Bayesian inference for Matérn repulsive processes

Vinayak Rao∗†, Ryan P. Adams‡ and David Dunson§

Abstract

In many applications involving point pattern data, the Poisson process assumption is unrealistic, with the data exhibiting a more regular spread. Such a 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 a number of extensions of the original Matérn type-III process for both the homogeneous and inhomogeneous cases. We also derive the probability density of this generalized Matérn process. This allows us to characterize the posterior distribution of the various latent variables, and leads to a novel and efficient Markov chain Monte Carlo algorithm involving data augmentation with a Poisson process. We apply our ideas to two datasets involving the spatial locations of trees.

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 ecology (Hill, 1973), geography (Kendall, 1939), neuroscience (Dayan and Abbott, 2005), epidemiology (Knox, 2004), sociology (Hansford-Miller, 1968), and astronomy (Peebles, 1974). The simplest and most popular model for such processes is the Poisson process (e.g., Kingman (1993), Daley and Vere-Jones (2008)); however, implicit in the use of a Poisson process is the assumption of independence of the event locations. This is a simplification that is unsuitable for many real applications. Instead, it is often of interest to account for interactions between nearby events. For instance, when modelling the distribution of trees in a geographical area, competition for light and other resources results in inter-event distance that is more spread out than the Poisson process (Strand, 1972). Other applications where modelling

∗Email:var11@stat.duke.edu
†Department of Statistical Science, Duke University, USA
‡School of Engineering and Applied Sciences, Harvard University, USA
§Department of Statistical Science, Duke University, USA
such interactions is important include the distribution of cities (Glass and Tobler, 1971), galaxies (Peebles, 1974), and infected agents in epidemiological studies (Jewell et al., 2009).

In this work, we focus on a class of repulsive point processes called the Matérn type-III process. This was first introduced in (Matérn, 1960, 1986), and involves thinning events of a ‘primary’ Poisson process that are ‘too close to each other’. Starting with the simplest instance of such a process (called a ‘hardcore’ process), we introduce various extensions that provide a flexible framework for modelling repulsive processes on general spaces. We derive the probability density of the resulting process, a characterization that allows us to identify the posterior distribution over the thinned events as a simple Poisson process. This allows us to provide a simple and efficient Markov chain Monte Carlo algorithm for posterior inference.

2 Repulsive point processes

On the real line, point processes can deviate from Poisson either by being more ‘bursty’ or more ‘refractory’. In higher dimensions, these are called ‘clustered’ and ‘repulsive’ point processes respectively. Our focus in this work is on the latter, characterized by being more regular (under-dispersed) than the Poisson process. The physical reasons for such repulsion could be competition for finite resources (in the case of cities or trees, for example), interaction between rigid objects (such as cells) or repulsive forces between particles. However, developing a flexible and tractable statistical framework to study such repulsion is not straightforward on spaces more complicated than the real line. For the latter, the ordering of points can be exploited to develop a powerful and convenient framework based on renewal processes (Daley and Vere-Jones, 2008).

One framework for modelling such interactions in higher dimensions is that of Gibbs processes (Daley and Vere-Jones, 2008). Such processes arose from the statistical physics literature to describe systems of interacting particles. A Gibbs process assigns a potential energy $U_S(\theta)$ to any configuration of events $S = \{s_1, \cdots, s_n\}$, defined most generally as

$$U_S(\theta) = \sum_{i=1}^{n} \sum_{1 \leq j_1 < \cdots < j_i \leq n} \psi_i(s_{j_1}, \cdots, s_{j_i}),$$  \hspace{1cm} (1)

where $\psi_i$ is an $i$th order potential term. Usually, interactions are limited to be pairwise, and by choosing these potentials appropriately, one can flexibly model different kinds of interactions. Then, the probability density of any configuration is proportional to its exponentiated negative energy. Letting $\theta$ represent the parameters that characterize the potential energy, we have

$$p(S | \theta) = \frac{\exp(-U(S; \theta))}{Z(\theta)}.$$  \hspace{1cm} (2)

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.
Another framework for modelling repulsion is that of determinantal point processes (see for example Hough et al. (2006) or Scardicchio et al. (2009)). While these processes are mathematically and computationally elegant, the repulsion induced in these models is not intuitive or easy to specify, and there is very little Bayesian work involving such models.

3 Matérn repulsive point processes

A simple and direct approach to constructing repulsive point processes was proposed in Matérn (1960) and is based on the idea of thinning events of a Poisson process. Matérn proposed three increasingly elaborate schemes, now called the Matérn type-I, type-II, and type-III hardcore point processes. The type-I process proceeds according to the following generative mechanism: sample a primary point process from a homogeneous Poisson process with some intensity (say \( \lambda \)), and then delete all points separated by a distance less than \( R \). While the simplicity of this scheme makes it amenable to theoretical analysis, the thinning strategy here is often too aggressive. In particular, as the number of primary Poisson events increases, the probability of a point falling within a radius \( R \) of some other point also increases, thereby increasing the probability of it being thinned. As \( \lambda \) increases, this latter effect begins to dominate, so that eventually the density of Matérn events begins to decrease with \( \lambda \).

The Matérn type-II process tries to rectify this. Rather than deleting both interacting points, we break symmetry by assigning each point an ‘age’. When there is a conflict between two points, the older point always wins. Observe that this construction implies that an event can be thinned because of the influence of an earlier point that was also thinned. This makes this procedure slightly unnatural; one might expect only surviving points to influence future events.

The Matérn type-III process does not have these limitations; as we describe more carefully in the next section, a newer event is thinned only if it falls within a radius \( R \) of an older event that was not thinned before. The resulting process has a number of desirable properties. In many applications, its thinning mechanism is more natural than that of 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. Another advantage is that this process can support higher event densities than the type-I and type-II processes with the same parameters; in fact, as \( \lambda \) increases, the average number of points in any area increases monotonically to the ‘jamming limit’, viz. the maximum density at which spheres of radius \( R \) can be packed in a bounded area (Møller et al., 2010). The monotonicity property with respect to the intensity of the primary process is also important in applications where we model inhomogeneity by allowing \( \lambda \) to vary with location (see section 7), since large values of \( \lambda \) imply high event 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 better understand the modelling assumptions implicit to this process.

A more severe impediment to the use of this process (and this is true for all Matérn processes) is that given a realization of a Matérn process, there do not exist efficient techniques for inference
over parameters such as $\lambda$ or the radius of interaction $R$. 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 $\lambda$ 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 too can be quite inefficient, with long waiting times until the sampler returns a perfect sample.

Somewhat surprisingly, despite being more complicated than the type-I and II processes, the information contained in a realization of the type-III process allows the development of an efficient MCMC sampler for posterior inference. Before describing this, we develop more general extensions of the Matérn type-III process, providing a flexible and practical framework for the Bayesian modelling of repulsive processes.

## 4 Generalized Matérn type-III processes

A Matérn type-III hardcore point process on a measurable space $(\mathcal{S}, \Sigma)$ is a repulsive point process parametrized by an intensity $\lambda$ and an interaction radius $R$. The process is obtained by thinning events of a homogeneous primary Poisson process $F$ that has intensity $\lambda$. 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 this takes values in the interval $[0, 1]$, which we call $\mathcal{T}$. The set of pairs $(f_i, t_i)$ forms a Poisson process $F^+\equiv(F, T^F)$ on the space $\mathcal{S} \times \mathcal{T}$ (whose intensity is still $\lambda$). We call the set of birth times $T^F$, and define $F^+\equiv(F, T^F)$. The set $T^F$ induces an ordering on the events in $F^+$, and thus on $F$. A secondary process $G^+_\equiv(G, T^G)$ is then obtained by traversing the elements of $F^+$ in this order and deleting all points within a distance $R$ of any earlier and undeleted point. We obtain the Matérn process $G$ by projecting $G^+$ onto $\mathcal{S}$.

Figure 1(left) shows the relevant events for the 1-dimensional case. The filled 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 the set of all locations in $\mathcal{S} \times \mathcal{T}$ that would be thinned by $(s^*, t^*)$. This is the set of all points whose $\mathcal{S}$-coordinate differs from $s^*$ by less than $R$, and whose $\mathcal{T}$-coordinate is greater than $t^*$. Letting $I$ be the indicator function, the shadow of $(s^*, t^*)$ at $(s, t)$ is given by

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

The shadow of a set $G^+$ is the set of all locations that would be thinned by any element of $G^+$, and
Figure 1: (left) The Matérn type-III hardcore point process on a one dimensional space $S$: The filled dots (projected onto $S$) represent the Matérn events, the empty dots being the thinned events. The shaded region is the Matérn shadow. (centre) and (right): Generalized Matérn type-III processes with varying radii and probabilistic deletion respectively

is given by

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

For notational convenience, we make the dependence of the shadow of $G^+$ on $R$ implicit in the equation above. The shaded area in figure [left] shows the shadow of all Matérn events, $G^+$. Note that all thinned events must lie in the shadow of the Matérn events, otherwise they couldn’t have been thinned. Similarly, Matérn events cannot lie in each others shadows; however, they can fall within the shadow of some thinned event.

The hardcore repulsive process can be extended in a number of ways, providing a flexible and tractable framework for modelling repulsive point processes. For instance, instead of requiring all Matérn events to have identical interaction radii, we can assign each one an independent radius drawn from some prior distribution $q(R)$. Such Matérn processes are called ‘softcore’ repulsive processes (Huber and Wolpert, 2009). In this case, the primary process can be viewed as a Poisson process on a space whose coordinates are location $S$, birth time $T$ and interaction radius $R$. Given a realization of this process $F^+ \equiv (F, T^F, R^F)$, we define a secondary point process $G^+ \equiv (G, T^G, R^G)$ by deleting all points that fall within the radius associated with an older, undeleted primary event. The set of locations $G$ constitute a sample from the softcore Matérn type-III process. Given the triplet $(G, T^G, R^G)$, we can once again calculate the shadow $H$, now defined as:

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

Figure [centre] illustrates this; note that the radii of the thinned events are irrelevant.

Another approach to soft repulsion is to probabilistically thin events of the primary Poisson process. This is a useful and flexible generalization of ideas present in the literature, and allows control over the strength of the repulsive effect (in addition to its span). The probability of deletion can be constant, or can depend on the distance of a point to a previous unthinned point, and a primary event is retained only if it is left unthinned by all surviving points with earlier birth times.
Write the deletion kernel associated with location \( s^* \) as \( K(\cdot, s^*) \); the probability of thinning an event located at \((s, t)\) is given by \( I(t > t^*)K(s, s^*) \). To keep this process efficient, one can use a deletion kernel with a compact support; figure [1][right] 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. More precisely, the shadow at any location \((s, t)\) is given by

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

Note that for this process, while the thinned events must still lie in the shadow \( \mathcal{H}(s, t; G^+) \), 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 times passes, and the thinning probability is given by \( K_1(t, t^*)K_2(s, s^*) \). While 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 point processes on the real line can be viewed as generalized Matérn type-III processes where the birth-times are observed. With probabilistic thinning, we recover the class of self-inhibiting point processes 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 \( \mathcal{H} \) has a special Markovian construction.

### 4.1 Probability density of the Matérn type-III point process

In this section, we calculate the probability density of the generalized Matérn process. In this paper, we will mainly use this result to develop a characterization of the various latent variables, thereby allowing the development of an efficient MCMC algorithm for posterior inference. However this result is interesting in its own right, and can serve as a starting point for a more theoretical study of the generalized Matérn process.

Consider a realization of a Matérn process on some compact, measurable space \((S, \Sigma)\). Since this process is obtained by thinning a primary Poisson process, it will be a finite point process if the primary process is finite. We restrict ourselves to this case.

For each finite \( n \), let \( S^n \) be the \( n \)-fold product space of \( S \), equipped with the usual product \( \sigma \)-algebra, \( \Sigma^n \). We shall refer to elements of \( S^n \) as \( S^n \). Thus \( S^n \equiv (s_1, \ldots, s_n) \) is a sequence in \( S \) of length \( n \). We define \( S^0 \) as a single point satisfying \( S^0 \times S = S \times S^0 = S \) and equip it with the trivial \( \sigma \)-algebra \( \Sigma^0 = \{\emptyset, S^0\} \). Then, define \( S^\cup \equiv \bigcup_{n=0}^{\infty} S^n \) as the resulting union space, which we equip with the \( \sigma \)-algebra \( \Sigma^\cup \equiv \{\bigcup_{n=0}^{\infty} A^n, \forall A^n \in \Sigma^n\} \). \( S^\cup \) is the space of all finite sequences in \( S \), and we refer to elements of this space (and thus realizations of our point processes) by uppercase letters without superscripts (eg. \( S \)).

Next, assume a measure \( \mu \) on \((S, \Sigma)\) (for Euclidean spaces, \( \mu \) is typically Lebesgue measure). Letting \( \mu^n \) be the \( n \)-fold product measure on the space \((S^n, \Sigma^n)\), assign any set \( B \in \Sigma^\cup \) the
measure
\[
\mu^\cup(B) = \sum_{n=1}^{\infty} \frac{1}{n!} \mu^n(B \cap S^n) = \sum_{n=1}^{\infty} \int_{B \cap S^n} \frac{1}{n!} \mu^n(dS^n). \tag{6}
\]

We associate point process realizations with random, finite, and unordered sets of points in \(S\); the factorial in the denominator of the measure above corrects for the fact that \(n\) sequences of length \(n\) map to the same unordered set.

Consider first a Poisson process with mean measure \(\Lambda\), and assume \(\Lambda\) admits a density \(\lambda\) with respect to \(\mu\). Let \(S\) be a realization (of length \(|S|\)) from this point process. We then have:

**Theorem 1** (Density of a Poisson process) A Poisson process on the space \(S\) with intensity \(\lambda(s)\) is a random variable taking values in \((\mathcal{S}^\cup, \Sigma^\cup)\) with probability density w.r.t. the measure \(\mu^\cup\) given by

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

For a proof, see for example Daley and Vere-Jones (2008). The density \(p(S)\) is called the Janossy density, and in the point process literature, is used to define a measure called the Janossy measure. The Janossy measure itself is not a probability measure; by contrast, we define the density with respect to the measure \(\mu^\cup\), ensuring that

\[
\int_{\mathcal{S}^\cup} p(S) \mu^\cup(dS) = 1. \tag{8}
\]

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 just the unit interval with the usual Borel \(\sigma\)-algebra. Let \(\mu\) be a measure on this product space; when \(S\) is a subset of the two-dimensional Euclidean space, \(\mu\) is just Lebesgue measure on \(\mathbb{R}^2\). 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** Consider a generalized Matérn type-III process with intensity \(\lambda\), and let \(G^+ = (G, T_G)\) be a sample from this process (augmented with the set of birth times). Let \(\mathcal{H}(s, t; G^+)\) be the shadow defined by \(G^+\), following the appropriate thinning scheme. Then, the density w.r.t. the measure \(\mu^\cup\) is given by

\[
p(G^+ | \lambda) = \exp \left( -\lambda \int_{\mathcal{S} \times \mathcal{T}} \left(1 - \mathcal{H}(s, t; G^+)\right) \mu(ds \, dt) \right) \lambda^{|G^+|} \times \prod_{g^+ \in G^+} \left(1 - \mathcal{H}(g^+; G^+)\right). \tag{9}
\]
We include a proof in the appendix. The product term in the expression above penalizes Matérn events that fall within the shadow of earlier events; in fact, for deterministic thinning, such a set will have zero probability. The exponentiated integral encourages the shadow to be large, which in turn implies that the events are spread out.

A similar result was derived in Huber and Wolpert (2009); they express the Matérn type-III density with respect to a homogeneous Poisson process with unit intensity. However, their result applied only to the Matérn hardcore process. Also, their proof technique 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 we introduced above.

We now have the following corollary of the previous theorem:

**Corollary 3** Let \( G^+ = (G, T^G) \) be a sample from an Matérn type-III process on the space \((S, \Sigma)\), augmented with its birth times. Let the primary intensity be \( \lambda \). Then, given \( G^+ \), the posterior distribution of the locations and birth times of the thinned set, \( \tilde{G}^+ = (\tilde{G}, T^{\tilde{G}}) \) is a Poisson process on \( S \times T \), with intensity \( \lambda \mathcal{H}(s, t; G^+) \).

**Proof 4** The joint probability density \( p(\tilde{G}^+, G^+) \) is the probability of the Poisson process \( F^+ = (\tilde{G}^+ \cup G^+) \) multiplied by the probability that the elements of \( F^+ \) are assigned the appropriate labels, ‘thinned’ or ‘not thinned’. It follows easily from Theorem 1 (see corollary 6 in the appendix) that:

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

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

\[
p(\tilde{G}^+ | G^+) = \frac{p(\tilde{G}^+, G^+)}{p(G^+)} = \exp\left(-\int_{S \times T} \lambda \mathcal{H}(s, t; G^+) \mu(dsdt)\right) \prod_{(\tilde{s}, \tilde{t}) \in \tilde{F}^+} \lambda \mathcal{H}(\tilde{s}, \tilde{t}; G^+).
\]

From Theorem 1, this is just the density of a Poisson process whose intensity is \( \lambda \mathcal{H}(s, t; G^+) \).

The result above 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 set one event at a time, we can jointly simulate all these 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 an MCMC sampler for the Bayesian model 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 secondary process, the thinned events are just Poisson distributed. Note that 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 $R$ (else it couldn’t have been thinned in the first place). However, two or more events can occur together. Clearly such a process is not Poisson, rather it possesses a clustered structure.

5 Bayesian modelling and inference for Matérn type-III processes

In the following, we model an observed set of points $G$ as a realization of a Matérn type-III process. The parameters governing this process are the intensity of the primary process, $\lambda$, and the parameters of the thinning kernel, $\theta$. For the hardcore repulsive process, $\theta$ is just the interaction radius $R$, while for the case of probabilistic thinning, $\theta$ might include an interaction radius $R$, as well as a thinning probability $p$ (with $p = 1$ recovering the hardcore model). For the softcore process, each Matérn event has its own interaction radius which we have to infer, and $\theta$ would be this sequence of radii. In this case, we might also assume that the distribution these radii are drawn from has unknown parameters.

Taking a Bayesian approach, we place priors on the unknown parameters. A natural prior for $\lambda$ is the conjugate Gamma density. The Gamma 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 softcore model, we model the radii as i.i.d. draws from a Gamma distribution (we can place appropriate hyperpriors on the unknown parameters). For simplicity, we leave out any hyperparameters in what follows, and writing $q$ for the prior on $\theta$, we have

$$\lambda \sim \text{Gamma}(a, b)$$  \hspace{1cm} (13)

$$\theta \sim q$$  \hspace{1cm} (14)

$$F^+ \equiv (F, T_F) \sim \text{Poisson Process}(\lambda)$$ \hspace{1cm} (15)

$$G^+ \equiv (G, T_G) \sim \text{Thin}(F^+, \theta)$$  \hspace{1cm} (16)

Note that $G^+$ includes the Matérn events $G$ as well as their birth times $T_G$; however, we only observe $G$. 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 kernel parameters $\theta$, we can calculate the shadow $\mathcal{H}_\theta(s, t; G^+)$ (here, we make explicit the dependence of the shadow on $\theta$). Sampling the thinned
events (call them \( \tilde{G}^+ \), so that \( F^+ = G^+ \cup \tilde{G}^+ \)) is now a straightforward application of Corollary 3: discard the old set, and simulate a new set of events from a Poisson process with intensity \( \lambda_{\mathcal{H}_\theta}(s, t; G^+) \).

We do this by applying the thinning theorem for Poisson processes (Lewis and Shedler (1979), see also Theorem 5). In particular, we first sample a homogeneous Poisson process with intensity \( \lambda \) on \((S \times T)\) and keep each point \((s, t)\) of this process with probability \( \mathcal{H}_\theta(s, t; G^+) \). The surviving set of points is a sample from the required Poisson process. For models with deterministic thinning, \( \mathcal{H}_\theta \) is a binary function, and the posterior is just a Poisson process with intensity \( \lambda \) 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 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

\[
p(T_G|G, F^+, \lambda, \theta) \propto \prod_{(s,t) \in G^+} (1 - \mathcal{H}(\tilde{s}, \tilde{t}; G^+)) \prod_{(\tilde{s}, \tilde{t}) \in \tilde{G}^+} \mathcal{H}(\tilde{s}, \tilde{t}; G^+) \tag{17}
\]

A simple Markov transition operator that maintains equation (17) as its stationary distribution is a Gibbs sampler that iteratively 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 intensity of the resulting 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 \([0, t_{\text{min}}]\), where \( t_{\text{min}} \) is the time of the oldest thinned event that does not lie in \( \mathcal{H}(s, t; (G \setminus g)^+) \) \( (t_{\text{min}} \) equals 1 if there is no such event).

While it is not hard 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. Figure 2 shows one such cycle of updating \( \tilde{G}^+ \) and \( T^G \).

5.3 Sampling the Poisson intensity

Having reconstructed the thinned events, it is easy to resample the Matérn parameters \( \lambda \). Note that the number of primary events \( |F| \) is Poisson distributed with intensity \( \lambda \). With a conjugate Gamma\((a, b)\) prior on the Poisson intensity \( \lambda \), the posterior is also Gamma distributed, with parameters \( a_{\text{post}} = a + |F|, \quad b_{\text{post}} = b + \frac{1}{\mu(S)} \).
5.4 Sampling the thinning kernel parameter $\theta$

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

$$ p(\theta|G^+, F^+, \lambda) \propto q(\theta) \prod_{(s,t) \in G^+} \left(1 - \mathcal{H}_\theta(s, t; G^+)\right) \prod_{(\tilde{s}, \tilde{t}) \in \tilde{G}^+} \mathcal{H}_\theta(\tilde{s}, \tilde{t}; G^+) $$

Again, sampling $\theta$ 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 $\theta$ result in different shadows $\mathcal{H}_\theta(\cdot; G^+)$, and the likelihood of $\theta$ is determined by the probability of the labels under the associated shadow. For the models we consider, this results in a simple piecewise-parametric posterior distribution.

For the Matérn hardcore process, $\theta = R$ is the interaction radius, whose posterior distribution is a truncated version of the prior. The lower bound for this truncation requires that no thinned event lies outside the new shadow, while the upper bound requires that no Matérn event lies inside the shadow. Sampling from this is straightforward. The same applies for the softcore model, only now, each Matérn event has its own interaction radius, and we can sequentially update them. 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 endpoints 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 a straightforward matter to sample a segment, and then conditionally simulate 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 $p$, these counts determine the Beta posterior.

6 Experiments

In this section, we consider two datasets, the classic redwood dataset (Ripley, 1977) and the Swedish pine tree dataset (Ripley, 1988); see figure 3. The former records the locations of trees belonging to the Californian giant redwood family, and consists of 62 trees located within an area.

---

1 Given the set of radii, updating any hyperparameters of $q$ is easy.
The redwood tree dataset (left) and the Swedish pine tree dataset (right).

Figure 3: The redwood tree dataset (left) and the Swedish pine tree dataset (right).

Figure 4: Redwood dataset: posterior distributions of Matérn intensity (left), interaction radius (centre) as well as the number of thinned events (right).

normalized to a square whose sides have length 5. The Swedish pine tree dataset consists of 71 trees, again located in a 5-by-5 square. Both datasets are available as part of R package spatstat (Baddeley and Turner, 2005).

We start by modelling the locations of the trees in both datasets as realizations of Matérn type-III hardcore point processes, placing Gamma(1, 1) priors on the intensity $\lambda$, and an improper (flat) prior on the radius $R$. We evaluated our sampler on this model for the two datasets, with all results obtained from MCMC runs with 10000 iterations, and a burn-in of 1000 samples. A Matlab implementation of our sampler on an Intel Core 2 Duo 3Ghz CPU took around half a minute to produce 10000 MCMC samples. To correct for correlations across MCMC iterations, and to assess mixing for our MCMC sampler, we used software from (Plummer et al., 2006) to estimate the effective sample sizes (ESS) of the various quantities; this gives the number of independent samples with the same information content as our MCMC output. Table 1 shows these values; these demonstrate that our sampler mixes rapidly.

The left and centre plots in figure 4 shows the posterior distributions over $\lambda$ and $R$ for the redwood dataset, while figure 5 shows these quantities for the Swedish dataset. Recall that the area of the square is 25, while the number of Matérn events is of the order 70 in both datasets. The fact that the posterior distributions over the intensities $\lambda$ concentrate around 2.5 to 3 suggests that the number of thinned events is small. That this is indeed the case can be seen from the rightmost plots in the two figures showing the posterior distributions over the number of points deleted due

\[ \text{The interaction radius cannot exceed the minimum separation of the Matérn observations.} \]
to Matérn thinning.

A small number of thinned events suggests a weak repulsive effect, and the reason for this is because the Matérn type-III hardcore model is too inflexible for this data. Observe that for the hardcore model, the posterior distribution of the interaction radius is bounded above by the minimum separating distance between all pairs of observations (otherwise one of these events would have deleted the other). In both datasets, we have at least one pair of events that has a very small separation. A small interaction radius results in a small shadow, and thus weak repulsion. Another limitation with the hardcore model can be seen for the redwood dataset. Though our model is homogeneous, this dataset is clearly not. We address this second issue in the next section. For the first problem, one approach is to use the softcore model (where each Matérn event has its own interaction radius). Again, we leave this for the next section, instead we use our extension with probabilistic thinning to model the data.

---

### Table 1: Effective sample sizes (per 1000 samples) for the Matérn type-III hardcore model

|                        | Redwood dataset | Swedish pine dataset |
|------------------------|-----------------|----------------------|
| Matérn interaction radius | 473.64          | 344.51               |
| Latent times (averaged across observations) | 998.9           | 989.47               |
| Primary Poisson intensity | 988.93          | 954.7                |

---

Figure 6: Swedish pine dataset: posterior distributions of interaction radius $R$ for thinning probability $p$ equal to 0 (left), and 0.75 (centre). The posterior over $p$ is shown in (right).
The details of this model are as follows: once again there is an interaction radius $R$ common to all Matérn events, and we place a Gamma$(1, 1)$ on this. There is also an unknown thinning probability $p \in [0, 1]$ characterizing the strength of the interactions, with all subsequent events within distance $R$ of an event $g$ having a probability $p$ of being thinned by $g$. We place a Beta$(1, 1)$ prior on $p$. We also place a Gamma$(1, 1)$ prior on the Poisson intensity $\lambda$.

To assess the effect of $p$, we first set it to different values, and look at the resulting distribution over $R$. We work with the Swedish pine tree data set. For $p = 1$, we recover the hardcore model, and get a posterior similar to figure 3(left). For $p = 0$, there is no repulsion, and as figure 4(left) shows, the posterior over $R$ reduces to the exponential prior. Finally, figure 6(centre) shows the posterior when $p = 0.75$. Observe that now we get a larger interaction radius than with the hardcore model. Figure 6(right) shows the actual posterior distribution over the thinning probability $p$ resulting from the Beta prior. This is peaked at 0.75, suggesting a strong repulsive effect, both in terms of the thinning probability and the interaction radius. This is confirmed by the large number of thinned events (around 125), and the large Poisson intensity $\lambda$ (around 8).

7 Nonstationary Matérn processes

It is useful in many applications to extend the previous models to allow various kinds of nonstationarity. A common requirement is to allow variation in the intensity of the Matérn events; this is clearly seen with the redwood dataset, and might arise due to variation in factors like soil fertility. A simple way to introduce such nonstationarity is via the intensity function $\lambda$ of the primary Poisson process: instead of requiring this to be homogeneous with a constant intensity $\lambda$, we allow $\lambda(s)$ to vary over $S$.

A flexible approach to modelling such a nonstationarity is to place a transformed Gaussian process prior on the intensity function $\lambda(s)$. Letting $K(\cdot, \cdot)$ be the covariance kernel of a Gaussian process, $\hat{\lambda}$ a positive scale parameter, and $\sigma(x) = (1 + \exp(-x))^{-1}$ the sigmoid transformation, set

\begin{align}
 l(\cdot) &\sim \mathcal{GP}(0, K), \\
\lambda(\cdot) &= \hat{\lambda} \sigma(l(\cdot))
\end{align}

The sigmoid transformation serves two purposes: it ensures that the intensity $\lambda(s)$ is nonnegative, and it provides a bound $\hat{\lambda}$ on the Poisson intensity. As shown in Adams et al. (2009), such a bound makes drawing events from the inhomogeneous primary process possible by a clever application of the thinning theorem. We briefly introduced the thinning theorem in section 5; we state it formally below.

Theorem 5 (Thinning theorem, Lewis and Shedler (1979)) Let $E$ be a random set sampled from a Poisson process with intensity $\lambda(s)$. For some nonnegative function $\lambda(s) \leq \hat{\lambda}(s) \forall s \in S$, assign each point $e \in E$ to set $F$ with probability $\frac{\lambda(e)}{\hat{\lambda}(e)}$. Then the random set $F$ is a draw from a Poisson process with intensity $\lambda(s)$.

\footnote{with differences arising from placing a Gamma prior on $R$ (rather than an improper prior).}
Returning to the nonstationary Matérn process, note from equation (20) that \( \hat{\lambda} \geq \lambda(s) \). Following the thinning theorem, one can obtain a sample from the inhomogeneous (rate \( \lambda(s) \)) Poisson process by thinning a random set \( E \) from a homogeneous (rate \( \hat{\lambda} \)) Poisson process. Importantly, the thinning theorem requires us to instantiate the random intensity \( \lambda(s) \) only on the elements of \( E \) (avoiding the 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 Gaussian process \( l(\cdot) \) on this set. 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 outlined previously to further thin \( F \), resulting in an inhomogeneous repulsive process \( G \). Observe that 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 \( F \). In algorithm 1, we outline the generative process for an inhomogeneous generalized Matérn type-III process.

**Algorithm 1** Algorithm to sample an inhomogeneous Matérn point process on space \( S \)

**Input:** A Gaussian process prior \( \mathcal{GP}(0, K) \) on the space \( S \), a constant \( \hat{\lambda} \) and the thinning kernel parameters \( \theta \).

**Output:** A sample \( G \) from the Matérn type-III process.

1: Sample \( E \) from a homogeneous Poisson process with intensity \( \hat{\lambda} \).
2: Instantiate the Gaussian process \( l(\cdot) \) on this set of points. Call vector \( l_E \).
3: Keep a point \( e \in E \) with probability \( \sigma(l(e)) \), otherwise thin it. The surviving set of points form the primary process \( F \).
4: Assign \( F \) a set of random birth times \( T^F \), independently and uniformly on \([0, 1]\).
5: Proceed through the 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.
6: The surviving set \( G \) forms the inhomogeneous Matérn type-III point process.

Like section 5, we place priors on \( \hat{\lambda} \) as well as \( \theta \). 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 is easy to see that under the posterior, the events thinned by the repulsive kernel \((G^+)^+\) are distributed as an inhomogeneous Poisson process with intensity \( \lambda(s) \mathcal{H}(s, t; G^+) \) (with the \( \lambda \) from corollary 3 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 \( \theta \) are identical to 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 5:

**Corollary 6** 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 $F$. Then given $F$, $F$ is a Poisson process with intensity $\hat{\lambda} - \lambda(s)$.

**Proof 7** A direct approach is to use the Poisson densities defined in Theorem 7. More intuitive is the following: by symmetry, the thinning theorem implies that the thinned events $\tilde{F}$ are distributed as a Poisson process with intensity $\hat{\lambda} - \lambda(s)$. Moreover, 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 the set (given its values on $F$), and keep element $\tilde{f}$ with probability $(1 - \lambda(f)/\hat{\lambda})$.

Having reconstructed the set $E = F \cup \tilde{F}$, we can update the values of the GP 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 wide variety of sampling algorithms to simulate such a latent Gaussian process, in our experiments we used the elliptical slice sampling from Murray et al. (2010). Finally, recalling that 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 algorithm 2.

### 7.2 Experiments

We return to the redwood and the Swedish pine datasets, now modelling them with the nonstationary Matérn type-III process described in the previous section. For these experiments, we model the repulsive thinning with a softcore Matérn process, where each Matérn event has its own radius, and where newer events falling within this radius are thinned deterministically. We assigned each Matérn observation an interaction radius drawn uniformly on the interval $(0, 5)$, and placed a Gamma$(1, 1)$ prior on the scaling parameter $\hat{\lambda}$. We modelled the inhomogeneous intensity function $\lambda(\cdot)$ using a zero-mean Gaussian process with a squared-exponential kernel, and placed lognormal hyperpriors on the GP hyperparameters. We used elliptical slice sampling (Murray et al., 2010) to
### Algorithm 2 MCMC update for inhomogeneous Matérn type-III process on a space $S$

**Input:**
- A set of Matérn events with birth times $G^+ \equiv (G,T^G)$,
- A set of thinned primary events $\tilde{G}^+ \equiv (\tilde{G}, T^{\tilde{G}})$ and thinned Poisson events $\tilde{F}$
- A GP realization $l_E$ on $E \equiv G \cup \tilde{G} \cup \tilde{F}$
- Parameters $\hat{\lambda}$ and $\theta$

**Output:**
- New sets $T^G_{\text{new}}, P^G_{\text{new}}, \tilde{G}^+_{\text{new}}, \tilde{F}_{\text{new}}$.
- A new instantiation of the GP on $G \cup \tilde{G}^+_{\text{new}} \cup \tilde{F}_{\text{new}}$.
- New values of $\hat{\lambda}$ and $\theta$

1. **Sample the Matérn thinned events $\tilde{G}^+$:**
   2. Discard the old Matérn - thinned event locations $\tilde{G}^+$. 
   3. Sample a set of events $A^+ \equiv (A, T^A)$ from a rate $\lambda$ Poisson process on $S \times T$.
   4. Sample $l_A|l_E$ (conditionally from a multivariate normal).
   5. Keep a point $a \in A$ with probability $\sigma(l(a)) \mathcal{H}(a; G^+)$, otherwise thin it.
   6. The surviving set of points form the new set of Matérn thinned events $\tilde{G}^+_{\text{new}}$.
   7. Discard GP evaluations on old Matérn events, and add the new ones to $l_E$.

8. **Sample the Poisson thinned events $\tilde{F}$:**
   9. Define $E_{\text{new}} \equiv G \cup \tilde{G}^+_{\text{new}} \cup \tilde{F}$
   10. Discard the old thinned primary Poisson events $\tilde{F}^+$. 
   11. Sample a set of events $B^+ \equiv (B, T^B)$ from a rate $\hat{\lambda}$ Poisson process on $S \times T$.
   12. Sample $l_B|l_{E_{\text{new}}}$ (conditionally from a multivariate normal).
   13. Keep a point $b \in B$ with probability $1 - \sigma(l(b))$, otherwise thin it. The surviving set of points form the new set of Poisson thinned events $\tilde{F}_{\text{new}}$.
   14. Define $E_{\text{new}} \equiv G \cup \tilde{G}^+_{\text{new}} \cup \tilde{F}_{\text{new}}$.
   15. **Sample the Matérn birth-times:**
   16. For each Matérn observation $g$, resample its birth time $T^G_g$ conditioned on all other variables.
   17. **Sample the GP values $l_{E_{\text{new}}}$:**
   18. We used elliptical slice sampling (Murray et al., 2010).
   19. **Sample the parameters $\hat{\lambda}$ and $\theta$:**
   20. Sample $\hat{\lambda}$ and $\theta$ given the remaining variables.

resample the GP values given all thinned events (step 20 in algorithm 2). The GP hyperparameters were resampled by slice-sampling (Murray and Adams, 2010). Again, all results were from MCMC runs with 10000 iterations, and a burn-in period of 1000.

Figure 7 shows the posterior mean and standard deviation of the intensity of the modulating function $\lambda(\cdot)$ for the redwood dataset. The intensity function deviates strongly from a constant, and
Figure 7: Posterior mean (left) and standard deviation (right) of the intensity of the primary process for the redwood tree dataset

the nonstationary process is clearly far more suitable for this dataset than the homogeneous Matérn process. The left plot in figure 8 displays the strength of the repulsion between events. Here, we divided $S$ into a 30 by 30 grid, and for each element of this grid, we plot the average number of primary Poisson events that were deleted due to Matérn thinning ($|\tilde{F}|$) divided by the area of the grid element. The plot thus shows the rate at which events are deleted due the Matérn repulsion. A high rate of Matérn thinning suggests that the process deviates strongly from the primary Poisson process. We see that while this rate of thinning is indeed high in the regions with clusters of trees, it is restricted to a relatively small portion of $S$. Because the trees are clustered together in regions where the intensity $\lambda(\cdot)$ is high, the interaction radii are small, as are the number of events deleted due to Matérn thinning. This is confirmed by figure 9 showing posterior distributions of the scale parameter $\hat{\lambda}$, the interaction radii $R^G$ (collected across all events) as well as the number of Matérn-thinned events. We see that the Matérn radii are fairly small, suggesting this dataset might be modelled quite well by an inhomogeneous Poisson process.

We have also estimated and plotted the rate at which events $\tilde{F}$ are thinned during the construction of the inhomogeneous Poisson process $F$ from $E$. The right plot in figure 8 shows this rate of deletion; by definition it is given by $(\hat{\lambda} - \lambda(\cdot))$. In regions where this quantity is high, a large number of the primary Poisson events were thinned.

Figures 10 and 11 shows the corresponding plots for the Swedish dataset. In this case, the intensity function is fairly constant, agreeing with the fact the the distribution of the trees is fairly uniform. However, the plot showing the rates of Matérn thinning suggests that this process is significantly more repulsive that a Poisson process. This is confirmed by plots for the number of thinned events and the posterior over the interaction radii (12). Thus, this dataset could be viewed as a sample from a homogeneous Matérn type-III softcore process.

Finally, we include plots showing the effective number of samples of various statistics produced per 1000 MCMC iterations (table 2). As a proxy for the effective sample size (ESS) of the GP intensity, at each MCMC iteration we evaluated the GP function on a regular 10 by 10 grid, and calculated the ESS of each component. The last row of table 2 reports the median of these values. Once again, the table demonstrates that our sampler mixes the relatively rapidly. The 1 in 20
Figure 8: Left: Posterior rate of deletions due to repulsion, and right: Posterior rate of deletions in sampling the primary process of the redwood tree dataset.

Figure 9: Redwood dataset: Posterior distribution of Matérn intensity (left), radius (centre) as well as the number of thinned events (right).

Figure 10: Posterior mean (left) and standard deviation (right) of the rate of the primary process of the Swedish pine tree dataset.
Figure 11: Left: Posterior rate of deletions due to repulsion, and right: Posterior rate of deletions in sampling the primary process of the Swedish pine tree datasets

Figure 12: Swedish pine tree dataset: Posterior distribution of Matérn intensity (left), radius (centre) as well as the number of thinned events (right).
Table 2: Effective sample sizes (per 1000 samples) for the inhomogeneous Matérn type-III softcore model

|                          | Redwood dataset | Swedish pine dataset |
|--------------------------|-----------------|----------------------|
| Average Matérn interaction radius | 314.13          | 370.01               |
| Latent times (averaged across observations) | 994.86          | 988.18               |
| Primary Poisson GP intensity (evaluated on a grid) | 52.6            | 627.19               |

effective sample size for the GP intensity for the redwood dataset is typical for Gaussian processes (Murray et al., 2010); the larger value for the Swedish dataset reflects the fact that it is relatively homogeneous.

8 Discussion

In this paper, we 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. Importantly, we developed an efficient MCMC sampling algorithm for posterior inference in these models. In our experiments, we applied our ideas to two datasets, the Swedish pine tree, and the redwood datasets. While we only considered events in a 2-dimensional space, it is easy to generalize to higher dimensions. This is useful in modelling, say, the distribution of galaxies in space, features in some feature space etc.

There are a number of interesting directions worth following. While we assumed the Matérn events were observed perfectly, there is often noise into this observation process. In this case, given the observed point process $G_{obs}$, we have to 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.

Our models assumed a 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 (for instance).

Finally, we are interested in applying our ideas to more realistic data, in particular to characterize series of neural spike trains. Another direction involves incorporating Matérn repulsiveness in bigger, hierarchical models, for instance to encourage diversity between cluster parameters in a mixture model.
9 Acknowledgements

This work was partially supported by the DARPA MSEE project. Vinayak Rao thanks Yee Whye Teh for useful discussions.

A Appendix

Theorem 2 Consider a Matérn type-III process with intensity $\lambda$, and let $G^+ = (G, T^G)$ be a sample from this process (augmented with the set of birth times). Let $\mathcal{H}_\theta(s, t; G^+)$ be the shadow defined by $G^+$, where $\theta$ parametrizes the Matérn thinning process. Then, the density w.r.t. the measure $\mu^j$ is given by

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

(21)

Proof 3 Recall the definition of $\mathcal{H}_\theta(s, t; G^+)$, the shadow of $G^+$:

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

(22)

Here, $K_\theta$ is the thinning kernel. Let $|G^+|$, the size of the set $G^+$ be $k$. $G^+$ is obtained by thinning $F^+$, a sample from a homogeneous Poisson process with intensity $\lambda$. Let the size of $F^+$ be $n \geq k$, and call the thinned points $\tilde{G}^+$, so that $F^+ = \tilde{G}^+ \cup G^+$. From Theorem 1, the density of $F^+$ w.r.t. the measure $\mu^j$ is

$$p(F^+) = \exp \left( -\lambda \mu(S \times T) \right) \lambda^n$$

(23)

Now, there are $\binom{n}{k}$ ordered versions of $F^+$ mapping to the Matérn sequence $G^+$, so that the conditional density of $G^+$ is given by

$$p(G^+ | F^+) = \frac{n!}{k!(n-k)!} \prod_{(s, t) \in \tilde{G}^+} \mathcal{H}_\theta(s, t; G^+) \prod_{(s, t) \in G^+} (1 - \mathcal{H}_\theta(s, t; G^+))$$

(24)

The first product in the equation above is the probability that all events in $G^+$ are not thinned by the Matérn shadow, while the second term is the probability that the remaining points are thinned. Then we have

$$p(F^+, G^+) = \frac{n!}{k!(n-k)!} \exp(-\lambda \mu(S \times T)) \lambda^n \prod_{(s, t) \in \tilde{G}^+} \mathcal{H}_\theta(s, t; G^+) \prod_{(s, t) \in G^+} (1 - \mathcal{H}_\theta(s, t; G^+))$$

(25)
Integrating out the locations of \( n - k \) thinned events in \( \hat{G}^+ \), we have

\[
p(G^+, |F^+| = n) = \frac{n!}{k!(n - k)!} \exp(-\lambda \mu(S \times T)) \lambda^k \prod_{(s,t) \in G^+} (1 - \mathcal{H}_\theta(s, t; G^+)) \left( \lambda \int_{S \times T} \mathcal{H}_\theta(s, t; G^+) \mu(dsdt) \right)^{n-k} \frac{n!}{k!}
\]

The \( n! / k! \) term in the denominator arises because of the factorial term in the base measure \( \mu^\cup \) (see equation (6), and note that now \( S \) is actually \( S \times T \)). In particular, for \( A \in \Sigma^k \), observe that

\[
\mu^\cup(A \times (S \times T)^k) = \frac{1}{n!} \mu^k(A) \mu^{n-k}(S \times T) = \left( \frac{k!}{n!} \mu^{n-k}(S \times T) \right) \mu^\cup(A)
\]

Finally, summing out \( n \), we have

\[
p(G^+) = \exp(-\lambda \mu(S \times T)) \lambda^k \prod_{(s,t) \in G^+} (1 - \mathcal{H}(s, t; G^+)) \sum_{n=k}^{\infty} \left( \lambda \int_{S \times T} \mathcal{H}(G^+, R) \mu(dsdt) \right)^{n-k} \frac{n!}{(n-k)!} \prod_{(s,t) \in G^+} \lambda(1 - \mathcal{H}_\theta(s, t; G^+))
\]

This is what we set out to prove.

We include the following corollary:

**Corollary 6** Define \( \hat{G} = F^+ \setminus G^+ \) as the set of thinned events. Then

\[
p(\hat{G}^+, G^+) = \exp(-\lambda \mu(S \times T)) \lambda^n \prod_{(s,t) \in G^+} \mathcal{H}_\theta(s, t; G^+) \prod_{(s,t) \in G^+} (1 - \mathcal{H}_\theta(s, t; G^+))
\]

**Proof 7** Observe that \((F^+, G^+)^n\) is of length \( n + k \), and observe that \((\hat{G}^+, G^+)\) has exactly the same information as \((F^+, G^+)^n\), but is now of length \( n \). Clearly,

\[
p(\hat{G}^+, G^+) = p(F^+, G^+)/\binom{n}{n-k}
\]

and the result follows from equation (25).
References

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

Adams, R. P., Murray, I., and MacKay, D. J. C. (2009). Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proceedings of the 26th International Conference on Machine Learning (ICML)*.

Andrieu, C. and Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, 37(2):697–725.

Baddeley, A. and Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42. ISSN 1548-7660.

Brown, E. N., Barbieri, R., Eden, U. T., and Frank, L. M. (2004). Likelihood methods for neural spike train data analysis. In Feng, J., editor, *Computational Neuroscience: A Comprehensive Approach*, volume 7, chapter 9, pages 253–286. CRC Press.

Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes*. Springer.

Dayan, P. and Abbott, L. F. (2005). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press.

Glass, L. and Tobler, W. R. (1971). Uniform distribution of objects in a homogeneous field: Cities on a plain. *Nature*, 233(5314):67–68.

Hansford-Miller, F. H. (1968). *A geographical and statistical analysis of religious nonconformity in England and Wales*. PhD thesis, University College London.

Hill, M. O. (1973). The intensity of spatial pattern in plant communities. *Journal of Ecology*, 61(1):pp. 225–235.

Hough, J. B., Krishnapur, M., Peres, Y., and Virg, B. (2006). Determinantal processes and independence. *Probab. Surv*.

Huber, M. L. and Wolpert, R. L. (2009). Likelihood-based inference for Matern type III repulsive point processes. *Advances in Applied Probability*, 41(4). 958–977.

Jewell, C. P., Kypraios, T., Neal, P., and Roberts, G. O. (2009). Bayesian Analysis for Emerging Infectious Diseases. *Bayesian Analysis*, 4(4):465–496.

Kendall, M. G. (1939). The geographical distribution of crop productivity in england. *Journal of the Royal Statistical Society*, 102(1):21–62.

Kingman, J. F. C. (1993). *Poisson processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York. Oxford Science Publications.
Knox, G. (2004). Epidemiology of childhood leukemia in northumberland and durham. *The Challenge of Epidemiology: Issues and Selected Readings*, 1(1):384–392.

Lewis, P. A. W. and Shedler, G. S. (1979). Simulation of nonhomogeneous Poisson processes with degree-two exponential polynomial rate function. *Operations Research*, 27(5):1026–1040.

Matérn, B. (1960). Spatial variation. *Meddelanden från Statens Skogsforkningsinstitut (Reports of the Forest Research Institute of Sweden)*, 49(5).

Matérn, B. (1986). *Spatial Variation*. Lecture notes in statistics. Springer-Verlag, second edition.

Mateu, J. and Montes, F. (2001). Likelihood inference for Gibbs processes in the analysis of spatial point patterns. *International Statistical Review / Revue Internationale de Statistique*, 69(1):pp. 81–104.

Møller, J., Huber, M. L., and Wolpert, R. L. (2010). Perfect simulation and moment properties for the matrn type iii process. *Stochastic Processes and their Applications*, 120(11):2142–2158.

Møller, J. and Waagepetersen, R. P. (2007). *Modern Statistics for Spatial Point Processes*. *Scandinavian Journal of Statistics*, 34(4):643–684.

Murray, I. and Adams, R. P. (2010). Slice sampling covariance hyperparameters of latent Gaussian models. In *Advances in Neural Information Processing Systems 23*.

Murray, I., Adams, R. P., and MacKay, D. J. (2010). Elliptical slice sampling. *JMLR: W&CP*, 9.

Peebles, P. J. E. (1974). The Nature of the Distribution of Galaxies. *Astronomy and Astrophysics*, 32:197.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1):7–11.

Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(2):pp. 172–212.

Ripley, B. D. (1988). *Statistical inference for spatial processes / B.D. Ripley*. Cambridge University Press, Cambridge [England] ; New York :.

Scardicchio, A., Zachary, C., and Torquato, S. (2009). Statistical properties of determinantal point processes in high-dimensional Euclidean spaces. ”*Phys. Rev. E”*, 79(4):Article 041108.

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