Bayesian Analysis of Sigmoidal Gaussian Cox Processes via Data Augmentation

Renaud Alie, David A. Stephens
Department of Mathematics and Statistics
McGill University
Montreal

Alexandra M. Schmidt
Department of Epidemiology, Biostatistics and Occupational Health
McGill University
Montreal

Abstract

Many models for point process data are defined through a thinning procedure where locations of a base process (often Poisson) are either kept (observed) or discarded (thinned). In this paper, we go back to the fundamentals of the distribution theory for point processes to establish a link between the base thinning mechanism and the joint density of thinned and observed locations in any of such models. In practice, the marginal model of observed points is often intractable, but thinned locations can be instantiated from their conditional distribution and typical data augmentation schemes can be employed to circumvent this problem. Such approaches have been employed in the recent literature, but some inconsistencies have been introduced across the different publications. We concentrate on an example: the so-called sigmoidal Gaussian Cox process. We apply our approach to resolve contradicting viewpoints in the data augmentation step of the inference procedures therein. We also provide a multitype extension to this process and conduct Bayesian inference on data consisting of positions of two different species of trees in Lansing Woods, Michigan. The emphasis is put on intertype dependence modeling with Bayesian uncertainty quantification.

1 Introduction

Spatial point processes describe the random behavior of point configurations in space. Several real-life phenomena have motivated the development of point process theory such as the position of stars in a galaxy [Babu and Feigelson 1996], the geographical position of trees [Wolpert and Ickstadt 1998] and the space-time locations of earthquakes [Ogata 1998]. The simplest model for point configurations is the Poisson point process (PPP) which is typically characterized through properties of its count function [Kingman 1992, Møller and Waagepetersen 2003]. PPPs come in finite and non-finite varieties, but only the finite case admits a constructive definition and a density. Cox models [Cox 1955] are hierarchical Poisson models with a stochastic intensity function. Even in the finite case, their density is usually not available in closed form. On the other hand, finite Markov point processes are described by properties of their density function [Møller and Waagepetersen 2003, van Lieshout 2019], which are traditionally taken with respect to the distribution of a PPP with unit intensity.

Other point process models are defined through a generative thinning procedure where only a subset of a base point process is observed. Examples include the three types of Matérn processes [Matérn 1960] where a deterministic thinning rule is applied to a base PPP. Matérn thinning prevents two points from lying within a predetermined distance of one another. Another important example is the thinning procedure of Lewis and Shedler [1979] to generate non-homogeneous PPPs. This one is probabilistic and operates independently across locations of the base process. It is usually the case that the marginal density of observed locations is not available in tractable form since one would need to integrate out the base point process along with the thinning procedure over every compatible configuration.

We focus on a Bayesian formulation of statistical inference for point process models, but some of the results we present also apply to other approaches where data augmentation is involved. Consider an observed point process \( Y \) which we postulate to be distributed according to a density \( f(Y|\theta) \) (wrt some measure \( \mu \)) indexed by a finite-dimensional parameter \( \theta \). We assign a prior density \( \pi(\theta) \) (wrt the Lebesgue measure \( \lambda \)) to the model parameters. This is a typical Bayesian specification where \( f(Y|\theta)\pi(\theta) \) is understood as the joint density of \( (Y, \theta) \) wrt to the product of...
μ and λ measures. In this context, one can use the observed $y$ and interpret $f(y|\theta)\pi(\theta)$ as, up to a normalizing constant, the density wrt to $\lambda$ of the regular probability distribution [Durrett, 2019 Section 4.1.3] of $\theta$ given $Y$. The situation is trickier, however, when considering data augmentation to circumvent some intractability in the model $f(Y|\theta)$. In such a case, a joint density $f(X,Y|\theta)$ is needed along with the prior, where $Y$ represents the observed data and $X$ the unobserved quantity. This joint model needs to be valid in the sense that it is a density wrt to some product measure $\nu \times \mu$ over the product of $X$ and $Y$ spaces. Proportionality arguments can then be used to interpret $f(X|Y,\theta) \propto f(X,Y|\theta)\pi(\theta)$ as the full conditional density (wrt to $\nu$) of $X$ given $Y$ and $\theta$. This fact is of direct importance in many computational approaches to the Bayesian analysis of point process data and models, which often utilize augmentation as a strategy.

We go back to the fundamentals of probability theory for point processes to resolve this obstacle and provide a construction of the posterior distribution that can be used in a general setting. Considering everything else fixed, any statement about the conditional distribution of thinned locations $X$ given the observed $Y$ should be obtainable from their joint density, provided they admit such. We present a colouring theorem that characterizes the joint density of thinned and observed locations for any thinning procedure applied to any point process that admits a density. This paper focuses on delicate aspects concerning computations for one class of point process models, the sigmoidal Gaussian Cox process (SGCP) [Adams et al., 2009]. We also examine, in Appendix [F] alternate derivations for some important results from the literature about the Matérn type III process. We illustrate how one specific result, the colouring theorem, streamlines most of the measure-theoretic details involved. Moreover, it provides a robust and unequivocal framework to state and verify any claim about joint, marginal and conditional distributions in models based on thinning procedures. The colouring theorem is not limited to binary categories (thinned and observed). It can be employed to derive the joint density of any number of point processes based on a categorical marking of a base process. This provides a powerful tool to construct multitype point processes with interesting dependence structure while retaining a tractable density. We use this result to provide a multitype extension of the SGCP along with the appropriate inference scheme that preserves the advantages of the original model: it does not involve the discretization or truncation approximations that are common to similar methods. We show how our multitype model and algorithms can be useful in investigating the interaction between different species of trees [Gerrard, 1969] from Lansing, Michigan: see Figure [1].

In the last decade or so, there has been a growing literature on non-homogeneous PPP modeling. Originally, Adams et al. [2009] proposed a data augmentation scheme to conduct inference about the intensity function of an observed point process $Y$. While the marginal PPP likelihood involves the integral of the intensity function, the complete data joint density $f(X,Y|\theta)$ of thinned ($X$) and observed ($Y$) locations is itself tractable. Imaginary thinned locations $X$ are instantiated at each step of the Markov chain Monte Carlo (MCMC) algorithm according to their full conditional. A space-time extension of this model is proposed in Gonçalves and Gamerman [2018]. The authors derive an alternative way to perform the data augmentation step by simulating the thinned locations exactly from their full conditional. As noted in their paper, their approach is incompatible with that of Adams et al. [2009]. In Rao et al. [2017], the authors use the same type of data augmentation but in the case of Matérn type III thinning of a non-homogeneous PPP. At the first step of the

![Figure 1: The location of maples and hickories in Lansing Woods, Michigan.](image-url)
Bayesian Analysis of SGCPs

hierarchy is the sigmoidal Gaussian Cox process (SGCP) model of [Adams et al., 2009]. The authors present yet another new procedure to instantiate the thinned point process from its exact full conditional. Even accounting for all the particularities associated with each model, the data augmentation step to be performed is the same in each of [Adams et al., 2009], [Rao et al., 2017] and [Gonçalves and Gamerman, 2018]. However, it can be shown that they each simulate from a different distribution. It is worth noting that the approach of [Gonçalves and Gamerman, 2018] has since been amended in a corrigendum [Gonçalves and Gamerman, 2023]. The authors adopt a new update strategy for the thinned point process, this one based on a retrospective sampling argument. Under a new unified formalism, our work provides a critical assessment of the four methods mentioned above.

Other comparable models based on latent random measures include the Poisson-gamma model of [Wolpert and Ickstadt, 1998] which has been extended to multitype modeling in [Kang et al., 2014]. The other notable example is the method of [Kottas and Sansó, 2007] in which the intensity measure is described by a scaled Dirichlet process mixture. This one has, hitherto, not been generalized to joint modeling of point patterns. Both methods include partial sum approximations of infinite mixtures. In contrast, our multitype adaptation of the SGCP preserves the advantages of the unitype version: it is exact up to Monte Carlo error. In the next section, we start by reviewing some basic notions of point process distribution theory. We describe and justify the type of dominating measure we shall use throughout as it will be necessary to discuss point processes in general spaces.

2 Point process modelling

2.1 Point Process Densities

The random locations of a point process take their values in some space \( S \). Usually, \( S \) corresponds to a real-life physical space: we have that \( S \) is a bounded subset of \( \mathbb{R}^d \) with \( d = 1, 2 \) or 3. Whenever \( d > 1 \), the locations exhibit no natural ordering. In this sense, it is usual to describe a point process \( X \) as a random countable subset of \( S \) [Møller and Waagepetersen, 2003]. We will concentrate on finite point processes (PPP) that have a finite number of locations with probability one. Therefore, we can assume that \( X \) is a random finite subset of \( S \) taking its values in \( \mathcal{N}_f(S) = \{ S \subset S : N(S) < \infty \} \), where we use \( N(S) \) to denote the cardinality of a set \( S \).

PPPs are parametrized by a positive intensity function \( \lambda : S \rightarrow [0, \infty) \). We write \( X \sim \text{PPP}(S, \lambda(\cdot)) \). In the finite case, we can construct such a process as follows:

1. The total number of points \( N(X) \) has the Poisson(\( \Lambda(S) \)) distribution, where \( \Lambda(S) = \int_S \lambda(s)ds \).
2. Conditional on \( N(X) \), the points are independently scattered over \( S \) according to the probability density \( \lambda(\cdot)\Lambda(S) \).

The process outlined above assumes the integral of \( \lambda(\cdot) \) over \( S \) is finite. For a measurable subset \( F \) of \( \mathcal{N}_f(S) \), we can write the distribution of \( X \) as

\[
P(X \in F) = \sum_{n \geq 0} \frac{\exp(-\Lambda(S))}{n!} \int_S \mathbb{1}_F(\{x_1, x_2, \ldots, x_n\}) \prod_{i=1}^{n} \lambda(x_i)dx_1dx_2\ldots dx_n. \tag{1}
\]

For completeness, the integral over \( S^0 \) should be understood as evaluating the inside function at the empty set \( \emptyset \).

Applying the monotone class theorem [Durrett, 2019, Section 5.2], we can extend the previous result from indicator functions to any positive and measurable function \( h : \mathcal{N}_f(S) \rightarrow [0, \infty) \):

\[
E[h(X)] = \sum_{n \geq 0} \frac{\exp(-\Lambda(S))}{n!} \int_S h(\{x_1, x_2, \ldots, x_n\}) \prod_{i=1}^{n} \lambda(x_i)dx_1dx_2\ldots dx_n. \tag{2}
\]

The unit rate PPP is defined by taking \( \lambda(\cdot) \) to be identically 1 everywhere on \( S \). If we define \( Y \sim \text{PPP}(S, 1) \), then \( Y \) is finite iff the space \( S \) is bounded (\( |S| < \infty \)). In such a case, we can derive the density of \( X \) with respect to the distribution of \( Y \) as

\[
f(\{x_1, x_2, \ldots, x_n\}) = \exp \left( (|S| - \Lambda(S)) \prod_{i=1}^{n} \lambda(x_i) \right)
\]

by careful manipulations of equations (1) and (2).

Taking the density of a point process with respect to the \( \text{PPP}(S, 1) \) distribution is a common approach in the literature. We argue in the following that being restricted to a bounded domain \( S \) can easily become hindering especially when considering marked point processes.
2.2 Counting-Scattering Measure

Taking the density of a FPP wrt to the unit rate PPP is analogous to taking the Radon-Nikodym derivative of an absolutely continuous distribution with respect to the uniform distribution, which is only valid if the support is bounded. In the unbounded \( S \) case, one could define the density wrt another Poisson process with intensity \( \rho(\cdot) \) chosen so that it is finite, i.e. \( \int_S \rho(s)ds < \infty \). This is analogous to taking the derivative of a continuous distribution with respect to an arbitrary distribution. However, the density of a FPP need not necessarily be taken wrt to a probability distribution. In the case of continuous random variables, we generally understand the term \textit{density} as the derivative wrt to the Lebesgue measure which is not a finite measure. Upon inspecting equation (1), we can see that the distribution of any finite PPP (regardless of whether \( S \) is bounded or not) is absolutely continuous wrt to the \( \sigma \)-finite measure

\[
\mu_{\pi}(F) = \sum_{n \geq 0} \int_{S^n} 1_F(\{x_1, x_2, \ldots, x_n\})d\lambda(x_1)d\lambda(x_2)\ldots d\lambda(x_n). \tag{3}
\]

Moreover, any FPP defined as in [Daley and Vere-Jones 2003, Section 5.3] from a probability mass function \( p_n (\sum_{n \geq 0} p_n = 1) \) on the total number of points and a family of symmetric densities \( \{\pi_n(\cdot), n \geq 1\} \) (each one over \( S^n \)) has density

\[
f(\{x_1, x_2, \ldots, x_n\}) = p_n \pi_n(\{x_1, x_2, \ldots, x_n\}) \tag{4}
\]

wrt \( \mu_{\pi} \). We refer to the process of assigning locations conditional on the number of points as \textit{scattering} and call \( \mu_{\pi} \) the \textit{counting-scattering} measure. The density wrt \( \mu_{\pi} \) is arguably more informative about the process than its counterpart taken wrt to the unit PPP because it relates directly to a simulation procedure. The density wrt (3) of the \( \text{PPP}(S, \lambda(\cdot)) \) distribution has the form

\[
f(\{x_1, x_2, \ldots, x_n\}) = \frac{\exp(-\Lambda(S))}{n!} \prod_{i=1}^{n} \lambda(x_i), \tag{5}
\]

provided it is finite.

The approach outlined above is inspired by Janossy measures [Jánossy 1950] and how they describe point process distributions. More recently, both [Rao et al. 2017] and Zhang [2017] defined a measure akin to (3). Distinctively, both papers consider FPPs to be ordered, albeit arbitrarily so. There is nothing inherently contradictory with this approach, but considering FPPs as sets is more in line with the established theory. Finally, the counting-scattering measure approach is also more easily adapted to more general spaces \( S \). The measure \( \mu_{\pi} \) as expressed in (3) is a hybrid between the counting and Lebesgue measures, but the scattering part of the distribution could use any \( \sigma \)-finite measure over a space \( S \) that may not be Euclidian. This will be useful in the next section when we consider FPPs with marks in general spaces.

2.3 Markings

Marks, unlike locations, do not generally lie in some physical space and therefore are not necessarily restricted to bounded spaces. Moreover, one could easily envision marks supported on a non-Euclidian space such as one with a discrete component. The unit rate PPP can be defined on more general spaces than subsets of \( \mathbb{R}^d \), but this is not common in the statistical literature. The counting-scattering measure is readily adapted to general spaces, however.

For finite point processes, we generally consider locations to precede marks. The latter are scattered conditionally on the former through a family of models

\[
\pi_n(m_1, \ldots, m_n|s_1, \ldots, s_n), \quad \text{for } n = 1, 2, 3, \ldots \tag{6}
\]

The conditional densities above should be invariant under permutations of indices \( \{1, 2, \ldots, n\} \) to preserve the natural symmetry of point processes. Fundamentally, the marks of a finite point process are themselves a FPP: a random number of locations in some mark space \( M \). For this reason, it is usual to consider the set of locations with corresponding marks as a single point process over the product space \( S \times M \). Along with location scattering distributions, the conditional densities in (6) specify the scattering models of this augmented point process:

\[
\pi_n(\{(s_1, m_1), \ldots, (s_n, m_n)\}) = \pi_n(m_1, \ldots, m_n|s_1, \ldots, s_n) \pi_n(s_1, \ldots, s_n), \tag{7}
\]

for any \( n \geq 1 \). From this point on, we assume the densities described in (7) to be respectively taken wrt to the product measure \( (\mu_S \times \mu_M)^n \) over measurable subsets of \((S \times M)^n\), where \( \mu_S \) (resp. \( \mu_M \)) is the dominating measure over the location (resp. mark) space.
As indicated above, there is no real distinction between points and marks in the product space representation. The two are equally important components of the same specification: a random finite subset of the product space \( S \times M \). As indicated in Section 2.2, we can write the density of this FPP as \( p_n \pi_n(\{(s_1, m_1), \ldots, (s_n, m_n)\}) \) without specifying whether a standard rate PPP is finite or even defined on such a space. This density is taken wrt to the counting-scattering measure defined by

\[
\mu(F) = \sum_{n \geq 0} \int_{(S \times M)^n} \mathbb{1}_F(\{(s_1, m_1), \ldots, (s_n, m_n)\}) \mu_S(ds_1) \mu_M(dm_1) \cdots \mu_S(ds_n) \mu_M(dm_n)
\]

for any measurable subset \( F \) of \( N_1(S \times M) \). How to construct a measure space over \( N_1(S \times M) \) is thoroughly discussed in [Zhang, 2017]. Essentially, one can define it from the measure spaces of \( \mu_S \) and \( \mu_M \) which are assumed to be provided. Marked point processes are akin to regular point processes except that they can take their values in general spaces rather than being confined to bounded subsets of \( \mathbb{R}^d \) where we typically observe locations. For this reason, it is helpful to avoid the restrictions associated with using the unit rate PPP as a dominating measure.

### 2.4 Colours

We define a colouring as a particular type of marking: one with a discrete and finite mark space \( M = \{0, 1, 2, \ldots, K\} \).

Although it is essentially the simplest type of marking one could imagine, colouring procedures are an interesting tool to construct multitype point processes. Indeed, by considering each colour as its own point configuration, we can study the joint, marginal and conditional distributions among the various types. A thinning procedure is a colouring with arbitrary colouring mechanisms. PPP are coloured independently of one another, then the point processes corresponding to each type are also Poisson.

Next, we present an important conceptual result that we term a colouring theorem in honor of the homonymous result in [Kingman, 1992]. The principle of a colouring theorem is to describe the marginal and joint structure of a multitype point process created by splitting a base point process into colours. The original version tells us that if the points of a marked point process are equally important components of the same specification: a random finite subset of the product space \( S \times M \) with respective sizes \( n_0, n_1, \ldots, n_K \).

To construct multitype point processes. Indeed, by considering each colour as its own point configuration, we can study the joint, marginal and conditional distributions among the various types. A thinning procedure is a colouring with arbitrary colouring mechanisms. PPP are coloured independently of one another, then the point processes corresponding to each type are also Poisson.

#### Theorem 1 (Colouring theorem)

Consider a FPP over the product space \( S \times \{0, 1, \ldots, K\} \) with density

\[
f(\{(s_1, c_1), \ldots, (s_n, c_n)\}) = p_n \pi_n(\{(s_1, c_1), \ldots, (s_n, c_n)\}).
\]

The dominating measure over the discrete mark space is the counting measure, hence we have

\[
\sum_{n \geq 0} \int_{S^n} \sum_{c \in \{0, 1, \ldots, K\}^n} f(\{(s_1, c_1), \ldots, (s_n, c_n)\}) \mu(ds_1) \cdots \mu(ds_n) = 1
\]

for some arbitrary measure \( \mu \) over subsets of \( S \). Let \( \mu_\ell \) denote the counting-scattering measure corresponding to \( \mu \) for FPPs over \( S \), that is

\[
\mu_\ell(F) = \sum_{n \geq 0} \int_{S^n} \mathbb{1}_F(\{(s_1, \ldots, s_n)\}) \mu(ds_1) \cdots \mu(ds_n), \quad \text{for measurable } F \subseteq N_1(S).
\]

Define the point processes \( X_0, X_1, \ldots, X_K \) as the locations in \( S \) with respective colours in \( \{0, 1, \ldots, K\} \). Let \( S_0, S_1, \ldots, S_K \) be finite subsets of \( S \) with respective sizes \( n_0, n_1, \ldots, n_K \). Then the joint density of \( X_0, X_1, \ldots, X_K \) wrt the product measure \( \mu_\ell^{K+1} \) has the form

\[
f(S_0, S_1, \ldots, S_K) = \binom{n}{n_0, n_1, \ldots, n_K} \int_{\bigcup_{k=0}^{K} S_k \times \{k\}}
\]

where \( n = \sum_{k=0}^{K} n_k \).

The proof can be found in Appendix A. In the finite case, the original colouring theorem for PPPs is recovered from Theorem II

**Remark 1.** It is possible to define a dominating measure such that the density of a FPP is of the form \( n! p_n \pi_n(\{s_1, \ldots, s_n\}) \). This changes the statement in the colouring theorem to

\[
f(S_0, S_1, \ldots, S_K) = f\left(\bigcup_{k=0}^{K} S_k \times \{k\}\right).
\]

In this case, the joint density is exactly equal to the density of the discretely marked point process, but it only holds for this particular choice.
We need to emphasize that, in general, the unitype density of the discretely marked point process is not the same thing as the multivariate density. The former is not taken wrt a product measure and therefore cannot be marginalized. This misconception has appeared in the literature and has led to some problematic arguments.

A result equivalent to Theorem 1 is stated in Møller and Waagepetersen [2003, Section 6.6.1], but no proof is given there. This arguably makes it more difficult to interpret when compared to our approach based on the counting-scattering measure. The implications of the result are important, yet under-appreciated, as we will discuss in Section 3. We will mostly limit ourselves to the particular case of independent colourings described in the following result.

**Corollary 1.** Let \( f \) be the density of the base point process \( X \) over \( S \). If the points of \( X \) are each independently coloured from a PMF \( p(c|s) \) for \( c = 0, 1, \ldots, K \) and \( s \in S \), then the joint density of the point processes \( X_0, X_1, \ldots, X_K \) corresponding to colours in \( \{0, 1, \ldots, K\} \) has the form

\[
f(S_0, S_1, \ldots, S_K) = \left(\begin{array}{c} n \\ n_0, n_1, \ldots, n_K \end{array}\right) f\left(\bigcup_{k=0}^{K} S_k\right) \prod_{k=0}^{K} \prod_{s \in S_k} p(k|s).
\]

In Appendix F we showcase the generality of Theorem 1 by applying it to the Matérn type III process and, doing so, handily recover some important results from this literature.

Theorem 1 provides a rigorous framework in which to state and verify statements about the marginal or conditional behaviour of thinned and observed locations in such contexts. It is simple enough, yet its implications are important to some of the statistical literature on point processes. In the next section, we showcase how the Gaussian Cox process introduced in Adams et al. [2009] and the data augmentation procedure described therein can be better understood through the lens of this colouring theorem. Extensions of this model are proposed in Rao et al. [2017], Gonçalves and Gamerman [2018], but some of the results across the three publications are inconsistent with one another. We use the framework described above to resolve those issues.

### 3 Unitype Sigmoidal Gaussian Cox Process

The sigmoidal Gaussian Cox Process (SGCP) was introduced in Adams et al. [2009]. It is a particular type of Cox process: a PPP model with a stochastic intensity function. The SGCP intensity is constructed by mapping a Gaussian process (GP) \( g(\cdot) \) to the \([0, 1]\) interval through a function \( \sigma(\cdot) \). Scaling is handled by a constant \( \lambda \). The global model hierarchy is presented below.

**Definition 1 (Sigmoidal Gaussian Cox Process).** The SGCP is the Cox process defined by

1. \( g \sim GP(m(\cdot), C(\cdot, \cdot)) \),
2. \( X_1 \sim \text{PPP}(\lambda \sigma \circ g(\cdot)) \).

The particular form of the mean \( m(\cdot) \) and covariance \( C(\cdot, \cdot) \) functions or their parametrization will not play a role in this section. The SGCP was originally presented in Adams et al. [2009] using the sigmoid \( \sigma(\cdot) = \expit(\cdot) := \exp(\cdot)/(1 + \exp(\cdot)) \) transformation. Gonçalves and Gamerman [2018] trade the expit link for a probit type of transformation \( \sigma(\cdot) = \Phi(\cdot) \) where \( \Phi \) is the standard normal CDF.

#### 3.1 Forward simulation

The usual difficulty with Gaussian Cox processes is that the PPP likelihood depends on the whole of \( g \) through the integral \( \int_S \lambda \sigma(g(s)) |ds| \) of the intensity function. It can be challenging to either conduct inference about the intensity function or integrate out \( g \) without relying on some finite-dimensional approximation. Yet, realizations of the SGCP, as defined above, can be simulated using a representation of the GP of random but finite dimension.

**Algorithm 1 (SGCP Simulation).** The following procedure generates a realization of the SGCP:

1. **Simulate a homogeneous PPP** \( X \) with intensity \( \lambda \),
2. **Instantiate the GP** \( g(\cdot) \) at the locations of \( X \) according to its finite dimensional Gaussian distribution,
3. **Keep every point** \( x \in X \) with probability \( \sigma(g(x)) \), otherwise discard \( x \).
The intuition behind the above algorithm is as follows. Suppose you could simulate and access the value of \( g(\cdot) \) at every location of \( S \). In any case, the intensity function \( \lambda g(\cdot) \) is bounded by \( \lambda \) and so step 2 of Definition 1 can be carried out by thinning a homogeneous PPP \( X \) of intensity \( \lambda \) (\( X \) does not depend on \( g(\cdot) \)). Once \( X \) is fixed, the thinning procedure operates without regard for the values of \( g(\cdot) \) at other locations, hence the rest of the GP values are superfluous. In that sense, the procedure outlined above is a retrospective sampler [Beskos et al., 2006; Papaspiliopoulos and Roberts, 2008]; the simulation of \( g(\cdot) \) is delayed until it is only needed in a finite-dimensional form to perform the thinning of \( X \). In Appendix B we demonstrate the validity of this procedure using only properties of Gaussian processes and point process distribution theory: the locations left at the end of Algorithm 1 are indeed exactly distributed according to the Gaussian Cox process of Definition 1.

Now consider \( \tilde{X}_0, \tilde{X}_1 \) as those point processes with GP distributed marks arising from the process \( g(\cdot) \), where \( \tilde{X}_0 \) corresponds to locations that were thinned and \( \tilde{X}_1 \) is its observed counterpart (those retained at the end of Algorithm 1). Both point processes consist of a finite number of values in \( S \times \mathbb{R} \). In particular, their distributions are not absolutely continuous wrt to the unit rate PPP distribution. Regardless, the joint density (in the sense described in Sections 2.2 and 2.3) of \( \tilde{X}_0, \tilde{X}_1 \) exists and is immediately derived by using Theorem 1. Indeed, following the steps of Algorithm 1 we can generate the \( \{0, 1\} \) coloured point process by first drawing a Poisson distributed number of points that are subsequently scattered uniformly on \( S \). The GP values can then be instantiated at those locations and finally each pair is coloured as observed with probability \( \sigma(g(\cdot)) \). Let \( (x, g, c) \) represent triplets in \( S \times \mathbb{R} \times \{0, 1\} \), the density of the marked point process described above is

\[
\begin{align*}
 f\left( \{ (x_1, g_1, c_1), \ldots, (x_n, g_n, c_n) \} \right) &= p_n \pi_n(x_1, \ldots, x_n) \pi_n(g_1, \ldots, g_n|x_1, \ldots, x_n) \\
 &\quad \pi_n(c_1, \ldots, c_n|x_1, g_1, \ldots, x_n, g_n) \\
 &= \exp\left(\frac{-\lambda |S|}{n!}N(g_1, \ldots, g_n|0, \Sigma(x_1, \ldots, x_n))\right) \\
 &\quad \prod_{i=1}^{n} \left[1 - \sigma(g_i)\right]^{1-c_i} \sigma(g_i)^{c_i}.
\end{align*}
\]

(10)

Notice that the dependence of the GP values on the locations acts through the covariance matrix \( \Sigma(x_1, \ldots, x_n) \) which is computed from the covariance function \( C(\cdot, \cdot) \) in the usual fashion. Also, recall that scattering models need to be equivalent under permutations of indices \( \{1, 2, \ldots, n\} \) as is the case here.

**Note:** Expression (10) is the density used in Gonçalves and Gamerman [2018] (they also use a counting-scattering type of dominating measure). However, this expression cannot be used to describe conditional distributions for thinned and observed locations because it is not a proper joint density (wrt to a product measure); it is the density of a single, unitype point process with a \( \{0, 1\} \) valued mark. This is a subtle but important distinction.

Theorem 1 tells us that the joint density of the thinned and observed point processes \( \tilde{X}_0 \) and \( \tilde{X}_1 \) is essentially the density of the \( \{0, 1\} \) marked point process multiplied by some combinatorial factor:

\[
\begin{align*}
 f(\tilde{S}_0, \tilde{S}_1) &= \binom{n_0 + n_1}{n_0, n_1} f(\{ \tilde{S}_0 \times \{0\} \} \cup \{ \tilde{S}_1 \times \{1\} \}) \\
 &= \exp\left(\frac{-\lambda |S|}{n_0! n_1!}N(\tilde{g}_0, \tilde{g}_1|0, \Sigma(\tilde{x}_0, \tilde{x}_1))\right) \prod_{i=1}^{n_0} \left[1 - \sigma(g_{0,i})\right] \prod_{j=1}^{n_1} \sigma(g_{1,j}).
\end{align*}
\]

(11)

In the final expression, \( \tilde{S}_k = \{ (x_{k,1}, g_{k,1}), \ldots, (x_{k,n_k}, g_{k,n_k}) \} \) is a finite subset of \( S \times \mathbb{R} \) while \( \tilde{x}_k = (x_{k,1}, \ldots, x_{k,n_k}) \) and \( \tilde{g}_k = (g_{k,1}, \ldots, g_{k,n_k}) \) are respectively the location and mark components (\( k = 0, 1 \)). Remember that the original objective was to obtain a tractable model for the augmented data \( \tilde{X}_0, \tilde{X}_1 \), and this is achieved: their joint density does not involve the integral of the intensity function. The key element of the colouring theorem in Section 2.4 is that deriving the joint density of the thinned and observed points becomes automatic.

Once the joint density of the thinned (\( \tilde{X}_0 \)) and observed (\( \tilde{X}_1 \)) locations has been established correctly, there is nothing controversial in using proportionality arguments to derive, up to a normalizing constant, the form of the conditional density of \( \tilde{X}_0|\tilde{X}_1 \). This is standard practice for deriving posterior distributions in Bayesian approaches. By removing every factor that does not involve \( \tilde{S}_0 \) in expression (11), we can write the density of the thinned locations (with associated GP marks) conditional on their observed counterpart:

\[
\begin{align*}
 f(\tilde{S}_0|\tilde{S}_1) \propto \frac{\lambda^{n_0}}{n_0!} N(\tilde{g}_0, \tilde{g}_1|0, \Sigma(\tilde{x}_0, \tilde{x}_1)) \prod_{i=1}^{n_0} \left[1 - \sigma(g_{0,i})\right].
\end{align*}
\]

(12)
The expression above is perfectly tractable. It could be employed in Metropolis-Hastings methods specifically designed for point processes such as the birth-death-move algorithm of Geyer and Møller [1994].

In expression (12), the GP values are intrinsically assumed to be known at the observed locations, that is $\tilde{g}_1$ needs to be given to use this conditional density. Those do not correspond to any observable quantities and also need to be simulated. In practice [Adams et al., 2009] [Rao et al., 2017] [Gonçalves and Gamerman, 2018], this is handled by alternating simulations of the thinned point process $\tilde{X}_0 | \tilde{X}_1$ from (12) and then all the GP values $\tilde{g}_0, \tilde{g}_1$ conditional on all the locations $\tilde{x}_0, \tilde{x}_1$, both thinned and observed. The thinned values $\tilde{g}_0$ of the GP are effectively sampled twice at each single step of the Markov chain.

Simulating the GP values conditional on locations is not a dimension-changing move and does not present any particular issue beyond standard applications of Metropolis-Hastings. This stage boils down to a latent GP model and one can choose the most suitable method to sample $\tilde{g}_0, \tilde{g}_1$ (along with covariance parameters) from an abundant literature. Examples include the elliptical slice sampler [Murray and Adams, 2010] and Hamiltonian Monte Carlo [Hensman et al., 2015]. See Filippone et al. [2013] for an extensive study of such algorithms.

The conditional in (12) is the Gaussian density evaluated at the GP values $\tilde{g}_0, \tilde{g}_1$. This necessitates $O(n^3)$ operations to compute, where $n = n_0 + n_1$ is the total number of observations (thinned and observed). Thus, the GP has to be instantiated at even more points than just the ones contained in the data and this number moves at every iteration. Moreover, if the observed points are configured in dense clusters, one might need a substantially larger number of thinned locations to fill in and make the base, uncoloured process homogeneous. When done repeatedly in an MCMC algorithm, the computational burden can accumulate quickly even if $n_1$ is only moderately large. Shiroti and Banerjee [2019] overcome this difficulty in the context of the SGCP using the nearest neighbor approximation [Datta et al., 2016] to factorize the Gaussian density.

In the next section, we focus on the point process on point process conditional (12) and assume everything else to be fixed. We discuss how to simulate the thinned point process from this density. Multiple methods have been proposed in the literature, some of which are inconsistent with one another, and we aim to provide a unifying clarification. See Appendix C for a full discussion.

### 3.2 Sampling the Thinned Point Process

Adams et al. [2009] employ birth-death-move proposals to simulate the thinned point process ($\tilde{X}_0$) conditional on the observed point process ($\tilde{X}_1$). The algorithm of Adams et al. [2009] can be interpreted as an application of existing trans-dimensional MCMC methodology. The acceptance ratios are identical to what would be obtained by applying the birth-death-move algorithm described in Geyer and Møller [1994]. They are also the same ratios as one would get upon using the more general reversible jump MCMC methodology [Green, 1995] (see example 3 in Tierney [1998]).

Inserting, deleting and displacing one point at a time can add up to a challenging computational task. However, previous computations can be stored efficiently to curb the complexity of evaluating the Gaussian conditional. Since only one location is changed at a time in the birth-death-move algorithm, the inverse of the covariance matrix can be determined after how many birth-death-move proposals can we reasonably consider the resulting point process to be a new realization of the thinned locations. If there are too few, then consecutive iterations are composed of essentially the same points.

The approach of Adams et al. [2009] is valid beyond any doubt in the sense that each update is reversible wrt the density in (12). It would nevertheless be preferable to simulate directly, if possible, from the distribution of the thinned locations $\tilde{X}_1$ conditional on the observed points. Gonçalves and Gamerman [2018] attempt direct sampling by first simulating the number of locations and then simulating their position in $\tilde{S}$. In their original publication, Gonçalves and Gamerman [2018] used a proportionality argument that omitted a term in the PMF of $n_0$. This step has been modified in Gonçalves and Gamerman [2023], which we discuss below.

### 3.3 Retrospective sampling

Conditional on the whole random field $g(\cdot)$, the thinned locations in $\tilde{S}$ are PPP($\lambda(1 - \sigma \circ g(\cdot))$) distributed and are independent of the observed locations. This is implied by the original Poisson colouring theorem [Lewis and Shedler, 1979], [Kingman, 1992]. In that case, the thinned locations can be simulated by thinning a Poisson process of homogeneous intensity $\lambda$. Consequently, only a finite representation of $g(\cdot)$ is required just as was the case in the retrospective sampler described by Algorithm 1.
This suggests a strategy to simulate from the conditional $\hat{X}_0|\hat{X}_1$. First, conceptually instantiate the GP from $g(\cdot)\mid \hat{X}_1$. Second, sample from $\hat{X}_0|g(\cdot)$ using Poisson thinning by uncovering the necessary values of $g(\cdot)$ in the process. The difficulty, as we will see, lies in determining the finite-dimensional distributions of the random field $g(\cdot)|\hat{X}_1$. On the one hand, $\hat{X}_1$ consists of locations in $S$ with associated GP marks. Intuitively speaking, we could conclude that $g(\cdot)|\hat{X}_1$ has the distribution of a GP conditional on some values being fixed. This is the argument invoked in [Rao et al., 2017, Section 7.1]. It implies the following retrospective sampling procedure to update the thinned point process:

1. Simulate a homogeneous PPP $X$ with intensity $\lambda$,

2. Instantiate the GP $g(\cdot)$ at the locations of $X$ according to its finite-dimensional Gaussian distribution conditional on the observed values $\hat{g}$, at the observed locations $\hat{x}$.

3. Keep each point of $\{(x, g(x)) : x \in X\}$ with probability $1 - \sigma(g(x))$.

The procedure above defines an SGCP-like Cox process with stochastic intensity $\lambda\{1 - \sigma(g_*(\cdot))\}$, where $g_*$ has the GP distribution of $g$ conditional on its value at the observed locations. Ultimately, the algorithm above is misconceived. Where the logic breaks down is that $g(\cdot)|\hat{X}_1$ is not simply a GP. There is more information in $\hat{X}_1$ than some values of $g(\cdot)$ at some locations of $S$. Those are values left after an independent thinning procedure with probabilities of being observed $\sigma \circ g(\cdot)$. In some sense, we would already expect the values of the GP marks in $\hat{X}_1$ to be high.

Another way to look at it is to consider the special case where the observed point process is empty. This situation is nonsensical from an applied perspective, but should nevertheless be handled by the general theory as the empty set is in the support of the observed point process $\hat{X}_1$. According to the procedure in [Rao et al., 2017], this would mean that the conditional $\hat{X}_0|\hat{X}_1 = \emptyset$ is the same as the marginal distribution of the thinned point process under the SGCP since there are no values upon which to condition the GP. This is somewhat suspicious as one would expect that 0 observations would be indicative of globally smaller values for the GP than under the marginal and, consequently, result in the conditional PMF for the number of thinned locations to be skewed towards higher values. In Appendix D, we use this special case as a counter-example to the overall claim.

The approach of [Gonçalves and Gamerman, 2023] is the same as the one in [Rao et al., 2017] with the exception that the GP values are simulated conditional on both the thinned and observed point processes.

**Algorithm 2.** Update procedure for the thinned point process $\hat{X}_0$ in [Gonçalves and Gamerman, 2023]:

1. Simulate a homogeneous PPP $X$ with intensity $\lambda$,

2. Instantiate the GP $g(\cdot)$ at the locations of $X$ according to its finite-dimensional Gaussian distribution conditional on the thinned and observed values $\hat{g}$, at the observed locations $\hat{x}$.

3. Keep each point of $\{(x, g(x)) : x \in X\}$ with probability $1 - \sigma(g(x))$.

This procedure is similar to the birth procedure in that new points are simulated uniformly on $S$ and are assigned a mark by conditioning the Gaussian distribution on all other observations. It is not an exact sample from the full conditional (12) since the current state of $\hat{X}_0$ plays a role in the update. The authors justify this method as a Gibbs sampler with the whole GP $g(\cdot)$ as one of the coordinates. The first step is to sample from $g(\cdot)|\hat{X}_0, \hat{X}_1$ which is performed retrospectively when sampling the Poisson process implied by $\hat{X}_0|g(\cdot), \hat{X}_1$. We would need to prove that the finite-dimensional distribution of $g(\cdot)$ conditional on the thinned and observed point process is indeed Gaussian as claimed. Recall that this was not the case when conditioning only on the observed points.

**Note:** In Appendix E, we offer a more direct argument that foregoes retrospective sampling entirely and instead operates at the level of the two finite point processes $\tilde{X}_0$ and $\tilde{X}_1$. We show that Algorithm 2 represents an update that is reversible wrt the conditional density in (12) and therefore leaves the target distribution invariant.

Rather than inserting or deleting one location at a time, the approach of [Gonçalves and Gamerman, 2023] updates the thinned locations with an entirely new point process at each iteration. It is also relatively straightforward to implement which makes it an enticing option when employed, as intended, as a data augmentation step in an inference algorithm. We apply the spirit of their algorithm in the extension we present in the next section.

### 4 Multitype Sigmoidal Gaussian Cox Process

In this section, we propose an extension to the SGCP that handles multiple types of points. The Lansing Woods data from Section 4 is publicly available from the spatstat package in R. We conduct MCMC-based Bayesian inference.
Bayesian Analysis of SGCPs

on the multitype point process consisting of the position of maples and hickories. This data set has been analyzed with a multitype log Gaussian Cox process (LGCP) model [Møller et al., 1998; Brix and Möller, 2001] in Waagepetersen et al. [2016] using a minimum contrast estimator. The sampling distribution of such an estimator can be complicated. The authors use the parametric bootstrap to quantify uncertainty in the estimation of correlation parameters. Our Bayesian approach offers an alternative way to measure uncertainty in such models. One advantage of multitype Gaussian Cox processes such as the LGCP or our multitype version of the SGCP is that inter-process dependence can be relegated to the cross-covariance function of the latent multivariate Gaussian random field. There is a vast literature on multivariate GPs, see Genton and Kleiber [2015] for a review of some of the more common modeling approaches.

4.1 Model Specification

We generalize the SGCP model with probit link that was used in Gonçalves and Gamerman [2018] to model multitype point patterns. We assume at the first level that the point processes $X_1, \ldots, X_K$ corresponding to the various types are independent and Poisson distributed conditional on some random and multivariate intensity function. Starting from a $K$-variate GP $g(\cdot)$, the function is defined by introducing a $K$-dimensional standard normal variable $Z$:

$$X_j \mid g(\cdot) \sim \text{PPP}(\lambda \sigma_j \circ g(\cdot)), \quad \text{where } \sigma_j(g) = P(\{\arg \max(g + Z) = j\} \cap \{\max(g + Z) > 0\}),$$

(13)

for $j = 1, 2, \ldots, K$. The $\sigma_j(\cdot)$ function maps a vector in $\mathbb{R}^K$ to a point on the $K$-dimensional probability simplex.

The $\sigma_j(\cdot)$ functions are defined by integrals of Gaussian densities over conic domains. They are cumbersome to compute whenever $K > 1$. For this reason, we adapt the auxiliary variable formulation of the probit multinomial model [McCulloch et al., 2000] to the spatial context by defining $Y(\cdot) = g(\cdot) + Z(\cdot)$. Here, $Z(\cdot)$ consists of $K$ independent white noise processes with unit variance. For $K = 1$, we recover the data augmentation scheme of Albert and Chib [1993]. We can redefine the model in terms of this new random field $Y(\cdot)$:

$$X_j \mid Y(\cdot) \sim \text{PPP}(\lambda \tau_j \circ Y(\cdot)), \quad \text{where } \tau_j(y) = 1(\{\arg \max(y) = j\} \cap \{\max(y) > 0\}),$$

(14)

where the functions $\tau_j(\cdot)$ are binary valued and at most one entry from $j = 1, 2, \ldots, K$ is non-null.

Adding a multivariate Gaussian random field distribution to $g(\cdot)$ completes the Cox process specification. We use the non-separable version of the linear model of coregionalization (LMC); see Schmidt and Gelfand [2003], Gelfand et al. [2004, 2005] for notable examples of its use in a Bayesian context. Such a process is constructed at each location from a linear transformation of independent random fields:

$$g(s) = \begin{bmatrix} g_1(s) \\ \vdots \\ g_K(s) \end{bmatrix} = A \begin{bmatrix} w_1(s) \\ \vdots \\ w_K(s) \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_K \end{bmatrix}, \quad \text{with } w_j(\cdot) \sim \text{GP}(0, C_j(| \cdot |)).$$

The independent GPs $w_j(\cdot)$ above are parametrized by some stationary and isotropic correlation functions $C_j(| \cdot |)$ while scaling is handled by the full rank matrix $A$. In our implementation, we use exponential correlation functions with distinct range parameter $\rho_j > 0$, i.e. $C_j(r) = \exp(-\rho_j r)$, $r \geq 0$. We add type-specific mean levels $\mu_j$, $j = 1, \ldots, K$. We chose the LMC as a multivariate spatial model over alternatives because its covariance structure can be exploited computationally to obtain likelihood evaluations that are linear in $K$ rather than cubic [Ali et al., 2024].

Same as in the univariate case described in Section 3, the product of PPP likelihoods implied by (14) involves the integral of the random intensity functions $\lambda r \circ Y(\cdot)$. On the other hand, we can interpret the model as arising from the colouring of a homogeneous PPP($\lambda$) process where a point at location $s \in S$ will be of species $j$ with probability $\sigma_j \circ g(\cdot)$ (equivalently if $\tau_j \circ Y(\cdot) = 1$) independently of other points. This point will be of none of the $K$ species if all the components of $Y(s)$ are negative. We consider the auxiliary PPP $X_0$ of intensity $\lambda \tau_0 \circ Y(\cdot)$ with $\tau_0(y) = 1(\{\max(y) < 0\})$. This point process consists of fictional thinned locations. Heuristically, $X_0$ fills in the gaps to ensure that the union $\bigcup_{j=1}^K X_j$ is homogeneous.

We can devise a retrospective sampling procedure analogous to Algorithm 1 to generate the point processes $X_0, X_1, \ldots, X_K$ (including the thinned locations) along with the $K$-variate GP values at each location. To summarize, we can simulate $X \sim \text{PPP}(\lambda)$ and instantiate $g(\cdot)$ at this finite number of locations in accordance with the multivariate normal distribution implied by the LMC. The values of the process $Y(\cdot)$ are obtained by adding to $g(\cdot)$ a standard $K$-variate normal random variable at each location in $X$. Each point $x \in X$ can then be classified in one of the $K + 1$ categories along with its GP marks $g(x)$ and $Y(x)$ according to the rule defined by the $\tau_j(\cdot)$ functions.

Such a procedure operates without regard for the values of the GPs $g(\cdot)$ and $Y(\cdot)$ at locations of $S$ not included in $X$. Define $\tilde{X}_j$ as the points in $x \in X$ of the categorical mark (colour) $j$ along with their 2 corresponding multivariate
Bayesian Analysis of SGCPs

GP values $g(x)$ and $Y(x)$ (each point has $2K$ marks). The density of the point processes $\tilde{X}_0, \tilde{X}_1, \ldots, \tilde{X}_K$ w rt the product of counting-scattering measures $\mu^K_{(K+1)}$ is obtained from the Colouring Theorem of Section 2.4

$$f(\tilde{S}_0, \tilde{S}_1, \ldots, \tilde{S}_K) = \frac{\exp(-\lambda |S|) \lambda^{2K} \exp(n_j)}{n_j!} \cdot N(\tilde{g}_0, \ldots, \tilde{g}_K|\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_K, \mu^K_{j=1}, A, \rho^K_{j=1}) \cdot \prod_{i=1}^{n_j} N(y_{0,i} | g_{0,i}, I) \cdot \max(y_{0,i}) < 0 \cdot \prod_{j=1}^{K} \cdot n_j \cdot \prod_{j=1}^{K} \cdot n_j \cdot N(y_{j,i} | g_{j,i}, I) \cdot \{ \arg \max(y_{j,i}) = j \} \cdot \max(y_{j,i}) > 0 \}.$$  \hspace{1cm} (15)

In this last expression, $\tilde{S}_j$ are finite subsets of $S \times \mathbb{R}^{2p}$ composed of physical locations $\tilde{x}_j = (x_{j,1}, \ldots, x_{j,n_j})$ along with marks $\tilde{g}_j = (g_{j,1}, \ldots, g_{j,n_j})$ and $\tilde{y}_j = (y_{j,1}, \ldots, y_{j,n_j})$ for $j = 1, \ldots, K$. Importantly, including the point process $\tilde{X}_0$ makes the complete data likelihood free of intractable integral terms.

4.2 Bayesian Inference

A fully Bayesian specification is completed by assigning a prior to the model parameters $\lambda, \mu^K_{j=1}, A, \rho^K_{j=1}$. We respectively assign gamma and normal priors to $\lambda$ and $\mu^K_{j=1}$. This is mainly to ensure conjugate full conditional updates for these quantities. The components of $A$ are each assigned a normal prior. Finally, we put a uniform prior over $[3, 30]$ on each of $\rho^K_{j=1}$ which roughly corresponds to a practical range (the distance at which correlation is equal to 0.05) between 0.1 and 1. We obtain a posterior sample by alternating updates that are reversible w rt the full conditionals of quantities of interest: the thinned point process $\tilde{X}_0$, the values $\tilde{g}_0, \tilde{g}_1, \ldots, \tilde{g}_K$ and $\tilde{y}_0, \tilde{y}_1, \ldots, \tilde{y}_K$ of the latent GPs and the model parameters $\lambda, \mu^K_{j=1}, A, \rho^K_{j=1}$.

First, we perform the data augmentation step where the thinned point process $\tilde{X}_0$ is updated along with its GP marks. This is done, in the spirit of Algorithm 2, by first sampling a homogeneous PPP $X^*$ over $S$ of intensity $\lambda$. We then instantiate the random fields $g(\cdot)$ and $Y(\cdot)$ conditional on the GP observations contained in the current state of $\tilde{X}_0$, $\tilde{X}_1$, $\ldots$, $\tilde{X}_K$. The new locations are those $x \in X^*$ for which $\max(Y(x)) < 0$. Those form the updated point process $\tilde{X}_0$ along with their $2K$ marks.

Next we sample all the GP values $\tilde{g}_0, \tilde{g}_1, \ldots, \tilde{g}_K$ and $\tilde{y}_0, \tilde{y}_1, \ldots, \tilde{y}_K$ in a fixed dimension update. The marks of the thinned point processes are effectively sampled twice in a single update of the Markov chain. The values of the auxiliary GP $Y(\cdot)$ are sampled in turn from their univariate full conditional distribution which, from [13], are truncated normal. Updating the values of $g(\cdot)$ boils down to instantiating random effects in a latent Gaussian model. We use the algorithm described in [Alie et al., 2024] rather than a straightforward multivariate normal update to avoid factorizing the $nK \times nK$ covariance matrix. Importantly, those potentially high dimensional quantities are all sampled from full conditionals. There are no proposal parameters that could otherwise be complicated to tune and tailor to any particular example.

As for the model parameters, we use regular Metropolis-Hastings to sample the range parameters $\rho_j$ in turn. The whole coregionalization matrix $A$ is updated using slice sampling [Neal, 2003] [Murray and Adams, 2010]. Finally, $\lambda$ and the mean vector $\mu$ are sampled from their conjugate full conditionals, respectively the gamma and multivariate Gaussian distributions.

In the next section, we present our joint analysis of the two point configurations from Lansing Woods (showcased in Figure 1). To this end, we ran the algorithm outlined above for 10,000 iterations, discarding the first 2,000 as burn-in.

4.3 Application

The multitype extension allows us to fit the intensity function of the hickory and maple species simultaneously (see Figure 2). At each iteration of the Markov chain, the $K$-variate Gaussian process $g(\cdot)$ can be instantiated on a fine grid conditionally on their values $\tilde{g}_0, \tilde{g}_1, \ldots, \tilde{g}_K$ at thinned and observed locations. The transformation [13] of $g(\cdot)$ that defines the intensity function is intractable. However, we can instantiate the $Y(\cdot)$ process on this grid by adding a $K$-dimensional standard normal at each location. We then compute the binary transformations $\tau_j(\cdot)$, multiply each one by the current value of $\lambda$ and ultimately average over all iterations of the Markov chain. Any desired resolution can be achieved although computations scale as the number of pixels cubed. This can technically be accomplished after generating the Markov chain using conditional properties of the multivariate normal distribution. However, it is more efficient to perform the calculation at each iteration so that the inverse covariance matrix at observed and thinned locations need not be recomputed or stored.
We measure intra- and inter-species interaction with the (cross) pair correlation function (PCF): \[ \exp \left( \gamma(s, t) \right) \] was devised to estimate the intensity function of a non-homogeneous PPP. In the multitype setting, the point processes are assumed independent at the first stage. In that sense, there is no added benefit in fitting both intensity functions jointly (unless we believed them to be dependent \textit{a priori}).

Rather than considering the transformed GP as a flexible prior for the intensity function of Poisson distributed point patterns, we instead consider the model to be the Gaussian Cox process implied by integrating out the GP values. The parameters of interest are the base intensity \( \lambda \) and those driving the LMC including the coregionalization matrix \( A \), the range parameters \( \rho_j \) and mean levels \( \mu_j \), \( j = 1, \ldots, K \). From the perspective of posterior sampling, nothing changes other than the fact that interpolating the intensity function on a fine grid might not be needed anymore. From this point of view, the particular pattern in which maple and hickories grow is irrelevant. It is not possible to discriminate between a Cox process and a non-homogeneous PPP from only one point configuration [Møller and Waagepetersen, 2003; Isham, 2010]. The difference lies in the nature of repetition. If we were to look at a different patch of forest, maples and hickories would certainly grow into a different pattern. Nevertheless, we would expect to observe the same seemingly repulsive interaction between the two species.

Figure 2 shows that the procedure is flexible enough to capture the particular growing pattern of the two types of trees through the functional form of their respective intensity. However, it tells us where maples and hickories grow, but it does not provide an answer as to why they grow in such a configuration. The univariate method of Adams et al. [2009] was devised to estimate the intensity function of a non-homogeneous PPP. In the multitype setting, the point processes imply points of the same type that cluster at short range. We are more interested in the cross-pair correlation parameters sampled from their posterior distribution to account for Bayesian model uncertainty.

We measure intra- and inter-species interaction with the (cross) pair correlation function (PCF):

\[
\gamma_{kl}(s, t) = \frac{E[\tau_k(Y(s))\tau_l(Y(t))]|A, \rho_{j=1}^K, \mu_{j=1}^K]}{E[\tau_k(Y(s))]|A, \rho_{j=1}^K, \mu_{j=1}^K]} \frac{E[\tau_l(Y(t))]|A, \rho_{j=1}^K, \mu_{j=1}^K]}{E[\tau_l(Y(t))]|A, \rho_{j=1}^K, \mu_{j=1}^K]}, \quad k, \ell = 1, 2, \ldots, K. \tag{16}
\]

Roughly speaking, \( \gamma_{kl}(s, t) \) is the probability of observing points of types \( k, \ell \) at infinitesimal balls centered around \( s \) and \( t \) normalized by the product of marginal probabilities of observing them in such areas. This quantity is independent of the scale of the intensity function (\( \lambda \) cancels out in this expression). The PCF as expressed in (16) is a property of the Cox process model. It is a function of model parameters thereby expectations are taken conditionally on them being fixed. This caveat is only necessary in the Bayesian paradigm where parameters also have distributions.

Globally, the model is stationary and isotropic provided the covariance structure of the driving GPs is (so is the case with the exponential kernels \( \exp(-\rho_j|\cdot|) \) we use in our LMC specification). In this case, \( \gamma_{kl}(s, t) \) is a function of the distance \( r = |s - t| \) between \( s \) and \( t \). The cross PCF (16) does not admit a simple parametric form as a function of \( r \). It can nevertheless be evaluated at any distance \( r > 0 \) and for any set of parameters \( A, \rho_{j=1}^K, \mu_{j=1}^K \) using a consistent Monte Carlo estimator. Reasonable precision is rapidly attained as we only rely on averages computed from repetitions of 2K-dimensional Gaussian vectors. Those are distributed as two points of the \( Y(\cdot) \) process separated by distance \( r > 0 \). We evaluate the pointwise estimate of the PCF at multiple distances that we average over multiple sets of parameters sampled from their posterior distribution to account for Bayesian model uncertainty.

The estimated PCFs for the maple and hickory example are illustrated in Figure 3. Stationary and isotropic Cox processes imply points of the same type that cluster at short range. We are more interested in the cross-pair correlation function \( \gamma_{12} \); this quantity being smaller than 1 implies a repulsive interaction between maples and hickories. The dependence seems to diminish and become negligible at distances bigger than 0.4. An expanded discussion on the dependence structures that can be induced by our multitype SGCP is featured in Appendix G.

Figure 2: The pointwise posterior mean intensity function \( \lambda \circ Y(\cdot), j = 1, 2 \) along with tree positions for maples (left) and hickories (right).
Bayesian Analysis of SGCPs

Figure 3: The pointwise posterior mean for the PCF $\gamma_{11}$ is illustrated in blue along with 90% credible intervals computed at each point. Similarly, $\gamma_{22}$ is shown in orange and the cross pair correlation $\gamma_{12}$ is in gray.

5 Discussion

In this paper, we presented a general approach to obtain the distribution of a multitype point process defined through any discrete marking (colouring) of a base point configuration. This has implications for point process models defined by thinning procedures: the joint density of thinned and observed locations can be derived from the colouring theorem of Section 2.4. Instantiating the thinned point process conditional on observed locations can help circumvent intractability in the marginal model and help carry inference of model parameters in a Bayesian framework.

Such data augmentation schemes have been employed in recent publications but in the absence of a unifying approach to the concept. For example, Rao et al. [2017], Gonçalves and Gamerman [2018] introduce alternative methods to simulate the thinned locations in the context of the SGCP, but those turn out to be incompatible with the original formulation of Adams et al. [2009]. The alternative formulation introduced in the 2023 corrigendum of Gonçalves and Gamerman [2018] is perhaps the most enticing approach from a conceptual and computational standpoint. However, apart from the now retracted version of Gonçalves and Gamerman [2018], none of the other proposals came with a formal proof of validity concerning the data augmentation update. The conceptual tools we presented in this paper allowed us to indubitably resolve these inconsistencies. Moreover, Theorem 1 can be a very useful tool to design new multitype point process models with interesting dependence structures. In this regard, we introduce a generalization of the SGCP to jointly model multiple point patterns. We showcase our new model on the Lansing Woods data and conduct inference on the cross-pair correlation function of maples and hickories.

Besides correcting some misconceptions from the literature, our work can empower future authors to use more involved models and design even more efficient sampling algorithms without having to consider the intricate measure-theoretic details involved in defining joint, conditional and marginal densities for point processes. Interesting avenues to explore include thinning procedures applied to more complex models than Poisson point processes. For example, independently thinned Markov point processes have been introduced in the seminal work of Baddeley et al. [2000]. The general inference framework we provided might be an interesting alternative to the two-step semi-parametric procedure they present.

Funding

Alie, Stephens and Schmidt acknowledge support from the Natural Sciences and Engineering Research Council of Canada (NSERC). Alie also acknowledges the support from the Fonds de recherche du Québec – Nature et technologies (FRQNT).

References

Ryan Prescott Adams, Iain Murray, and David JC MacKay. Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In Proceedings of the 26th Annual International Conference on Machine Learning, pages 9–16, 2009.
Bayesian Analysis of SGCPs

James H Albert and Siddhartha Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, 1993.

Renaud Alie, David A Stephens, and Alexandra M Schmidt. Computational considerations for the linear model of coregionalization. *arXiv preprint arXiv:2402.08877*, 2024.

Gutti Jogesh Babu and Eric D Feigelson. Spatial point processes in astronomy. *Journal of statistical planning and inference*, 50(3):311–326, 1996.

Adrian J Baddeley, Jesper Møller, and Rasmus Waagepetersen. Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350, 2000.

Alexandros Beskos, Omiros Papaspiliopoulos, and Gareth O Roberts. Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6):1077–1098, 2006.

Anders Brix and Jesper Møller. Space-time multi type log Gaussian Cox processes with a view to modelling weeds. *Scandinavian Journal of Statistics*, 28(3):471–488, 2001.

David R Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society: Series B (Methodological)*, 17(2):129–157, 1955.

D. Daley and David Vere-Jones. *An Introduction to the Theory of Point Processes. Vol. I: Elementary Theory and Methods*. 2nd ed. Springer, 2003.

Abhirup Datta, Sudipto Banerjee, Andrew O Finley, and Alan E Gelfand. Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812, 2016.

Rick Durrett. *Probability: Theory and Examples*, volume 49. Cambridge University Press, 2019.

Maurizio Filippone, Mingjun Zhong, and Mark Girolami. A comparative evaluation of stochastic-based inference methods for Gaussian process models. *Machine Learning*, 93(1):93–114, 2013.

Alan E Gelfand, Alexandra M Schmidt, Sudipto Banerjee, and CF Sirmans. Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2):263–312, 2004.

Alain E Gelfand, Sudipto Banerjee, and Dani Gamerman. Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics: The official journal of the International Environmetrics Society*, 16(5):465–479, 2005.

Marc G Genton and William Kleiber. Cross-covariance functions for multivariate geostatistics. *Statistical Science*, 30(2):147–163, 2015.

Douglas James Gerrard. *Competition quotient: a new measure of the competition affecting individual forest trees*. Research Bulletin 20, Agricultural Experiment Station, Michigan State University, Lansing, 1969.

Charles J Geyer and Jesper Møller. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian journal of statistics*, pages 359–373, 1994.

Flávio B Gonçalves and Dani Gamerman. Exact Bayesian inference in spatiotemporal Cox processes driven by multivariate Gaussian processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):157–175, 2018.

Flávio B Gonçalves and Dani Gamerman. Corrigendum: Exact bayesian inference in spatiotemporal cox processes driven by multivariate gaussian processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(1):176–176, 2023.

Peter J Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

James Hensman, Alexander G Matthews, Maurizio Filippone, and Zoubin Ghahramani. MCMC for variationally sparse Gaussian processes. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

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

Valerie Isham. Spatial point process models. In *Handbook of Spatial Statistics*. CRC press, 2010.

Lajos Jánossy. On the absorption of a nucleon cascade. In *Proceedings of the Royal Irish Academy. Section A: Mathematical and Physical Sciences*, volume 53, pages 181–188. JSTOR, 1950.

Jian Kang, Thomas E Nichols, Tor D Wager, and Timothy D Johnson. A bayesian hierarchical spatial point process model for multi-type neuroimaging meta-analysis. *The annals of applied statistics*, 8(3):1800, 2014.
The following lemma encapsulates most of the technicalities involved in proving the colouring theorem of Section 2.4.

**Lemma 1.** Let $S_0, S_1, \ldots, S_K$ be any partition of the space $S$. The following holds for any positive function $h : \mathcal{N}(S) \rightarrow [0, \infty)$:

$$
\int h(S) \mu(S) dS = \int_{\mathcal{N}(S_K)} \cdots \int_{\mathcal{N}(S_1)} \int_{\mathcal{N}(S_0)} \left( \sum_{k=0}^{K} N_k \right) \mu(S_0) \mu(S_1) \cdots \mu(S_K),
$$

where we use $N_k = N(S_k)$ to denote the cardinality of the set $S_k$. 

**Proof of Theorem**

The following lemma encapsulates most of the technicalities involved in proving the colouring theorem of Section 2.4. It tells us how to decompose an integral over the set $\mathcal{N}(S)$ of all finite subsets of $S$ into an integral over finite subsets of smaller spaces $S_0, S_1, \ldots, S_K$.

**Lemma 1.** Let $S_0, S_1, \ldots, S_K$ be any partition of the space $S$. The following holds for any positive function $h : \mathcal{N}(S) \rightarrow [0, \infty)$:

$$
\int h(S) \mu(S) dS = \int_{\mathcal{N}(S_K)} \cdots \int_{\mathcal{N}(S_1)} \int_{\mathcal{N}(S_0)} \left( \sum_{k=0}^{K} N_k \right) \mu(S_0) \mu(S_1) \cdots \mu(S_K),
$$

where we use $N_k = N(S_k)$ to denote the cardinality of the set $S_k$. 

**Proof of Theorem**

The following lemma encapsulates most of the technicalities involved in proving the colouring theorem of Section 2.4. It tells us how to decompose an integral over the set $\mathcal{N}(S)$ of all finite subsets of $S$ into an integral over finite subsets of smaller spaces $S_0, S_1, \ldots, S_K$.

**Lemma 1.** Let $S_0, S_1, \ldots, S_K$ be any partition of the space $S$. The following holds for any positive function $h : \mathcal{N}(S) \rightarrow [0, \infty)$:

$$
\int h(S) \mu(S) dS = \int_{\mathcal{N}(S_K)} \cdots \int_{\mathcal{N}(S_1)} \int_{\mathcal{N}(S_0)} \left( \sum_{k=0}^{K} N_k \right) \mu(S_0) \mu(S_1) \cdots \mu(S_K),
$$

where we use $N_k = N(S_k)$ to denote the cardinality of the set $S_k$. 

**Proof of Theorem**

The following lemma encapsulates most of the technicalities involved in proving the colouring theorem of Section 2.4. It tells us how to decompose an integral over the set $\mathcal{N}(S)$ of all finite subsets of $S$ into an integral over finite subsets of smaller spaces $S_0, S_1, \ldots, S_K$.

**Lemma 1.** Let $S_0, S_1, \ldots, S_K$ be any partition of the space $S$. The following holds for any positive function $h : \mathcal{N}(S) \rightarrow [0, \infty)$:

$$
\int h(S) \mu(S) dS = \int_{\mathcal{N}(S_K)} \cdots \int_{\mathcal{N}(S_1)} \int_{\mathcal{N}(S_0)} \left( \sum_{k=0}^{K} N_k \right) \mu(S_0) \mu(S_1) \cdots \mu(S_K),
$$

where we use $N_k = N(S_k)$ to denote the cardinality of the set $S_k$. 

**Proof of Theorem**

The following lemma encapsulates most of the technicalities involved in proving the colouring theorem of Section 2.4. It tells us how to decompose an integral over the set $\mathcal{N}(S)$ of all finite subsets of $S$ into an integral over finite subsets of smaller spaces $S_0, S_1, \ldots, S_K$.

**Lemma 1.** Let $S_0, S_1, \ldots, S_K$ be any partition of the space $S$. The following holds for any positive function $h : \mathcal{N}(S) \rightarrow [0, \infty)$:

$$
\int h(S) \mu(S) dS = \int_{\mathcal{N}(S_K)} \cdots \int_{\mathcal{N}(S_1)} \int_{\mathcal{N}(S_0)} \left( \sum_{k=0}^{K} N_k \right) \mu(S_0) \mu(S_1) \cdots \mu(S_K),
$$

where we use $N_k = N(S_k)$ to denote the cardinality of the set $S_k$. 

**Proof of Theorem**

The following lemma encapsulates most of the technicalities involved in proving the colouring theorem of Section 2.4. It tells us how to decompose an integral over the set $\mathcal{N}(S)$ of all finite subsets of $S$ into an integral over finite subsets of smaller spaces $S_0, S_1, \ldots, S_K$.

**Lemma 1.** Let $S_0, S_1, \ldots, S_K$ be any partition of the space $S$. The following holds for any positive function $h : \mathcal{N}(S) \rightarrow [0, \infty)$:

$$
\int h(S) \mu(S) dS = \int_{\mathcal{N}(S_K)} \cdots \int_{\mathcal{N}(S_1)} \int_{\mathcal{N}(S_0)} \left( \sum_{k=0}^{K} N_k \right) \mu(S_0) \mu(S_1) \cdots \mu(S_K),
$$

where we use $N_k = N(S_k)$ to denote the cardinality of the set $S_k$. 

**Proof of Theorem**

The following lemma encapsulates most of the technicalities involved in proving the colouring theorem of Section 2.4. It tells us how to decompose an integral over the set $\mathcal{N}(S)$ of all finite subsets of $S$ into an integral over finite subsets of smaller spaces $S_0, S_1, \ldots, S_K$.

**Lemma 1.** Let $S_0, S_1, \ldots, S_K$ be any partition of the space $S$. The following holds for any positive function $h : \mathcal{N}(S) \rightarrow [0, \infty)$:

$$
\int h(S) \mu(S) dS = \int_{\mathcal{N}(S_K)} \cdots \int_{\mathcal{N}(S_1)} \int_{\mathcal{N}(S_0)} \left( \sum_{k=0}^{K} N_k \right) \mu(S_0) \mu(S_1) \cdots \mu(S_K),
$$

where we use $N_k = N(S_k)$ to denote the cardinality of the set $S_k$. 

**Proof of Theorem**

The following lemma encapsulates most of the technicalities involved in proving the colouring theorem of Section 2.4. It tells us how to decompose an integral over the set $\mathcal{N}(S)$ of all finite subsets of $S$ into an integral over finite subsets of smaller spaces $S_0, S_1, \ldots, S_K$.

**Lemma 1.** Let $S_0, S_1, \ldots, S_K$ be any partition of the space $S$. The following holds for any positive function $h : \mathcal{N}(S) \rightarrow [0, \infty)$:

$$
\int h(S) \mu(S) dS = \int_{\mathcal{N}(S_K)} \cdots \int_{\mathcal{N}(S_1)} \int_{\mathcal{N}(S_0)} \left( \sum_{k=0}^{K} N_k \right) \mu(S_0) \mu(S_1) \cdots \mu(S_K),
$$

where we use $N_k = N(S_k)$ to denote the cardinality of the set $S_k$. 

**Proof of Theorem**

The following lemma encapsulates most of the technicalities involved in proving the colouring theorem of Section 2.4. It tells us how to decompose an integral over the set $\mathcal{N}(S)$ of all finite subsets of $S$ into an integral over finite subsets of smaller spaces $S_0, S_1, \ldots, S_K$.

**Lemma 1.** Let $S_0, S_1, \ldots, S_K$ be any partition of the space $S$. The following holds for any positive function $h : \mathcal{N}(S) \rightarrow [0, \infty)$:

$$
\int h(S) \mu(S) dS = \int_{\mathcal{N}(S_K)} \cdots \int_{\mathcal{N}(S_1)} \int_{\mathcal{N}(S_0)} \left( \sum_{k=0}^{K} N_k \right) \mu(S_0) \mu(S_1) \cdots \mu(S_K),
$$

where we use $N_k = N(S_k)$ to denote the cardinality of the set $S_k$.
Bayesian Analysis of SGCPs

Proof. It is sufficient to demonstrate the case for a binary partition \( S_0, S_1 \). The general case can then be obtained by further partitioning those sets. For a positive function \( h \), the integral with respect to the counting-scattering measure can be computed as

\[
\int h(S) \mu_t(dS) = \sum_{n \geq 0} \int_{S^n} h(\{x_1, x_2, \ldots, x_n\}) \mu(dx_1) \mu(dx_2) \cdots \mu(dx_n)
\]

(17)

where \( \mu \) is a measure over subsets of \( S \). This is the extension of equation \( \text{[3]} \) from indicator to positive functions by the monotone class theorem.

For every \( n \geq 0 \), the inside integral on the RHS of (17) can be split into the \( 2^n \) parts defined by whether each of the coordinates belong to \( S_0 \) or \( S_1 \):

\[
\sum_{k \in \{0, 1\}^n} \int_{S_{k_0}} \cdots \int_{S_{k_n}} h(\{x_1, x_2, \ldots, x_n\}) \mu(dx_1) \mu(dx_2) \cdots \mu(dx_n).
\]

(18)

Recall that \( h \) is a set function and the arbitrary ordering \( 1, 2, \ldots, n \) is irrelevant to its value. It is symmetric wrt to the integral coordinates \( x_1, x_2, \ldots, x_n \).

Therefore, to evaluate any of the summands in (18), it matters only to know how many coordinates among the \( n \) belong to \( S_0 \) and, accordingly, how many belong to its complement \( S_1 \). We can thus catalog each term in summation (18) by first noting how many \( m \in \{0, 1, \ldots, n\} \) of its coordinates are integrated over \( S_0 \) and then sum the \( \binom{n}{m} \) equivalent components:

\[
\sum_{m=0}^{n} \binom{n}{m} \int_{S_0^m} \int_{S_1^{n-m}} h(\{x_1, x_2, \ldots, x_n\}) \mu(dx_1) \mu(dx_2) \cdots \mu(dx_n).
\]

(19)

Substituting (19) for the integral on the RHS of (17), we obtain

\[
\int h(S) \mu_t(dS) = \sum_{n \geq 0} \sum_{m=0}^{n} \binom{n}{m} \int_{S_0^m} \int_{S_1^{n-m}} h(\{x_1, x_2, \ldots, x_n\}) \mu(dx_1) \mu(dx_2) \cdots \mu(dx_n).
\]

We can re-index the double sum on the RHS using \( n_0 = m \) and \( n_1 = n - m \) and get

\[
\sum_{n_0 \geq 0} \sum_{n_1 \geq 0} \binom{n_0 + n_1}{n_0, n_1} \int_{S_0^{n_0}} \int_{S_1^{n_1}} h(\{x_1, x_2, \ldots, x_{n_0+n_1}\}) \mu(dx_1) \mu(dx_2) \cdots \mu(dx_{n_0+n_1}).
\]

This last expression is equivalent to the double integral

\[
\int_{N_i(S_1)} \int_{N_i(S_0)} \binom{N_0 + N_1}{N_0, N_1} h(S_0 \cup S_1) \mu_t(dS_0) \mu_t(dS_1),
\]

which completes the proof.

We now prove the colouring theorem of Section 2.4. Recall that this result relates the density of a point process \( \tilde{X} \) over domain \( \mathcal{S} \) with mark space \( \{0, 1, \ldots, K\} \) to the multitype process \( X_0, X_1, \ldots, X_K \) consisting of the locations in \( \mathcal{S} \) with those respective marks.

Proof of Theorem 7. Take any sets \( F_0, F_1, \ldots, F_K \subseteq N_t(S) \), it is sufficient to verify that the distribution

\[
P(X_0 \in F_0, X_1 \in F_1, \ldots, X_K \in F_K)
\]

(20)

is equal to the integral

\[
\int \cdots \int_{F_K} \cdots \int_{F_1} \left( \sum_{k=0}^{K} \binom{N_k}{N_0, N_1, \ldots, N_K} f(\bigcup_{k=0}^{K} S_k \times \{k\}) \mu_t(dS_0) \mu_t(dS_1) \cdots \mu_t(dS_K),
\]

where \( f \) is the density of the base process \( \tilde{X} \) wrt to \( \tilde{\mu}_t \): the counting-scattering measure constructed from the product of \( \mu \) with the counting measure, i.e.

\[
\tilde{\mu}_t = \sum_{n \geq 0} \int_{\mathcal{S}^n} \sum_{c \in \{0, 1, \ldots, K\}^n} 1_{\tilde{F}}(\{(s_1, c_1), \ldots, (s_n, c_n)\}) \mu(ds_1) \cdots \mu(ds_n).
\]
We need to express the distribution \( P \). We first transform equation (20) into a statement \( \tilde{P} \) about the distribution of the base point process for some carefully chosen \( \tilde{F} \subseteq N_{\lambda}(S \times \{0, 1, \ldots, K\}) \). For all \( k = 0, 1, \ldots, K \), we need to have \( \tilde{X} \cap \{S \times \{k\}\} = S_k \times \{k\} \) for some finite set \( S_k \in F_k \). Equivalently, it means that for

\[
\tilde{F} := \left\{ \bigcup_{k=0}^{K} S_k \times \{k\} : S_k \in F_k, k = 0, 1, \ldots, K \right\},
\]

we have \( \tilde{X} \in \tilde{F} \iff X_0 \in F_0, X_1 \in F_1, \ldots, X_K \in F_k \).

By definition of \( f \), we have

\[
P(\tilde{X} \in \tilde{F}) = \int \mathbf{1}_{\tilde{F}}(\tilde{\mathbf{S}}) f(\tilde{\mathbf{S}}) \mu_i(\text{d}\tilde{\mathbf{S}})
\]

which, for the partition \( \tilde{S}_0 = S \times \{0\}, \tilde{S}_1 = S \times \{1\}, \ldots, \tilde{S}_K = S \times \{K\} \), can be decomposed as

\[
\int_{N_i(\tilde{S}_0)} \cdots \int_{N_i(\tilde{S}_1)} \int_{N_i(\tilde{S}_0)} \left( \sum_{k=0}^{K} N_k \right) \mathbf{1}_{\tilde{F}} \left( \bigcup_{k=0}^{K} \tilde{S}_k \right) f \left( \bigcup_{k=0}^{K} \tilde{S}_k \right) \mu_i(\text{d}\tilde{S}_0) \mu_i(\text{d}\tilde{S}_1) \cdots \mu_i(\text{d}\tilde{S}_K)
\]

according to Lemma [1]. We can verify that an integral over the set of finite subsets of \( \tilde{S}_k = S \times \{k\} \) for any fixed \( k = 0, 1, \ldots, K \) can be expressed as an integral over \( N_i(S) \) as

\[
\int_{N_i(\tilde{S}_k)} h(\tilde{S}_k) \mu_i(\text{d}\tilde{S}_k) = \int h(S_k \times \{k\}) \mu_i(\text{d}S_k)
\]

for any integrable \( h \), where \( \mu_i \) is the counting-scattering measure over subsets of \( S \). Applying this fact to equation (22), we obtain

\[
P(\tilde{X} \in \tilde{F}) = \int \left( \sum_{k=0}^{K} \left( N_{0, N_1, \ldots, N_K} \right) \mathbf{1}_{\tilde{F}} \left( \bigcup_{k=0}^{K} S_k \times \{k\} \right) f \left( \bigcup_{k=0}^{K} S_k \times \{k\} \right) \mu_i(\text{d}S_0) \mu_i(\text{d}S_1) \cdots \mu_i(\text{d}S_K).
\]

Finally, per definition (21), we have that \( \bigcup_{k=0}^{K} S_k \times \{k\} \) will be in \( \tilde{F} \) if \( S_k \in F_k \) for all \( k = 0, 1, \ldots, K \). Substituting \( \mathbf{1}_{\tilde{F}} \) for \( \mathbf{1}_{\tilde{F}}(\bigcup_{k=0}^{K} S_k \times \{k\}) \) in (23) completes the proof.

\[
\square
\]

**B Validity of Algorithm [1]**

In this section, we demonstrate the equivalence between the marginal distribution of observed locations in \( S \) implied by Algorithm [1] and the Cox process of Definition [1]. The starting point is the density [1] of \( \tilde{X}_0, \tilde{X}_1 \), which are respectively the thinned and observed point processes with values in \( N_{i}(S \times \mathbb{R}) \). We will use \( \tilde{X}_1 \) to denote the observed locations without the Gaussian process marks.

We need to express the distribution \( P(X_1 \in F) \) for some arbitrary \( F \subseteq N_{i}(S) \) in terms of the joint distribution of \( \tilde{X}_0, \tilde{X}_1 \). The thinned locations can take any value but the observed locations along with GP marks need to be in a set \( \tilde{F} \) comprised of any finite set \( \{(x_1, g_1), \ldots, (x_n, g_n)\} \) such that \( \{x_1, \ldots, x_n\} \in F \) and \( \{g_1, \ldots, g_n\} \in \mathbb{R}^n \). We are looking for

\[
P(X_1 \in F) = P(\tilde{X}_0 \in N_i(S \times \mathbb{R}), \tilde{X}_1 \in \tilde{F})
\]

\[
= \sum_{n_0 \geq 0} \sum_{n_1 \geq 0} \int_{\mathbb{R}^{n_1}} \int_{\mathbb{R}^{n_0}} \mathbf{1}_{\tilde{F}}(\{(x_1,1, \ldots, x_1,n_1)\}) \frac{\exp(-\lambda(S))\lambda^{n_0+n_1}}{n_0!n_1!} \int_{\mathbb{R}^{n_1}} \int_{\mathbb{R}^{n_0}} N(\tilde{g}_0, \tilde{g}_1|0, \Sigma(\tilde{x}_0, \tilde{x}_1)) \prod_{i=1}^{n_0} \{1 - \sigma(\tilde{g}_0,i)\} \prod_{j=1}^{n_1} \sigma(\tilde{g}_{1,j})d\tilde{g}_0 d\tilde{g}_1 d\tilde{x}_0 d\tilde{x}_1.
\]

(24)
Importantly, we used the fact that

$$\mathds{1}_F \{ \{ (x_{1,1}, g_{1,1}), \ldots, (x_{1,n_1}, g_{1,n_1}) \} \} = \mathds{1}_F \{ \{ x_{1,1}, \ldots, x_{1,n_1} \} \}$$

since, by definition of $\tilde{F}$, it only matters that locations $\{x_{1,1}, \ldots, x_{1,n_1}\}$ are in $F$. The GP marks can take any value. Using the defining property of Gaussian processes, we can replace the integral over $\mathbb{R}^{n_0+n_1}$ in (24) by the expectation

$$E \left[ \prod_{i=1}^{n_0} \{1 - \sigma(g(x_{0,i}))\} \prod_{j=1}^{n_1} \sigma(g(x_{1,j})) \right]$$

with $g \sim \text{GP}(0, C(\cdot, \cdot))$.

By Fubini’s theorem for positive functions, we can relegate this expectation to the outside and choose to first compute the integral over the thinned locations:

$$\sum_{n_0 \geq 0} \int_{\mathcal{S}^{n_0}} \frac{\mathcal{L}^{n_0}}{n_0!} \prod_{i=1}^{n_0} \{1 - \sigma(g(x_{0,i}))\} d\tilde{x}_0 = \exp \left( \lambda \int_{\mathcal{S}} 1 - \sigma(g(s)) ds \right).$$

Returning to equation (24), we are left with

$$P(X_1 \in F) = E \left[ \sum_{n_1 \geq 0} \int_{\mathcal{S}^{n_1}} \mathds{1}_F \{ \{ x_{1,1}, \ldots, x_{1,n_1} \} \} \frac{\exp(-\lambda \int_{\mathcal{S}} \sigma(g(s)) ds) \lambda^{n_1}}{n_1!} \prod_{j=1}^{n_1} \sigma(g(x_{1,j})) d\tilde{x}_1 \right].$$

This is what we set out to prove as the above equation is simply the expectation over the GP function $g$ of the PPP($\lambda \sigma(g(\cdot))$) distribution. This is exactly the marginal implied by Definition 1.

### C Sampling the Thinned Process: Algorithms

**Adams et al. [2009]**: The authors employ birth-death-move proposals to simulate the thinned point process ($\tilde{X}_0$) conditional on the observed point process ($\tilde{X}_1$). When a birth is proposed, a new location $x_{\text{birth}}$ in $\mathcal{S} \times \mathbb{R}$ is simulated by first choosing a random point $x_{\text{birth}}$ uniformly in $\mathcal{S}$. Its associated mark $g_{\text{birth}}$ is then drawn from the Gaussian distribution conditioned on both thinned and observed marks $\tilde{g}_0, \tilde{g}_1$ at locations $\tilde{x}_0, \tilde{x}_1$. This move is accepted with probability

$$\min \left(1, \frac{\lambda |\mathcal{S}|}{n_0+1} \frac{1-b(n_0+1)}{b(n_0)} \left(1 - \sigma(g_{\text{birth}})\right), \right),$$

where $b(n_0)$ is the probability that a birth is proposed if there are $n_0$ thinned locations at the current state of the Markov chain. When a death is proposed, a random point $(x_{\text{death}}, S_{\text{death}})$ is chosen uniformly among the current thinned locations and its removal is accepted with probability

$$\min \left(1, \frac{n_0}{\lambda |\mathcal{S}|} \frac{b(n_0-1)}{1-b(n_0)} \frac{1}{1 - \sigma(g_{\text{death}})} \right).$$

The locations can also be displaced. The acceptance rate in this case can be deduced from the regular, fixed dimension Metropolis-Hastings procedure.

**Gonçalves and Gamerman [2018]**: The authors attempt direct sampling by first simulating the number of locations and then simulating their position in $\mathcal{S}$. To do this, we need to derive both the PMF for the total number of locations and the scattering distributions for $X_0 | X_1$: the point process with density

$$f(\tilde{S}_0 | \tilde{S}_1) = \frac{c^{\lambda_{n_0}}}{n_0!} \mathcal{N}(\tilde{g}_0, \tilde{g}_1 | 0, \Sigma(\tilde{x}_0, \tilde{x}_1)) \prod_{i=1}^{n_0} \{1 - \sigma(g_{0,i})\},$$

where $c$ is a constant that does not depend on $\tilde{S}_0$. The scattering density over $(\mathcal{S} \times \mathbb{R})^{n_0}$ for a given $n_0$ is proportional to $\mathcal{N}(\tilde{g}_0, \tilde{g}_1 | 0, \Sigma(\tilde{x}_0, \tilde{x}_1)) \prod_{i=1}^{n_0} \{1 - \sigma(g_{0,i})\}$ and can be simulated from by using a rejection sampling algorithm. However, the normalizing constant of this scattering density is a function of $n_0$. It means that the PMF for the total number $n_0$ of thinned locations is of the form

$$p_{n_0} \propto \lambda_{n_0} \int_{(\mathcal{S} \times \mathbb{R})^{n_0}} \mathcal{N}(\tilde{g}_0, \tilde{g}_1 | 0, \Sigma(\tilde{x}_0, \tilde{x}_1)) \prod_{i=1}^{n_0} \{1 - \sigma(g_{0,i})\} d\tilde{x}_0 d\tilde{g}_0,$$

which is highly intractable.
D Conceptual Mistake in [Rao et al. 2017, Section 7.1]

The thinning algorithm proposed by [Rao et al. 2017, Section 7.1] (see Section 3.3) does not simulate from the distribution (described by (12)) of thinned locations $\tilde{X}_0$ conditional on observed points $\tilde{X}_1$. To see this, we consider the simplest counterexample.

Consider the absurd but nonetheless valid situation where the observed point process is empty, i.e. $\tilde{X}_1 = \emptyset$. Under these circumstances, there are no observed GP values to condition upon in step 2 and therefore the marks are distributed in accordance with the random field $g \sim \text{GP}(0, C(\cdot, \cdot))$. In that case, the conditional distribution of $\tilde{X}_0 | \tilde{X}_1 = \emptyset$ would be the same as the marginal of $\tilde{X}_0$. We could ask: what is the probability of observing $\tilde{X}_0 = \emptyset$ given that $\tilde{X}_1 = \emptyset$? If the procedure outlined in [Rao et al. 2017] is valid, it would mean that

$$f(\emptyset, \emptyset) = \int f(\emptyset, \tilde{S}_1) \mu_{\tilde{S}_1}(d\tilde{S}_1) \int f(\tilde{S}_0, \emptyset) \mu_{\tilde{S}_0}(d\tilde{S}_0).$$

(25)

In this last expression, $f(\emptyset, \emptyset)$ is the joint density $\{1\}$ of $\tilde{X}_0$, $\tilde{X}_1$ evaluated at empty sets, and the integral terms are the marginal probabilities of being empty for the thinned and observed point processes under the SGCP. We demonstrate that equality (25) does not hold.

The RHS of equation (25) has the form

$$\sum_{n_0 \geq 0} \sum_{n_1 \geq 0} \exp(-2\lambda|S|) \frac{\lambda^{n_0+n_1}}{n_0!n_1!} \int_{S^{n_0+n_1}} \int_{\mathbb{R}^{n_0+n_1}} \mathcal{N}(\tilde{g}_0|0, \Sigma(\tilde{x})) \mathcal{N}(\tilde{g}_1|0, \Sigma(\tilde{x}_1)) \prod_{i=1}^{n_0} \{1 - \sigma(g_{0,i})\} \prod_{j=1}^{n_1} \sigma(g_{1,j}) d\tilde{g}_0 d\tilde{g}_1 d\tilde{x}_0 d\tilde{x}_1.$$

By the defining property of the GP $g$, this last expression is equal to

$$\sum_{n_0 \geq 0} \sum_{n_1 \geq 0} \exp(-2\lambda|S|) \frac{\lambda^{n_0+n_1}}{n_0!n_1!} \int_{S^{n_0+n_1}} E \left[ \prod_{i=1}^{n_0} \{1 - \sigma(g_{0,i})\} \right] E \left[ \prod_{j=1}^{n_1} \sigma(g(x_{1,j})) \right] d\tilde{x}_0 d\tilde{x}_1.$$

By Fubini’s theorem for positive functions, we can pull the sum and integral over the thinned point process inside the first expectation and obtain

$$E \left[ \sum_{n_0 \geq 0} \frac{\lambda^{n_0}}{n_0!} \left\{ \int_S 1 - \sigma(g(x)) dx \right\}^{n_0} \right] = E \left[ \exp \left( \lambda \int_S 1 - \sigma(g(x)) dx \right) \right],$$

where we used the series representation of the exponential function. Likewise, we can also put the integral and sum over the observed point process inside the second expectation:

$$E \left[ \sum_{n_1 \geq 0} \frac{\lambda^{n_1}}{n_1!} \left\{ \int_S \sigma(g(x)) dx \right\}^{n_1} \right] = E \left[ \exp \left( \lambda \int_S \sigma(g(x)) dx \right) \right].$$

Finally, we can bound the product of the two expectations below by employing Jensen’s inequality on each one. By doing so, we obtain

$$\exp(-\lambda|S|) < \int f(\emptyset, \tilde{S}_1) \mu_{\tilde{S}_1}(d\tilde{S}_1) \int f(\tilde{S}_0, \emptyset) \mu_{\tilde{S}_0}(d\tilde{S}_0),$$

where we used the linear property of expectations inside the exponential function. The LHS of the inequality above is exactly $f(\emptyset, \emptyset)$, and therefore this debunks equation (25). This is in agreement with the intuition in Section 3.3: the probability of the thinned locations being empty given that the observed point process is empty should be lower than under the marginal distribution of the thinned locations.

E Validity of Algorithm 2

Algorithm 2 is a thinning algorithm that is almost identical to Algorithm 1, the original algorithm for exact simulations of the SGCP. The difference is that Gaussian process marks are instantiated conditional on the values of $\tilde{g}_0, \tilde{g}_1$ at respective locations $\tilde{x}_0, \tilde{x}_1$. This GP information corresponds to what is contained in the current state of the thinned ($\tilde{X}_0$) and observed ($\tilde{X}_1$) point processes.
Bayesian Analysis of SGCPs

The joint density (wrt to the product measure $\mu \times \mu$) of the new thinned and observed point processes resulting from Algorithm 2 has a form that is very similar to (11):

$$f'(\tilde{S}_0, \tilde{S}_1) = \frac{\exp(-\lambda |S|) \lambda^{n_0+n_1'}}{n_0!n_1'} \frac{\mathcal{N}(\tilde{g}_0, \tilde{g}_1 | 0, \Sigma(\tilde{x}_0, \tilde{x}_1))}{\mathcal{N}(\tilde{g}_0, \tilde{g}_1 | 0, \Sigma(\tilde{x}_0, \tilde{x}_1))} \prod_{i=1}^{n_0'} (1 - \sigma(g_{0,i})) \prod_{j=1}^{n_1'} \sigma(g'_{1,j}). \quad (26)$$

where $\tilde{S}_k = \{(x'_{k,1}, g'_{k,1}), \ldots, (x'_{k,n_k}, g'_{k,n_k})\}$ is a finite subset of $S \times \mathbb{R}$ while $\tilde{x}'_k = (x'_{k,1}, \ldots, x'_{k,n_k})$ and $\tilde{g}'_k = (g'_{k,1}, \ldots, g'_{k,n_k})$ are respectively the vectorized form of the location and mark components ($k = 0, 1$). Conditional on the current thinned point process, the update density is obtained by integrating out the observed coordinate $\tilde{S}_1$ from (26):

$$q(\tilde{S}_0 \to \tilde{S}_0') = \int f'(\tilde{S}_0', \tilde{S}_1') \mu_1(d\tilde{S}_1').$$

From (12), we have

$$f(\tilde{S}_0 | \tilde{S}_1)q(\tilde{S}_0 \to \tilde{S}_0') \propto \int \frac{\lambda^{n_0+n_1'} \mathcal{N}(\tilde{g}_0, \tilde{g}_1 | 0, \Sigma(\tilde{x}_0, \tilde{x}_1))}{n_0!n_1'} \prod_{i=1}^{n_0'} (1 - \sigma(g_{0,i})) \prod_{j=1}^{n_1'} \sigma(g'_{1,j}) \mu_1(d\tilde{S}_1').$$

The expression above is symmetric wrt $\tilde{S}_0$ and $\tilde{S}_0'$ and therefore the update is reversible wrt the conditional density (12).

F Matérn Type III Process

In their article, [Rao et al. 2017] showcase another use of data augmentation in the context of Bayesian inference for point processes. The model they discuss is the Matérn type III process: one of the three repulsive processes described in Matérn [1960]. In their basic form, the Matérn processes do not allow points to fall within a distance $R$ of one another which leads to regular or underdispersed realizations. Each one is described by a thinning procedure applied to a base PPP. We focus on the type III process as it has a tractable distribution for the thinned locations conditional on the observed points. This is one of the main results in [Rao et al. 2017]. We re-derive it here using the theory developed in Section 2 to illustrate how the colouring theorem streamlines most of the intricate details involved.

The thinning procedure involved in generating a Matérn type III process proceeds as follows. First, we simulate a homogeneous PPP of intensity $\lambda > 0$ on $S$. Then, each point is independently assigned a time mark from the uniform distribution on $[0, 1]$. In the product space representation, this amounts to a homogeneous PPP on $S \times [0, 1]$ with intensity $\lambda$. Finally, browsing through the points in order of time, each point is thinned if it lies within a radius $R$ of an earlier non-deleted point. The base point process is Poisson, but the locations are not independently thinned. We use the general version of the colouring theorem presented in Section 2.4.

The fact that $S$ is bounded entails edge effects in the Matérn type III process. Indeed, those points near the border have less space around them and are therefore overall less likely to be thinned. In practice, we observe points in an arbitrary bounded window contained in a larger, conceptually unbounded, space where the process still unfolds. There is no magical border that would prevent a point in $S$ from being thinned by a location outside of it. In the case of the Matérn type III process, there is no easy fix to this phenomenon as the distance at which points can interact is unbounded. For example, a point near the border of $S$ could be thinned by an outside point within distance $R$ of it, but only if this point had not been thinned itself by another within distance $R$ and so on. Perfect simulation algorithms (without edge effects) of the Matérn type III process and other details are discussed in [Møller et al. 2010]. We focus on the simplistic case as it is set up in [Huber and Woltjer 2009] and [Rao et al. 2017] to demonstrate the results therein as particular applications of the colouring theorem.

We proceed exactly as in Section 3 for the SGCP. This approach applies in general to any thinning procedure over any PPP with a known density including the other two types of Matérn processes. We start by deriving the density of the base, $\{0, 1\}$ marked point process from the generative procedure. From it, we obtain the joint density for the thinned and observed point processes by applying Theorem 1. Once we have the proper joint density, we can discuss conditional and marginal densities without any ambiguity by using standard definitions. The generative procedure of the discretely marked point process corresponding to Matérn type III thinning can be condensed into the following two steps.

1. Simulate a homogeneous PPP $\tilde{X}$ with intensity $\lambda$ over $S \times [0, 1]$,
2. Cycling through all the points \((s, t) \in \bar{X}\) in order of time, assign label \(c = 0\) (thinned) to \((s, t)\) if it lies within distance \(R\) of an earlier point that was not thinned. Otherwise, \(c = 1\) (observed).

The density of the (untyped) FPP over \(S \times [0, 1] \times \{0, 1\}\) implied by the above procedure is the product of a homogenous PPP density with the proper scattering distribution \(\pi_n(c_1, \ldots, c_n|s_1, t_1, \ldots, s_n, t_n)\) for some \(n \geq 1\) with \((c_1, \ldots, c_n) \in \{0, 1\}^n\). This latter function accounts for how points in \(S \times [0, 1]\) are assigned the label 0 (thinned) or 1 (observed). Those are allocated sequentially in order of time in a deterministic fashion. Consider the permutation \(\sigma\) in which times are sorted in strictly ascending order, that is \(t_{\sigma(1)} < t_{\sigma(2)} < \cdots < t_{\sigma(n)}\). Browsing through the points in \(\sigma\) order, the \(i^{th}\) point is kept \((c_{\sigma(i)} = 1)\) if

\[
\prod_{j=1}^{i-1} \left\{1 - \mathbb{1}(||s_{\sigma(i)} - s_{\sigma(j)}|| < R)\right\}^{c_s(j)} = 1,
\]

where the exponent \(c_s(j)\) accounts for the fact that a point can only be thinned by previous observed points (we use \(0^0 = 1\) as a convention). The label scattering distribution can thus be factored in this \(\sigma\) order as

\[
\prod_{i=1}^{n} \left(1 - \prod_{j=1}^{i-1} \left\{1 - \mathbb{1}(||s_{\sigma(i)} - s_{\sigma(j)}|| < R)\right\}^{c_s(j)}\right)^{1-c_s(i)}\left(\prod_{j=1}^{i-1} \left\{1 - \mathbb{1}(||s_{\sigma(i)} - s_{\sigma(j)}|| < R)\right\}^{c_s(j)}\right)^{c_s(i)} \right) = 1.
\]

(27)

The value of this density whenever such an order \(\sigma\) cannot be defined (in the case of ties) is not important because the dominating measure allocates zero mass to such cases.

Defining the label scattering distribution in terms of ascending time order is intuitive given the particular generative procedure of the Matérn type III process. However, working with the global ordering \(\sigma\) is impeding when our goal is to discuss the thinned and observed point process separately. We aim for an equivalent form of equation (27) that can be evaluated in any order. We follow [Rao et al. 2017] in defining the shadow \(\mathcal{H}(s, t, s^*, t^*)\) of a location \((s^*, t^*) \in S \times [0, 1]\) at another location \((s, t)\) as the function that takes value 1 if \((s, t)\) would be thinned by \((s^*, t^*)\) and 0 otherwise:

\[
\mathcal{H}(s, t, s^*, t^*) = \mathbb{1}(t > t^*)\mathbb{1}(||s - s^*|| < R).
\]

(28)

Note that a point does not lie in its own shadow. Using this higher level concept, we can now let the innermost products in expression (27) run over all points (instead of only those that precede \(i\) in \(\sigma\) order) and let the first indicator function in (28) ensure that latter points do not contribute in determining the thinning configuration of earlier points:

\[
\prod_{i=1}^{n} \left(1 - \prod_{j=1}^{n} \mathcal{H}(s_i, t_i, s_j, t_j)\right)^{c_i} \left(\prod_{j=1}^{n} \mathcal{H}(s_i, t_i, s_j, t_j)\right)^{c_i} = 1.
\]

(29)

Since the thinning indicators are now evaluated for every pair of points, it does not matter in which order we compute each product. That is why we removed any mention of \(\sigma\).

As it stands, the procedure for thinning the homogeneous PPP on \(S \times [0, 1]\) operates in a deterministic fashion: we can rule exactly which points are to be labeled as thinned or observed from the locations and times. One of the most interesting contributions of [Rao et al. 2017] is to extend the Matérn Type III process to probabilistic thinning. This can be achieved by replacing the second indicator in (28) by a \([0, 1]\) valued kernel \(K(\cdot; s^*)\) centered at \(s^*\). Location \((s, t)\) shall be thinned by \((s^*, t^*)\) with probability

\[
\mathcal{H}(s, t, s^*, t^*) = \mathbb{1}(t > t^*)K(s; s^*).
\]

Now each previously unthinned location \(s^*\) has a chance to thin a point at \(s\), but there is no guarantee even if the two are very close.

We can consider the deterministic version as a special case of the probabilistic one. This generalization does not change the form of the label scattering distribution in (29) and every result in what follows can be expressed in terms of the \([0, 1]\) valued shadow function \(\mathcal{H}\), however we choose to define it. Various other extensions can also readily be handled such as location-dependent and/or random radius \(R\), but we stay clear of those for simplicity.
In any case, having obtained the expression for the scattering PMF of the thinned \((c = 0)\) and observed \((c = 1)\) labels (conditional on locations and times), we can now write the overall density of the base, \(\{0, 1\}\) marked point process as

\[
f(\{(x_1, t_1, c_1), \ldots, (x_n, t_n, c_n)\}) = p_n \pi_n(x_1, \ldots, x_n) \pi_n(t_1, \ldots, t_n) 
\]

\[
= \exp(-\lambda |S|) \lambda^n \prod_{i=1}^n \left[ 1 - \prod_{j=1}^n \left( 1 - H(s_i, t_i, s_j, t_j) \right)^{c_j} \right]^{-c_i} \prod_{j=1}^n \left( 1 - H(s_i, t_i, s_j, t_j) \right)^{c_j}.
\]

This is the density of the point process consisting of locations, times and thinning labels.

Now, let \(\tilde{X}_0\) represent the thinned physical locations along with time stamps and let \(\tilde{X}_1\) be its observed counterpart. Together, they form a multitype point process over \(S \times [0, 1]\). From Theorem 1, the joint density of \(\tilde{X}_0, \tilde{X}_1\) is the density of the marked point process evaluated at the proper thinning labels times a combinatorial factor:

\[
f(\tilde{S}_0, \tilde{S}_1) = \frac{\exp(-\lambda |S|) \lambda^{n_0 + n_1}}{n_0! n_1!} \prod_{i=1}^{n_0} \left[ 1 - \prod_{j=1}^{n_1} \left( 1 - H(s_{0,i}, t_{0,i}, s_{1,j}, t_{1,j}) \right) \right]^{n_0} \prod_{j=1}^{n_1} \left( 1 - H(s_{1,j}, t_{1,j}) \right)^{n_1} \prod_{i=1}^{n_0} \left( 1 - h(s_{0,i}, t_{0,i}, \tilde{S}_1) \right) \prod_{i=1}^{n_1} \left( 1 - h(s_{1,i}, t_{1,i}, \tilde{S}_1) \right),
\]

where \(\tilde{S}_k = \{(s_{k,1}, t_{k,1}) \ldots (s_{k,n_k}, t_{k,n_k})\}\) are finite subsets of \(S \times [0, 1]\) \((k = 0, 1)\). In the last equality, we introduce the compact notation \(h(s, t; \tilde{S}_1) = 1 - \prod_{j=1}^{n_1} \{ 1 - H(s, t, s_{1,j}, t_{1,j}) \}\). This function maps each location in \(S \times [0, 1]\) to a probability in \([0, 1]\) and can be understood as the shadow of all observed locations taken together.

We can compare this joint density with the marginal for the observed points. The likelihood of the observed locations was first derived in Huber and Wolpert [2009, Theorem 2.1] for the Matérn type III process. In their article, they use an accept-reject construction of this process to demonstrate this theorem. We can recover their result by integrating out the thinned point process from the joint density using standard calculations:

\[
f(\tilde{S}_1) = \int_{\mathcal{N}(S \times [0, 1])} f(\tilde{S}_0, \tilde{S}_1) \mu(d\tilde{S}_0)
\]

\[
= \sum_{n_0 \geq 0} \int_{(S \times [0, 1])^{n_0}} f(\{(s_1, t_1), \ldots, (s_{n_0}, t_{n_0})\}, \tilde{S}_1) \, ds_1 dt_1 \ldots ds_{n_0} dt_{n_0}
\]

\[
= \exp(-\lambda |S|) \lambda^{n_1} \prod_{i=1}^{n_1} \left( 1 - h(s_{1,i}, t_{1,i}, \tilde{S}_1) \right) \sum_{n_0 \geq 0} \frac{\lambda^{n_0}}{n_0!} \left\{ \int_{S \times [0, 1]} h(s, t; \tilde{S}_1) \, dsdt \right\}^{n_0}
\]

\[
= \exp(-\lambda \int \{ 1 - h(s, t; \tilde{S}_1) \} \, dsdt) \lambda^{n_1} \prod_{i=1}^{n_1} \left( 1 - h(s_{1,i}, t_{1,i}; \tilde{S}_1) \right).
\]

In the last equality, we use the power series representation of the exponential and the fact that \(|S| = \int dsdt\). This is reminiscent of a PPP density, but it is not the case as \(h(\cdot, \cdot; \tilde{S}_1)\) is a function of each and every location while PPPs have independent scattering. Just like the PPP density however, this last expression involves the integral of the shadow function which itself depends on model parameters such as the radius of thinning \(R\) in the deterministic version. In contrast, the joint distribution \(\mathcal{N}(S \times [0, 1])\) is better conditioned for inference.

Now, let us assume that time marks are available at the observed locations in \(S\) and we wish to instantiate the thinned point process \(\tilde{X}_0\) conditional on \(\tilde{X}_1\) in a data augmentation step. This could be useful for inference schemes where other quantities of interest can be updated elsewhere such as a Gibbs sampling algorithm. In the case of the Matérn type III process, this conditional distribution corresponds to that of a PPP. Indeed, with the observed locations fixed, the thinned locations do not interact with one another as noted in [Rao et al., 2017]. This fact can be verified by noting that
Bayesian Analysis of SGCPs

this conditional density is proportional (as a function of \( \tilde{s}_0 \)) to the joint of \( \tilde{X}_0, \tilde{X}_1 \):

\[
f(\tilde{s}_0|\tilde{s}_1) \propto f(\tilde{s}_0, \tilde{s}_1) \propto \frac{\lambda_{s_0}}{\lambda_{s_0}} \prod_{i=1}^{n_0} h(s_{0i}, t_{0i}; \tilde{s}_1).
\]

(32)

With \( \tilde{s}_1 \) fixed, we can see that the density in (32) is, up to a normalizing constant, that of a non-homogeneous PPP over \( S \times [0, 1] \) with intensity \( \lambda h(s, t; \tilde{s}_1) \) (recall the form of the PPP density in equation (5)). One can use a Metropolis-Hastings-based algorithm to sample from this conditional, but a perfect simulation of (32) can be obtained more efficiently by using the thinning procedure of Lewis and Shedler[1979].

The same result is obtained by using the standard definition \( f(\tilde{s}_0|\tilde{s}_1) = f(\tilde{s}_0, \tilde{s}_1)/f(\tilde{s}_1) \) which is how Rao et al. [2017] reach the same conclusion. To them however, the joint density of \( \tilde{s}_0, \tilde{s}_1 \) is the same thing as the density of the base, \( \{0, 1\} \) marked point process. However, this is only true for the particular choice of dominating measure they use (see Remark 1) and it is unclear whether the authors are aware of that. For example, the proof they provide would have led to an incoherent result had they used the unit rate PPP distribution as a dominating measure. Nevertheless, the approach we showcase here of working with the joint density gives a principled way of discussing conditional and marginal densities.

G Inter-Process Dependence Structure

This section explores the pairwise dependence structures that can be induced by the multitype SGCP model introduced in Section 4. We present, for two sets of GP parameter values, the resulting cross pair correlation function. In each case, we showcase how this dependence structure translates in a realization of the point processes and how the MCMC procedure we provide is able to capture this behavior. The relationship between the pair correlation function and the underlying vector valued GP \( g(\cdot) \) is given by

\[
\gamma_{k\ell}(|s - t|) = \frac{E[\sigma_k(g(s))\sigma_{\ell}(g(t))]}{E[\sigma_k(g(s))]E[\sigma_{\ell}(g(t))]}, \quad k, \ell = 1, 2, \ldots, K, \tag{33}
\]

where the transformations are given by

\[
\sigma_k(g) = P(\{\arg \max(g + Z) = k\} \cap \{\max(g + Z) > 0\}) \tag{34}
\]

for a random variable \( Z \sim N(0, I) \). The PCF in (33) is dependent on the GP covariance parameters \( A, \rho_{j=1}^K \) and the GP mean levels \( \mu_{j=1}^K \). It is invariant with respect to the base intensity level \( \lambda \). A PCF \( \gamma_{k\ell}(\cdot) > 1 \) implies an attractive interaction between processes \( k \) and \( \ell \) for \( k \neq \ell \) whereas \( \gamma_{k\ell}(\cdot) < 1 \) corresponds to a repulsive relationship. Both cases relate directly to the cross covariance function of the positive random fields \( \sigma_k(g)(\cdot) \) and \( \sigma_{\ell}(g)(\cdot) \) with positive (resp. negative) values corresponding to an attractive (resp. repulsive) interaction among processes \( k \) and \( \ell \).

If we consider the multivariate GP \( g(\cdot) \) to have high mean parameters \( \mu_{j=1}^K \), then from equation (34) we have \( \sigma_k(g) \approx P(\arg \max(g + Z) = k) \). In that context, there is a sort of competitive interplay between the random fields \( \sigma_k(g)(\cdot) \) and \( \sigma_{\ell}(g)(\cdot) \) and they might exhibit negative dependence even if the individual components \( g_k(\cdot) \) and \( g_\ell(\cdot) \) are positively correlated. On the other hand, if the mean levels \( \mu_{j=1}^K \) are lower, then we have \( \sigma_k(g) \approx P(g_k + Z_k > 0) \) and in that case the covariance among processes \( \sigma_k(g)(\cdot) \) and \( \sigma_{\ell}(g)(\cdot) \) better reflects the dependence among the underlying processes \( g_k(\cdot) \) and \( g_\ell(\cdot) \). In other words, lower mean parameters for the \( K \)-dimensional GP \( g(\cdot) \) leads to a more flexible pairwise dependence structure among the modeled point processes.

We study two examples that correspond to the attractive and repulsive inter-dependence structures described above. In both cases, we set an equal mean level for the two processes \( \mu_1 = \mu_2 = -1 \) and employ the same spatial range parameters \( \rho_1 = 5 \) and \( \rho_2 = 10 \). In the first example (illustrated in Figure 4), we induce a positive cross-correlation structure among processes \( g_1(\cdot) \) and \( g_2(\cdot) \) by employing the coregionalization matrix \( A = [1, 0.5]^{1}, [0.5, 1]^{1} \) when simulating the multitype SGCP. In the second example (Figure 5), we showcase a negative dependence structure with \( A = [[1, -0.5]^{1}, [0.5, -1]^{1}] \). We run the MCMC algorithm described in Section 4.2 for 10,000 iterations (discarding the first 2,000) for both illustrated realizations and present the 90% credible intervals around the true cross pair correlation function.
Figure 4: Positively Correlated Point Processes. On the left: We observe that both processes cluster in the same areas. On the right: The true cross PCF is illustrated in black and the shaded area in gray corresponds to the pointwise 90% credible intervals computed from posterior samples of the model parameters.

Figure 5: Negatively Correlated Point Processes. On the left: We observe that both processes cluster in different areas. On the right: The true cross PCF is illustrated in black and the shaded area in gray corresponds to the pointwise 90% credible intervals computed from posterior samples of the model parameters.