



CENTRE FOR **STOCHASTIC GEOMETRY**  
AND ADVANCED **BIOIMAGING**



Approximate Bayesian inference for a spatial point process model  
exhibiting regularity and random aggregation

**Ninna Vihrs, Jesper Møller and Alan E. Gelfand**

No. 05, April 2020

# Approximate Bayesian inference for a spatial point process model exhibiting regularity and random aggregation

Ninna Vihrs<sup>1</sup>, Jesper Møller<sup>1</sup> and Alan E. Gelfand<sup>2</sup>

<sup>1</sup>Department of Mathematical Sciences, Aalborg University

<sup>2</sup>Department of Statistical Science, Duke University

## Abstract

In this paper, we propose a doubly stochastic spatial point process model with both aggregation and repulsion. This model combines the ideas behind Strauss processes and log Gaussian Cox processes. The likelihood of this model is not expressible in closed form, but it is easy to simulate under the model. We therefore explain how to use approximate Bayesian computation (ABC) for statistical inference both for this specific model but also for spatial point process models in general. We suggest a method for model validation and comparison based on posterior predictions and global envelopes. We illustrate the ABC procedure and model comparison approach using both simulated point patterns and a real data example.

*Keywords:* Approximate Bayesian computation (ABC); doubly stochastic process; log Gaussian Cox process; model comparison; posterior prediction; Strauss process

## 1 Introduction

Spatial point patterns are usually divided into three cases: regularity/repulsiveness, complete spatial randomness, and aggregation/clustering. There is a wide selection of point process models suitable for these situations, see e.g. the overview in Lavancier & Møller (2016, Section 1) and the references therein. However, some point patterns show repulsiveness between the points at small scale and aggregation at a larger scale, see Lavancier & Møller (2016) for a detailed discussion. In this regard, Lavancier & Møller (2016) suggested a model for this situation obtained by a dependent thinning of a repulsive point process. It is also possible to construct certain Gibbs point processes with this behaviour, see e.g. Goldstein et al. (2015).

## 1.1 The log Gaussian Cox Strauss process

In this paper, we present a model for regularity at small scale and aggregation at larger scale which is a combination of a pairwise interaction point process and a log Gaussian Cox process. It is constructed by the following two steps.

First, we consider a pairwise interaction point process defined as follows. Let  $\mathbf{X}$  be a spatial point process viewed as a finite random subset of a given bounded region  $W \subset \mathbb{R}^2$  (we think of  $W$  as an observation window). Then  $\mathbf{X}$  is a pairwise interaction point process if  $\mathbf{X}$  follows a density (with respect to the unit rate Poisson process on  $W$ ) of the form

$$f(\mathbf{x} \mid \psi, \varphi) = \frac{1}{C_{\psi, \varphi}} \prod_{i=1}^n \psi(x_i) \prod_{i < j} \varphi(\|x_i - x_j\|) \quad (1.1)$$

for all point patterns  $\mathbf{x} = \{x_1, \dots, x_n\} \subset W$  with  $0 \leq n < \infty$  (if  $n = 0$  then  $\mathbf{x} = \emptyset$  is the empty point pattern), where the notation means the following:  $\psi : W \rightarrow [0, \infty)$  is a so-called first order interaction function;  $\varphi : [0, \infty) \rightarrow [0, \infty)$  is a so-called second order interaction function;  $\|\cdot\|$  denotes usual Euclidean distance; and  $C_{\psi, \varphi} = 1/f(\emptyset \mid \psi, \varphi)$  is the normalising constant which is required to be positive and finite. Usually,  $\varphi(\cdot) \leq 1$ , in which case the density is well defined and results in a model for repulsion between the points. The first order interaction function may be used to model systematic aggregation of points.

Second, we consider a doubly stochastic construction, by replacing  $\psi$  with a random function  $\Psi$  in order to introduce random aggregation to the model. This is an extension of a Cox process (the case  $\varphi = 1$ , cf. Cox, 1955), and such a model was considered in Berthelsen & Møller (2008) when  $\Psi$  is the stochastic intensity function of a shot noise Cox process. Instead, we use the random intensity function of a log Gaussian Cox process (LGCP, see Møller et al., 1998), which is a popular model for random aggregation. Specifically, we let

$$\Psi(u) = \exp(Z(u)), \quad u \in W, \quad (1.2)$$

where  $\mathbf{Z} := \{Z(u)\}_{u \in W}$  is a Gaussian random field (GRF) with constant mean  $\mu \in \mathbb{R}$  and exponential covariance function

$$c(u, v) = \sigma^2 \exp(-\|u - v\|/s), \quad u, v \in W.$$

Here,  $\sigma^2 \geq 0$  is the variance and  $s > 0$  is a scale parameter. For  $\sigma^2 > 0$ , the flexible stochastic process  $\Psi(u)$  may account for aggregation caused by unobserved covariates. Note that  $\Psi(u) = \exp(\mu)$  if  $\sigma^2 = 0$ .

For the second order interaction function in (1.1), Berthelsen & Møller (2008) used a piecewise linear function, whereas we will use the much simpler second order interaction function of a Strauss process (Strauss, 1975; Kelly & Ripley, 1976). This gives us a density for  $\mathbf{X}$  (with respect to the unit rate Poisson process) of the form

$$f(\mathbf{x} \mid \theta) = \mathbb{E} \left[ \frac{1}{C_{\theta}(\mathbf{Z})} \prod_{i=1}^n \exp(Z(x_i)) \prod_{i < j} \gamma^{1[\|x_i - x_j\| \leq R]} \right], \quad (1.3)$$

where  $\theta = (\mu, \sigma^2, s, \gamma, R)$  is the parameter vector. Here, the expectation is with respect to the GRF;  $C_\theta(\mathbf{Z})$  is the normalising constant obtained by conditioning on  $\mathbf{Z}$ ;  $1[\cdot]$  denotes the indicator function; and we use the convention  $0^0 = 1$ . The parameter  $R > 0$  is called the interaction radius and the parameter  $\gamma \in [0, 1]$  controls the repulsion between points. This model for  $\mathbf{X}$  will be referred to as an LGCP-Strauss process.

The model includes some well-known special cases:

- (a) Conditioned on  $\mathbf{Z}$ ,  $\mathbf{X}$  is an inhomogeneous Strauss process.
- (b) If  $\sigma^2 = 0$ ,  $\mathbf{X}$  is a usual Strauss process. If in addition  $\gamma = 0$ ,  $\mathbf{X}$  is a hard core Gibbs process with hard core parameter  $R$ ; or if in addition  $\gamma = 1$ ,  $\mathbf{X}$  is a homogeneous Poisson process on  $W$  with intensity  $\exp(\mu)$ .
- (c) If  $\gamma = 1$ ,  $\mathbf{X}$  is an LGCP.

The following coupling result becomes useful when interpreting the meaning of  $\gamma$  and when we later discuss simulation of the LGCP-Strauss process. To stress the dependence of  $\gamma$ , we write  $\mathbf{X} = \mathbf{X}_\gamma$ . Then, using a dependent thinning technique (Kendall & Møller, 2000) it follows that there exists a coupling of the LGCP-Strauss processes  $\mathbf{X}_\gamma$  for all  $\gamma \in [0, 1]$  such that  $\mathbf{X}_\gamma \subseteq \mathbf{X}_{\gamma'}$  whenever  $0 \leq \gamma < \gamma' \leq 1$ . In particular, the special case of the LGCP  $\mathbf{X}_1$  (item (c) above) dominates any of the LGCP-Strauss processes  $\mathbf{X}_\gamma$ . The intensity of  $\mathbf{X}_1$  is  $\exp(\mu + \sigma^2/2)$  (Møller et al., 1998), so  $\exp(\mu + \sigma^2/2)|W|$  provides an upper bound on the expected number of points in  $\mathbf{X}_\gamma$ . Here,  $|W|$  denotes the area of  $W$ .

Note that if we do not have any of the above special cases (a)–(c), both the intensity and other moment characteristics of  $\mathbf{X}$ , the density (1.3), and the Papangelou conditional intensity (see e.g. Møller & Waagepetersen, 2004) are not expressible in closed form. Therefore, in general, usual approaches for estimation based on likelihood, pseudo-likelihood, composite likelihood, and minimum contrasts (see the review in Møller & Waagepetersen, 2017) are not feasible for the LGCP-Strauss process. This makes statistical inference challenging.

## 1.2 Objectives and outline

In this paper, we show how to use approximate Bayesian computation (ABC) to make statistical inference for spatial point process models in general and provide further details for the LGCP-Strauss process. In brief, ABC is a flexible method for approximate inference in a Bayesian framework, which does not require the likelihood to be expressible in closed form. Instead, it is based on the ability to make simulations under the assumed model, which are then compared to the observed data by using summary statistics.

In previous work on ABC in the setting of spatial point process models, Shirota & Gelfand (2017) explained how ABC can be used for Strauss process models and determinantal point process models. For the Strauss process model they estimated the interaction radius using maximum profile pseudo likelihood and then kept the interaction radius fixed at this estimate during the ABC procedure. Further, Soubeyrand et al. (2013) presented an ABC method using functional summary statistics such as the pair correlation function, which they exemplified for a Thomas process model

and a marked point process model. Finally, Stoica et al. (2017) presented an ABC method applicable to spatial point process models with a continuously differentiable likelihood function.

In contrast to Shirota & Gelfand (2017), the method we use for statistical inference is based entirely on ABC, and unlike Shirota & Gelfand (2017) and Soubeyrand et al. (2013) we do not fix any of the unknown parameters during the ABC procedure. Furthermore, we provide a detailed discussion of the choice of summary statistics for ABC when making statistical inference for the LGCP-Strauss process. This discussion is equally relevant if using ABC in connection with other point process models which are special cases of the LGCP-Strauss process such as the LGCP or Strauss process. We also suggest a method for model validation and comparison based on posterior predictions and global envelopes. We use this in a simulation study to assess the quality of ABC results for LGCP-Strauss processes and to investigate whether realisations of the LGCP-Strauss process can be distinguished from LGCPs and Strauss processes.

The remainder of this paper is organized as follows. In Section 2, our chosen method for ABC model fitting is specified. Section 3 presents simulated examples of LGCP-Strauss processes with corresponding ABC analyses. Section 4 contains a real data example using a point pattern of oak trees which suffer from frost shake. Section 5 concludes with a brief summary and paths for future work.

The open source software R (R Core Team, 2019) is used for all statistical computations. Most plots are created with the R-package `ggplot2` (Wickham, 2016) and some of the functionalities of the R-package `spatstat` (Baddeley et al., 2015) are used to handle spatial point patterns.

## 2 ABC for spatial point process models

### 2.1 The general case

Consider a spatial point process  $\mathbf{X}$  defined on a bounded region  $W \subset \mathbb{R}^2$  and which follows a parametric model with parameter vector  $\theta$ . Assume a realisation  $\mathbf{x}_{\text{obs}}$  of  $\mathbf{X}$  is observed. Our chosen procedure for ABC is specified in Algorithm 1 below. It is inspired by Shirota & Gelfand (2017) and is based on the semi-automatic approach by Fearnhead & Prangle (2012). Shirota & Gelfand (2017) used a Markov chain Monte Carlo method for the ABC sampling whereas we choose ABC rejection sampling, because of its simplicity and ability to be run in parallel.

In Algorithm 1,  $n(\mathbf{x})$  is the number of points in a point pattern  $\mathbf{x}$ , and in the first and last for loop we demand that  $n(\mathbf{x}) > m$  for each simulated  $\mathbf{x}$ . This is not strictly necessary for ABC, but it is a way to insure that summary statistics are not calculated for point patterns with very few points. Most summary statistics for spatial point patterns can only be calculated or considered reliable if there is a reasonable number of points in the point pattern. In the examples of Sections 3 and 4,  $m = 10$  was found to be sufficient.

In the second for loop of Algorithm 1, the linear models approximating the posterior means  $E[\theta_i \mid \mathbf{x}]$ ,  $i = 1, \dots, p$ , are fitted in a two-step procedure, which is a special case of a relaxed Lasso (Meinshausen, 2007): First, a model is fitted

---

**Algorithm 1:** Procedure for ABC

---

**Input** : Data  $\mathbf{x}_{\text{obs}}$ , a prior distribution  $\pi(\theta)$  for  $\theta = (\theta_1, \dots, \theta_p)$ , a procedure for simulating from the likelihood  $\pi(\mathbf{x} \mid \theta)$ , a summary statistic  $T(\mathbf{x}) = (T_1(\mathbf{x}), \dots, T_d(\mathbf{x}))$ , positive integers  $k_{\text{pilot}}$  and  $k_{\text{ABC}}$ , and a non-negative integer  $m$ .

**Output:** A sample  $\theta^{\text{ABC},1}, \dots, \theta^{\text{ABC},k_{\text{ABC}}}$  from the ABC approximate posterior distribution.

Calculate  $T_{\text{obs}} = T(\mathbf{x}_{\text{obs}})$ .

**Pilot run:**

**For**  $i = 1, \dots, k_{\text{pilot}}$

**repeat**

        | sample  $\theta^{\text{pilot},i} \sim \pi(\theta)$  and  $\mathbf{x}^{\text{pilot},i} \sim \pi(\mathbf{x} \mid \theta^{\text{pilot},i})$

**until**  $n(\mathbf{x}^{\text{pilot},i}) > m$ .

**For**  $j = 1, \dots, p$

    based on the sample  $\{(\theta^{\text{pilot},i}, \mathbf{x}^{\text{pilot},i})\}_{i=1}^{k_{\text{pilot}}}$ , fit a linear model for the posterior mean

$$\mathbb{E}[\theta_j \mid \mathbf{x}] \approx \theta_j(\mathbf{x}) := \alpha^j + \beta^{jT} (T(\mathbf{x}) - T_{\text{obs}})$$

    where  $\mathbf{x}$  is a realisation of  $\mathbf{X}$ ,  $\alpha^j \in \mathbb{R}$ , and  $\beta^j = (\beta_1^j, \dots, \beta_d^j) \in \mathbb{R}^d$ . Let  $\hat{\theta}_j(\mathbf{x})$  be the estimate of  $\theta_j(\mathbf{x})$  when  $\alpha^j$  and  $\beta^j$  are replaced by the estimates  $\hat{\alpha}^j$  and  $\hat{\beta}^j$ .

Define the distance measure

$$\chi(\mathbf{x}, \mathbf{x}_{\text{obs}}) = \sum_{j=1}^p \frac{(\hat{\theta}_j(\mathbf{x}) - \hat{\theta}_j(\mathbf{x}_{\text{obs}}))^2}{\text{var}(\hat{\theta}_j)} = \sum_{j=1}^p \frac{(\hat{\theta}_j(\mathbf{x}) - \hat{\alpha}^j)^2}{\text{var}(\hat{\theta}_j)}$$

where  $\text{var}(\hat{\theta}_j)$  is the empirical variance of  $\{\hat{\theta}_j(\mathbf{x}^{\text{pilot},i})\}_{i=1}^{k_{\text{pilot}}}$ . Choose  $\varepsilon$  as the empirical 1% percentile of  $\{\chi(\mathbf{x}^{\text{pilot},i}, \mathbf{x}_{\text{obs}})\}_{i=1}^{k_{\text{pilot}}}$ .

**ABC rejection sampling:**

**For**  $i = 1, \dots, k_{\text{ABC}}$

**repeat**

**repeat**

            | sample  $\theta^{\text{ABC},i} \sim \pi(\theta)$  and  $\mathbf{x}^{\text{ABC},i} \sim \pi(\mathbf{x} \mid \theta^{\text{ABC},i})$

**until**  $n(\mathbf{x}^{\text{ABC},i}) > m$ .

**until**  $\chi(\mathbf{x}^{\text{ABC},i}, \mathbf{x}_{\text{obs}}) < \varepsilon$

---

with Lasso regression, where the penalty term is chosen based on a cross-validation argument using the ‘one-standard-error rule’ (see e.g. Hastie et al., 2015, Chapter 2). Let  $\hat{\beta}_j^{i,\text{Lasso}}$ ,  $j = 1, \dots, d$ , be the resulting estimate of  $\beta_j^i$  and set  $T^{i,\text{Lasso}}(\mathbf{x}) = \{T_j(\mathbf{x}) \mid \hat{\beta}_j^{i,\text{Lasso}} \neq 0, j = 1, \dots, d\}$ . Second, the summary statistics in  $T^{i,\text{Lasso}}$  are used as predictors in a linear model fitted with ordinary least squares, which results in the final model for approximating  $E[\theta_i \mid \mathbf{x}]$ .

## 2.2 The special case of LGCP-Strauss process models

Consider again an LGCP-Strauss process  $\mathbf{X}$  on the observation window  $W$  with density (1.3), which depends on the unknown parameter vector  $\theta = (\mu, \sigma^2, s, \gamma, R)$ . One requirement for ABC is the ability to simulate data under this model for a given  $\theta$ . We do this in two steps: First, a realisation  $\mathbf{z}$  of  $\mathbf{Z}$  is simulated (see e.g. Schlather (1999)). In R, this can be done with the function `RFsimulate` from the R-package `RandomFields` (Schlather et al., 2015, 2019). Second, a realisation of  $\mathbf{X}$  given  $\mathbf{Z} = \mathbf{z}$  is simulated using an MCMC algorithm, namely a birth-death Metropolis-Hastings algorithm (Geyer & Møller, 1994, specifically, a birth is proposed with probability  $1/2$  and otherwise a death is proposed; for a birth proposal, the new point is generated from a density proportional to  $\exp(\mathbf{z})$ ; and for a death proposal, the point to die is selected uniformly from the current point pattern).

Another requirement for ABC is the selection of appropriate summary statistics  $T_j$ ,  $j = 1, \dots, d$  for Algorithm 1. Since the parameter  $\mu$  especially affects the number of points in a point pattern generated by an LGCP-Strauss process, we include the number of observed points as a summary statistic. Further, the  $L$ -function  $L(r) = \sqrt{K(r)/\pi}$ , where  $r > 0$  denotes inter-point distance and  $K$  is Ripley’s  $K$ -function (Ripley, 1976, 1977), summarises many important aspects of the second order moment properties of  $\mathbf{X}$ . Since  $L(r) = r$  for a Poisson process, one usually considers  $T(r) := L(r) - r$ . If  $L(r) < r$  ( $L(r) > r$ ), this indicates that  $\mathbf{X}$  is regular/repulsive (aggregated/clustered) at inter-point distances  $r$ . A simulation study suggested that for realisations of an LGCP-Strauss process, the empirical estimate of  $L(r) - r$  often has a global minimum when  $r$  is close to the interaction radius  $R$ , at least when there is strong to moderate repulsion in the model. In this regard, the bottom row of Figure 1 shows some empirical  $L$ -functions associated with realisations of LGCP-Strauss processes. We therefore also include summary statistics related to the  $L$  function (see (b)–(c) below).

Furthermore, the parameters  $\sigma^2$  and  $s$  of the LGCP-Strauss process affect the clustering in the process. To supply some summary statistics describing this, assume for ease of exposition that  $W$  is a square with side length  $h$ . Then we split  $W$  into  $q^2$  squares  $W_{i,j}$  of side length  $h/q$ ,  $i, j = 1, \dots, q$ , and let  $n(W_{i,j})$  be the number of points falling in  $W_{i,j}$ . We choose summary statistics which describe how  $n(W_{i,j})$  varies (see (d) below) and which are calculated for a user-specified finite range of  $q$ -values.

Specifically, for a point pattern  $\mathbf{x}$  (either  $\mathbf{x}_{\text{obs}}$  or one of the simulated point patterns in Algorithm 1), we chose the following summary statistics.

- (a)  $\log(n(\mathbf{x}))$ .

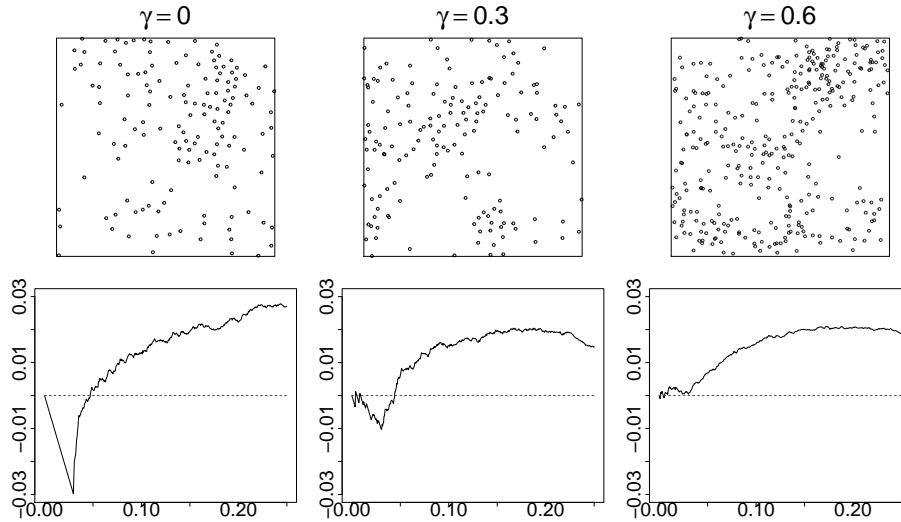
- (b)  $\max(\hat{L}(r) - r)$ ,  $\min(\hat{L}(r) - r)$ , and  $\arg\min(\hat{L}(r) - r)$ , where  $\hat{L}$  is a non-parametric estimate of the  $L$ -function evaluated over a user-specified finite range of  $r$ -values.
- (c)  $\hat{L}(r) - r$  evaluated at  $m$  equally spaced values of  $r$  between 0 and  $0.2h$ .
- (d)  $\max_{i,j=1,\dots,q}(\{n(W_{i,j})/n(\mathbf{x})\})$ ,  $\min_{i,j=1,\dots,q}(\{n(W_{i,j})/n(\mathbf{x})\})$ , and  $\log(\hat{\text{var}}(\{n(W_{i,j})/n(\mathbf{x})\}_{i,j=1}^q))$ , where again  $\hat{\text{var}}$  means empirical variance.

We have chosen these specific forms of the summary statistics based on some numerical experiments. In the examples of Sections 3 and 4,  $m = 40$  and  $q = 2, \dots, 5$ . This means that the vector of summary statistics  $T$  has dimension equal to 56.

### 3 Simulated examples

#### 3.1 Simulated data and prior specification

The top panels of Figure 1 show examples of simulated realisations of the LGCP-Strauss processes on the unit square for three different values of  $\gamma$ ; the remaining parameters are specified in the caption. The bottom panels of Figure 1 show the corresponding estimated  $L$ -functions using Ripley's isotropic edge correction (Ripley, 1977). As expected, the point patterns exhibit both regularity and aggregation, with an increasing degree of regularity at small to moderate distances as  $\gamma$  decreases, but a similar degree of aggregation at large distances.



**Figure 1:** Top panels: simulated LGCP-Strauss processes on the unit square, where the parameters are  $\mu = 5$ ,  $\sigma^2 = 2$ ,  $R = 0.03$ ,  $s = 0.3$ , and  $\gamma$  is as specified at the top of each column. Bottom panels: corresponding empirical  $L$ -function minus the identity (solid line) and the theoretical  $L$ -function for a Poisson process minus the identity (dashed line).

The ABC procedure in Algorithm 1 requires specification of (proper) prior distributions for the parameters in order to draw prior samples of the parameters. These samples are then used to simulate LGCP-Strauss processes with the given parameters. The more points a simulated point pattern has, the more computationally



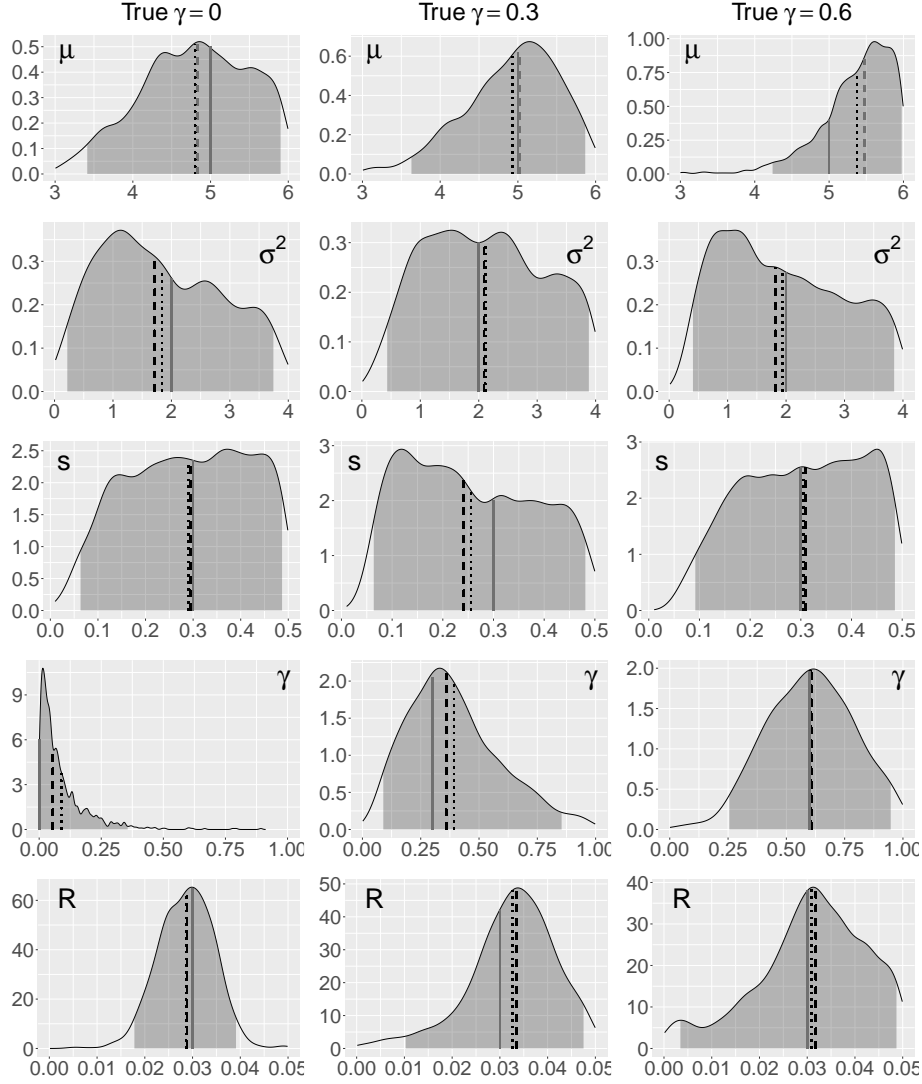
expensive the simulation procedure will be (see below). Therefore, we choose the prior distributions in such a way that these simulated point patterns will not yield unreasonably many points compared to the number of points in our observed point patterns. Specifically, when considering the prior distributions of  $\mu$  and  $\sigma^2$ , recall that  $\exp(\mu + \sigma^2/2)$  is an upper bound on the expected number of points in the unit square. So, for the examples in this section, independent uniform prior distributions are chosen for  $\mu$  on the interval  $(3, 6)$ ,  $\sigma^2$  on  $(0, 4)$ ,  $s$  on  $(0.01, 0.5)$ ,  $\gamma$  on  $(0, 1)$ , and  $R$  on  $(0, 0.05)$ .

In order to use the MCMC algorithm when a realisation  $\mathbf{z} = \{z(u)\}_{u \in W}$  of the GRF is given (see Section 2.2), it is necessary to choose a burn-in which can be used for all simulations in the ABC procedure. In order to choose this burn-in, we considered 30 samples of the parameters drawn from the prior distributions; used the MCMC algorithm for all these samples; and considered trace plots of the number of points and  $R$ -close pairs. Figure 7 shows these plots for three different prior samples for illustration. It seems that the higher the number of points, the slower the convergence. The burn-in should be high enough for the MCMC algorithm to have converged given any prior sample, but increasing the burn-in will also increase the computation time. Considering all 30 examples, 20 000 appears to be an appropriate overall burn-in. We furthermore choose to initiate the MCMC algorithm at the empty point pattern or at a realisation of an inhomogeneous Poisson process on  $W$  with intensity function  $\exp(z(u))$  (these initial states are extreme because of the coupling result mentioned in Section 1.1). The simulations in the top panels of Figure 1 and all following simulations are iteration 20 001 of the MCMC algorithm initiated at the empty point pattern.

### 3.2 Posterior results

We used Algorithm 1 on the three point patterns in Figure 1 with  $k_{\text{pilot}} = 10\,000$  and  $k_{\text{ABC}} = 1000$ , (the same choice as in Shirota & Gelfand, 2017). Figure 2 shows kernel density estimates of the resulting (approximate) marginal posterior distributions of the parameters, using a Gaussian kernel and a bandwidth chosen with the method by Sheather & Jones (1991). From Figure 2 we see the following. In all cases, the posterior mean and median are close and in most cases they agree with the true parameter value. As the true value of  $\gamma$  increases, the posterior distribution of  $\mu$  becomes more and more left skew. The approximate posterior distributions for  $\sigma^2$  and  $s$  look rather similar to their prior uniform distributions. When the true value of  $\gamma$  is 0, the posterior distribution of  $\gamma$  is very concentrated near 0, but it becomes more and more symmetric around the true value of  $\gamma$  as this increases. The spread of the posterior distribution of  $R$  seems to increase as the true value of  $\gamma$  increases. Overall, the ABC procedure seems to be most successful for estimating  $\gamma$  and  $R$ . Note that when fitting a Strauss process to a point pattern, Shirota & Gelfand (2017) first estimated  $R$  by maximum pseudo-likelihood and then used this value of  $R$  in their ABC procedure; in contrast we found no need to fix  $R$  when fitting an LGCP-Strauss process with our ABC procedure.

In geostatistics, it is known that the scale and variance parameters of an exponential covariance function for a GRF are unidentifiable (see e.g. Zhang, 2004). This



**Figure 2:** Estimated marginal ABC posterior densities for the parameters of the LGCP-Strauss model used for the point patterns in Figure 1. Each row represents a parameter (stated at the corner of each plot), and each column represents one of the three point patterns, as indicated by the true value of  $\gamma$ . The area between the 2.5% and 97.5% percentiles is shaded. The dashed, dotted, and solid lines indicate the median, mean, and true value, respectively.

might explain why the ABC procedure is not so successful when it comes to identifying these parameters. Therefore, we made the same analysis as in Figure 2 when  $s = 0.3$  is given. However, the posterior marginal distributions of the remaining parameters (not shown) looked very similar to those in Figure 2.

We now investigate how the ABC procedure for fitting an LGCP-Strauss process works when the data is generated from some of the special cases of this process. For this purpose, we simulated a realisation of an LGCP with parameters  $\mu = 5$ ,  $\sigma^2 = 2$ , and  $s = 0.3$ , and a realisation of a Strauss process with parameters  $\mu = 5$ ,  $\gamma = 0.3$ , and  $R = 0.03$ . Notice that when simulating under an LGCP, there is no need to employ the MCMC algorithm described at the beginning of Section 2.2. We used

the faster method implemented in the function `rLGCP` from the package `spatstat` (Baddeley et al., 2015). We used the same ABC procedure as above for fitting an LGCP-Strauss process to these point patterns and the posterior results can be seen in Figure 3. For the point pattern generated from an LGCP, the true value of  $\mu$  seems to be identified well when fitting the LGCP-Strauss process. The posterior marginal distribution of  $\gamma$  is rather concentrated near 1, and a plot of the posterior samples of  $\gamma$  and  $R$  (not shown) shows that very small values of  $\gamma$  appear together with very small values of  $R$ . This indicates that the fitted LGCP-Strauss process is close to the special case of an LGCP, which is the true model. Again, it seems to be difficult to identify  $\sigma^2$  and  $s$ .

For the point pattern generated from a Strauss process, the marginal posterior distribution for  $\sigma^2$  is very concentrated near zero, which is the true value. The true value of  $\gamma$  seems to be well identified whereas the median and mean of the marginal posterior distributions of  $\mu$  and  $R$  are somewhat higher than the true values. It appears that for this example, the ABC procedure is not as successful for  $R$  as it was for the point patterns in Figure 1. For the Strauss process,  $s$  should be irrelevant, and a plot of the posterior samples of  $s$  and  $\sigma^2$  (not shown) shows that  $s$  is irrelevant for small values of  $\sigma^2$ .

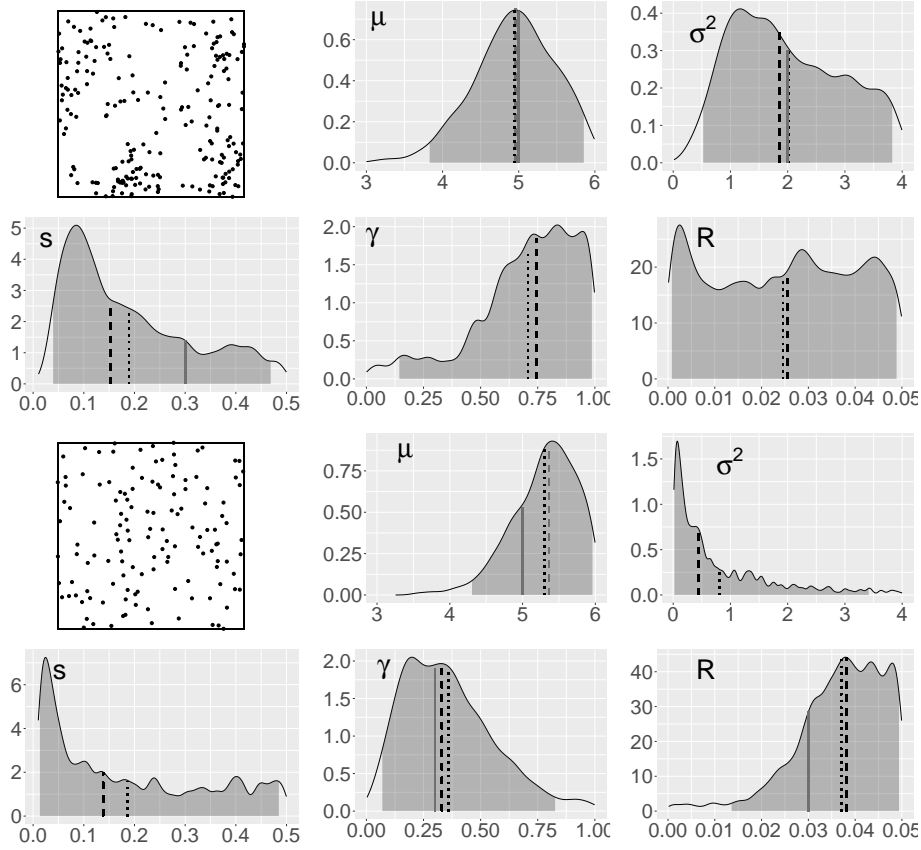
### 3.3 Model checking and comparison

We are interested in whether the point patterns in Figure 1 can be distinguished from realisations of an LGCP and a Strauss process, so for comparison we also fitted an LGCP and a Strauss process to each point pattern, using the ABC procedure in Algorithm 1. We used the same summary statistics as for the LGCP-Strauss process and the same prior distributions on the relevant parameters (that is, the parameters  $\mu$ ,  $\sigma^2$ , and  $s$  when fitting the LGCP, and the parameters  $\mu$ ,  $\gamma$ , and  $R$  when fitting the Strauss process). Again, when simulating under an LGCP, we used the faster method implemented in `spatstat`.

For model checking and comparison we suggest to make global envelope tests based on posterior predictions as follows. For each ABC realisation of  $\theta$ , a realisation  $\mathbf{x}$  of the process in question given  $\theta$  is simulated. For each  $\mathbf{x}$ , a functional summary statistic is estimated. These empirical curves are then used to construct global envelopes and corresponding tests based on extreme rank lengths (Myllymäki et al., 2017; Mrkvička et al., 2018, note that we only used 1000 simulations instead of the recommended 2499, because the ABC procedure is rather time consuming). The R-package `GET` (Myllymäki et al., 2017) was used for this purpose.

In order to compare the fitted LGCP-Strauss, LGCP, and Strauss process models, we used 95% global envelopes based on posterior predictions and the empirical  $L$ - and  $J$ -function, with  $J(r) = (1 - G(r))/(1 - F(r))$  where  $F$  is the empty space function and  $G$  is the nearest-neighbour distribution function (see van Lieshout & Baddeley, 1996). We also tried to use the  $F$ - and  $G$ -functions for model validation but these functional summary statistics were unable to distinguish between the models.

Figure 4 shows 95% combined global envelopes for the  $L$ - and  $J$ -function, meaning that, under the LGCP-Strauss process, the probability that both empirical curves are within their respective envelopes is approximately 95%. To combine the envelopes



**Figure 3:** The first six panels show a realisation of an LGCP on the unit square and the estimated marginal ABC posterior distributions for the parameters of the LGCP-Strauss process. The last six panels show a realisation of a Strauss process on the unit square and the estimated marginal ABC posterior distributions for the parameters of the LGCP-Strauss process. For the density plots, the parameters are stated at the top corner; the area between the 2.5% and 97.5% percentiles is shaded; the dashed and dotted lines indicate the median and mean, respectively and the solid line indicate the true value, when relevant.

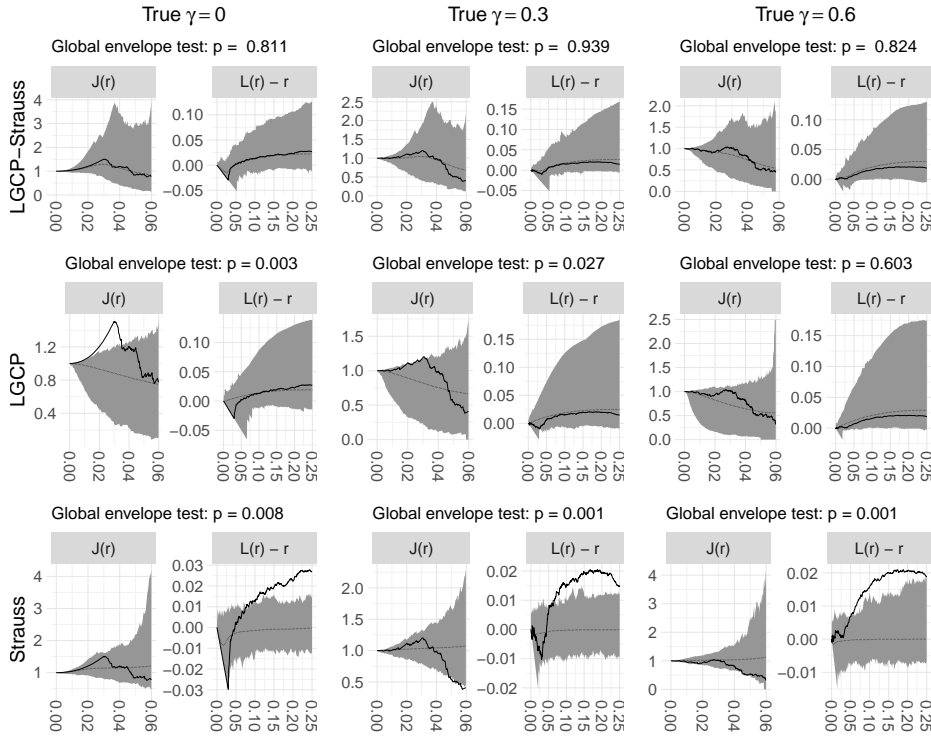
we have used the two-step combining procedure described in Myllymäki & Mrkvička (2019). Note that the  $J$ -function can only be estimated reliably for all simulations in the interval  $(0, 0.6)$ , whereas the  $L$ -function can be estimated reliably on a larger interval.

In all cases, the  $p$ -values of the global envelope tests are highest in the situation of the LGCP-Strauss process, which may indicate that they provide the best fit to data. The LGCP is rejected in the cases where  $\gamma = 0$  and  $\gamma = 0.3$  because the empirical  $J$ -functions in these cases are above the 95% global envelopes at small inter-point distances. This indicates that the point patterns are more regular at small inter-point distances than what would be expected under the fitted LGCPs. For the case  $\gamma = 0.6$  (the case with weakest inhibition), the LGCP cannot be rejected. Notice that the  $p$ -values of these tests are increasing as  $\gamma$  increases which is in agreement with the fact that the LGCP-Strauss process approaches the special case of an LGCP.

The Strauss process model is rejected in all three cases because the empirical  $L$ -function clearly shows that the point patterns are more clustered at moderate to

large inter-point distances than what can be modelled with a Strauss process. In the case  $\gamma = 0.3$ , the empirical  $J$ -function also shows this, but for the remaining two cases the  $J$ -function is contained completely within the envelopes.

Overall, it appears that the  $J$ -function is best at criticizing the LGCP and the  $L$ -function is best at criticizing the Strauss process. The latter may have something to do with the fact that the  $J$ -function can only be estimated on a relatively small interval. So, it is less likely to capture the aggregation, which happens on a larger scale, than the  $L$ -function which can be estimated on a bigger interval. When we use the  $L$ -function for model validation we keep in mind that it was also used in the ABC procedure which might lead us to conclude that the model fits better to data than it actually does.



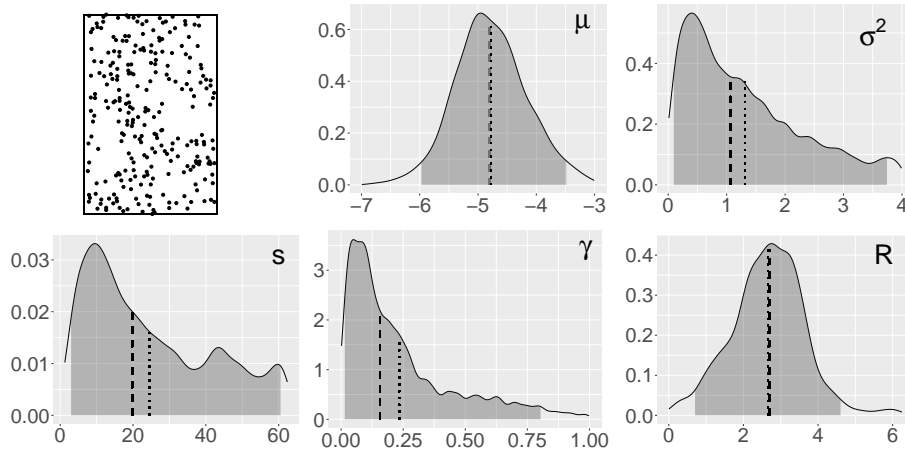
**Figure 4:** Combined global envelopes based on the empirical  $J$ - and  $L$ -function for LGCP-Strauss, LGCP, and Strauss processes fitted with ABC to the three point patterns in Figure 1. The choice of the fitted model is stated to the left of each row and each column represents a different point pattern, as indicated by the true value of  $\gamma$ . The solid curves are the empirical summary statistics for the observed point patterns and the dotted curves are the means obtained from 1000 posterior predictions. Each shaded area indicates a 95% global envelope based on the extreme rank length. At the top of each plot, the  $p$ -value of the corresponding global envelope test is stated.

In order to investigate how this model validation and comparison works when the LGCP-Strauss process is overfitting, we also fitted an LGCP to the first point pattern in Figure 3 and a Strauss process model to the second point pattern in Figure 3 and compare them to the fitted LGCP-Strauss process models (the global envelopes are not shown). For the realisation of an LGCP, the  $p$ -values of the 95% combined global envelope test for the fitted LGCP-Strauss and LGCP were 0.933 and 0.766,

respectively. Since the data is generated from an LGCP, both models should fit the data equally well, so the higher  $p$ -value for the LGCP-Strauss process is probably a result of the fact that it is overfitting. For the realisation of a Strauss process model, the  $p$ -values of the 95% combined global envelope test for the fitted LGCP-Strauss and Strauss process were 0.376 and 0.829, respectively. In this example, the  $p$ -values do not reveal the fact that the LGCP-Strauss process is overfitting.

## 4 Data example

The first panel in Figure 5 shows the locations of 256 oak trees which suffer from frost shake (frost shake refers to cracks in the trunk of the tree) in a  $125\text{ m} \times 188\text{ m}$  rectangular region of Allogny in France. This data set is part of the Allogony data set from the R-package `ads` (Pélissier & Goreaud, 2015). The data set of the oaks was analysed in Lavancier & Møller (2016) using a parametric point process model with regularity on the small scale and aggregation on the large scale. The LGCP-Strauss model should also be able to model this behaviour.

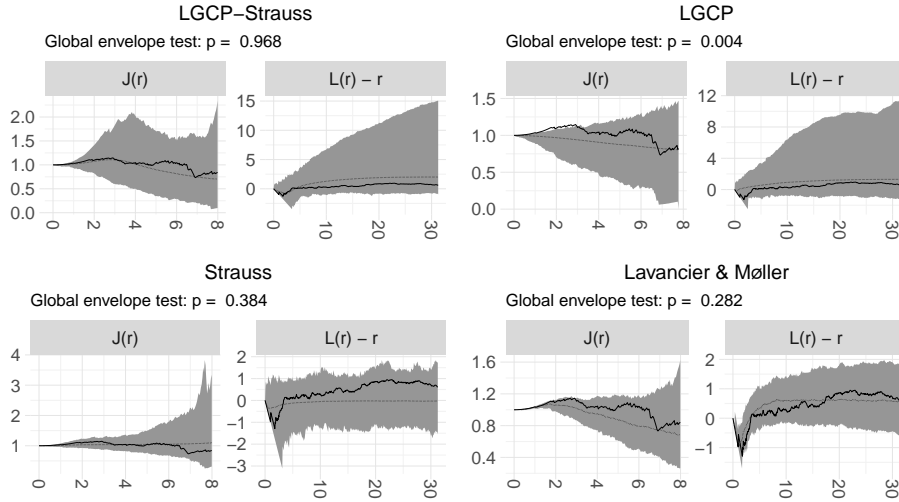


**Figure 5:** The first panel shows the frost shake oak point pattern dataset where the observation window is a  $125\text{ m} \times 188\text{ m}$  rectangle. The other panels show the estimated marginal ABC posterior distributions for the five parameters, with each parameter stated at the top corner of each panel. The area between the 2.5% and 97.5% percentiles is shaded. The dashed and dotted lines indicate the median and mean, respectively.

We used Algorithm 1 on this oak data set. Here, independent uniform prior distributions are chosen for  $\mu$  on the interval  $(-7, -3)$ ,  $\sigma^2$  on  $(0, 4)$ ,  $s$  on  $(1.25, 62.5)$ ,  $\gamma$  on  $(0, 1)$ , and  $R$  on  $(0, 6.25)$ . Notice that the observation window for the oak data is much larger than the ones in Section 3.1, and the prior distributions are chosen to take this into account. Furthermore, when calculating the summary statistics for the ABC procedure,  $W_{i,j}$ ,  $i, j = 1, \dots, q$ , are now rectangular sets of the same size (see Section 2.2). Trace plots as those in Figure 7 (supplied in an appendix) suggested that 20 000 iterations of the MCMC algorithm is a sufficient burn-in for this example. Again, a pilot sample of 10 000 simulations was used and the resulting ABC posterior sample consists of 1000 draws from the approximate posterior distribution.

The marginal posterior distributions, which are estimated from the ABC sample, can be seen in Figure 5. They are all clearly different from their uniform priors. The posterior distributions of  $\mu$  and  $R$  look approximately normal, whilst the posterior distributions of  $\sigma^2$ ,  $s$ , and  $\gamma$  are right skew. Note that the posterior distribution of  $\gamma$  indicates strong repulsion between the points. The posterior distribution of  $\sigma^2$ , particularly its heavy tail, suggests some aggregation among the splited oaks.

The first plot in Figure 6 shows 95% combined global envelopes for the fitted LGCP-Strauss process as described in Section 3. The overall behaviour of the observed point pattern seems to be captured well by the LGCP-Strauss process, and the  $p$ -value is very high. For comparison, the remaining plots in Figure 6 show the corresponding 95% envelopes for a fitted LGCP, a fitted Strauss process model, and the model fitted in Lavancier & Møller (2016), respectively. The LGCP and Strauss process models are fitted with the ABC procedure in Algorithm 1, and the envelopes are based on posterior predictions. The model in Lavancier & Møller (2016) was not fitted in a Bayesian setup, so simulations under this model are not posterior predictions. For simulating this model, we used the technique suggested in Lavancier & Møller (2016). The combined global envelopes indicate that the LGCP model provides a poor fit to data. The tests conclude that both the Strauss process model and the model fitted in Lavancier & Møller (2016) fit well. However, the  $p$ -values are lower than the corresponding  $p$ -value for the LGCP-Strauss process, indicating that the later may provide a better fit.



**Figure 6:** Combined global envelopes based on the empirical  $J$ - and  $L$ -function for different fitted models (as indicated at the top of each plot). The solid curves correspond to the splited oak point pattern and the dotted curves are the means obtained from 1000 simulations (posterior predictions for the models fitted with ABC). The shaded area indicate a 95% global envelope based on the extreme rank length. At the top of each plot, the  $p$ -value of the corresponding global envelope test is stated.

## 5 Summary and future work

We have proposed a novel spatial point process model which enables capturing of regularity through pairwise interactions and aggregation through a Gaussian process realization. This doubly stochastic spatial point process generalizes both the customary log Gaussian Cox process and the customary Gibbs process. Because the likelihood is intractable for this model we have developed model fitting through an ABC method. We have provided both simulation investigation and a real data application in order to reveal the behaviour of process realizations and also our ability to fit the model and do full inference for given point pattern realizations.

Future work can consider marked point patterns or so-called multi-type versions of our model (see e.g. Møller & Waagepetersen, 2004). Such multi-type modelling may allow attraction or inhibition within types but also introduce attraction or inhibition between types. A different direction would consider space-time versions. That is, a realization of the process is seen as a spatial point pattern by *integrating* over a window of time.

## Acknowledgements

The research of the first two authors was supported by The Danish Council for Independent Research | Natural Sciences, grant DFF – 7014-00074 ‘Statistics for point processes in space and beyond’. The second author was also supported by the ‘Centre for Stochastic Geometry and Advanced Bioimaging’, funded by grant 8721 from the Villum Foundation.

## References

- Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. Boca Raton: Chapman and Hall/CRC Press.
- Berthelsen, K. K. & Møller, J. (2008). Non-parametric Bayesian inference for inhomogeneous Markov point processes. *Australian & New Zealand Journal of Statistics*, 50, 257–272.
- Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society: Series B (Methodological)*, 17, 129–157.
- Fearnhead, P. & Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74, 419–474.
- Geyer, C. J. & Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, 21, 359–373.
- Goldstein, J., Haran, M., Simeonov, I., Fricks, J., & Chiaromonte, F. (2015). An attraction–repulsion point process model for respiratory syncytial virus infections. *Biometrics*, 71, 376–385.

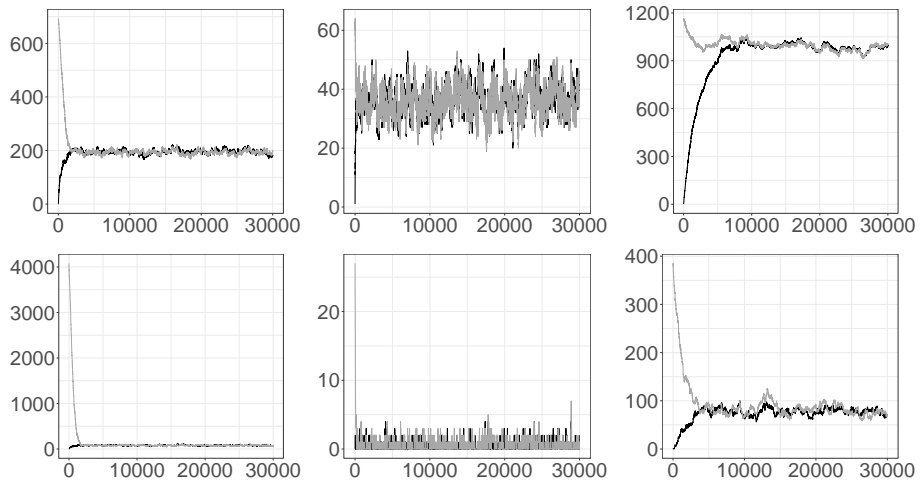


- Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Boca Raton: Chapman and Hall/CRC.
- Kelly, F. P. & Ripley, B. D. (1976). A note on Strauss’s model for clustering. *Biometrika*, 63, 357–360.
- Kendall, W. & Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*, 32, 844–865.
- Lavancier, F. & Møller, J. (2016). Modelling aggregation on the large scale and regularity on the small scale in spatial point pattern datasets. *Scandinavian Journal of Statistics*, 43, 587–609.
- Lavancier, F. & Møller, J. (2016). Modelling aggregation on the large scale and regularity on the small scale in spatial point pattern datasets. *Scandinavian Journal of Statistics*, 43, 587–609.
- Meinshausen, N. (2007). Relaxed Lasso. *Computational Statistics & Data Analysis*, 52, 374–393.
- Møller, J., Syversveen, A. R., & Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25, 451–482.
- Møller, J. & Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman and Hall/CRC.
- Møller, J. & Waagepetersen, R. P. (2017). Some recent developments in statistics for spatial point patterns. *Annual Review of Statistics and Its Application*, 4, 317–342.
- Mrkvička, T., Myllymäki, M., Jílek, M., & Hahn, U. (2018). A one-way ANOVA test for functional data with graphical interpretation. Available at arXiv:1612.03608.
- Myllymäki, M. & Mrkvička, T. (2019). GET: Global envelopes in R. arXiv preprint arXiv:1911.06583.
- Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., & Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 381–404.
- Pélissier, R. & Goreaud, F. (2015). ads package for R: A fast unbiased implementation of the  $K$ -function family for studying spatial point patterns in irregular-shaped sampling windows. *Journal of Statistical Software*, 63, 1–18.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13, 255–266.
- Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39, 172–192.

- Schlather, M. (1999). *An introduction to positive definite functions and to unconditional simulation of random fields*. Technical Report st 99-10, Department of Mathematics and Statistics, Lancaster University.
- Schlather, M., Malinowski, A., Menck, P. J., Oesting, M., & Strokorb, K. (2015). Analysis, simulation and prediction of multivariate random fields with package RandomFields. *Journal of Statistical Software*, 63, 1–25.
- Schlather, M., Malinowski, A., Oesting, M., Boecker, D., Strokorb, K., Engelke, S., Martini, J., Ballani, F., Moreva, O., Auel, J., Menck, P. J., Gross, S., Ober, U., Ribeiro, P., Ripley, B. D., Singleton, R., Pfaff, B., & R Core Team (2019). RandomFields: Simulation and analysis of random fields. R package version 3.3.6.
- Sheather, S. J. & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53, 683–690.
- Shirota, S. & Gelfand, A. E. (2017). Approximate Bayesian computation and model assessment for repulsive spatial point processes. *Journal of Computational and Graphical Statistics*, 26, 646–657.
- Soubeyrand, S., Carpentier, F., Guiton, F., & Klein, E. K. (2013). Approximate Bayesian computation with functional statistics. *Statistical Applications in Genetics and Molecular Biology*, 12, 17–37.
- Stoica, R. S., Philippe, A., Gregori, P., & Mateu, J. (2017). ABC Shadow algorithm: A tool for statistical analysis of spatial patterns. *Statistics and Computing*, 27, 1225–1238.
- Strauss, D. J. (1975). A model for clustering. *Biometrika*, 62, 467–475.
- van Lieshout, M. N. M. & Baddeley, A. J. (1996). A nonparametric measure of spatial interaction in point patterns. *Statistica Neerlandica*, 50, 344–361.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99, 250–261.

## A Trace plots for accessing the burn-in for the simulation algorithm

Figure 7 shows trace plots of the number of points and  $R$ -close pairs for the MCMC algorithm when simulating an LGCP-Strauss process for different draws of the parameter vector  $\theta$  from its prior distribution which is described in Section 3. For each prior sample of  $\theta$ , a realisation  $\mathbf{z}$  of the GRF was simulated, and the MCMC algorithm was used to simulate the LGCP-Strauss process given  $\mathbf{Z} = \mathbf{z}$ . This analysis was used to choose an appropriate burn-in in Section 3.



**Figure 7:** Trace plots of the number of points (top) and  $R$ -close pairs (bottom) for 30 000 iterations of the MCMC algorithm for simulating an LGCP-Strauss process on the unit square with parameter vector  $\theta$  drawn from the prior distribution. Each column of images represent a different sample of  $\theta$  and a corresponding realization  $\mathbf{z} = \{z(u)\}_{u \in W}$  of the GRF  $\mathbf{Z}$ . For each column, the MCMC algorithm was initiated at the empty point pattern (black line) or a realisation of an inhomogeneous Poisson process with intensity function  $\exp(z(u))$  (grey line).