\mathcal{S} \times \mathcal{T})$ and keep each point (s, t) of this process with probability $\mathscr{H}_\theta(s, t; G^+)$. The surviving points form a sample from the required Poisson process. For models with deterministic thinning, $\mathscr{H}_\theta$ is a binary function, and the posterior is just a Poisson process with intensity λ restricted to the shadow. Note that this step eliminates the need for any birth–death steps and provides a simple and global way to vary the number of thinned events from iteration to iteration.

5.2. Sampling the birth times of the Matérn events

From equation (10), we see that the birth times T_G of the Matérn events have density

$$p(T^G|G, F^+, \lambda, \theta) \propto \prod_{(s,t)\in G^+} \{1 - \mathscr{H}(s, t; G^+)\} \prod_{(s,t)\in \tilde{G}^+} \mathscr{H}(s, t; G^+). \tag{17}$$

A simple Markov operator that maintains equation (17) as its stationary distribution is a Gibbs sampler that updates the birth times one at a time. For each Matérn event $g \in G$, we look at all primary events (thinned or not) within distance r^g (where r^g is the interaction radius associated with g). The birth times of these events segment the unit interval into a number of regions, and the birth time t^g of g is uniformly distributed within each interval (since, as t^g moves over an interval, the shadow at all primary events remains unchanged). As t^g moves from one segment to the next, one of the primary events moves into or out of the shadow of g. The probability of any interval is then proportional to the probability that these neighbouring events are assigned their labels 'thinned' or 'not thinned' under the shadow that results when g is assigned to that interval; this is easily calculated for each thinning mechanism. For instance, in the Matérn process with deterministic thinning, the birth time t^g is uniformly distributed on the interval $[0, t_{\min}]$, where $t_{\min}$ is the time of the oldest event that is thinned only by g ($t_{\min} = 1$ if there is no such event).

Although it is not difficult to develop more global moves, we found it sufficient to sweep through the Matérn events, sequentially updating their birth times. This, together with jointly updating all thinned event locations and birth times, was enough to ensure that the chain mixed rapidly.

5.3. Sampling the Poisson intensity

Having reconstructed the thinned events, it is easy to resample the intensity λ. Note that the number of primary events $|F|$ is Poisson distributed with intensity λ. With a conjugate gamma(a, b) prior on the Poisson intensity λ, the posterior is also gamma distributed, with parameters $a_{\text{post}} = a + |F|$ and $1/b_{\text{post}} = 1/b + 1/\mu(\mathcal{S})$.

5.4. Sampling the thinning kernel parameter θ

Like the birth times T^G, the posterior distribution of θ follows from equation (10). For a prior $q(\theta)$, the posterior is just

$$p(\theta|G^+, F^+, \lambda) \propto q(\theta) \prod_{(s,t)\in G^+} \{1 - \mathscr{H}_\theta(s, t; G^+)\} \prod_{(s,t)\in \tilde{G}^+} \mathscr{H}_\theta(s, t; G^+). \tag{18}$$

Again, sampling θ is equivalent to sampling a latent variable in a two-class classification problem, with the Matérn events and thinned events corresponding to the two classes. Different values of θ result in different shadows $\mathscr{H}_\theta(\cdot; G^+)$, and the likelihood of θ is determined by the probability of the labels under the associated shadow. For the models that we consider, this results in a simple piecewise parametric posterior distribution.

For the Matérn hard-core process, $\theta = R$ is the interaction radius, whose posterior distribution is a truncated version of the prior. The lower bound requires that no thinned event lies outside the new shadow, whereas the upper bound requires that no Matérn event lies inside the shadow. Sampling from this is straightforward. The same applies for the soft-core model, only, now, each Matérn event has its own interaction radius, and we can conditionally update them. Given the set of radii, updating any hyperparameters of q is easy. Finally, for the model with probabilistic thinning, we have $\theta = (p, R)$. To simulate R, we segment the positive real line into a finite number of segments, with the end points corresponding to values of R when a primary event moves into the shadow of a Matérn event. Over each segment, the likelihood remains constant. It is straightforward to sample a segment, and then to simulate conditionally a value of R within that segment. To simulate p, we simply count how many opportunities to thin events were taken or missed and, with a beta prior on p, the posterior follows easily.

6. Experiments

We start with the classic Swedish pine tree data set (Ripley, 1988), which is available as part of R package `spatstat` (Baddeley and Turner, 2005). This consists of the locations of 71 trees which we normalized to lie in a 10×10 square (Fig. 2(a)), and which we modelled as a realization of a Matérn type III hard-core point process. We placed a gamma$(1, 1)$ prior on the intensity λ (noting there are about 100 points in a 10×10 square), and a flat prior on the radius R (noting that R cannot exceed the minimum separation of the Matérn events). All results were obtained from 10000 iterations of our MCMC sampler after discarding the first 1000 samples as burn-in. A MATLAB implementation of our sampler on an Intel Core 2 Duo 3-Ghz central processor unit took around 1 min to produce 10000 MCMC samples. To correct for correlations across MCMC iterations, and to assess mixing for our MCMC sampler, we used software from

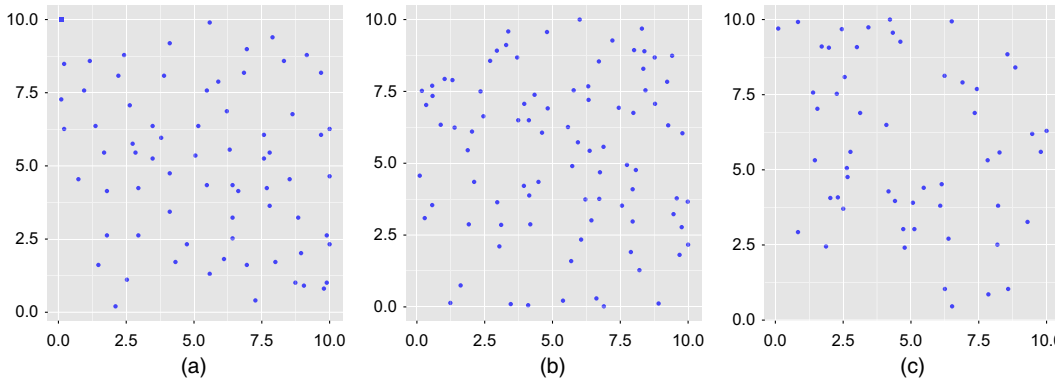

Fig. 2. (a) Swedish pine tree data set, and nerve fibre entry points for (b) mild and (c) moderate neuropathy

Table 1. Effective sample sizes (per 1000 samples) for the Matérn type III process

	Sample size for Swedish pine data set (hard-core model)	*Sample size for nerve fibre data (probabilistic thinning)*
Matérn interaction radius	344.51	121
Latent times (averaged across observations)	989.4	880.1
Primary Poisson intensity	954.7	680.2

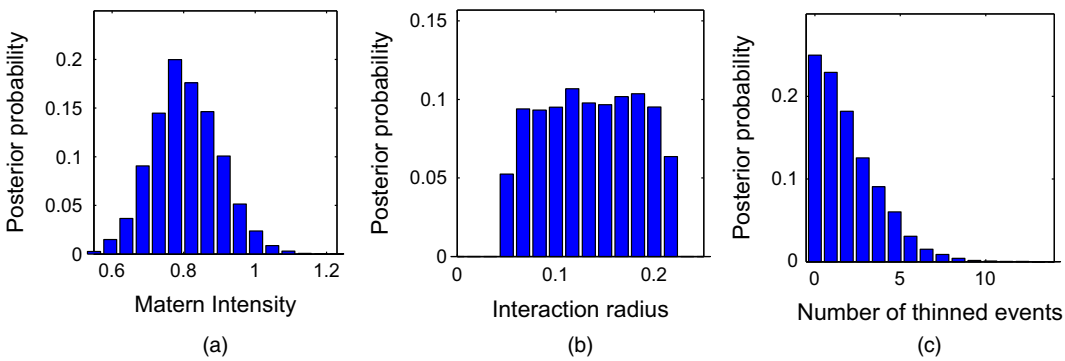

Fig. 3. Swedish tree data set: Matérn hard-core posteriors of (a) intensity, (b) thinning radius and (c) number of thinned events

Plummer *et al*. (2006) to estimate the effective sample sizes of the various quantities; this gives the number of independent samples with the same information content as the MCMC output. Table 1 shows these values, suggesting that our sampler mixes rapidly.

Figs 3(a) and 3(b) show the posterior distributions over λ and R for the Swedish pine data set. Recall that the area of the square is 100, and the data set has 71 Matérn events. The fact that the posterior over the intensity λ concentrates around 0.8 suggests that the number of thinned events is small. This is confirmed by Fig. 3(c) which shows that the posterior distribution over the

number of thinned points is less than 10. Although the data in Fig. 2 are clearly underdispersed, few thinned events suggests a weak repulsion. This mismatch is a consequence of occasional nearby events in the data, and the fact that the hard-core model requires all events to have the same radius. Consequently, the interaction radius must be bounded by the minimum interevent distance. Fig. 3(b) shows that, under the posterior, the interaction radius is bounded around 0.2, which is much less than the typical interevent distance.

Fig. 4 quantifies this lack of fit, showing Besag's L-function $L(r)$ (Besag, 1977). For convenience, by $L(r)$ we mean $L(r) - r$, which measures the excess number of events (relative to a Poisson process) within a distance r of an arbitrary event. We estimated this (and other statistics) by using the `spatstat` package. A Poisson process has $L(r) = 0$, whereas $L(r) < 0$ indicates greater regularity than Poisson at distance r. The full magenta curves in Fig. 4 show the empirical $L(r)$ as a function of distance r. We see that the pine tree data are much more regular than Poisson distributed, especially over distances between 0.5 and 1.5, where a clear repulsive trend is seen. The blue envelope in Fig. 4(a) shows the posterior predictive distribution of the L-function from the type III hard-core process. For this, at each MCMC iteration we generated

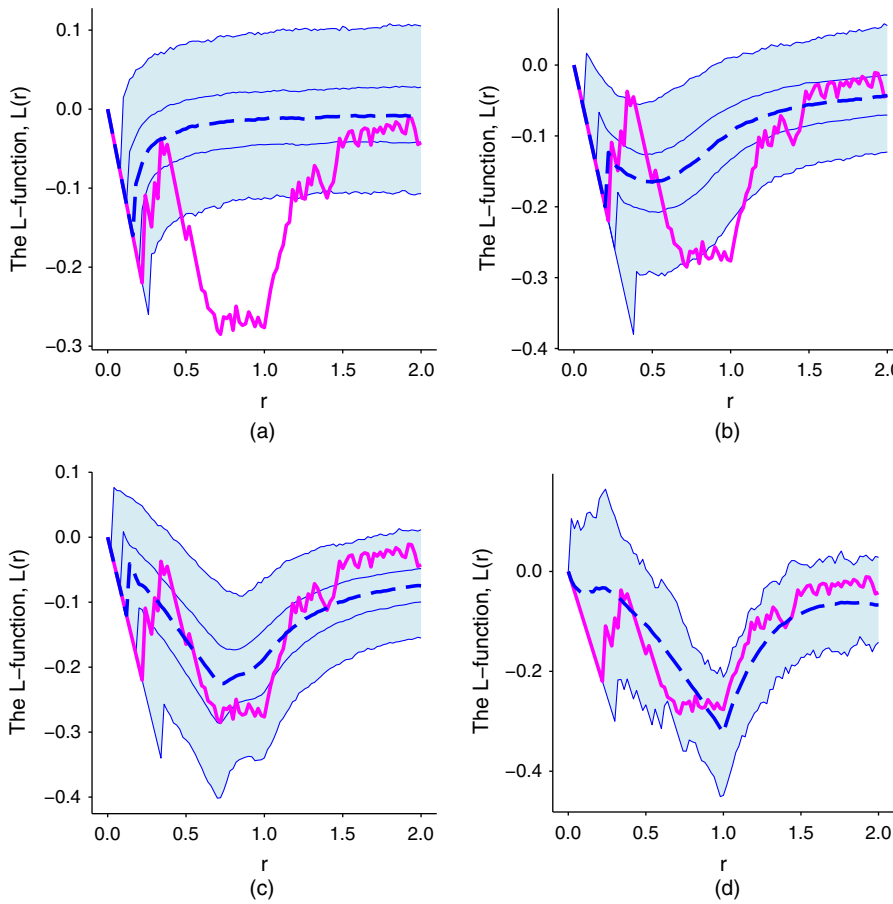

Fig. 4. Posterior predictive values of L-functions for the Swedish pine tree data set (——, data; – – –, model predictions): (a) Matérn hard-core model; (b) soft-core model; (c) probabilistic thinning; (d) predictive values for a Strauss process fit

a new realization of the hard-core process with the current parameter values, and estimated $L(r)$ for that. The broken curve is the mean $L(r)$ across MCMC samples, and the two bands show 25–75% and 2.5–97.5% values. Although the data are repulsive, the predictive intervals for the hard-core process in Fig. 4 do not match the trend that is observed in the data, corresponding instead to a Poisson process with a small R.

We next fit the data by using the soft-core model (where each Matérn event has its own interaction radius). We allowed the radii to take values in the interval $[r_L, r_U]$ and placed conjugate priors on these limits of this interval. Fig. 4(b) shows the posterior predictive intervals for the L-functions and, although we do see some of the trends in the data, the fit is still not adequate. The problem now is that typical interevent distances in the data are around 0.5–1; however, the presence of a few nearby points requires a small lower bound r_L. The model then predicts many more nearby points than are observed in the data.

One remedy is to use some other distribution over the radii (e.g. a truncated Gaussian distribution), so that, rather than driving the Matérn parameters, occasional nearby events can be viewed as atypical. Instead, we use our new extension with probabilistic thinning. We use the simplest version, with an interaction radius R that is common to all Matérn events, and a thinning probability p. We place a unit mean gamma$(1, 1)$ and a flat beta$(1, 1)$ prior on these. As before, we place a gamma$(1, 1)$ prior on the Poisson intensity λ.

Fig. 4(c) shows the results from our MCMC sampler. The predicted L-function values now mirror the empirical values much better, and in general we find that probabilistic thinning strikes a better balance between simplicity and flexibility. The on-line supplementary material also includes plots for another statistic, the J-function (van Lieshout and Baddeley, 1996), agreeing with these conclusions. Fig. 5 plots the posteriors over the interaction radius, the thinning probability and the number of thinned events. The first and last are much larger than from the hard-core model and coupled with a thinning probability of around 0.6 confirm that the data arise from a repulsive process.

We also estimated point process Fano factors by dividing the space into 25 2×2 squares, and counting the number of events falling in each. Our estimate was then the standard deviation of these counts divided by the mean. A Poisson distribution would have this equal to 1, with values less than 1 suggesting regularity. Fig. 5(d) shows the empirical Fano factor (the red circle), as well as posterior predictive values (again obtained by calculating the Fano factors of synthetic data sets generated by each MCMC iteration). The plots confirm both that the data are repulsive and that the model captures this aspect of repulsiveness.

As a comparison, we include also a fit from a Gibbs-type process (a Strauss process, in particular) in Fig. 4(d). These predictions of the L-function were produced from an approximate maximum likelihood estimation fit (in particular, a maximum pseudolikelihood estimate by using routines from `spatstat`). For this data set, the fit is comparable with that of the Matérn process with probabilistic thinning (though the L- and J-functions are only just contained in the 95% prediction band). Our next section considers a data set where the Strauss process fares much worse.

6.1. Nerve fibre data

For our next experiment, we consider a data set of nerve fibre entry locations that were recorded from patients suffering from diabetes (Waller *et al.*, 2011). Fig. 2 plots these data for two patients at two stages of neuropathy: mild and moderate. As noted in Waller *et al.* (2011), there is an increased clustering moving from mild to moderate, and this is confirmed by Fig. 6, which shows the empirical values for the L-function (full magenta curves). Also included are posterior pre-

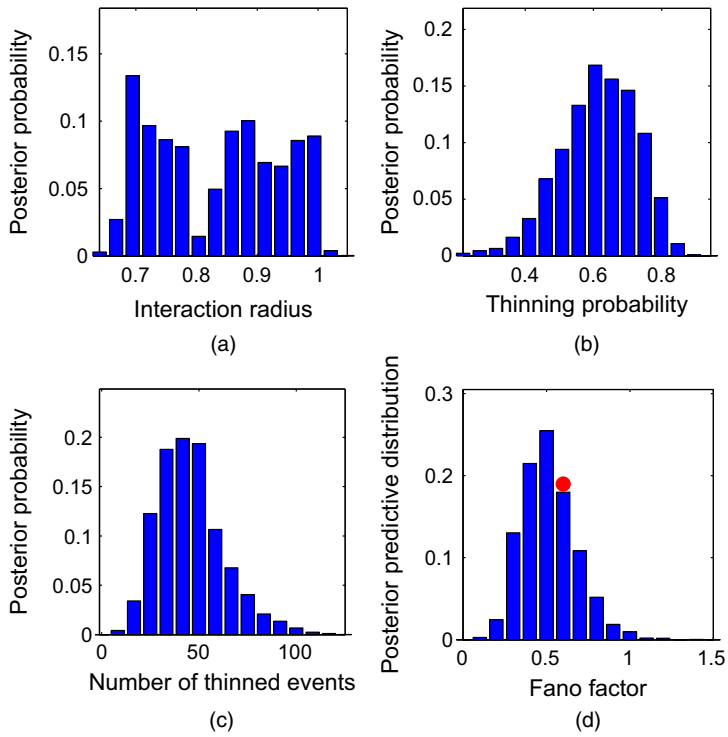

Fig. 5. Swedish tree data set: posterior distributions of (a) interaction radius R, (b) thinning probability p and (c) number of thinned events, and (d) empirical (●) as well posterior predictive distribution over the Fano factor

dictive values obtained from 10 000 posterior samples from the Matérn model with probabilistic thinning. Figs 6(a) and 6(c) for mild neuropathy show a repulsive effect (especially a dip in the L-function around the 0.5-mark), and this is captured by the model. By contrast, the moderate case exhibits repulsion and clustering at different scales, and the model settles with a Poisson process whose predictive intervals include such behaviour.

To show the difference in the fits for the two conditions, at each MCMC iteration we calculate a statistic measuring repulsive influence: the thinning probability multiplied by the squared interaction radius. Fig. 7 plots this for the two cases, and we see that there is more repulsion for the mild case. Another approach is to fix the thinning probability (to say 0.75), and to plot the posterior distributions over the interaction radius. We obtained similar results but have not reported them here. Distributions like these are important diagnostics towards the automatic assessment of neuropathy.

Fig. 6 shows fits for a Strauss process: this is much worse, with the estimate essentially reducing to a Poisson process. The failure of this model is partly because it is not as flexible as our Matérn model with probabilistic thinning; another factor is the use of a maximum likelihood estimation estimate as opposed to a fully Bayesian posterior. Although it is possible to use a more complicated Gibbs-type process, inference (especially Bayesian inference) becomes correspondingly more difficult. Furthermore, as we demonstrate next, it is easy to incorporate non-stationarity in Matérn processes; doing this for a Gibbs-type process is not at all straightforward. In the on-line supplementary material, we include similar results for the J-function as well. Waller *et al.* (2011) obtained maximum likelihood estimates of the interaction radius R of

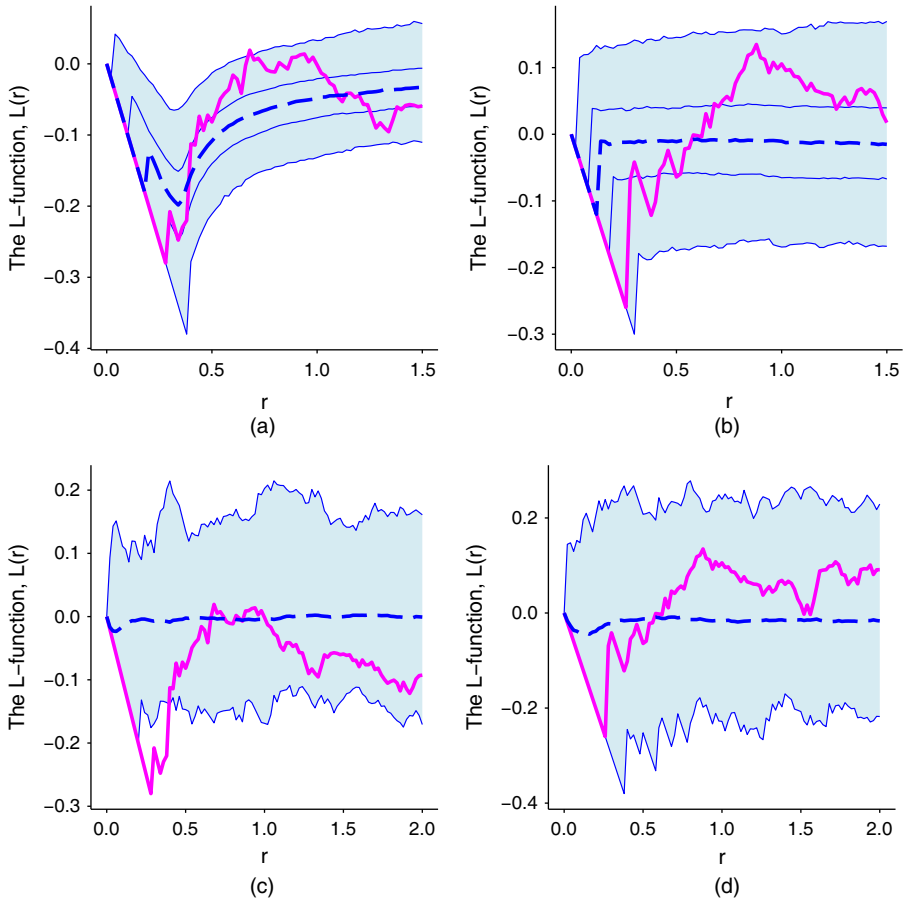

Fig. 6. Posterior predictive values of the *L*-functions for the Matérn model with probabilistic thinning for (a) mild and (b) moderate neuropathy and (c), (d) corresponding predictive values for Strauss process fits (——, data; — —, model predictions)

a Matérn type II hard-core process. Our Bayesian approach can be naturally extended to allow variability among patients through hierarchical modelling. In the next section, we show how it can incorporate spatial inhomogeneity as well.

7. Non-stationary Matérn processes

In many settings, it is useful to incorporate non-stationarity into repulsive models. For example, factors like soil fertility, rainfall and terrain can affect the density of trees at different locations. We shall consider a data set of locations of Greyhound bus stations: to allocate resources efficiently, these will be underdispersed. At the same time, we expect larger intensities in urban than in more remote areas. In such situations, it is important to account for non-stationarity while estimating repulsiveness.

A simple approach is to allow the intensity of the primary Poisson $\lambda(s)$ to vary over $\mathcal{S}$. A flexible model of such a non-stationary intensity function is a transformed Gaussian process (GP). Let $\mathcal{K}(\cdot,\cdot)$ be the covariance kernel of a GP, $\hat{\lambda}$ a positive scale parameter and $\sigma(x)=\{1+\exp(-x)\}^{-1}$ the sigmoidal transformation. Then $\lambda(\cdot)$ is a random function defined as

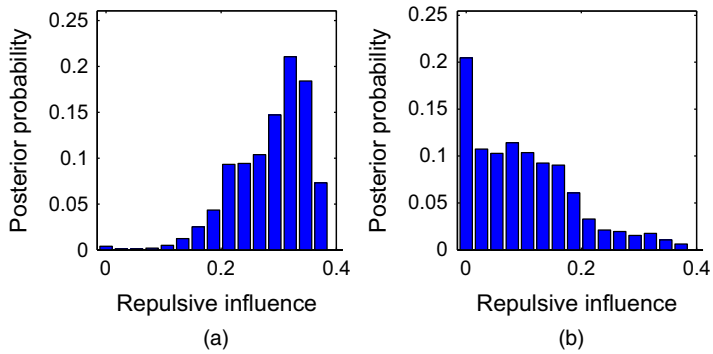

Fig. 7. Posterior distribution over thinning area times thinning probability for (a) mild and (b) moderate neuropathy

$$\lambda(\cdot) = \hat{\lambda}\sigma\{l(\cdot)\}, \qquad l(\cdot) \sim \mathcal{GP}(0, \mathcal{K}) \tag{19}$$

This model is closely related to the log-Gaussian Cox process, with a sigmoid link function instead of an exponential. The sigmoid transformation serves two purposes: to ensure that the intensity $\lambda(s)$ is non-negative, and to provide a bound $\hat{\lambda}$ on the Poisson intensity. As shown in Adams *et al.* (2009), such a bound makes drawing events from the primary process possible by a clever application of the thinning theorem. We introduced the thinning theorem in Section 5; we state it formally below.

Theorem 3 (thinning theorem; Lewis and Shedler (1979)). Let E be a sample from a Poisson process with intensity $\hat{\lambda}(s)$. For some non-negative function $\lambda(s) \leqslant \hat{\lambda}(s), \forall s \in \mathcal{S}$, assign each point $e \in E$ to F with probability $\lambda(e)/\hat{\lambda}(e)$. Then F is a draw from a Poisson process with intensity $\lambda(s)$.

From equation (19) we have $\hat{\lambda} \geqslant \lambda(s)$. Following the thinning theorem, we can obtain a sample from the rate $\lambda(s)$ inhomogeneous Poisson process by thinning a random sample E from a rate $\hat{\lambda}$ homogeneous Poisson process. This allows us to instantiate the random intensity $\lambda(s)$ only on the elements of E, avoiding any need to evaluate integrals of the random function $\lambda(s)$. We then have the following retrospective sampling scheme: sample a homogeneous Poisson process with intensity $\hat{\lambda}$, and instantiate the GP $l(\cdot)$ on this sequence. Keeping each element e with probability $\sigma\{l(e)\}$, we have an *exact* sample from the inhomogeneous primary process. Call this F, and call the thinned events $\tilde{F}$. We can then use any of the Matérn thinning schemes that were outlined previously to thin F further, resulting in an inhomogeneous repulsive process G. There are now two stages of thinning: the first is an application of the Poisson thinning theorem to obtain the inhomogeneous primary process F from the homogeneous Poisson process E and, the second, the Matérn thinning to obtain G from F. The following algorithm 1 outlines the generative process of an inhomogeneous Matérn type III process on $\mathcal{S}$.

The *input* is a GP prior $\mathcal{GP}(0, \mathcal{K})$ on the space $\mathcal{S}$, a constant $\hat{\lambda}$ and the thinning kernel parameters θ and the *output* is a sample G from the non-stationary Matérn type III process.

Step 1: sample E from a homogeneous Poisson process with intensity $\hat{\lambda}$.
Step 2: instantiate the GP $l(\cdot)$ on these points. Call vector l_E.
Step 3: keep a point $e \in E$ with probability $\sigma\{l(e)\}$; otherwise thin it. The surviving points form the primary process F.
Step 4: assign F a set of random birth times T^F, independently and uniformly on $[0, 1]$.

Step 5: proceed through elements of F in order of birth. At each element, evaluate the shadow of the previous surviving elements, and keep it or thin it as appropriate.
Step 6: the surviving points G form the inhomogeneous Matérn type III point process.

As in section 5, we place priors on $\hat{\lambda}$ as well as θ. We also place hyperpriors on the hyperparameters of the GP covariance kernel.

7.1. Inference for the inhomogeneous Matérn type III process

Proceeding as in Section 4.1, it follows that the events $\tilde{G}^+$ thinned by the repulsive kernel are conditionally distributed as an inhomogeneous Poisson process with intensity $\lambda(s)\,\mathscr{H}(s,t;G^+)$ (with the λ from corollary 1 replaced by $\lambda(s)$). From the thinning theorem, simulating from such a process is a simple matter of thinning events of a homogeneous, rate $\hat{\lambda}$ Poisson process, exactly as outlined in the previous section. Having reconstructed the inhomogeneous primary Poisson process, the update rules for the birth times T_G and the thinning kernel parameters θ are identical to those for the homogeneous case.

The only new idea involves updating the random function $\lambda(s)$ (more precisely, the latent GP $l(s)$, and the scaling factor $\hat{\lambda}$). To do this, we first instantiate $\tilde{F}$, the events of E thinned in constructing the inhomogeneous primary process F. For this step, Adams *et al.* (2009) constructed a Markov transition kernel, involving a set of birth–death moves that updated the number of thinned events, as well as a sequence of moves that perturbed the locations of the thinned events. This kernel was set up to have as equilibrium distribution the required posterior over the thinned events. Instead, similar in spirit to our idea of jointly simulating the thinned Matérn events, it is possible to produce a conditionally independent sample of $\tilde{F}$. Instead of a number of local moves that perturb the current setting of $\tilde{F}$ in the Markov chain, we can discard the old thinned events and jointly produce a new sample given the rest of the variables. The required joint distribution is given by the following corollary of theorem 3.

Corollary 2. Let F be a sample from a Poisson process with intensity $\lambda(s)$, produced by thinning a sample E from a Poisson process with intensity $\hat{\lambda}$. Call the thinned events $\tilde{F}$. Then, given F, $\tilde{F}$ is a Poisson process with intensity $\hat{\lambda} - \lambda(s)$.

Proof. A direct approach uses the Poisson densities that were defined in theorem 1. More intuitive is the following fact: by symmetry, the thinning theorem implies that the thinned events $\tilde{F}$ are distributed as a Poisson process with intensity $\hat{\lambda} - \lambda(s)$. By the independence property of the Poisson process, F and $\tilde{F}$ are independent, and the result follows.

Applying this result to simulate the thinned events $\tilde{F}$ is a straightforward application of the thinning theorem: sample from a homogeneous Poisson process with intensity $\hat{\lambda}$, conditionally instantiate the function $\lambda(\cdot)$ on these points (given its values on F) and keep element $\tilde{f}$ with probability $1 - \sigma(\tilde{f})$.

Having reconstructed the sequence $E = F \cup \tilde{F}$, we can update the values of the GP that was instantiated on E. Recall that an element $e \in E$ is assigned to F with probability $\sigma\{l(s)\}$; otherwise it is assigned to $\tilde{F}$. Thus updating the GP reduces to updating the latent GP in a classification problem with a sigmoidal link function. There are a variety of sampling algorithms to simulate such a latent GP; in our experiments we used the elliptical slice sampling from Murray *et al.* (2010). Finally, as the number of elements of E is Poisson distributed with rate $\hat{\lambda}$, a gamma prior results in a gamma posterior as in Section 5.3. We describe our overall sampler in detail in the following algorithm 2 (MCMC update for inhomogeneous Matérn type III process on $\mathcal{S}$).

The *inputs* are Matérn events with birth times $G^+ \equiv (G, T^G)$, thinned primary events $\tilde{G}^+ \equiv (\tilde{G}, T^{\tilde{G}})$ and thinned Poisson events $\tilde{F}$, a GP realization l_E on $E \equiv G \cup \tilde{G} \cup \tilde{F}$, and parameters

$\hat{\lambda}$ and θ. The *outputs* are new realizations T^G_{new}, $\tilde{G}^+_{\text{new}}$, $\tilde{F}_{\text{new}}$, $\hat{\lambda}$ and θ, and a new instantiation of the GP on $G \cup \tilde{G}_{\text{new}} \cup \tilde{F}_{\text{new}}$.

Step 1: sample the Matérn thinned events $\tilde{G}^+$.

(a) Discard the old Matérn thinned event locations $\tilde{G}^+$.
(b) Sample events $A^+ \equiv (A, T^A)$ from a rate $\hat{\lambda}$ Poisson process on $\mathcal{S} \times \mathcal{T}$.
(c) Sample $l_A | l_E$ (conditionally from a multivariate normal distribution).
(d) Keep a point $a \in A$ with probability $\sigma\{l(a)\}\mathcal{H}(a; G^+)$; otherwise thin it.
(e) The surviving points form the new Matérn thinned events $\tilde{G}^+_{\text{new}}$.
(f) Discard GP evaluations on old Matérn events, and add the new events to l_E.

Step 2: sample the Poisson thinned events $\tilde{F}$.

(a) Define $E_{\text{new}} \equiv G \cup \tilde{G}_{\text{new}} \cup \tilde{F}$.
(b) Discard the old thinned primary Poisson events $\tilde{F}^+$.
(c) Sample events $B^+ \equiv (B, T^B)$ from a rate $\hat{\lambda}$ Poisson process on $\mathcal{S} \times \mathcal{T}$.
(d) Sample $l_B | l_{E_{\text{new}}}$ (conditionally from a multivariate normal).
(e) Keep a point $b \in B$ with probability $1 - \sigma\{l(b)\}$; otherwise thin it. The surviving points form the new Poisson thinned events $\tilde{F}_{\text{new}}$.
(f) Define $E_{\text{new}} \equiv G \cup \tilde{G}_{\text{new}} \cup \tilde{F}_{\text{new}}$.

Step 3: sample the Matérn birth times. For each Matérn event g, update birth time T^G_g given all other variables.
Step 4: sample the GP values $l_{E_{\text{new}}}$. We used elliptical slice sampling (Murray *et al.*, 2010).
Step 5: sample the parameters $\hat{\lambda}$ *and* θ. Sample $\hat{\lambda}$ and θ given the remaining variables.

7.2. Experiments

We consider a data set of the locations of 79 Greyhound bus stations in three states in the south-east USA (North Carolina, South Carolina and Georgia) (obtained from `http://www.poi-factory.com/`). Fig. 8(a) plots these locations along with the inferred intensity function from our non-stationary Matérn process. We modelled the intensity function $\lambda(\cdot)$ by using a sigmoidally transformed GP with zero mean and a squared exponential kernel. We placed log-normal hyperpriors on the GP hyperparameters ((shape, scale) as $(1, 1)$), and a gamma$(10, 10)$ prior on the scale parameter $\hat{\lambda}$ (we allowed larger variance since there are two levels of thinning). To ease comparison across two cases, we fixed the thinning probability to 0.75 and placed a gamma$(1, 1)$ prior on the interaction radius R (which we expect to be less than 1° of latitude or longitude). The contours in Fig. 8(a) correspond to the posterior mean of $\lambda(\cdot)$, obtained from 10 000 MCMC samples following algorithm 2. We used elliptical slice sampling (Murray *et al.*, 2010) to resample the GP values (step 4 in algorithm 2), and the GP hyperparameters were resampled by slice sampling (Murray and Adams, 2010).

We see a strong non-stationarity, driven mainly by the Atlantic Ocean to the lower right, and the absence of recordings to the top left (in Tennessee). Not surprisingly, the intensity function takes large values over the three states of interest. Fig. 8(b) shows the same analysis, now restricted to North Carolina (with observations in the other states discarded). Here we see more clearly a fine structure in the intensity function with two peaks: one in the research triangle area (Raleigh–Durham–Chapel Hill) to the west, and the other corresponding to the Charlotte metropolitan area.

Fig. 9 plots the posterior distributions over the thinning radius R for both cases (recall that, since we used a fixed thinning probability of 0.75, R is a direct measure of the repulsive effect).

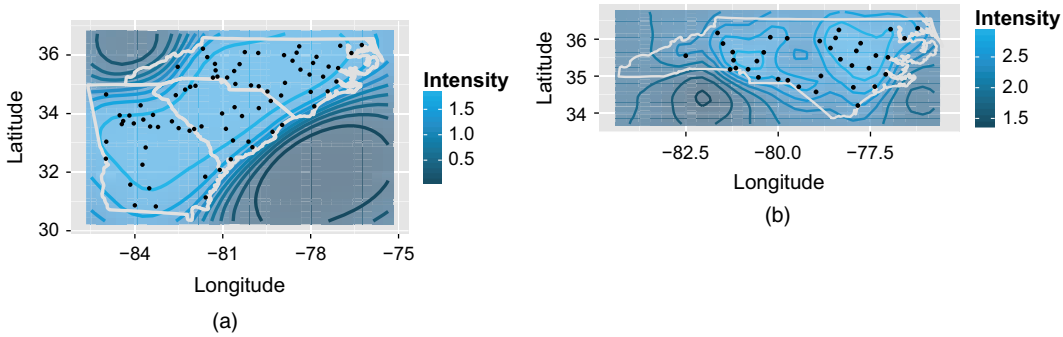

Fig. 8. Posterior mean of the intensity for the Greyhound bus station data set: (a) all three states; (b) North Carolina

In both cases, we see that this distribution is peaked around 0.15–0.2°, with the North Carolina data set having a slightly clearer repulsive effect. For the grouped analysis, we required all three states to share the same unknown R while having their own non-stationarity: more generally, one can use a hierarchical model over R to shrink or cluster the radii across different states. The resulting posterior over R can help to understand variations in pricing, reliability and delays, as well as assist towards future developmental work (Şahin and Süral, 2007).

Fig. 10 compares fits for a non-stationary Matérn process and a Poisson process, using the inhomogeneous L-function (Baddeley *et al.*, 2000). We see that the Poisson process (Fig, 10(b)) fails to capture the empirical values over distances up to around 0.2°. This agrees with Fig. 9 that repulsion occurs over these distances. The Matérn process (Fig. 10(a)) provides a much better fit. We also include plots for the the non-stationary J-function in the on-line supplementary material. This did not indicate much deviation from a Poisson process, though, as Baddeley *et al.* (2000) pointed out, this does not imply that the point process is Poisson.

8. Discussion

We have described a Bayesian framework for modelling repulsive interactions between events of a point process based on the Matérn type III point process. Such a framework allows flexible and intuitive repulsive effects between events, with parameters that are interpretable and realistic. We developed an efficient MCMC sampling algorithm for posterior inference in these models and applied our ideas to data sets of locations of trees, nerve fibres and bus station locations.

Some interesting directions are worth following. Although we considered events in only a two-dimensional space, it is easy to generalize to higher dimensions to model, say, the distribution of galaxies in space or features in some feature space. As with all repulsive point processes, very high densities lead to inefficiency (in our case because of very high primary process intensities). It is important to understand the theoretical and computational limitations of our model in this regime better. Whereas we assumed that the Matérn events G were observed perfectly, there is often noise in this observation process. In this case, given the observed point process G_{obs}, we must instantiate the latent Matérn process G. Simulating the locations of the events in G would require incremental updates and, if we allow for missing or extra events, we would need a birth–death sampler as well. A direction for future study is to see how these steps can be performed efficiently. Having instantiated G, all other variables can be simulated as outlined in this paper. A related question concerns whether our ideas can be extended to develop efficient samplers for Matérn type I and II processes as well.

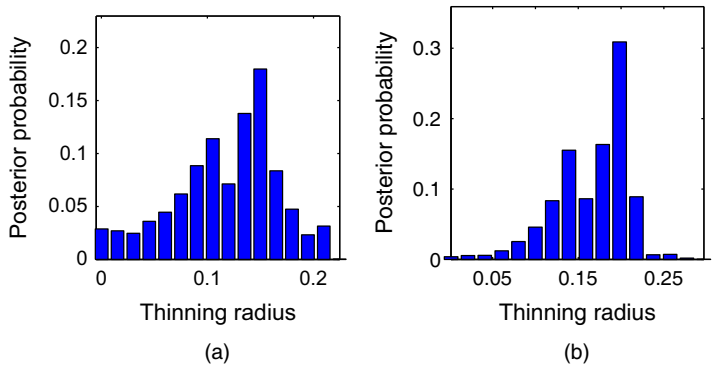

Fig. 9. Posterior over the thinning radius for (a) all three states and (b) North Carolina

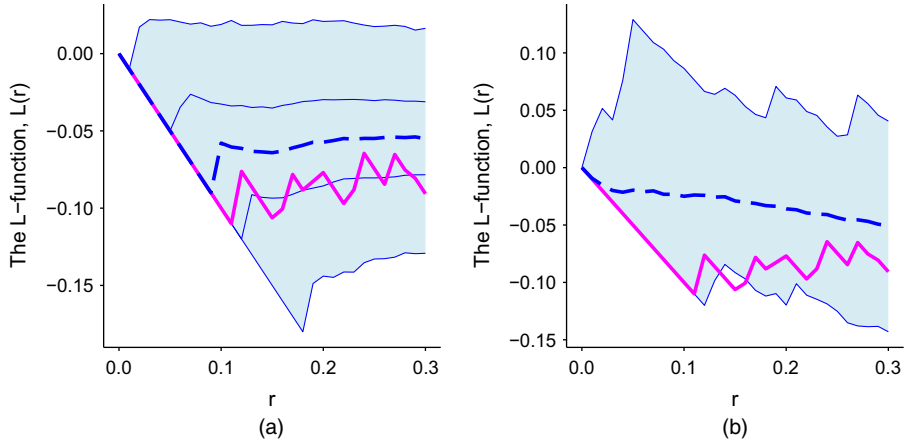

Fig. 10. Inhomogeneous L-function for the Greyhound bus station data set (——, data; – – –, envelope median): (a) posterior predictive values for the non-stationary Matérn process with probabilistic thinning; (b) fit of an inhomogeneous Poisson process

Our models assumed homogeneity in the repulsive properties of the Matérn events. An interesting extension is to allow, say, the interaction radius or the thinning probability to vary spatially (rather than the Poisson intensity $\lambda(\cdot)$). Similarly, one might assume a clustering of these repulsive parameters; this is useful in situations where the Matérn observations represent cells of different kinds. Finally, it is of interest to apply our ideas to hierarchical models that do not necessarily represent point pattern data, for instance to encourage diversity between cluster parameters in a mixture model.

Acknowledgements

This work was partially supported by the Defense Advance Research Projects Agency's 'Mathematics of sensing, exploitation and execution' programme and by National Science Foundation grant IIS-1421780. We thank Aila Särkkä for providing us with the nerve fibre data. The Greyhound bus station data set was obtained from `http://www.poi-factory.com/`.

Appendix A: Proof of theorem 2

As described in Section 4.1, $G^+ = (G, T^G)$ is a sequence in the union space $(\mathcal{S} \times \mathcal{T})^\cup$. Its elements are

ordered by the last dimension, so T^G is an increasing sequence. Let $|G^+|$, the length of G^+, be k. G^+ is obtained by thinning $F^+ = (F, T^F)$, which is a sample from a homogeneous Poisson process with intensity λ. Let the size of F^+ be $n \geqslant k$, and call the thinned points $\tilde{G}^+$. From theorem 1, the density of F^+ with respect to the measure $\mu^\cup$ is

$$p(F^+) = \exp\{-\lambda\mu(\mathcal{S}\times\mathcal{T})\}\lambda^n. \tag{20}$$

Recall the definition of $\mathcal{H}_\theta(s,t;G^+)$, the shadow of G^+ (with K_θ the thinning kernel):

$$\mathcal{H}_\theta(s,t;G^+) = 1 - \prod_{(s^*,t^*)\in G^+}\{1 - I(t>t^*)\,K_\theta(s^*,s)\}. \tag{21}$$

We traverse sequentially through F^+, assigning element i to the Matérn process G^+ or the thinned set $\tilde{G}^+$ with probability determined by the shadow $\mathcal{H}_\theta(\cdot;G^+)$, and

$$p(G^+,\tilde{G}^+|F^+) = \prod_{(s,t)\in\tilde{G}^+}\mathcal{H}_\theta(s,t;G^+)\prod_{(s,t)\in G^+}\{1-\mathcal{H}_\theta(s,t;G^+)\}. \tag{22}$$

In our notation above, the shadow $\mathcal{H}_\theta$ thins or keeps points to form $G^+ = (G, T^G)$ and $\tilde{G}^+ = (\tilde{G}, T^{\tilde{G}})$, but $\mathcal{H}_\theta$ also depends on G^+. There is no circularity: the shadow of a later point cannot affect an earlier point. The joint probability density with respect to $\mu^\cup$ is

$$\begin{aligned} p(G^+,\tilde{G}^+,F^+) &= p(G^+,F^+) = p(G^+,\tilde{G}^+) \\ &= \exp\{-\lambda\mu(\mathcal{S}\times\mathcal{T})\}\lambda^n\prod_{(s,t)\in\tilde{G}^+}\mathcal{H}_\theta(s,t;G^+)\prod_{(s,t)\in G^+}\{1-\mathcal{H}_\theta(s,t;G^+)\}. \end{aligned} \tag{23}$$

Note that $\mu^\cup(\mathrm{d}F^+) = \mu(\mathrm{d}F^+) = \mu(\mathrm{d}G^+)\,\mu(\mathrm{d}\tilde{G}^+) = \mu^\cup(\mathrm{d}G^+)\,\mu(\mathrm{d}\tilde{G}^+)$, where we abuse the notation slightly by letting μ be Lebesgue measure of dimensionality determined by its argument. Then, integrating out the $\mathcal{S}$-locations of $n-k$ thinned elements,

$$p(G^+,T^{\tilde{G}}) = \exp\{-\lambda\mu(\mathcal{S}\times\mathcal{T})\}\lambda^k\prod_{(s,\,t)\in G^+}\{1-\mathcal{H}_\theta(s,t;G^+)\}\prod_{t^{\tilde{g}}\in T^{\tilde{G}}}\left\{\lambda\int_{\mathcal{S}}\mathcal{H}_\theta(s,t^{\tilde{g}};G^+)\,\mu(\mathrm{d}s)\right\}. \tag{24}$$

We now integrate out the values of $T^{\tilde{G}}$, noting that it is an ordered sequence of $n-k$ elements in $[0,1]$. We are left with $p(G^+, |T^{\tilde{G}}| = n-k)$, the joint probability of a sequence G^+ and that there are $n-k$ thinned events:

$$\begin{aligned} p(G^+,|T^{\tilde{G}}| = n-k) = \exp\{-\lambda\mu(\mathcal{S}\times\mathcal{T})\}\lambda^k\prod_{(s,t)\in G^+}\{1-\mathcal{H}_\theta(s,t;G^+)\} \\ \times\frac{\left\{\lambda\displaystyle\int_{\mathcal{S}\times\mathcal{T}}\mathcal{H}_\theta(s,t;G^+)\,\mu(\mathrm{d}s\,\mathrm{d}t)\right\}^{n-k}}{(n-k)!}. \end{aligned} \tag{25}$$

Finally, summing over values of $|T^{\tilde{G}}|$,

$$p(G^+) = \exp\left[-\lambda\int_{\mathcal{S}\times\mathcal{T}}\{1-\mathcal{H}_\theta(s,t;G^+)\}\mu(\mathrm{d}s\,\mathrm{d}t)\right]\prod_{(s,t)\in G^+}\lambda\{1-\mathcal{H}_\theta(s,t;G^+)\}. \tag{26}$$

References

Adams, R. P. (2009) Kernel methods for nonparametric Bayesian inference of probability densities and point processes. *PhD Thesis*. University of Cambridge, Cambridge.

Adams, R. P., Murray, I. and MacKay, D. J. C. (2009) Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proc. 26th Int. Conf. Machine Learning* (eds L. Bottou and M. Littman), pp. 9–16. Montreal: Omnipress.

Affandi, R. H., Fox, E., Adams, R. P. and Taskar, B. (2014) Learning the parameters of determinantal point process kernels. In *Proc. 31st Int. Conf. Machine Learning* (eds E. P. Xing and T. Jebara), pp. 1224–1232.

Andrieu, C. and Roberts, G. O. (2009) The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, **37**, 697–725.

Baddeley, A. J., Møller, J. and Waagepetersen, R. (2000) Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statist. Neerland.*, **54**, 329–350.

Baddeley, A. and Turner, R. (2005) Spatstat: an R package for analyzing spatial point patterns. *J. Statist. Softwr.*, **12**, 1–42.
Besag, J. E. (1977) Discussion on 'Modelling spatial patterns' (by B. D. Ripley). *J. R. Statist. Soc.* B, **39**, 193–195.
Brown, E. N., Barbieri, R., Eden, U. T. and Frank, L. M. (2004) Likelihood methods for neural spike train data analysis. In *Computational Neuroscience: a Comprehensive Approach*, vol. 7 (ed. J. Feng), ch. 9, pp. 253–286. Boca Raton: CRC Press.
Daley, D. J. and Vere-Jones, D. (2008) *An Introduction to the Theory of Point Processes*. New York: Springer.
Green, P. J. (2003) Trans-dimensional Markov chain Monte Carlo. In *Highly Structured Stochastic Systems* (eds P. J. Green, N. L. Hjort and S. Richardson). Oxford: Oxford University Press.
Hill, M. O. (1973) The intensity of spatial pattern in plant communities. *J. Ecol.*, **61**, 225–235.
Hough, J. B., Krishnapur, M., Peres, Y. and Virág, B. (2006) Determinantal processes and independence. *Probab. Surv.*, **3**, 206–229.
Huber, M. L. and Wolpert, R. L. (2009) Likelihood-based inference for Matérn type III repulsive point processes. *Adv. Appl. Probab.*, **41**, 958–977.
Kendall, M. G. (1939) The geographical distribution of crop productivity in England (with discussion). *J. R. Statist. Soc.*, **102**, 21–62.
Knox, G. (2004) Epidemiology of childhood leukemia in Northumberland and Durham. In *The Challenge of Epidemiology: Issues and Selected Readings*, vol. 1, pp. 384–392. Washington DC: Pan American Health Organization.
Lewis, P. A. W. and Shedler, G. S. (1979) Simulation of nonhomogeneous Poisson processes with degree-two exponential polynomial rate function. *Ops Res.*, **27**, 1026–1040.
van Lieshout, M. N. M. and Baddeley, A. J. (1996) A nonparametric measure of spatial interaction in point patterns. *Statist. Neerland.*, **50**, 344–361.
Matérn, B. (1960) Spatial variation. *Medd. Statns Skogsforskinst.*, **49**, no. 5.
Matérn, B. (1986) *Spatial Variation*, 2nd edn. Berlin: Springer.
Mateu, J. and Montes, F. (2001) Likelihood inference for Gibbs processes in the analysis of spatial point patterns. *Int. Statist. Rev.*, **69**, 81–104.
Møller, J., Huber, M. L. and Wolpert, R. L. (2010) Perfect simulation and moment properties for the Matérn type-III process. *Stoch. Processes Appl.*, **120**, 2142–2158.
Møller, J. and Waagepetersen, R. P. (2007) Modern statistics for spatial point processes. *Scand. J. Statist.*, **34**, 643–684.
Murray, I. and Adams, R. P. (2010) Slice sampling covariance hyperparameters of latent Gaussian models. In *Advances in Neural Information Processing Systems*, vol. 23 (eds J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel and A. Culotta), pp. 1723–1731. Vancouver: Curran Associates.
Murray, I., Adams, R. P. and MacKay, D. J. (2010) Elliptical slice sampling. In *Proc. 13th Int. Conf. Artificial Intelligence and Statistics* (eds Y. W. Teh and M. Titterington), vol. 9, pp. 541–548.
Peebles, P. J. E. (1974) The nature of the distribution of galaxies. *Astron. Astrophys.*, **32**, 197.
Plummer, M., Best, N., Cowles, K. and Vines, K. (2006) CODA: convergence diagnosis and output analysis for MCMC. *R News*, **6**, 7–11.
Ripley, B. D. (1988) *Statistical Inference for Spatial Processes*. Cambridge: Cambridge University Press.
Şahin, G. and Süral, H. (2007) A review of hierarchical facility location models. *Comput. Ops Res.*, **34**, 2310–2331.
Scardicchio, A., Zachary, C. and Torquato, S. (2009) Statistical properties of determinantal point processes in high-dimensional Euclidean spaces. *Phys. Rev.* E, **79**, article 041108.
Schervish, M. J. (1995) *Theory of Statistics*. New York: Springer.
Strand, L. (1972) A model for strand growth. In *Proc. 3rd Conf. Advisory Group of Forest Statisticians*, pp. 207–216. Paris: Institut National de la Recherche Agronomique.
Waller, L. A., Särkkä, A., Olsbo, V., Myllymäki, M., Panoutsopoulou, I. G., Kennedy, W. R. and Wendelschafer-Crabb, G. (2011) Second-order spatial analysis of epidermal nerve fibers. *Statist. Med.*, **30**, 2827–2841.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Bayesian inference for Matérn repulsive processes (Supplementary material)'.