



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



Jesper Møller, Mohammad Ghorbani and Ege Rubak

**Mechanistic spatio-temporal point process models for marked point processes, with a view to forest stand data**

No. 12, August 2014

# Mechanistic spatio-temporal point process models for marked point processes, with a view to forest stand data

Jesper Møller, Mohammad Ghorbani and Ege Rubak

Department of Mathematical Sciences, Aalborg University  
jm@math.aau.dk, ghorbani@math.aau.dk, rubak@math.aau.dk

## Abstract

We show how a spatial point process, where to each point there is associated a random quantitative mark, can be identified with a spatio-temporal point process specified by a conditional intensity function. For instance, the points can be tree locations, the marks can express the size of trees, and the conditional intensity function can describe the distribution of a tree (i.e., its location and size) conditionally on the larger trees. This enable us to construct parametric statistical models which are easily interpretable and where likelihood-based inference is tractable. In particular, we consider maximum likelihood based inference and tests for independence between the points and the marks.

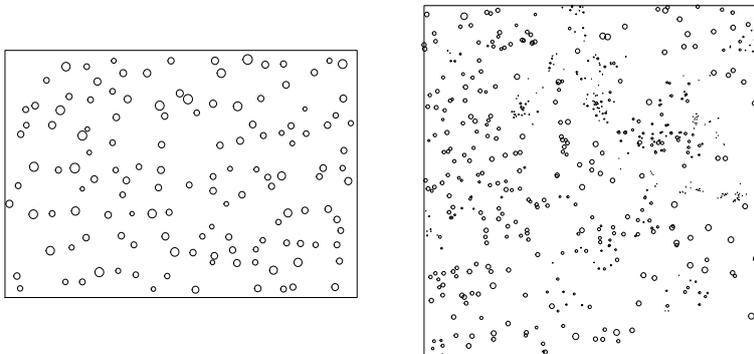
*Keywords:* conditional intensity; likelihood ratio statistic; independence between points and marks; maximum likelihood; model checking; quantitative marks.

## 1 Introduction

Marked point process (MPP) models have frequently been used for analyzing forest stand datasets which typically consist of the locations of trees in a given observation area and a list of marks associated with each tree, e.g., the height, the diameter at breast height (DBH), and the species of the tree (Pommerening, 2002). Illian et al. (2008) discussed different model classes for the marks given the points, including the random-field model (i.e., when the marks are generated by sampling a random field  $M$  at the points, where  $M$  is independent of the points) and the special case of independent marking (i.e., when in addition the marks are independent, identically distributed, and independent of the points); see also Stoyan and Walder (2000), Schlater et al. (2004), and Schoenberg (2004). As Guan and Afshartous (2007) remarked: ‘Generally speaking, however, the literature on modeling dependent marked point processes is still limited and worth further investigation.’ In fact most statistical techniques for analyzing marked point pattern datasets have mainly been based

on various non-parametric summary statistics describing the second order properties of the points and the marks (see e.g., Illian et al. (2008), Baddeley (2010), and Diggle (2013)), and often the focus has been on testing for the random-field model (Schlatter et al., 2004; Guan and Afshartous, 2007) or independent marking (see e.g., Stoyan and Stoyan (1994), Illian et al. (2008), and Myllymäki et al. (2013)).

In this paper, we focus for specificity and simplicity on point pattern tree datasets with one quantitative mark for each tree which results from the growth of the tree (our approach can easily be extended to a general MPP with several marks, including a quantitative mark which results from a dynamic process so that the mark is either an increasing or a decreasing function of time, and with covariate information (e.g., terrain elevation and soil quality) included, but we leave such extensions to be discussed in future work). For instance, the quantitative mark can express the size of the tree, whether it be the height, the DBH, or some other measure of size which grows over time. Figure 1 shows two examples. Our idea is simply to identify such a MPP with a spatio-temporal process, where the dynamics is specified by a *conditional intensity function*  $\lambda^*(t, x)$  which represents the infinitesimal expected rate of events at time  $t$  and location  $x$ , given all the observations up to time  $t$ —in the terminology of e.g., Diggle (2013), this is a *mechanistic model*. For instance, for a forest stand dataset, considering the DBH as a convenient measure of size, a large/small tree corresponds to a small/large time, and we use the conditional intensity function to model the distribution of a present tree (i.e., its location and DBH) conditionally on the larger trees. Section 2 formalizes this idea. The main advantages of our approach is that the model is typically easily interpretable and likelihood-based inference will often be tractable as demonstrated in Section 3.



**Figure 1:** Left: The centres of the discs specify the locations of 134 Norwegian spruce trees in a  $56 \times 38$  metre sampling region in Saxonia, Germany, where the radius of a disc is two times the DBH (see Section 3.1 for more information on this dataset). Right: Positions of 584 longleaf pine trees in a  $200 \times 200$  metre sampling region in southern Georgia (USA), where the radius of a disc is the DBH in 1987 minus the DBH in 1979 (see Section 3.2 for more information on this dataset).

There exist a number of parametric models (and further models can easily be developed) for temporal and spatio-temporal conditional intensity functions, including the Hawkes process (Hawkes, 1971) and its spatio-temporal extensions to epidemic

type aftershock sequence (ETAS) models (Ogata, 1988), and the self-correcting or stress-release process (Isham and Westcott, 1979) and its spatio-temporal extensions (Rathbun, 1996). These extensions have mainly been used for modelling earthquake datasets. See the reviews in Ogata (1998) and Daley and Vere-Jones (2008). Section 3 develops new models fitted to the datasets in Figure 1, using an inhibitory (self-correcting) model for the spruce dataset and a clustering (self-exciting) model for the pine dataset. In particular, we find maximum likelihood estimates, test for independence between the points and times based on the likelihood ratio statistic, and discuss various diagnostics for model checking.

## 2 Marked point processes specified by a conditional intensity

### 2.1 The general model

For specificity we refer to forest stand data: Consider a finite MPP  $\{(U_i, M_i) : i = 0, \dots, N\}$ , where  $U_i$  is the random location of the  $i$ th tree,  $M_i = M(U_i)$  is its size which for specificity we let be the DBH,  $M$  is a real-valued random field, and  $N$  (the number of tree) is a non-negative discrete random variable. We assume that the DBHs are increasingly ordered continuous random variables such that  $0 \leq M_0 < \dots < M_N \leq \tau < \infty$  (the practical problem of ties is discussed at the end of this section), where we treat  $\tau > 0$  (the maximal possible DBH) as an unknown parameter. Furthermore, the points  $U_i$  are located in a given planar sampling region  $W$  of finite area  $|W|$ .

For our purpose it is useful to relate the MPP to an infinite spatio-temporal point process  $\{(T_1, X_1), (T_2, X_2), \dots\}$  (we can view this as another MPP where actually the ‘points’ are strictly increasing times  $0 < T_1 < T_2 < \dots$  and the ‘marks’  $X_1, X_2, \dots$  are points in  $W$ ; this is one reason for writing  $(T_i, X_i)$  instead of  $(X_i, T_i)$ , though we use the common terminology ‘spatio-temporal’): We shall condition on  $(U_N, M_N)$ , the tree with the largest DBH. Then we let the times

$$T_1 = M_N - M_{N-1}, \dots, T_N = M_N - M_0,$$

be the gaps between the largest DBH and the other DBHs in decreasing order. The tree locations corresponding to the times  $T_1, \dots, T_N$  are

$$X_1 = U_{N-1}, \dots, X_N = U_0,$$

respectively. For the largest tree,  $T_0 = M_N - M_N = 0$  and  $X_0 = U_N$ . For the continuation of the spatio-temporal point process, we assume  $T_{N+1} > \tau$ . The actual model for  $(X_{N+1}, T_{N+2}, X_{N+2}, \dots)$  will not play any importance in this paper, since we shall only exploit the simple one-to-one correspondence between  $((U_0, M_0), \dots, (U_N, M_N))$  and  $((U_N, M_N), (T_1, X_1), \dots, (T_N, X_N))$ .

As generic notation for realizations, we use small letters and write e.g.,  $(U_i, M_i) = (u_i, m_i)$  and  $(T_i, X_i) = (t_i, x_i)$ . It will always be understood without mentioning that if we consider a realization  $((U_0, M_0), \dots, (U_N, M_N)) = ((u_0, m_0), \dots, (u_n, m_n))$ , then

$(u_0, \dots, u_n) \in W^{n+1}$ ,  $0 \leq m_0 < \dots < m_n \leq \tau$ , and the corresponding realization of the spatio-temporal point process up to time  $\tau$  is  $((T_0, X_0), \dots, (T_N, X_N)) = ((t_0, x_0), \dots, (t_n, x_n))$  with  $(t_0, \dots, t_n) = (m_n - m_n, \dots, m_n - m_0)$  and  $(x_0, \dots, x_n) = (u_n, \dots, u_0)$ . Note that having  $(U_N, M_N) = (u_n, m_n)$  is not implying that we know that  $N = n$ . However, considering a realization  $((U_0, M_0), \dots, (U_N, M_N)) = ((u_0, m_0), \dots, (u_n, m_n))$ , this is of course specifying that  $N = n$ .

Now, conditional on  $(U_N, M_N)$  we assume that the spatio-temporal point process is specified by a conditional intensity function

$$\lambda^*(t, x) = \lambda^*(t, x | \mathcal{F}_t), \quad t > 0, x \in W,$$

where the star indicates that  $\lambda^*(t, x)$  depends on the history  $\mathcal{F}_t$  at time  $t$ , i.e., trees ‘appearing’ before time  $t$ ; formally,  $\mathcal{F}_t$  is the  $\sigma$ -algebra generated by  $M_N$  and those  $(T_i, X_i)$  with  $T_i < t$  (for technical details, see Daley and Vere-Jones (2003)). Heuristically, denoting  $N(A)$  the number of  $(T_i, X_i)$  falling in a set  $A \subset (0, \infty) \times W$ ,

$$\lambda^*(t, x) dt dx = E(N(dt \times dx) | \mathcal{F}_t).$$

The essential assumption for this type of model to be reasonable for forest stand data is that the distribution of a present tree and its DBH is determined by conditioning on those trees which are bigger (i.e., conditioning on their locations and corresponding DBHs). More formally, for  $i = 1, 2, \dots$ , conditional on that  $\mathcal{F}_t$  is specified by a realization  $(M_N, (T_0, X_0), \dots, (T_{i-1}, X_{i-1})) = (m_n, (t_0, x_0), \dots, (t_{i-1}, x_{i-1}))$  with  $t_0 = 0$ , the density function for  $(T_i, X_i)$  is

$$\begin{aligned} & p(t_i, x_i | m_n, (t_0, x_0), \dots, (t_{i-1}, x_{i-1})) \\ &= \lambda^*(t_i, x_i) \exp\left(-\iint_{(t_{i-1}, t_i) \times W} \lambda^*(t, x) dt dx\right), \quad t_i > t_{i-1}, x_i \in W. \end{aligned} \quad (2.1)$$

To ensure that (2.1) is indeed a density function, we require that  $\lambda^*(t, x)$  is a non-negative measurable function such that the integral in (2.1) is finite and the integral

$$\iint_{(t_{i-1}, \infty) \times W} \lambda^*(t, x) dt dx$$

is infinite when the history  $\mathcal{F}_t$  is unchanged for  $t \geq t_{i-1}$ . In addition, to ensure that  $N = \sup\{i \geq 0 : T_i \leq \tau\}$  is finite, we assume that the time process  $(T_0, T_1, \dots)$  is not explosive on the time interval  $[0, \tau]$ .

From formula (2.1) we obtain the joint density for a realization  $((U_0, M_0), \dots, (U_{N-1}, M_{N-1})) = ((u_0, m_0), \dots, (u_{n-1}, m_{n-1}))$  conditional on  $(U_N, M_N) = (u_n, m_n)$  (if  $n = 0$ , we interpret  $((u_0, m_0), \dots, (u_{n-1}, m_{n-1}))$  as  $\emptyset$ , the empty marked point configuration):

$$\begin{aligned} & p((u_0, m_0), \dots, (u_{n-1}, m_{n-1}) | (u_n, m_n)) \\ &= \exp\left(-\iint_{(0, \tau) \times W} \lambda^*(t, x) dt dx\right) \prod_{i=1}^n \lambda^*(t_i, x_i) \end{aligned} \quad (2.2)$$

where the dominating measure is  $\nu_\tau = \sum_{n=0}^{\infty} \mu_n$ , with  $\mu_n$  being  $n$ -fold Lebesgue measure on  $W \times [0, \tau]$  (where  $\mu_0$  is the Dirac measure on  $\emptyset$ ). In other words,  $\exp(-\tau|W|)\nu_\tau$  is the distribution for a unit rate Poisson process on  $W \times [0, \tau]$  when the DBHs have been ordered (see e.g., Møller and Waagepetersen (2004)).

## 2.2 Independence

Recall that the DBH  $M_i = M(U_i)$  is the value of the random field  $M$  at the tree location  $U_i$ , cf. the beginning of Section 2.1. In the so-called random-field model (Takahata, 1994; Mase, 1996),  $M$  is assumed to be independent of the spatial point process for the tree locations. This simplifies the statistical analysis greatly, since the tree locations and the DBHs can be investigated separately by using standard techniques for spatial point processes and for geostatistical data (Schlatter et al., 2004). Formal tests for the hypothesis that a given dataset is generated by the random-field model are discussed in Schlatter et al. (2004) and in Guan and Afshartous (2007). Note that the hypothesis does not warrant an independence among the DBHs nor among the tree locations. The special case of the random-field model where the DBHs are independent, identically distributed, and independent of the tree locations is called the independently marked point process model. A test of this more restrictive hypothesis is described in Stoyan and Stoyan (1994); see also Illian et al. (2008) and Myllymäki et al. (2013). We remark that all tests mentioned above are based on non-parametric summary statistics and stationarity and isotropy conditions are imposed; in addition Schlatter et al. (2004) required the marginal distribution of the marks to be normal.

In our dynamic setting it is natural to consider the null hypothesis of conditional independence between the times  $(T_1, \dots, T_N)$  and the points  $(X_1, \dots, X_N)$  given the information  $(U_N, M_N)$  about the largest tree, i.e., the gaps of DBHs  $(M_N - M_{N-1}, \dots, M_N - M_0)$  are independent of the tree locations  $(U_{N-1}, \dots, U_0)$  given  $(U_N, M_N)$ . This null hypothesis of independence is equivalent to  $\lambda^*(t, x)$  being of the form

$$\lambda^*(t, x) = \lambda^*(t)h_t^*(x), \quad t > 0, \quad x \in W, \quad (2.3)$$

where

- (i)  $\lambda^*(t) = \lambda^*(t \mid \mathcal{G}_t)$  is a conditional intensity for the temporal point process  $T_1, T_2, \dots$  which may depend on its history  $\mathcal{G}_t$  before time  $t$  (recall that the temporal point process is required to be non-explosive),
- (ii)  $h_t^*(x) = h^*(x \mid m_n, (x_i; t_i < t))$  is a density function on  $W$  which only depends on  $m_n$  and the ordered set  $(x_i; t_i < t)$  of points appearing before time  $t$  (with the ordering given by the times; notice that  $x_0$  is included in  $(x_i; t_i < t)$ ).

When we later discuss simulation and model checking, we use the temporal integrated intensity

$$\Lambda^*(t) = \int_0^t \lambda^*(s) \, ds, \quad t > 0,$$

and the fact that  $S_i = \Lambda^*(T_i)$ ,  $i = 1, 2, \dots$ , form a unit rate Poisson process on  $(0, \infty)$ . Note that (ii) implies that conditional on  $(U_N, M_N) = (x_0, m_n)$ ,  $N = n$ , and the ordering of the points, the density for the ordered point process  $(X_1, \dots, X_n)$  with respect to Lebesgue measure on  $W^n$  is the so-called Janossy density

$$p(x_1, \dots, x_n \mid x_0, m_n, n) = \exp(-|W|) \prod_{i=1}^n h_i^*(x_i) \quad (2.4)$$

where, with a slight abuse of notation,  $h_i^*(x_i) = h_{t_i}^*(x_i)$ . Note also that Schoenberg (2004) considered a more restrictive hypothesis than (2.3), requiring that the tree locations are independent, and he developed a non-parametric test for this hypothesis.

In Section 2.3 below we discuss maximum likelihood inference when a parametric model for  $\lambda^*(t, x)$  has been specified. Then, if the null hypothesis given by (2.3) is a submodel, we propose to test the null hypothesis by a likelihood ratio test as exemplified in Section 3. Based on general experience, we expect the likelihood ratio test to be more powerful than the non-parametric tests discussed above, but we leave an investigation of this for future work.

## 2.3 Likelihoods

Suppose a realization  $((U_0, M_0), \dots, (U_N, M_N)) = ((u_0, m_0), \dots, (u_n, m_n))$  has been observed. It follows from (2.2) that the largest gap between the DBHs,  $\hat{\tau} = t_n = m_n - m_0$ , is the maximum likelihood estimate (MLE) for  $\tau$ . If data has been collected with a known lower bound  $m_{\min} \geq 0$  on the DBH of trees, it may be more reasonable to use the estimate  $m_n - m_{\min}$  for  $\tau$ .

Suppose also that a parametric model  $\lambda_\theta^*(t, x)$  for the conditional intensity has been specified in terms of an unknown parameter  $\theta$  which is variation independent of  $\tau$ . Throughout this paper, we refer to this as the *full model*. By (2.2) and using the MLE  $\hat{\tau} = t_n$ , the spatio-temporal log-likelihood conditional on  $(U_N, M_N) = (u_n, m_n)$  is

$$L(\theta) = \sum_{i=1}^n \log \lambda_\theta^*(t_i, x_i) - \iint_{(0, t_n) \times W} \lambda_\theta^*(t, x) dt dx. \quad (2.5)$$

The situation simplifies if we assume independence as in (2.3) so that

$$\lambda_\theta^*(t, x) = \lambda_{\theta_1}^*(t) h_{\theta_2, t}^*(x), \quad t > 0, \quad x \in W, \quad (2.6)$$

with  $\theta = (\theta_1, \theta_2)$  where  $\theta_1$  and  $\theta_2$  are assumed to be variation independent. Throughout this paper, we refer to this as the *reduced model of independence*. Under this model,

$$L(\theta) = L(\theta_1) + L(\theta_2)$$

where we can separately treat the temporal log-likelihood

$$L(\theta_1) = \sum_{i=1}^n \log \lambda_{\theta_1}^*(t_i) - \Lambda_{\theta_1}^*(\hat{\tau}) \quad (2.7)$$

and the spatial log-likelihood for the spatial point process

$$L(\theta_2) = \sum_{i=1}^n \log h_{\theta_2, i}^*(x_i) \quad (2.8)$$

(omitting the constant  $\log \exp(|W|) = |W|$  from (2.4)). When we specify

$$h_{\theta_2, i}^*(x_i) = \tilde{h}_{\theta_2, i}^*(x_i) / c_{\theta_2, i}^*$$

by an unnormalized density  $\tilde{h}_{\theta_2,i}^*(x_i) = \tilde{h}_{\theta_2}^*(x_i | m_n, (x_0, \dots, x_{i-1}))$  with normalizing constant  $c_{\theta_2,i}^* = c_{\theta_2,i}^*(m_n, (x_0, \dots, x_{i-1}))$ , then (2.8) becomes

$$L(\theta_2) = \sum_{i=1}^n [\log \tilde{h}_{\theta_2,i}^*(x_i) - \log c_{\theta_2,i}^*]. \quad (2.9)$$

In practice a problem arises when ties occur in DBHs due to discretization in the data. We follow Diggle et al. (2010) in jittering tied DBHs.

### 3 Examples

For the models and data considered in this section, we let  $(u_n, m_n)$  and  $(t_i, x_i)$ ,  $i = 1, \dots, n$ , denote the observed events. Further,  $W$  refers to one of the sampling regions in Figure 1.

When determining the MLE of  $\theta$ , we used the `NLOpt` library (Johnson, 2010) as implemented in the R package `nloptr`. First, the DIRECT-L method (Gablonsky and Kelley, 2001) was used to obtain a global optimum  $\bar{\theta}$ . (As a check we also used the `DEoptim` function of the R package `RcppDE` which gave similar results.) Afterwards, to polish the optimum to a greater accuracy, we used  $\bar{\theta}$  as a starting point for the local optimization ‘bound-constrained by quadratic approximation’ (BOBYQA) algorithm (Powell, 2009) and obtained a final estimate  $\hat{\theta}$ .

Considering the reduced model of independence, the temporal integrated intensity appearing in the temporal log-likelihood (2.7) will be expressible on closed form, while the normalizing constants in the spatial log-likelihood (2.9) have to be approximated by numerical methods. Also the three-dimensional integral in the spatio-temporal log-likelihood (2.5) have to be approximated by numerical methods. For these approximations we used the Cuhre method of the `Cuba` library (Hahn, 2005) as implemented in the `cuhre` function of the R package `R2Cuba` specifying its arguments such that the absolute errors are small and in the case of the spatio-temporal log-likelihood approximately equal under the full and reduced models (this becomes important when we later consider likelihood ratios).

For plotting the datasets and results we have used the R package `spatstat` (Baddeley and Turner, 2005).

#### 3.1 Norwegian spruces

The spruce dataset in Figure 1 (left panel) was first analyzed in Fiksel (1984, 1988) by fitting parametric models for unmarked Gibbs point processes (see also Stoyan et al. (1995), Møller and Waagepetersen (2004), and Illian et al. (2008)). Penttinen et al. (1992), Illian et al. (2008), and Grabarnik et al. (2011) accepted the hypothesis of independent marking using tests based on non-parametric summary statistics, and Penttinen et al. (1992) constructed a MPP model under this hypothesis. Goulard et al. (1996) and Møller and Waagepetersen (2004) fitted parametric Gibbs MPP models where the points and marks are dependent. We propose instead a parametric model for the conditional intensity whereby the reduced model versus the full model can be tested using the likelihood ratio statistic.

We consider the following structure:

$$\lambda_{\theta}^*(t, x) = \lambda_{\theta_1}^*(t) h_{\theta_2, t}^*(x) g_{\theta_3}^*(t, x) \quad (3.1)$$

where  $\lambda_{\theta_1}^*(t)$  and  $h_{\theta_2, t}^*(x)$  are as in (i)-(ii) in Section 2.2,  $g_{\theta_3}^*(t, x) = g_{\theta_3}^*(t, x | \mathcal{F}_t)$  may depend on  $\mathcal{F}_t$ , and  $\theta = (\theta_1, \theta_2, \theta_3)$ . Specifically, we assume the self-correcting model (Isham and Westcott, 1979)

$$\lambda_{\theta_1}^*(t) = \exp(\alpha_1 + \beta_1 t - \gamma_1 N(t))$$

with  $\theta_1 = (\alpha_1, \beta_1, \gamma_1) \in \mathbb{R} \times [0, \infty) \times [0, \infty)$  and  $N(t) = \#\{i \geq 0 : t_i < t\}$  being the number of trees just before time  $t$ ; for each integer  $i > 0$ , the density  $h_{\theta_2, t_i}^*(x_i) = h_{\theta_2, i}^*(x_i)$  is given by

$$h_{\theta_2, i}^*(x_i) = \frac{1}{c_{\theta_2, i}^*} \prod_{j: j < i} \phi_{\theta_2}(\|x_i - x_j\|)$$

which is inspired by Figure 9.2 in Møller and Waagepetersen (2004), where  $\theta_2 = (\alpha_2, \beta_2) \in [0, \infty) \times [0, \infty)$ ,

$$\phi_{\theta_2}(r) = \mathbf{1}[r \leq \alpha_2] (r/\alpha_2)^{\beta_2} + \mathbf{1}[r > \alpha_2], \quad r \geq 0, \quad (3.2)$$

$$c_{\theta_2, i}^* = \int_W \prod_{j: j < i} \phi_{\theta_2}(\|x - x_j\|) dx$$

is the normalizing constant, and where  $\mathbf{1}[\cdot]$  denotes the indicator function (the function (3.2) is similar to a pairwise-interaction function in Diggle and Gratton (1984) with zero hardcore); and

$$g_{\theta_3}^*(t, x) = \exp\left(-\alpha_3 \sum_{t_i < t} \mathbf{1}[\|x - x_i\| \leq \beta_3, t - t_i \geq \gamma_3]\right)$$

where  $\theta_3 = (\alpha_3, \beta_3, \gamma_3) \in [0, \infty)^3$  with  $\beta_3 = \gamma_3 = 0$  if  $\alpha_3 = 0$ . This spatio-temporal point process is well-defined and finite on  $[0, s] \times W$  for every  $s \in (0, \infty)$ , since for any  $t \in [0, s]$ ,  $\lambda_{\theta_1}^*(t) \leq \rho(s)$  where  $\rho(s) = \exp(\alpha_1 + \beta_1 s)$  is a constant, and since  $h_{\theta_2, t}^*$  is a density and  $g_{\theta_3}^* \leq 1$ .

The parameters have the following interpretation. Clearly,  $\alpha_1$  and  $\beta_1$  specify a log-linear and increasing trend in time, however, if  $\beta_1 > 0$  and  $\gamma_1 > 0$ , then a large difference between  $N_t$  and the target  $\beta_1/\gamma_1$  implies that  $\lambda^*(t)$  compensates to force this difference back towards zero. The density  $h_{\theta_2, t}^*(x)$  specifies around each larger tree with location  $x_i$  (i.e.,  $t_i < t$ ) an inhibitive circular region  $D(x_i, \theta_2)$ ; this inhibitive interaction is weakened linearly as the distance  $\|x - x_i\|$  grows; and there is no interaction if  $\|x - x_i\| \geq \alpha_2$ . The reduced model of independence is the case  $\alpha_3 = 0$ . If  $\alpha_3 > 0$ , then the  $g_{\theta_3}^*(t, x)$ -term specifies around every  $x_i$  with  $t - t_i \geq \gamma_3$  (i.e., the DBH of the tree located at  $x_i$  has to be at least  $\gamma_3$  units larger) a spatio-temporal source of inhibition given by a circular influence zone  $D(x_i, \beta_3)$ .

The spatio-temporal point process can be simulated on  $[0, \hat{\tau}] \times W$  as follows. Notice that under the self-correcting model,

$$S_i = \Lambda_{\theta_1}^*(T_i) = \frac{\exp(\alpha_1)}{\beta_1} \sum_{j=1}^i \exp(-\gamma_1 j) [\exp(\beta_1 T_j) - \exp(\beta_1 T_{j-1})], \quad i = 1, 2, \dots, \quad (3.3)$$

form a unit rate Poisson process on  $(0, \infty)$ , cf. Section 2.2. Conversely,

$$T_1 = \frac{1}{\beta_1} \log \{1 + \beta_1 \exp(\gamma_1 - \alpha_1) S_1\}$$

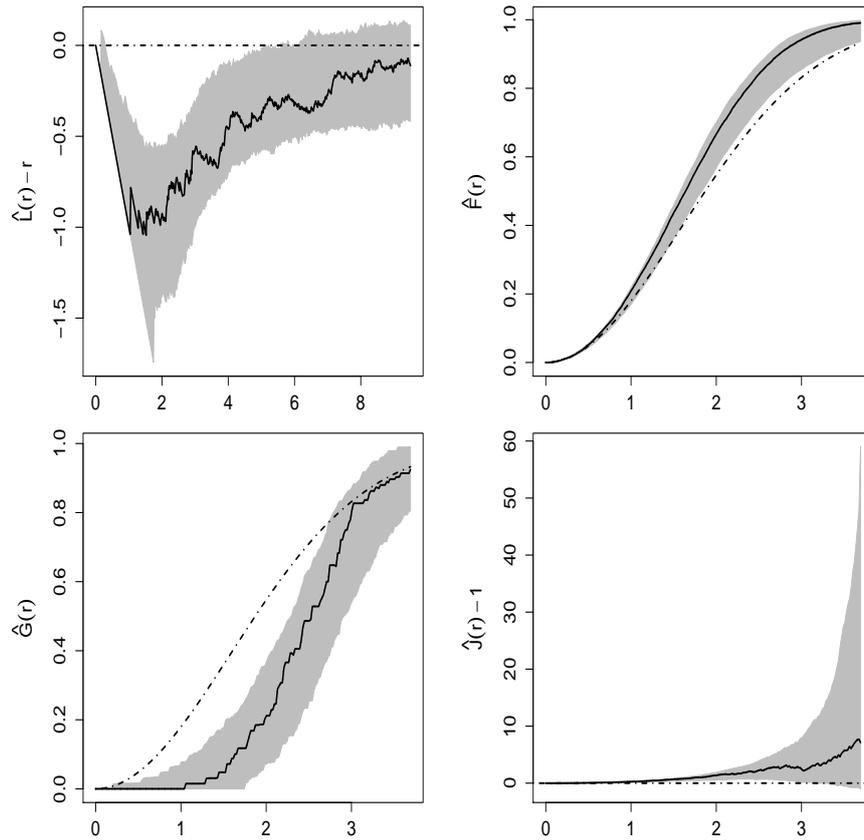
and

$$T_i = \frac{1}{\beta_1} \log \left\{ \exp(\beta_1 T_{i-1}) + \beta_1 \exp(\gamma_1 i - \alpha_1) S_i - \sum_{j=1}^{i-1} \exp(\gamma_1(i-j)) [\exp(\beta_1 T_j) - \exp(\beta_1 T_{j-1})] \right\}, \quad i = 2, 3, \dots$$

Thereby we easily obtain a simulated realization  $t_1 < \dots < t_n$ , say, under the self-correcting model restricted to  $[0, \hat{\tau}]$ . Further, assuming for the moment that  $\alpha_3 = 0$ , for each  $i = 1, \dots, n$ , we generate a point  $x_i$  from the density  $h_{\hat{\theta}_2, i}^*(x_i)$ ; here, we simply use rejection sampling, with a uniform proposal distribution on  $W$ . Finally, if in fact  $\alpha_3 > 0$ , we make a thinning in accordance to the ordering in time so that each  $(t_i, x_i)$  is kept with probability  $g_{\hat{\theta}_3}^*(t_i, x_i)$ , where the history is given by  $(U_N, M_N) = (u_n, m_n)$  and the kept events  $(t_j, x_j)$ ,  $j < i$ , by this thinning procedure; the kept events then form a simulation of the spatio-temporal point process.

For the reduced model of independence (the case  $\alpha_3 = 0$ ), since (3.3) specifies  $\Lambda_{\hat{\theta}_1}^*(\hat{\tau}) = \Lambda_{\hat{\theta}_1}^*(t_n)$ , the temporal log-likelihood (2.7) can easily be calculated. The MLE under the full model is given by  $\hat{\alpha}_1 = 5.52$ ,  $\hat{\beta}_1 = 21.72$ ,  $\hat{\gamma}_1 = 0.02$ ,  $\hat{\alpha}_2 = 2.17$ ,  $\hat{\beta}_2 = 3.11$ ,  $\hat{\alpha}_3 = 0.37$ ,  $\hat{\beta}_3 = 2.81$ , and  $\hat{\gamma}_3 = 0.05$ . A rather similar MLE under the reduced model of independence is given by  $\hat{\alpha}_1 = 5.40$ ,  $\hat{\beta}_1 = 20.01$ ,  $\hat{\gamma}_1 = 0.02$ ,  $\hat{\alpha}_2 = 2.86$ , and  $\hat{\beta}_2 = 2.25$ .

For the likelihood ratio statistic  $Q$  for testing the null hypothesis  $\alpha_3 = 0$  against the alternative hypothesis  $\alpha_3 > 0$ , the value of  $-2 \log Q$  compared with a  $\chi^2$ -distribution with  $8 - 5 = 3$  degrees of freedom provides a  $p$ -value of about 79%. Recall that if  $\alpha_3 = 0$ , then  $S_i, i = 1, 2, \dots$ , given by (3.3) form a unit rate Poisson process on  $[0, \infty)$ . Thus for further testing the null hypothesis we also consider the one-sample Kolmogorov-Smirnov test for  $\Lambda_{\hat{\theta}_1}^*(t_i) - \Lambda_{\hat{\theta}_1}^*(t_{i-1})$ ,  $i = 1, \dots, n$ , being a sample from a unit rate exponential distribution; the  $p$ -value is about 9%. Furthermore, Figure 2 shows non-parametric estimates of four functional summary statistics (the  $L$ ,  $F$ ,  $G$ , and  $J$ -functions, see e.g., Møller and Waagepetersen (2004)) for the spruces locations together with so-called 95% simultaneous rank envelopes (Myllymäki et al., 2013) obtained by 2499 simulations under the fitted reduced model of independence, so that the estimated probability for one of the curves of the non-parametric estimates is outside the corresponding envelope is 5% if  $\alpha_3 = 0$ . The 95% envelopes cover the non-parametric estimates and the deviation from the theoretical curve for a homogeneous Poisson process indicates inhibition between the tree locations. Finally, the estimated  $p$ -value for the reduced model of independence using the combined global rank envelope test (Myllymäki et al., 2013) is between 64.6% and 65.6%. In conclusion, the spruces dataset is reasonable well described by the reduced model of independence.



**Figure 2:** Non-parametric estimates of functional summary statistics for the spruces locations (solid lines) together with 95% simultaneous rank envelopes (shaded areas) calculated from 2499 simulations of the fitted reduced model of independence. For comparison the theoretical curves for a homogeneous Poisson process are shown (dot-dashed lines). Top left:  $(L(r) - r)$ -function. Top right:  $F$ -function. Bottom left:  $G$ -function. Bottom right:  $(J(r) - 1)$ -function.

### 3.2 Longleaf pines

The pines dataset in Figure 1 (right panel) were collected and analyzed by Platt et al. (1988); see also Rathbun and Cressie (1994) for a detailed description of the data. Rathbun and Cressie (1994) considered a larger dataset, including information about annual mortality and ‘disturbance paths’, and they divided the trees into various time and size groups, which were analyzed individually using different types of spatial models. Cressie (1993), Stoyan and Stoyan (1996), Mecke and Stoyan (2005), Tanaka et al. (2008), and Ghorbani (2012) fitted different Neyman-Scott point process models for the pine locations. To the best of our knowledge, a parametric MPP model for the pines dataset in Figure 1 has so far not been suggested and analyzed.

We consider a kind of marked Hawkes process where (a) trees ‘live, grow, and produce offspring in a random fashion’ and (b) ‘a large tree is likely to have a greater influence on the growth of a small tree than a small tree has on a large tree’ (the quotations in (a)-(b) are from Platt et al. (1988) and (b) may explain the observation in Chiu et al. (2013) that trees close together tend to have smaller diameters than the typical tree): Let

$$\lambda_{\theta}^*(t, x) = \mu + \sum_{t_i < t} \alpha q_{\gamma}(t | t_i) q_{\sigma}(x | x_i) \exp(-\beta(t - t_i) / \|x - x_i\|) \quad (3.4)$$

where  $\theta = (\mu, \alpha, \gamma, \sigma, \beta)$  with  $\mu \geq 0$ ,  $0 \leq \alpha < 1$ ,  $(\sigma, \gamma) \in (0, \infty)^2$ ,  $\beta \geq 0$ ,  $q_{\gamma}(t | t_i) = \mathbf{1}[t - t_i \leq \gamma] / \gamma$  is the uniform density on  $(t_i, t_i + \gamma)$ , and  $q_{\sigma}(x | x_i)$  is the bivariate Cauchy density function with scale parameter  $\sigma$  and restricted to  $W$ . Since  $W$  is rectangular, the normalizing constant of this truncated bivariate Cauchy distribution is expressible on closed form (Nadarajah and Kotz, 2007).

This spatio-temporal point process can be interpreted as an immigrant-offspring process, whereby it can easily be simulated, since

- the immigrants form a Poisson process on  $(0, \infty) \times W$  with constant intensity  $\mu$  (we also consider  $(T_0, X_0) = (0, x_0)$  as an immigrant);
- each immigrant or offspring  $(T_i, X_i)$  generates a Poisson processes on  $(0, \infty) \times W$ , where the intensity function associated to  $(T_i, X_i) = (t_i, x_i)$  is given by the term after the sum in (3.4) for  $(t, x) \in (t_i, \infty) \times W$  and it is zero otherwise—for simulation of this Poisson process, we first simulate a Poisson process where the number of events is Poisson distributed with parameter  $\alpha$ , the times are i.i.d. with density  $q_{\gamma}(t | t_i)$ , the locations are i.i.d. with density  $q_{\sigma}(x | x_i)$ , and the times and locations are independent, and second we make an independent thinning where the retention probability is given by the exponential term in (3.4);
- thus there is a cluster associated to each immigrant  $(T_i, X_i)$ , where the cluster is given by  $(T_i, X_i)$  and its first, second,  $\dots$  generation offspring processes;
- given the immigrants, these clusters are independent;

and hence the temporal process can be viewed as a branching process which is seen to be non-explosive.

Suppose  $\beta = 0$ . This is the reduced model of independence and the temporal process is a Hawkes process with conditional intensity

$$\lambda_{(\mu, \alpha, \gamma)}^*(t) = \mu|W| + \alpha \sum_{t_i < t} q_\gamma(t | t_i)$$

and integrated intensity

$$\Lambda_{(\mu, \alpha, \gamma)}^*(t) = \mu|W|t + \alpha N(t - \gamma) + \alpha \sum_{i: t - \gamma < t_i < t} (t - t_i) / \gamma$$

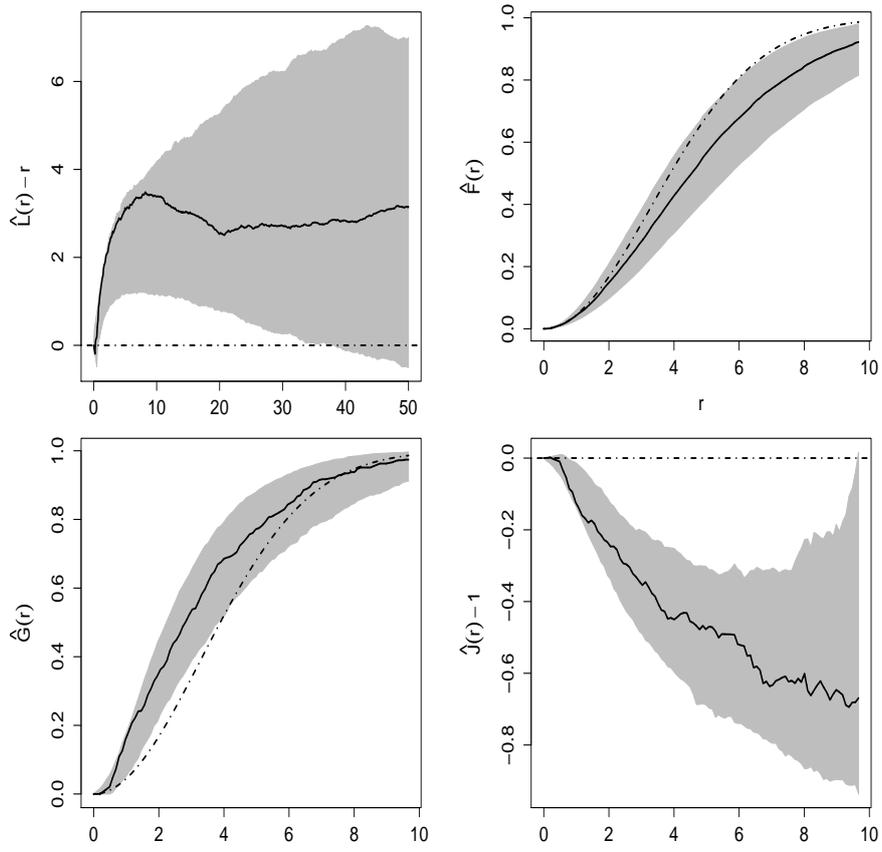
where we set  $N(t) = 0$  whenever  $t < 0$ . Thus the temporal log-likelihood (2.7) is easily handled. Note that the mean number of points in each cluster is  $1/(1 - \alpha)$  (see e.g., Section 2.2 in Møller and Rasmussen (2005)), and so by ignoring edge effects, the estimated expected number of points is

$$\hat{\mu}|W|\hat{\tau}/(1 - \hat{\alpha}) \tag{3.5}$$

when using our parameter estimates given below.

The MLE under the full model is given by  $\hat{\mu} = 4.950 \times 10^{-5}$ ,  $\hat{\alpha} = 0.999$ ,  $\hat{\gamma} = 5.051$ ,  $\hat{\sigma} = 3.669$ , and  $\hat{\beta} = 0.375$ . Under the reduced model of independence, the MLE is given by  $\hat{\mu} = 4.601 \times 10^{-5}$ ,  $\hat{\alpha} = 0.953$ ,  $\hat{\gamma} = 5.078$ ,  $\hat{\sigma} = 3.984$ , and hence using (3.5),  $\hat{\mu}|W|\hat{\tau}/(1 - \hat{\alpha}) = 2893.50$  is providing an unreasonable high estimate for the expected number of longleaf pines. Indeed, considering the likelihood ratio statistic  $Q$  for the null hypothesis  $\beta = 0$  versus the alternative hypothesis  $\beta > 0$ , the value of  $-2 \log Q$  evaluated in a  $\chi^2$ -distribution with one degree of freedom gives a highly significant  $p$ -value of  $6 \times 10^{-4}$ . Also a one-sample Kolmogorov-Smirnov test for the null hypothesis based on the times (similar to the one considered in Section 3.1) is providing a highly significant  $p$ -value of  $2.2 \times 10^{-16}$ .

Performing for the fitted full model a one-sample Kolmogorov-Smirnov test based on the times, where the temporal integrated intensity has to be calculated by numerical methods, the  $p$ -value is about 60%. Figure 3 is similar to Figure 2 but for the longleaf pines and the fitted full model. The figure indicates a reasonable fit and a more clustered behaviour than expected under a homogeneous Poisson process model. The estimated  $p$ -value for the fitted full model using the combined global rank envelope test (Myllymäki et al., 2013) is between 15.9% and 17.6%. In conclusion, although  $\hat{\alpha}$  is close to the boundary of the parameter space, the longleaf pines dataset is reasonable well described by the full model while the model of independence should clearly be rejected.



**Figure 3:** As Figure 2 but for the longleaf pine trees and the fitted full model.

## Acknowledgements

Supported by the Danish Council for Independent Research | Natural Sciences, grant 12-124675, “Mathematical and Statistical Analysis of Spatial Data”, and by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation. We thank Jakob G. Rasmussen and Alaviyeh Sajjadi for helpful discussion.

## References

- Baddeley, A. (2010). Multivariate and marked point processes, *in* A. E. Gelfand, P. J. Diggle, P. Guttorp and M. Fuentes (eds), *Handbook of Spatial Statistics*, CRC Press, Boca Raton, pp. 299–337.
- Baddeley, A. and Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns, *Journal of Statistical Software* **12**: 1–42.
- Chiu, S. N., Stoyan, D., Kendall, W. S. and Mecke, J. (2013). *Stochastic Geometry and Its Applications*, third edn, John Wiley & Sons.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*, second edn, Wiley, New York.
- Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*, second edn, Springer-Verlag, New York.
- Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure*, second edn, Springer-Verlag, New York.
- Diggle, P. J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, Chapman & Hall/CRC, Boca Raton.
- Diggle, P. J. and Gratton, R. J. (1984). Monte Carlo methods of inference for implicit statistical models (with discussion), *Journal of the Royal Statistical Society Series B* **46**: 193–227.
- Diggle, P. J., Kaimi, I. and Abellana, R. (2010). Partial-likelihood analysis of spatio-temporal point-process data, *Biometrics* **66**: 347–354.
- Fiksel, T. (1984). Estimation of parameterized pair potentials of marked and non-marked Gibbsian point processes, *Elektronische Informationsverarbeitung und Kybernetik* **20**: 270–278.
- Fiksel, T. (1988). Estimation of interaction potentials of Gibbsian point processes, *Statistics* **19**: 77–86.
- Gablonsky, J. and Kelley, C. (2001). A locally-biased form of the DIRECT algorithm, *Journal of Global Optimization* **21**: 27–37.
- Ghorbani, M. (2012). Cauchy cluster process, *Metrika* **76**: 697–706.

- Goulard, M., Särkkä, A. and Grabarnik, P. (1996). Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method, *Scandinavian Journal of Statistics* **23**: 365–379.
- Grabarnik, P., Myllymäki, M. and Stoyan, D. (2011). Correct testing of mark independence for marked point patterns, *Ecological Modelling* **222**: 3888–3894.
- Guan, Y. and Afshartous, D. R. (2007). Test for independence between marks and points of marked point processes: a subsampling approach, *Environmental and Ecological Statistics* **14**: 101–111.
- Hahn, T. (2005). Cuba - a library for multidimensional numerical integration, *Computer Physics Communications* **168**: 78–95.
- Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes, *Biometrika* **58**: 83–90.
- Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*, John Wiley & Sons, Chichester.
- Isham, V. and Westcott, M. (1979). A self-correcting point process, *Stochastic Processes and their Applications* **8**: 335–347.
- Johnson, S. G. (2010). The NLOpt nonlinear-optimization package. <http://ab-initio.mit.edu/nlopt>.
- Mase, S. (1996). The threshold method for estimating total rainfall, *Annals of the Institute of Statistical Mathematics* **48**: 201–213.
- Mecke, K. and Stoyan, D. (2005). Morphological characterization of point patterns, *Biometrical Journal* **47**: 473–488.
- Møller, J. and Rasmussen, J. G. (2005). Perfect simulation of Hawkes processes, *Advances in Applied Probability* **37**: 629–646.
- Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*, Chapman & Hall/CRC, Boca Raton.
- Myllymäki, M., Mrkvička, T., Seijo, H. and Grabarnik, P. (2013). Global envelope tests for spatial processes, *arXiv:1307.0239[stat.ME]*.
- Nadarajah, S. and Kotz, S. (2007). A truncated bivariate Cauchy distribution, *Bulletin of the Malaysian Mathematical Sciences Society* **30**: 185–193.
- Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes, *Journal of the American Statistical Association* **83**: 9–27.
- Ogata, Y. (1998). Space-time point-process models for earthquake occurrences, *Annals of the Institute of Statistical Mathematics* **50**: 379–402.
- Penttinen, A., Stoyan, D. and Henttonen, H. M. (1992). Marked point processes in forest statistics, *Forest Science* **38**: 806–824.
- Platt, W. J., Evans, G. W. and Rathbun, S. L. (1988). The population dynamics of a long-lived conifer (*Pinus palustris*), *The American Naturalist* **131**: 491–525.

- Pommerening, A. (2002). Approaches to quantifying forest structures, *Forestry* **75**: 305–324.
- Powell, M. J. D. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. Research report NA2009/06, Department of Applied Mathematics and Theoretical Physics, Cambridge, England.
- Rathbun, S. L. (1996). Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes, *Journal of Statistical Planning and Inference* **51**: 55–74.
- Rathbun, S. L. and Cressie, N. (1994). A space-time survival point process for a longleaf pine forest in southern Georgia, *Journal of the American Statistical Association* **89**: 1164–1174.
- Schlater, M., Riberio, P. and Diggle, P. J. (2004). Detecting dependence between marks and locations of marked point processes, *Journal of Royal Statistical Society Series B* **66**: 79–93.
- Schoenberg, F. P. (2004). Testing separability in spatial-temporal marked point processes, *Biometrics* **60**: 471–481.
- Stoyan, D., Kendall, W. S. and Mecke, J. (1995). *Stochastic Geometry and Its Applications*, second edn, Wiley, Chichester.
- Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields*, Wiley, Chichester.
- Stoyan, D. and Stoyan, H. (1996). Estimating pair correlation functions of planar cluster processes, *Biometrical Journal* **38**: 259–271.
- Stoyan, D. and Walder, O. (2000). On variograms in point process statistics, II: Models for markings and ecological interpretation, *Biometrical Journal* **42**: 171–187.
- Takahata, H. (1994). Nonparametric density estimations for a class of marked point processes, *Yokohama Mathematical Journal* **41**: 127–152.
- Tanaka, U., Ogata, Y. and Stoyan, D. (2008). Parameter estimation and model selection for Neyman-Scott point processes, *Biometrical Journal* **50**: 43–57.