Bayesian inference for Matern repulsive processes

By Vinayak Rao, Ryan P. Adams, and David D. Dunson

Report by James Buenfil

Department of Statistics, University of Washington Seattle, WA, 98195, USA

1 Introduction

1.1 Spatial Statistics

Spatial statistics is a broad field where the goal is to perform statistical analysis while taking into account spatial knowledge. There are 3 primary sub-fields that frequently overlap: Geo-statistics, areal or lattice data, and spatial point processes van Lieshout [2019]. This paper lies firmly within the field of point processes.

1.2 Point processes

There are several ways in which a point process can be defined, for instance using counting functions or random measures. In this report we will use the definition of a finite point process described in section 5.3 of Daley et al. [2003], as this is the theoretical framework the authors of the paper use. This definition is convenient because it allows us to understand how realizations of the process can be produced.

Definition (Finite Point Process). A finite point process X on $\mathcal{S} \subset \mathbb{R}^d$ is a random set satisfying the following three conditions:

(a) The random set contains a finite number of elements of $\mathcal{S}$.

(b) The cardinality of X, $|X|$, is distributed according to some discrete distribution defined by $\{p_n\}_{n=0,\ldots\infty}$ where $\sum_{n=0}^{\infty} p_n$.

(c) Given $|X| = N$, the points of X are distributed according to some joint distribution on N points $\Pi_n(\cdot)$ defined on Borel sets of $\mathcal{S}^n = \mathcal{S} \times ... \times \mathcal{S}$

1.3 Applications

Spatial point processes are applied in many scientific domains including astronomy, ecology, demography, geography, seismology, and neuroscience Baddeley et al. [2006]. Given a subset of points in Euclidean space $\mathbb{R}^d$, we can model these points as a realization of a point process. The point process is determined by the aforementioned discrete and joint distributions which are used to sample from it, and in modeling we specify these distributions. These distributions are governed by parameters, which a practitioner is interested in estimating.

As an example, consider the problem of diagnosing patients with diabetic neuropathy, which is a type of nerve damage which can occur in diabetes patients. It was observed that the set of locations where epidermal nerve fibers enter the epidermis from the dermis (layers of human skin) exhibited different distributions in patients with and without diabetic neuropathy, where in the patients with neuropathy, the nerve fibers appear more clustered. Therefore, given an arbitrary patient, it is of interest if diagnosis can be performed simply by looking at the differences in spatial distribution.

For example, if a point process model contains some parameter which characterizes repulsiveness in the data, then we would expect that the estimated parameters for two sets of nerve fiber locations from patients with and without diabetic neuropathy would differ. This problem was investigated in Waller et al. [2011], and we also consider their datasets.

To motivate the model studied in this paper we introduce the simplest point process, the Poisson process, and consider its limitations considered.

1.4 Poisson Point Process

While there are many characterizations of the Poisson point process, we define it in terms of the definition of the general point process given above. We define the inhomogeneous Poisson process of which the homogeneous process is a special case. First we give preliminary definitions:

Definition (intensity function): Given a set $\mathcal{S} \subseteq \mathbb{R}^d$, let the intensity function $\lambda : \mathcal{S} \to [0, \infty)$ be locally integrable, i.e. $\int_B \lambda(s)\mu(\mathrm{d}s) < \infty$ for all bounded $B \subseteq \mathcal{S}$, where μ denotes the Lebesgue measure.

Definition (mean measure): Let the mean measure be defined by

$$\Lambda(B) = \int_B \lambda(s)\mu(\mathrm{d}s), \quad B \subseteq \mathcal{S}.$$

For our purposes $\mathcal{S}$ will be compact so all subsets B of $\mathcal{S}$ are bounded.

Definition (Inhomogeneous Poisson Process): The discrete distribution is determined by $p_n = \frac{\Lambda(\mathcal{S})^n}{n!}e^{-\Lambda(\mathcal{S})}$, and the joint distribution Π_n is the distribution determined by i.i.d sampling n points according to density function $\lambda(s)/\Lambda(\mathcal{S})$.) We then write $F \sim \text{Poisson}(\mathcal{S}, \lambda(s))$.

When the intensity $\lambda(s)$ is a constant λ over $\mathcal{S}$, the Poisson point process is called a homogeneous Poisson point process with parameter λ.

1.5 Limitations of the Poisson Point Process

One particular way the homogeneous Poisson point process can fail to model data is that because the data exhibits more clustering (points are clustered together more often) or regularity (repulsiveness or space between points). There are also cases where both clustering or repulsiveness are present, for example, imagine clusters of points which as clusters are widely spaced apart. In this case, on a small scale the points are clustered, but on a larger scale they are regular. The fact that deviations from the homogeneous process can occur at different scales is of central importance when trying to model spatial phenomena.

We are primarily interested in modeling spatial data which exhibits repulsiveness, and our model of choice is the Matern type III process.

1.6 Matern type III process

Given a realization of point process, there are many ways to modify the given points in order to produce a realization of another point process. One such way is called thinning. Given a realization of a point process, $(x_1, ...x_N)$, we can define a thinned process by deleting points of the original process according to some method.

Thinning is very general, and includes random thinning methods. One simple way to thin a process is simply to, for every point in the realization, delete it with fixed probability $p \in [0, 1]$.

Bertil Matern introduced three particular ways of thinning point processes in his original thesis for his Doctorate Matérn [2013]. The focus of the paper is the third type, which we define now:
Definition: Given a realization of a point process F, with $|F| = N$, assign each event $f \in F$ a random mark, chosen uniformly from the interval $[0, 1]$. These marks induce an ordering on F. Fix a parameter $R > 0$, the interaction radius. Traversing through the events of F in order from smallest to largest, when considering an event f, delete all points which have both mark less than f, and are also within distance R of f.

In particular, Matern defined his thinning methods when the original process that we started with (also known as the primary process) is a homogeneous Poisson process with some parameter λ. The thinned process is called the Matern type III process.
The thinning process described above is known as "hard-core" Matern type III thinning.

Now that we have defined the Matern type III process and its defining parameters, we can summarize the paper.

2 Bayesian inference for Matern repulsive processes

2.1 High level summary of paper

Poisson processes are simple models for spatial point process data, however, they can fail to capture clustering or repulsiveness in data. Thinning a point process means that some criterion is used on each point in a process which decides if that point should be removed from the process. A Matern type III process is a point process which thins a Poisson process so that repulsiveness can be better modeled. Observing real spatial data, we can model it as a realization of a Matern type III process. A Matern process is governed by parameters, and thus inference over these parameters is of interest. In this paper a Bayesian approach to modeling is adopted, and prior distributions are placed on the model parameters. Through Gibbs sampling, we obtain samples from the posterior distribution of these parameters.

The main theoretical tool developed for this purpose is the derivation of the conditional density of the Matern thinned events given the Matern events (Corollary 1). This provides the conditional distributions necessary in the Gibbs sampling.

Three types of Matern thinning are discussed. The traditional and most restrictive method, "hard-core" thinning and a generalization known as "soft-core" thinning are discussed. A novel generalization of hard-core thinning is introduced known as probabilistic thinning.

To produce a Matern process, a Poisson process is thinned. Rather than thin a homogeneous poisson process with constant intensity, an inhomogeneous process can be thinned instead, which is more flexible. Methods for sampling from the posterior distributions of the parameters which describe the inhomogeneous Matern type III process are also introduced.

Methods are applied to three data sets. Analysis of how well these methods are able to model the data sets is performed through Besags's L and J functions, the Fano factor, and effective Monte Carlo Markov Chain (MCMC) sample sizes. All three types of thinning are compared. The inhomogeneous method is used only in conjunction with probabilistic thinning. Comparison with another point process, the Strauss process, is done as well.

2.2 Other approaches to modeling repulsive point processes

There are several other point processes commonly used to model spatial data. Cox processes are Poisson processes where the intensity itself is a stochastic process Cressie [2015]. The inhomogeneous approach considered here is very related. Gibbs Processes allow for interaction between points and are well studied. However, inference is not simple as the normalizing constant of the density is generally unknown. Markov Point processes (introduced in Ripley and Kelly [1977]) model repulsiveness in an explicit way involving equivalence classes of neighborhoods of points. Homogeneity is defined differently for Markov Point processes but inhomogeneous variants can be defined as well Jensen and Nielsen [2001]. There have also been density based approaches, using gaussian mixed models to model spatial data Binelli [2017] Matern type I and II processes have also been studied, a recent work is Teichmann et al. [2013] which generalized Matern type I and II processes by introducing probabilistic thinning for these processes. This is done in a very similar way to this paper.

3 Generalizations of the Matern Repulsive Process

We have already defined the Matern type III process with "hard-core" thinning, but now we generalize to "soft-core" and "probabilistic" thinning. We also introduce notation which will be used throughout. In this section, all Poisson processes are assumed to be homogeneous so that they are parameterized by a single paramter λ.

3.1 Hard-core

Recall that the hard-core Matern type III process on a measurable space $(\mathcal{S}, \Sigma)$ is a repulsive point process parameterized by an intensity λ and an interaction radius R which is obtained by thinning the events of a homogeneous primary Poisson process F with intensity λ. This thinning is done in the following way:

Each event $f \in F$ of the primary process is independently assigned a random mark t, the time of its birth. We assume that this takes values in the interval $\mathcal{T} = [0, 1]$. Call the collection of birth times T^F, and define $F^+ = (F, T^F)$ as the collection of (location, birth time) pairs, then F^+ is a Poisson process on $\mathcal{S} \times \mathcal{T}$ whose intensity is still λ. Any process augmented with birth times is called an augmented process.

T^F induces an ordering on the events in F^+, and a secondary process $G^+ = (G, T^G)$ is obtained by traversing F^+ in this order and deleting all points within a distance R of any earlier undeleted point. G is then a realization of a type III hard-core Matern process.

3.2 Shadow of Process

Next we define the shadow of the augmented Matern process G^+ with radius R. The shadow is a function on $\mathcal{S} \times \mathcal{T}$ which is 1 if its input (s, t) would be thinned by the events in G^+ using radius R, and 0 otherwise. In other words, the shadow of G^+ evaluated at (s, t) is 1 is there exists $(s_j, t_j) \in \mathcal{S} \times \mathcal{T}$ such that $d(s, s_j) < R$ and $t_j < t$, and 0 otherwise.

Letting I be the indicator function, we have that:

$$\mathcal{H}\left(s, t; G^+\right) = 1 - \prod_{(s^*, t^*) \in G^+} \{1 - I(t > t^*) I(d(s, s^*) < R)\}.$$

Note that if we evaluate the shadow on any event from the primary process which was thinned to produce G^+, we obtain 1, otherwise they would not have been thinned. Similarly, evaluating the shadow of $G^+/(s^*, t^*)$ on (s^*, t^*) gives 0, as Matern events can not have thinned each other.
If the reader is familiar with type I and II Matern processes, then they should notice that thinned events in those processes are still able to thin other events, while in the type III process, once an event is thinned, it has no role in determining whether other points are thinned or not. This is a kind of "independence" between thinned events, and in Corollary 1 this is formalized. As this property does not hold for Matern type I and II processes, our methods and theorems don't apply to them.

3.3 Soft-Core

Soft-core thinning is a generalization of hard-core thinning where each Matern event has its own interaction radius, which is drawn from some distribution $q(R)$. To produce a Matern sample, we start from a realization $F^+ = (F, T^F, R^F)$ where R^F contains the radii of the events (i.i.d. drawn from $q(R)$), and (F, T^F) is drawn in the same way as the hard-core case. Then the Matern and thinned events are found the same way: we delete all points that fall within the radius that is associated with an older, undeleted primary event $f \in F$.

Denoting the augmented Matern process as $G^+ = (G, T^G, R^G)$, the shadow of G^+ can be defined in an analogous way to the shadow in the hard-core case.

3.4 Probabilistic Thinning

Probabilistic thinning is exactly the same as hard-core thinning, except points may not thin each other. A parameter $p \in [0, 1]$ controls the thinning probability. The shadow takes on intermediate values of p rather than 1 if a point is in the shadow of a process.

4 Theorems

In this section we state the key theorems which justify our methods.

We begin with a result on the density of the Poisson point process. First, we define the framework for studying the Matern process. We want to define the densities of the point processes of interest, and in particular we derive the density of the thinned events given the Matern events, as this is the cornerstone our sampling methods.

4.1 Framework

Let $\mathcal{S}$ be a compact subset of $\mathbb{R}^d$, and let Σ be the Borel σ-algebra of $\mathcal{S}$. For two points $s_1, s_2 \in \mathcal{S}$, define $s_1 > s_2$ if $s_{1d} > s_{2d}$. The last co-ordinate defines a partial ordering on $\mathcal{S}$, and we associate it with the birth time of an event (we imagine points living spatially in a subset of $\mathbb{R}^{d-1}$, and which also have birth times). A realization of a Matern process on $\mathcal{S}$ is obtained by thinning a Poisson process; the Poisson process in this context is called the primary process. With probability 1, there is unique ordering of the Poisson elements according to the partial ordering that we have just defined. This allows us to associate realizations of point processes with unique, increasing sequences of points in $\mathcal{S}$. Below, we define this space of point process realizations.

For each n, let $\mathcal{S}^n$ be the n-fold product space with the product σ-algebra, Σ^n. Elements of $\mathcal{S}^n$ are referred to as S^n. Define the disjoint union space $\tilde{S}^{\cup} \equiv \cup_{n=0}^{\infty} \mathcal{S}^n$ (where $\mathcal{S}^0$ is a point corresponding to an empty sequence) and assign it the σ-algebra $\tilde{\Sigma}^{\cup} \equiv \{\cup_{n=0}^{\infty} A^n, A^n \in \Sigma^n\}$. $\tilde{S}^U$ is the space of all finite sequences in $\mathcal{S}$: to see this, if we consider any element of $\tilde{S}^U$, it must belong to some $\mathcal{S}^n$, and hence be a collection of n points in $\mathbb{R}^{d-1}$ along with n associated birth times.

Now define $(\mathcal{S}^{\cup}, \Sigma^{\cup})$ as its restriction to increasing sequences in $\mathcal{S}$ (i.e. for every n, we only consider elements of $\mathcal{S}^n$ which satisfy $s_{1d} < s_{2d}, ...s_{nd}$). 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 $(\mathcal{S}^n, \Sigma^n)$. Define for any set $B \in \Sigma^{\cup}$ the measure

$$\mu^{\cup}(B) = \sum_{n=0}^{\infty} \mu^n \left(B \cap \mathcal{S}^n\right) = \sum_{n=0}^{\infty} \int_{B \cap \mathcal{S}^n} \mu^n \left(\, \mathrm{d}S^n\right),$$

where $B \cap \mathcal{S}^n$ is considered a subset of $\mathcal{S}^n$ (not as a subset of the disjoint union space).

We now have a measure space $(\mathcal{S}^{\cup}, \Sigma^{\cup}, \mu^{\cup})$. As we will see in the proof of Theorem 1 below, this measure space has a natural connection to the definition of a point process given in the introduction (Section 1.2), and in particular to the Janossy density (see Section 5.3 of Daley et al. [2003]).

4.2 Theorem 1

Now, let S be a realization of a Poisson process with intensity λ with mean measure Λ admits a density. 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 $\lambda(s)$ is a random variable taking values in $(\mathcal{S}^{\cup}, \Sigma^{\cup})$ with probability density with respect to the measure $\mu^{\cup}$ given by

$$p(S) = \exp\{-\Lambda(\mathcal{S})\} \prod_{j=1}^{|S|} \lambda\left(s_j\right),$$

where $S \in \mathcal{S}^{\cup}$, and $s_1, ...s_{|S|} \in \mathcal{S}$.

The proof of this result can be found in the Appendix.

4.3 Theorem 2

In theorem 1 we found the density of the Poisson process with respect to the measure defined. Now we wish to find the density of the augmented Matern type III process with respect to the measure $\mu^{\cup}$.

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

$$\begin{aligned} p\left(G^{+} \mid \lambda\right) = & \exp\left[-\lambda \int_{\mathcal{S}\times\mathcal{T}} \left\{1 - H\left(s, t; G^{+}\right)\right\} \mu(\mathrm{d}s\ \mathrm{d}t)\right] \lambda^{|G^{+}|} \\ & \times \prod_{g^{+}\in G^{+}} \left\{1 - H\left(g^{+}; G^{+}\right)\right\} \end{aligned}$$

The proof of this result is found in the Appendix.

The intuition behind this result is that the exponentiated term encourages the shadow of the process to be large, while the product term encourages (or requires, in the non-probabilistic thinning case) that Matern events do not lie in each other's shadows.

The idea of the proof is to begin with the density of the primary Poisson process (which we have thanks to Theorem 1) as well as the conditional density of the thinned and Matern events given the primary Poisson process (which follows from the definition of the thinning process). Using Bayes rule we can obtain the joint distribution of thinned, Matern, and primary process events, which we can integrate to obtain the density of just the augmented Matern events.

4.4 Corollary 1

Now we come to the main result:

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

A proof is found in the original paper. The idea behind this result is that we can again apply Bayes' rule to the joint density of the thinned and Matern events (which was found in the proof of Theorem 2). Then applying the result of Theorem 2, we can derive the conditional density of the thinned events given the Matern events.

This result is key to our modeling approach because as the thinned events are conditionally an inhomogeneous Poisson process, it allows for the thinned events to be sampled all at once, given the Matern events. This is the sense in which the thinned events are "indepedent" from one another, in that they can be sampled all at once.

Thus far, we have discussed how to sample from a homogeneous Poisson process, but not an inhomogeneous one. The next theorem allows us to sample from inhomogeneous Poisson processes by simply thinning a homogeneous one.

4.5 Theorem 3

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

5 Bayesian Modeling

Now we fully describe our modeling approach. In the original paper, in conjunction with homogeneous Poisson process (i.e. the primary process is a homogeneous Poisson Process) MCMC methods using the three types of thinned described thus far are implemented. For the purposes of space, and because describing one thinning method is enough to understand the fundamental ideas, I choose to focus on probabilistic thinning.

5.1 Model

We model an observed sequence of points G as a realization of a Matern type III process. The parameters governing this are the intensity of the primary process, λ, and the thinning probability p. Note that we do not observe the birth times T^G.

We take a Bayesian approach and place priors on the unknown parameters. We use a Gamma prior for the intensity λ (with shape and rate parameters), and a uniform prior q on the interaction radius R. We use a Beta prior on the thinning probability p.

The full model is below:

$$\lambda \sim \text{Gamma}(a_\lambda, b_\lambda),$$

$$R \sim q(R),$$

$$p \sim \text{Beta}(a_p, b_p),$$

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

$$G^+ \equiv \left(G, T^G\right) \sim \text{MaternThinning}\left(F^+, R, p\right)$$

From this model, we want to produce a MCMC which has stationary distribution equal to the conditional joint distribution $P(\lambda, p, R|G)$. We use Gibbs sampling to produce this Markov chain. However, it is not straightforward to sample from the conditional distributions $P(R|\lambda, G)$ and $P(\lambda|R, G)$. Rather than use Gibbs sampling on just $P(\lambda, p, R|G)$, we instead use the augmented posterior $P(F^+, T^G, \lambda, R, p|G)$. The algorithm is to iterate the following five steps in sequence, each time using the updated versions of the parameters.

I note that the original authors did not detail many of the following steps for the soft-core and probabilistic thinning cases, and in particular there was no mention of having to use the Metropolis-Hastings algorithm due to smaller than machine-epsilon values (although this is what was done in the code). I note that I implemented all three ways of thinning.

5.1.1 1. Sampling from $P(F^+|T^G, \lambda, R, G, p)$

Given the augmented Matern events, sampling the primary process is equivalent to sampling the thinned events. Since we also have R and p, we can calculate the shadow of G^+. By Corollary 1, the thinned events follow an inhomogeneous Poisson process with intensity $\lambda\mathcal{H}\left(s, t; G^+\right)$. Applying Theorem 3, we can 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 $\mathcal{H}\left(s, t; G^+\right)$. Note that while the shadow is a function, we only need its values on the points of the primary process F^+. The surviving points are the sampled thinned events.

5.1.2 2. Sampling from $P(T^G|F^+, \lambda, R, G, p)$

From 2, we have that

$$P\left(T^G \mid G, F^+, \lambda, \theta\right) \propto \prod_{(s,t)\in G^+} \left\{1 - \mathcal{H}\left(s, t; G^+\right)\right\} \prod_{(s,t)\in \tilde{G}^+} \mathcal{H}\left(s, t; G^+\right). \tag{1}$$

Since T^G is a sequence of times, we can use Gibbs sampling again in order to sample the times of the Matern events sequentially, conditioning on all other times. Fix a Matern event g, we will describe how to sample its time. We do this through a series of observations:

Observation 1. If we look at all primary events (thinned or not) within distance R of g, the birth times of these events segment the unit interval into a number of regions. The birth time t_g of g is uniformly distributed within each interval, since as t_g changes over an interval, the shadow at all primary events remains unchanged.

Observation 2. As t_g moves from one segment to the next, one of the primary events moves into or out of the shadow of g. From 1 we see that the probability of any interval is 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.

Observation 3. In observation 1 we looked at all primary events within distance R of g, but we really only need to consider the thinned events within distance R. This is because, no matter how t_g is changed, for every $h \in G$, the total numbers of Matern events that come before any h will remain the same, because the interaction radius R of every event is the same (a Matern event which is born earlier than another Matern event has a chance to thin it, and therefore determines the probability of the event being thinned).

Observation 4. Thinned events need to have been thinned by at least one Matern event. This means that the birth time of g must be before any thinned events which are thinned by no other Matern events.

Observation 5. For each interval being considered, there are a fixed number of thinned events which come before and after g. For those that come before, in order to calculate the probabilities that they were thinned (1), we need to count the number of Matern events within radius R of it which came before each thinned event. Similarly, for each thinned event which is born after t_g, we do the same, but add one to the counts, because g is also born before these events. With these counts, we can compute the RHS of 1 for the given interval, which allows us to compute the probabilities mentioned in observation 2. (normalize the computed quantities by their sum)

Observation 6. Once we have the probabilities of each interval, we simply sample a multinomial distribution using the computed probabilities, and then sample t_g uniformly from the selected interval.

5.1.3 3. Sampling from $P(\lambda|F^+, T^G, R, G, p)$

Sampling λ is straightforward because we know that the number of primary events $|F^+|$ is Poisson distributed. Since we have placed a Gamma(a, b) prior on λ, which is the conjugate prior, it's straightforward to derive that the posterior of λ is a Gamma distribution with parameters $a_{\text{pos}} = a + |F^+|$, $b_{\text{pos}} = b + \mu(\mathcal{S})$.

5.1.4 4. Sampling from $P(R|F^+, T^G, \lambda, G, p)$

Like the birth times, the posterior distribution of R follows from 2, and with a prior q on R it is of the form

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

We again describe the sampling process through a series of observations:

Observation 1. From the form of the posterior, we need to know the thinning probabilities for both the Matern and the thinned events.

Observation 2. Like in the birth times case, the posterior of R is constant over different intervals of R, whose endpoints are determined by the unique distances between Matern and thinned events. However, if we proceed in the same way as the birth times and try to calculate the probabilities of the intervals directly, in practice we find that they are too small.

Observation 3. Instead of sampling from the posterior in that more precise way, we can use the Metropolis-Hastings algorithm to sample the distribution. This circumvents the issue of having extremely small values because throughout the calculation we can take logarithms which make them much more reasonable sized, and we can compare logarithms of values to accept or reject new points. The downside is that we potentially need to perform many iterations in order to produce a valid sample.

5.1.5 5. Sampling from $P(p|F^+, T^G, \lambda, G, R)$

The posterior distribution of p is the same as that of sampling R, except the prior q is replaced by our Beta prior on p.

The same problems as sampling R are encountered with sampling p, and a solution is to again use the Metropolis-Hastings algorithm. The calculations are very similar to the case of sampling R, except instead of proposing a new R each iteration, a new p is proposed instead. In both cases of p and R the proposal distribution can be chosen to be the prior on the respective parameter.

6 Inhomogeneous Methods

The inhomogeneous methods are largely the same as the homogeneous versions. The main theorems continue to hold in the case of an inhomogeneous primary Poisson process, so the methods can still be used. The intensity of the primary inhomogeneous Poisson process is modeled as output of a Gaussian process after plugging it into the sigmoid function, so that it lies between 0 and 1. Then, these intensities are multiplied by a constant $\hat{\lambda}$. Theorem 3 can be applied to sample from this inhomogeneous process, because by construction the intensity values are bounded by $\hat{\lambda}$. There are only a couple of new ideas for applying their methods, for example elliptical slice sampling Murray and Adams [2010] is used to resample the Gaussian process values at every iteration.

7 Evaluation of methods

In modeling point processes, since we do not know the true parameters, model success is often evaluated by simulating new realizations of processes using the parameters estimated by our model, and then inspecting visually if these new realizations share similar 1st and 2nd order statistics. One such 2nd order statistic is the L function.

7.1 L-function

The K function is introduced in Ripley [1977], and an empirical estimate commonly used to analyze the clustering/repulsiveness of spatial point process data. Wilschut et al. [2015] contains a nice example of how it is used in practice. The empirical estimate of K is defined as

$$\widehat{K}(t) = \lambda^{-1} \sum_{i \neq j} \frac{I\left(d_{ij} < t\right)}{n},$$

where the d_{ij} are the distances between points of the process for which it is being evaluated (where there are n points). Since in practice λ is unknown, a simple estimate is $n/\mu(\mathcal{S})$. A related quantity is the empirical L function, defined as

$$\widehat{L}(t) = \left(\frac{\widehat{K}(t)}{\pi}\right)^{1/2}.$$

(From here on the empirical L function is referred to as the L function.) The important property of the L function is that $E\left[\hat{L}(r)\right] = r$ for realizations of homogeneous Poisson processes. Because of this, plots of $\hat{L}(r) - r$ for spatial data are used to measure deviations from homogeneity for different values of r (since this is on average 0 for a homogeneous process).

However, it should be noted that the L function does not fully characterize the process: see Baddeley and Silverman [1984]. In summary, there exist very different looking point processes which have the same (theoretical) L function.

The expectation 0 L function described above holds when the Poisson process is sampled over all of $\mathbb{R}^d$, but in practice it is biased due to so called "boundary effects"(Wiegand and A. Moloney

[2004] contains a review of boundary effect correction methods). The bias grows with r, and when I calculated the L function, I did not use any boundary correction techniques. The maximum r I consider is 2 and the region of interest $\mathcal{S} = [0, 10] \times [0, 10]$, and for homoegeneous Poisson processes I simulated using values of λ consistent with their average values estimated from datasets, the bias was visually not too large.

7.2 Fano factor

The Fano factor for finite univariate datasets is simply defined by the variance of the data divided by its mean. It can be applied to point process data by dividing up the region of interest into a grid of rectangles, and then calculating the Fano factor over the dataset obtained by counting the number of datapoints within each rectangle. For a homogeneous Poisson process, the distribution of the number of points in each rectangle is the same (Poisson distributed), and because the Poisson distribution has the property that its variance is equal to its mean, the Fano factor will be 1. A Fano factor calculated for spatial data with a value that is lower than one can suggest repulsiveness.

7.3 Other ways of model evaluation which were not considered

Perry et al. [2006] contains an excellent review of different approaches to modeling spatial data, in particular in plant ecology.

8 Application to real data

8.1 Swedish Trees

I applied the hardcore and probablistic thinning approaches to the Swedish Trees dataset.

8.1.1 Hard-core

For λ, the Gamma prior parameters were chosen to be 1 and 1. Below we see the empirical L function plotted along with the quantiles (generated from the parameters estimated at each iteration of MCMC). 10000 MCMC iterations were performed. Also included are the empirical L functions found for probabilistic thinning (same experimental setup). A uniform prior was used for p. See the extensions section below for details on the L function computation.

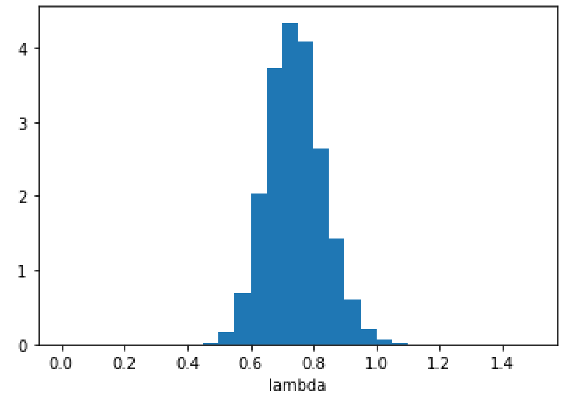


Figure 1: 1a

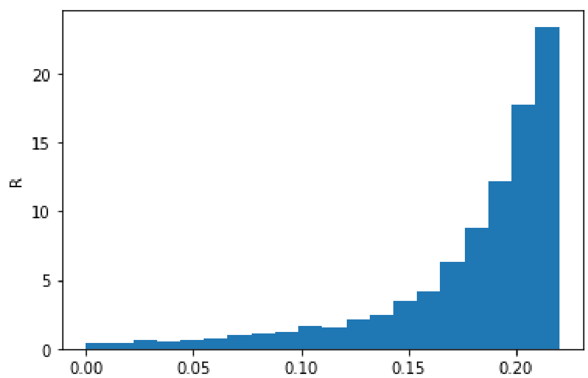


Figure 2: 1b

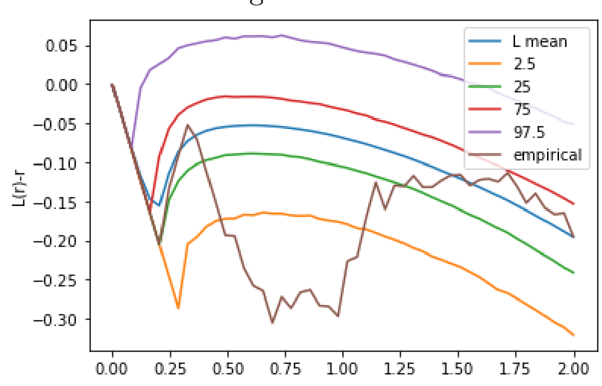


Figure 3: 1a

Figure 4: Posterior plots of lambda, R and empirical L function for Swedish Trees dataset (hard-core thinning)

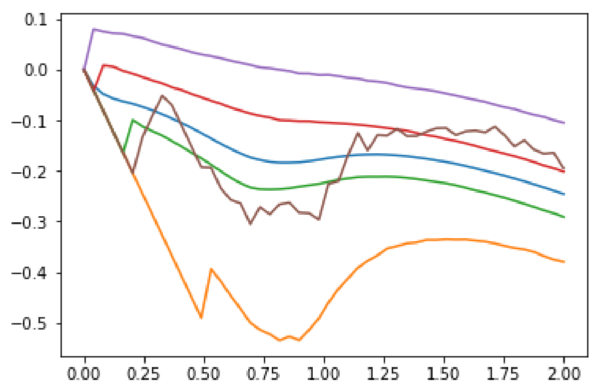


Figure 5: empirical L function for Swedish Trees dataset (probabilistic thinning)

9 Limitations

Analysis of proposed methods is limited in that it was primarily performed by the visual inspection of graphs which show the extent to which models fit real data. Analysis on simulated data, where true parameters and data generating processes are known, was not performed. A prior sensitivity analysis is not performed, and the convergence of the proposed Markov chains is not well-studied.

9.1 Issues with implementation in code online

In the appendix, I include several minor errors found in their code. Here, I present more significant errors.

In hard-core thinning for Swedish trees portion of code, there is a conceptual issue with the radius update (section 5.4). This issue leads to an incorrect posterior distributions over R. In computing the lower bound for R, a minimum is computed, and in this minimum the value 0 is incorrectly minimized over. Instead, the minimization should occur over all the non-zero values. There is a subsequent maximization over these minima, but all minima are 0, and the lower bound is found to be 0. This explains the difference in my posterior plots from theirs. If I alter my code in this way, then I recover the more uniform posterior radius plot found in the paper (it looks uniform because every iteration, R is just sampled from a uniform distribution).

In the code online, the updating of Matern birth times (section 5.2) for the soft-core thinning case is performed conceptually incorrectly. Although Gibbs sampling is being performed over the times, after updating the time of one event, when simulating the next time, the previous time is not updated until after simulating all times. It's not clear how this error affects posterior inference practically.

This same issue is present in the updating of the Matern radii (soft-core thinning), where the nested Gibbs sampling is done incorrectly. When updating a radius through sampling (the distribution from which we sample will depend on the other radii which we condition on), the previous radii updates are not used. It is computationally more efficient, but incorrect conceptually. It's not clear how this error affects posterior inference practically.

10 Extensions

10.1 Simulated Data

One basic extension is to apply their methods not to real data, but to simulated data which was thinned in the way the methods assume. If the true parameters can be inferred from the posterior distributions, then this provides confidence that these methods are usable because they can at least model data produced under the model.
To this effect, I carried this out for the hard-core thinning case. Setting $\lambda = 1, R = 1$, I produced a Matern type III process, which I then applied the methods of the paper to.

I performed 10000 MCMC iterations, letting the first 1000 be burn-in iterations. The prior distribution on R was uniform on $[0, 2]$. The prior distribution on λ was $Gamma(1, 1)$. Computing the mean over the non burn-in samples, the average λ was found to be .873, an underestimate. The average R was found to be .983, which is much closer.

Below is a plot of the empirical L function $(L(r) - r)$ for the simulated process, as well as 2.5, 25, 75, and 97.5 quantiles of L functions produced over the non burn-in iterations. To produce these L functions, at each iteration of the MCMC sampling, a Matern type III process was simulated using the currently estimated values of R and λ. We can see that the empirical L function is comfortably within the range of the posterior predictive intervals, and shares the same shape.

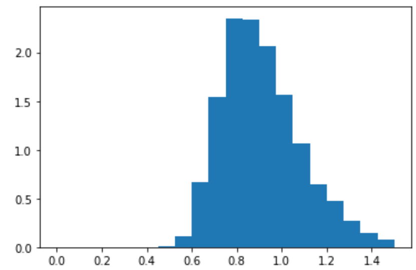
h 

Figure 6: 1a

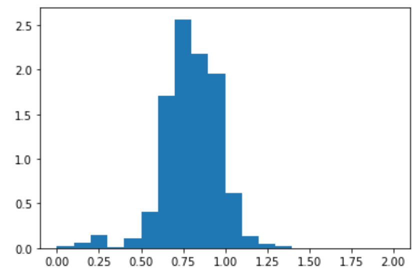


Figure 7: 1b

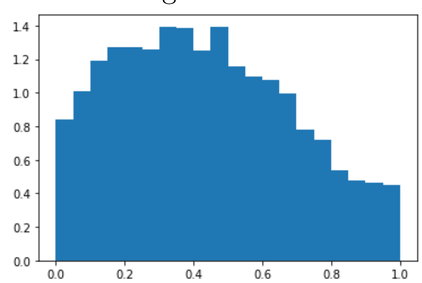


Figure 8: 1a

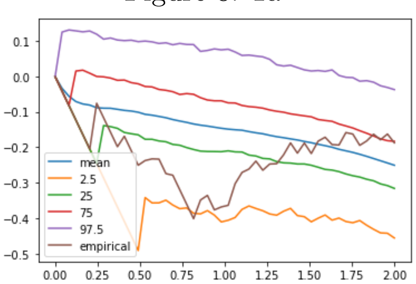


Figure 9: 1a

Figure 10: Posterior plots of lambda, R, p, and empirical L function for Simulated data (probabilistic thinning)

11 Appendix

11.1 Proof of Theorem 1

It follows from the definitions here that a Poisson process takes values in $(\mathcal{S}^{\cup}, \Sigma^{\cup})$, because with probability 1 we can order the points by their birth times, and because we can view finite subsets of $\mathcal{S}$ as elements of $\mathcal{S}^n \subseteq \mathcal{S}^{\cup}$.

To prove the second claim, we use the concepts of the Jannosy measure and density. First we write down the Jannosy density of the inhomogeneous Poisson point process.

Recall from Section 1.4 that we can see the Poisson Process as a point process in the sense of the definition given in Section 1.2: The discrete distribution is determined by $p_n = \frac{\Lambda(\mathcal{S})^n}{n!} e^{-\Lambda(\mathcal{S})}$ (Poisson probabilities), and the joint distribution Π_n is the distribution determined by i.i.d sampling n points according to density function f.)

By the definition of the inhomogeneous Poisson process, the density $f(s)$ can be written as $f(s) = \lambda(s)/\Lambda(\mathcal{S})$

To derive the Jannosy density of this process, we use the fact that the Poisson process is an independently and identically distributed cluster process (see Section 5.1 of Daley et al. [2003]). In Section 5.3 example 5.3(a) of Daley et al. [2003], the Jannosy density of independently and identically distributed cluster processes is derived to be:

$$j_n(s_1, ..., s_n) = p_n n! f(s_1)...f(s_n)$$

and plugging in our definitions of f and p_n,

$$j_n(s_1, ..., s_n) = \frac{\Lambda(\mathcal{S})^n}{n!} e^{-\Lambda(\mathcal{S})} n! \prod_{j=1}^{n} \left(\frac{\lambda(s_j)}{\Lambda(\mathcal{S})} \right) = e^{-\Lambda(\mathcal{S})} \prod_{j=1}^{n} \lambda(s_j)$$

Now we derive the stated result, that for any $B \in \Sigma^{\cup}$,

$$P(B) = \int_B p(S) \mu^{\cup}(\mathrm{d}S)$$

where $p(S)=e^{-\Lambda(\mathcal{S})}\prod_{j=1}^{n}\lambda(s_j)$.

Starting with the LHS, by definition of $\mu^{\cup}$,

$$P(B)=\sum_{n=0}^{\infty}p_n\Pi_n(B\cap\mathcal{S}^n)n!.$$

A subtle point here is that although $B\subseteq\mathcal{S}^{\cup}$ and therefore consists of increasing sequences only, by definition of the Poisson point process and the way that we view realizations of the Poisson point process as elements of $\mathcal{S}^{\cup}$, the $n!$ term appears above. This is because in considering the probability of an event B where the points of any $S\in B$ are in increasing order, according to the way we sample the Poisson process on $\mathbb{R}^d$ and then order them according to their last coordinate, there are $n!$ ways in which we can reorder the n elements so that the same element $S\in B$ resulted. Now on the RHS, by definition of $\mu^{\cup}$,

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

Therefore, to equate the LHS and RHS, it suffices to show terms in the sums are equal for every n. Using the definition of the Janossy measure and density we have that

$$p_n\Pi_n(B\cap\mathcal{S}^n)n!=J_n(B\cap\mathcal{S}^n)=\int_{B\cap\mathcal{S}^n}j_n(S)\mu^n(\mathrm{d}S)$$

From our derivation of j_n above, we see that this is exactly $p(S)$ and the claim is shown.

11.2 Proof of Theorem 2

Throughout the proof, λ is conditioned on, but we omit this from the notation. $G^+=(G,T^G)$ 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 the length of G^+, $|G^+|$, be k. $G^+$is obtained by thinning $F^+=\left(F,T^F\right)$, which is a sample from a homogeneous Poisson process with intensity λ. Let the size of F^+be $n\geq k$, and call the thinned points $\tilde{G}^+$. From Theorem 1, the density of F^+ with respect to the measure $\mu^{\cup}$ is

$$P\left(F^+\right)=\exp\{-\lambda\mu(\mathcal{S}\times\mathcal{T})\}\lambda^n.$$

The shadow of G^+ with thinning kernel K_θ is defined as

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

Next we want to calculate the conditional joint density of the thinned and Matern events given the primary process. Note that with the primary process given, the positions of the thinned and Matern events can only take on the values of the positions of the primary process. Similarly, their birth times are fixed. What is not fixed are the labels thinned or not thinned. By definition of the shadow and Matern thinning, we have:

$$P\left(G^+, \tilde{G}^+ \mid F^+\right) = \prod_{(s,t)\in\tilde{G}^+} \mathcal{H}_\theta\left(s,t;G^+\right) \prod_{(s,t)\in G^+} \left\{1 - \mathcal{H}_\theta\left(s,t;G^+\right)\right\}$$

because for a given labeling to occur, each thinned event must have been thinned by a Matern event, and each Matern event must not have been thinned by other Matern events, the probabilities of these occurrences are exactly given by the shadow (by definition), and they are all independent because as the sequential thinning process takes place, only the shadow of the previous Matern events affects the thinning probabilities of the current event.

The joint probability density with respect to $\mu^\cup$ is

$$P\left(G^+, \tilde{G}^+, F^+\right) = P\left(G^+, \tilde{G}^+ | F^+\right) P\left(F^+\right)$$

$$= \exp\{-\lambda\mu(\mathcal{S}\times\mathcal{T})\}\lambda^n \prod_{(s,t)\in\tilde{G}^+} \mathcal{H}_\theta\left(s,t;G^+\right) \prod_{(s,t)\in G+} \left\{1 - \mathcal{H}_\theta\left(s,t;G^+\right)\right\} \tag{2}$$

Also note that $P\left(G^+, \tilde{G}^+, F^+\right) = P\left(G^+, \tilde{G}^+\right)$. Integrating out the $\mathcal{S}$ locations of $n-k$ thinned elements,

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

Next we integrate out the values of $T^{\tilde{G}}$, which is an ordered sequence of $n-k$ elements in $[0,1]$.

What remains is $P\left(G^{+},\left|T^{\bar{G}}\right|=n-k\right)$, the joint probability of a sequence $G^{+}$and the event that there are $n-k$ thinned events:

$$P\left(G^{+},\left|T^{\bar{G}}\right|=n-k\right)=\exp\{-\lambda\mu(\mathcal{S}\times\mathcal{T})\}\lambda^{k}\prod_{(s,t)\in G^{+}}\left\{1-\mathcal{H}_{\theta}(s,t;G^{+})\right\}\times\frac{\left\{\lambda\int_{\mathcal{S}\times\mathcal{T}}\mathcal{H}_{\theta}\left(s,t;G^{+}\right)\mu(\mathrm{d}s\ \mathrm{d}t)\right\}^{n-k}}{(n-k)!}.$$

The $n-k$ power appears because this is the number of thinned events and the number of terms in the product; the $(n-k)!$ appears because we not only integrating with respect to the times themselves, but also with respect to their ordering. Summing this over the values of $\left|T^{\bar{G}}\right|$, which range from 0 to ∞,

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

In obtaining the last line, we used the Taylor expansion of the exponential function, and the fact that $\int_{\mathcal{S}\times\mathcal{T}}\mu(\mathrm{d}s\ \mathrm{d}t)=\mu(\mathcal{S}\times\mathcal{T})$.
This completes the proof.

11.3 Other Proofs

Corollary 1 is the fundamental result which allows the usage of Gibbs sampling because it allows for easy sampling of the thinned events given the Matern events, and once we have both of those, we have the full primary process. Conditioning on the full primary process makes sampling other parameters straightforward in the other Gibbs sampling steps.

11.4 Minor errors in code

In hard-core thinning for Swedish trees, there is a nonsensical prior used over parameter λ, which Gamma(10,.1), where .1 is the rate parameter. This is nonsensical because this is suggesting that on average, there should be 100 points per unit area, and since we're in a volume 100 window, this is saying we expect that (on average) there were 10000 primary points which were then thinned to obtain only 71 Matern events, with maximum interaction radius around .22 (since this is the minimum distance between Matern events). Even if the interaction radius were equal to the maximum, we would expect to have observed on the order of $100/(\pi * .22^2) \approx 657 >> 71$ total Matern events. I believe this is a typo in their code as in the paper they say they use shape and rate parameters equal to 1.

When calculating the Fano factor, they incorrectly use the standard deviation instead of the mean. This can lead to incorrect interpretation of data because computing the Fano factor in this way means that it will not be on average 1 for a homogeneous Poisson process.

References

Adrian Baddeley, Pablo Gregori, Jorge Mateu, Radu Stoica, and Dietrich Stoyan. *Case studies in spatial point process modeling*, volume 185. Springer, 2006.

AJ Baddeley and Bernard Walter Silverman. A cautionary example on the use of second-order methods for analyzing point patterns. *Biometrics*, pages 1089–1093, 1984.

José Javier Quinlan Binelli. Repulsive processes: Theory and applications. 2017.

Noel Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015.

Daryl J Daley, David Vere-Jones, et al. *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer, 2003.

Eva B Vedel Jensen and Linda Stougaard Nielsen. A review on inhomogeneous markov point processes. *Lecture Notes-Monograph Series*, pages 297–318, 2001.

PA W Lewis and Gerald S Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval research logistics quarterly*, 26(3):403–413, 1979.

Bertil Matérn. *Spatial variation*, volume 36. Springer Science & Business Media, 2013.

Iain Murray and Ryan P Adams. Slice sampling covariance hyperparameters of latent gaussian models. *Advances in neural information processing systems*, 23, 2010.

George LW Perry, Ben P Miller, and Neal J Enright. A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. *Plant ecology*, 187(1):59–82, 2006.

Brian D Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(2):172–192, 1977.

Brian D Ripley and Frank P Kelly. Markov point processes. *Journal of the London Mathematical Society*, 2(1):188–192, 1977.

Jakob Teichmann, Felix Ballani, and Karl Gerald van den Boogaart. Generalizations of matérn's hard-core point processes. *Spatial Statistics*, 3:33–53, 2013.

Maria Nicolette Margaretha van Lieshout. *Theory of Spatial Statistics: A Concise Introduction*. Chapman and Hall/CRC, 2019.

Lance A Waller, Aila Särkkä, Viktor Olsbo, Mari Myllymäki, Ioanna G Panoutsopoulou, William R Kennedy, and Gwen Wendelschafer-Crabb. Second-order spatial analysis of epidermal nerve fibers. *Statistics in Medicine*, 30(23):2827–2841, 2011.

Thorsten Wiegand and Kirk A. Moloney. Rings, circles, and null-models for point pattern analysis in ecology. *Oikos*, 104(2):209–229, 2004.

Liesbeth I Wilschut, Anne Laudisoit, Nelika K Hughes, Elisabeth A Addink, Steven M de Jong, Hans AP Heesterbeek, Jonas Reijniers, Sally Eagle, Vladimir M Dubyanskiy, and Mike Begon. Spatial distribution patterns of plague hosts: point pattern analysis of the burrows of great gerbils in kazakhstan. *Journal of Biogeography*, 42(7):1281–1292, 2015.