

# INHOMOGENEITY IN SPATIAL COX POINT PROCESSES – LOCATION DEPENDENT THINNING IS NOT THE ONLY OPTION

MICHAELA PROKEŠOVÁ

Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 18675 Praha, Czech Republic

e-mail: prokesov@karlin.mff.cuni.cz

(Accepted June 30, 2010)

## ABSTRACT

In the literature on point processes the by far most popular option for introducing inhomogeneity into a point process model is the location dependent thinning (resulting in a second-order intensity-reweighted stationary point process). This produces a very tractable model and there are several fast estimation procedures available. Nevertheless, this model dilutes the interaction (or the geometrical structure) of the original homogeneous model in a special way. When concerning the Markov point processes several alternative inhomogeneous models were suggested and investigated in the literature. But it is not so for the Cox point processes, the canonical models for clustered point patterns. In the contribution we discuss several other options how to define inhomogeneous Cox point process models that result in point patterns with different types of geometric structure. We further investigate the possible parameter estimation procedures for such models.

Keywords: Cox processes, inhomogeneity, moment estimation, parameter estimation, spatial point processes.

## INTRODUCTION

In this paper we want to discuss the problem of introducing inhomogeneity into spatial Cox point processes. In contrast with the Markov point processes, where several models with different types of inhomogeneities (or different mechanisms how to introduce inhomogeneity into the homogeneous model) were introduced (see, *e.g.*, Jensen and Nielsen, 2001 for the overview), looking through the papers about Cox point processes we see that the only used type of inhomogeneous model is the location dependent thinning (resulting in the so-called second order intensity reweighted point process, Baddeley *et al.*, 2000).

There are two reasons for this situation. The first is the simplicity and tractability of moment estimation procedures for the parameters of the location dependent thinning model whereas for other kind of inhomogeneities it is not clear what to do since the models are highly “nonlinear” in nature. We will show in the sequel that certain approximations can deal with this problem and it is possible to derive reasonable parameter estimators also for the more complicated inhomogeneities.

The second reason is that by Markov point processes with strong negative interactions (*e.g.*, hardcore interaction) it is clearly visible that different types of inhomogeneities produce point patterns with very different geometrical structure. Which means that for fitting the data with such strong interactions (like,

*e.g.*, cell tissues or advanced materials like sinter filters with gradient structure – see, *e.g.*, Hahn *et al.*, 2003 for examples) it is necessary to have a correct inhomogeneous model. On the other side, the type of data usually modelled by Cox point processes (like different plant communities in ecology) do not usually exhibit interactions strong enough to make the misfit we get by fitting the location dependent thinning type inhomogeneity obvious at first sight. Thus the misfit is pragmatically ignored in the applications.

Let us start by explaining what we mean by different geometrical structure. In Fig. 1, we see examples of realizations of four different inhomogeneous hardcore point processes with the same intensity function.

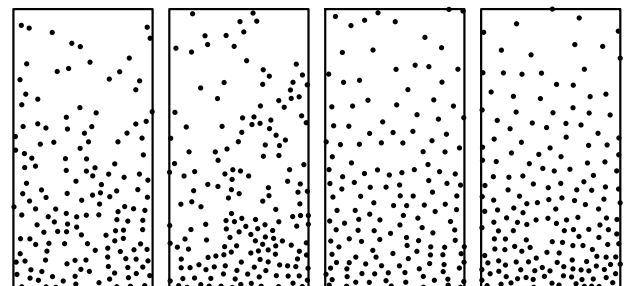


Fig. 1. Examples of different inhomogeneous Markov point processes with the same intensity function. For details see the text.

The inhomogeneity was introduced into the model by (from left to right) an inhomogeneous first-order potential (Ogata and Tanemura, 1986), location dependent thinning (Baddeley *et al.*, 2000), a 1-1 spatial transformation of the homogeneous point pattern (Nielsen and Jensen, 2004) and by location dependent scaling (Hahn *et al.*, 2003).

When examining the point pattern produced by location dependent thinning we see that this model dilutes the interaction (or the geometrical structure) of the original homogeneous model in a special way – the repulsive interactions so obvious in the area with high intensity are hardly recognizable in the area with low intensity – the point pattern there looks almost like a Poisson point process. All the other models preserve more interactions in the area with low intensity, *i.e.*, more from the geometrical structure of the homogeneous version of the hard-core model. The other extreme is the point pattern produced by location dependent scaling which keeps the same geometrical structure in any part of the observation window independently of the intensity function.

The fact, that the situation is similar also for the Cox point processes is illustrated in Fig. 2 where we can see realizations of several shot-noise Cox processes with the same first order intensity function. All three processes were derived from the same homogeneous Thomas process but different types of inhomogeneity were used.

Here we can see the location dependent thinning type inhomogeneity in the middle panel and again we observe the dilution of the clusters in the area with low intensity. The point pattern in the low intensity area looks not like a cluster process but almost like a Poisson process. The other two point patterns are obviously different – the left one having the same clusters but distributed inhomogeneously in the space, the right one showing a location dependent scaling type of behaviour where the point patterns in areas with different intensity look like scaled copies of the same homogeneous template process. The clustered nature of the point pattern is preserved also in the areas of low intensity. We will discuss the exact models from which the point patterns were generated in detail later in the paper.

Nevertheless the differences between the different inhomogeneities become less pronounced when the interactions are weaker (*i.e.*, the clustering is weaker). Fig. 3 shows what happens if we double the scale of the clusters (of the homogeneous template) and keep the rest of the model parameters the same as in Fig. 2. Here we still can see the differences in the geometrical structure in the areas of low intensity but the difference is not as obvious as in Fig. 2.

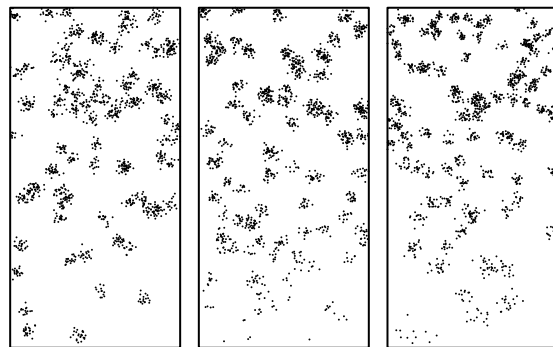


Fig. 2. *Examples of different inhomogeneous shot-noise Cox point processes with strong interactions and the same intensity function. For details see the text.*

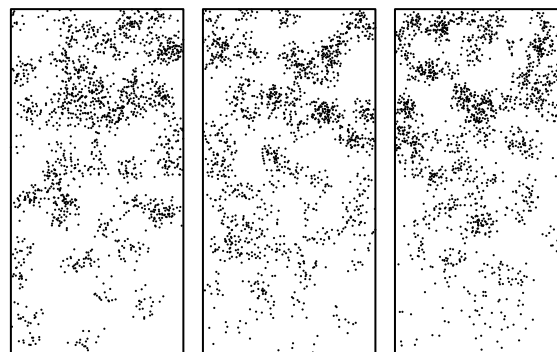


Fig. 3. *Examples of different inhomogeneous shot-noise Cox point processes with moderate interactions and the same intensity function. For details see the text.*

Concerning the estimation of the models we already said that the reason, why the location dependent thinning models are so popular, is the simplicity of moment estimation in such models. The higher order intensity functions  $\rho^{(k)}(x_1, \dots, x_k)$ ,  $k \geq 2$  are namely equal to the product

$$\rho^{(k)}(x_1, \dots, x_k) = \rho_H(x_1, \dots, x_k) \prod_{i=1}^k \rho(x_i), \quad (1)$$

of the  $k$ -th order intensity function of some homogeneous point process and the product of the first order intensity function of the thinned process  $\rho(x_i) = \rho^{(1)}(x_i)$  for the points  $x_i$  from the point configuration  $\mathbf{x} = (x_1, \dots, x_k)$ . And this fact makes it possible to estimate first the inhomogeneity – *i.e.*, the (first order) intensity function  $\rho$ , and then conditionally on this estimate to proceed with the estimation of the interaction parameters from the higher order intensity functions in the same way as in the homogeneous case.

However this strategy is not applicable to the other types of inhomogeneities. There the inhomogeneity enters the model in a more “nonlinear” way, thus we do not have the product structure for the higher order intensity functions.

But we can use other properties of the models for a 2-step estimation of the inhomogeneity parameters and subsequently the interaction parameters. For example in the transformation model we can estimate the transformation from the intensity function, transform the point pattern back by the estimated transformation and then estimate the interaction parameters like in the homogeneous case (Nielsen and Jensen, 2004). Or for the locally scaled model we can estimate the local scaling function from the intensity function and then conditionally on this estimate of the inhomogeneity parameters (parametrizing the local scaling function) use the pseudolikelihood methods to estimate the interaction parameters of the model (Prokešová *et al.*, 2006). Inspired by some of these ideas we derive two step moment estimation procedures for several kinds of inhomogeneous Cox processes.

## PRELIMINARIES

We first recall the basic notions and introduce our notation. For more detailed information see the standard references Daley and Vere-Jones (1988) and Stoyan *et al.* (1995).

Let  $\mathcal{X} \subseteq \mathbb{R}^d$  and  $X$  be a point process on  $\mathcal{X}$ . For a Borel set  $A$  in  $\mathbb{R}^d$   $|A|$  will denote the volume of  $A$  and  $|X \cap A|$  the number of points  $X$  has in  $A$ . For an  $R > 0$  we will denote  $B(0, R)$  the ball centered in the origin with radius  $R$ .

For any given point  $u \in \mathbb{R}^d$  let  $du$  be the infinitesimal region that contains the point  $u$ . Following Diggle (2003) we can define the (first-order) intensity function  $\rho$  of  $X$  by

$$\rho(u) = \lim_{|du| \rightarrow 0} \left( \frac{E|X \cap du|}{|du|} \right), \quad (2)$$

*i.e.*, the mean number of points from  $X$  occurring in  $du$ ; and the second-order intensity function  $\rho^{(2)}(u, v)$  by

$$\rho^{(2)}(u, v) = \lim_{|du|, |dv| \rightarrow 0} \left( \frac{E|(X \cap du)(X \cap dv)|}{|du||dv|} \right). \quad (3)$$

When  $X$  is simple (it does not have multiple points) then the intuitive meaning of  $\rho^{(2)}(\cdot, \cdot)$  is that  $\rho^{(2)}(u, v)|du||dv|$  is the approximate probability that  $du$  and  $dv$  each contain a point from  $X$ , where

$u \neq v$ . Higher order intensity functions are defined analogically.

If  $X$  is stationary (*i.e.*, its distribution is invariant with respect to the simultaneous shifts of all the points in  $X$ ), then  $\rho(u) = \rho = \text{const}$ , and

$$\begin{aligned} \rho^{(k)}(u, v_1, \dots, v_{k-1}) &= \\ \rho^{(k)}(0, v_1 - u, \dots, v_{k-1} - u). \end{aligned} \quad (4)$$

Thus all the  $k$ -th order intensity functions can be reduced to equivalent functions of only  $(k - 1)$  arguments.

In the stationary case two important summary statistics, which are often used in the applications, are defined by means of the second order intensity function  $\rho^{(2)}$ . The so-called pair correlation function (sometimes called simply the  $g$ -function) is defined by

$$g(u, v) = \frac{\rho^{(2)}(u, v)}{\rho^2}, \quad (5)$$

and because of the reducibility (Eq. 4) of  $\rho^{(2)}$  it is equivalent to a function of one argument

$$g(u, v) = g(u - v), \quad u, v \in \mathbb{R}^d. \quad (6)$$

The fact that the correlation function of a stationary point process depends only on the difference of the locations  $u$  and  $v$  and not on the locations themselves is an important property of the stationary point processes. In the sequel we will denote the  $g$ -function of a stationary process as a function of just one argument, as is usual in the literature. For the Poisson point process (*i.e.*, a process with no interactions) the  $g$ -function is identically equal to 1.  $g$ -function values larger than 1 indicate positive correlation between the occurrence of the points of the point process in the locations  $u, v$ , smaller values than 1 indicate negative correlations.

The second important statistic is the  $K$ -function defined by

$$\begin{aligned} K(r) &= \int_{\|u\| < r} \rho^{(2)}(0, u) du / \rho^2 \\ &= \int_{\|u\| < r} g(u) du, \quad r > 0, \end{aligned}$$

The  $K$ -function corresponds to the mean number of points of the point pattern in a ball  $B(x, r)$  centered in a random point  $x$  of the point pattern divided by the intensity. Thus  $K(r) = \pi r^2$  for the Poisson point process in  $\mathbb{R}^2$ , larger/smaller values of  $K$ -function than  $\pi r^2$  indicating respectively clustering/inhibition in the point process at range  $r$ .

## COX PROCESSES

Let  $\mathcal{X} \subseteq \mathbb{R}^d$  and  $\{\Lambda(u), u \in \mathcal{X}\}$  be a non-negative random field. The well-known definition states that a point process  $X$  on  $\mathcal{X}$  is a Cox point process with the driving field  $\Lambda$  if conditionally on  $\Lambda = \lambda$ ,  $X$  is a Poisson point process with the (first order) intensity function  $\lambda$ . For the intensity functions then holds

$$\rho^{(k)}(v_1, \dots, v_k) = E \prod_{i=1}^k \Lambda(v_i). \quad (7)$$

A Cox point process is stationary if and only if the driving field is stationary.

The two most often used classes of spatial Cox point processes are the log Gaussian Cox processes (Møller *et al.*, 1998) and the shot noise Cox processes (Møller, 2003). Even though these can be generalized and unified in the framework of Lévy based Cox processes (Hellmund *et al.*, 2008) for ease of exposition we stick to the two classical models.

Let  $\{\Psi(u) : u \in \mathcal{X}\}$  be a Gaussian random field. Then the Cox process  $X$  with the driving field  $\Lambda(u) = \exp(\Psi(u))$  is called the log Gaussian Cox process. Since the distribution of the Gaussian field  $\Psi$  is completely determined by the mean value field  $m(u) = E\Psi(u)$  and the covariance function  $c(u, v) = \text{Cov}(\Psi(u), \Psi(v))$  so is the distribution of  $X$  and it holds

$$\rho(u) = \exp\left(m(u) + \frac{1}{2}c(u, u)\right), \quad (8)$$

$$\rho^2(u, v) = \rho(u)\rho(v)\exp(c(u, v)). \quad (9)$$

The other class of shot noise Cox processes is defined by the driving field  $\Lambda$  of the form

$$\Lambda(u) = \sum_{(r,w) \in \Phi} rk(w, u). \quad (10)$$

Here  $k$  is a kernel, *i.e.*, a measurable, nonnegative function on  $\mathbb{R}^d \times \mathbb{R}^d$  such that  $k(w, \cdot)$  is a probability density for all  $w$ , and  $\Phi$  is a Poisson point process on  $(0, \infty) \times \mathcal{X}$  with a locally integrable intensity measure  $\zeta$ . Thus the shot noise Cox process has the same distribution as the superposition  $\bigcup_{(r,w) \in \Phi} X_{(r,w)}$  of independent Poisson processes  $X_{(r,w)}$  with intensity functions  $rk(w, \cdot)$ . The shot noise Cox process is stationary if the measure  $\zeta$  is translation invariant with respect to shifts in the spatial coordinate  $w \in \mathcal{X}$  and if the kernel  $k$  is invariant under simultaneous translations in both coordinates, thus being just a function of the difference  $k(w, u) = k(w - u)$ .

The first and second order intensity function are given by

$$\rho(u) = \int \int rk(w, u)\zeta(dr, dw) \quad (11)$$

$$\begin{aligned} \rho^{(2)}(u, v) &= \rho(u)\rho(v) \\ &+ \int \int r^2k(w, u)k(w, v)\zeta(dr, dw). \end{aligned} \quad (12)$$

*Example.* The simplest example of a shot noise Cox process is the so-called modified Thomas process (Thomas, 1949) (in the sequel we will call it shortly a Thomas process). In  $\mathbb{R}^2$  it is given by the choice

$$\zeta(dr, dw) = \mu \delta_v(dr) dw, \quad (13)$$

where  $\delta_v$  denotes the Dirac measure in  $v$ ,  $dw$  integration with respect to the Lebesgue measure and  $k(x) = \exp(-\|x\|^2/(2\sigma^2))/(2\pi\sigma^2)$  the bivariate normal kernel with scale parameter  $\sigma > 0$ . Thus it corresponds to a superposition of clusters  $\bigcup_{w \in \Phi'} X_w$  where  $\Phi'$  is a stationary Poisson process with intensity  $\mu$  (mother intensity) and each mother  $w$  generates a Poisson cluster  $X_w$  of daughter points with mean number of daughter points in the cluster equal to  $v$  and probability distribution function of the location of the daughter points relative to its parent is a bivariate radially symmetric normal distribution. We can easily derive that  $\rho(u) = \mu v$  and

$$\begin{aligned} \rho^{(2)}(u, v) &= \\ &\mu^2 v^2 \left(1 + \frac{1}{4\pi\sigma^2\mu} \exp\left(-\frac{\|u-v\|^2}{4\sigma^2}\right)\right). \end{aligned} \quad (14)$$

## LOCATION DEPENDENT THINNING

Now we review the most popular type of inhomogeneity in spatial point processes and the moment estimation methods available for this type of (parametric) models.

Let  $\mathcal{X} \subseteq \mathbb{R}^d$  and  $X$  be a homogeneous point process on  $\mathcal{X}$  (by homogeneous we mean stationary - for  $\mathcal{X} = \mathbb{R}^d$ , or a restriction of a stationary point process to  $\mathcal{X} \subset \mathbb{R}^d$ ) and let  $p : \mathcal{X} \rightarrow [0, 1]$  be a function on  $\mathcal{X}$ . Let  $Y$  be the point process

$$Y = \{x \in X : R(x) \leq p(x)\}, \quad (15)$$

where  $R(x) \sim \text{Uniform}[0, 1]$  are i.i.d. random variables independent of  $X$ . Then  $Y$  is a location dependent thinning of  $X$  with retention probabilities  $p(x)$ . For  $Y$  then holds

$$\rho_Y^{(k)}(v_1, \dots, v_k) = \rho_X^{(k)}(v_1, \dots, v_k) \prod_{i=1}^k p(v_i), \quad (16)$$



which implies Eq. 1 and it is possible to define consistently an inhomogeneous version of the  $g$ -function

$$g_i(u, v) = g_i(u - v) = \frac{\rho^{(2)}(u, v)}{\rho(u)\rho(v)}, \quad (17)$$

which depends only on  $(u - v)$  and the inhomogeneous  $K$ -function

$$K_i(r) = \int_{\|u\| < r} g_i(u) du, \quad r > 0, \quad (18)$$

which is equal to the  $K$ -function of the original unthinned process  $X$ . Like for the stationary processes the inhomogeneous  $g$ -function  $g_i$  will be denoted as a function of only one argument.

Let us remark here, that processes derived by location dependent thinning are not the only ones for which Eq. 17 defines consistently the inhomogeneous  $g$ -function. The construction works for a broader class of processes called second-order intensity-reweighted stationary (SOIRS) point processes. These were introduced in Baddeley *et al.* (2000), as was the inhomogeneous  $g$ -function. Nevertheless the SOIRS processes do not generally fulfill Eq. 1 for  $k > 2$ .

Let us now suppose that the point process model is parametrized by a vector parameter  $\psi$  consisting of two subvectors  $\psi = (\beta, \theta)$ , where  $\beta$  parametrizes the thinning function  $p_\beta(x)$  (inhomogeneity parameter) and  $\theta$  parametrizes the original homogeneous point process  $X_\theta$  with fixed constant intensity.

*Example.* Inhomogeneous Thomas process with log-linear intensity function (Waagepetersen, 2007): Let  $\mathcal{X} \subseteq \mathbb{R}^2$ ,  $X_\theta$  be a Thomas process with mother intensity  $\mu$ , mean number of daughter points in a cluster equal to  $\nu$  and  $k$  the bivariate normal kernel with scale parameter  $\sigma$ . And let  $z(u) \in \mathbb{R}^k$  be a vector of covariates that are recorded in the location  $u \in \mathcal{X}$  and let us define the retention probability

$$p(u) = \frac{1}{M} \exp(z(u)\beta^T), \quad (19)$$

where  $M = \max_{u \in \mathcal{X}} (\exp(z(u)\beta^T))$ . Then

$$\rho(u) = \frac{\mu\nu}{M} \exp(z(u)\beta^T), \quad (20)$$

$$\rho^{(2)}(u, v) = \rho(u)\rho(v)g_i(u - v), \quad (21)$$

and  $g_i(u - v) = g_\theta(u - v)$  is equal to the  $g$ -function of the homogeneous Thomas process

$$1 + \frac{1}{4\pi\sigma^2\mu} \exp\left(-\frac{\|u - v\|^2}{4\sigma^2}\right). \quad (22)$$

After the reparametrization  $\beta_0 = \log(\mu\nu/M)$  we have

$$\rho(u) = \rho_\beta(u) = \exp(\beta_0 + z(u)\beta^T). \quad (23)$$

Our inhomogeneity parameter is then  $\beta = (\beta_0, \dots, \beta_k)$  and interaction parameter  $\theta = (\mu, \sigma)$ .

The estimation in such models then proceeds as follows: we estimate the inhomogeneity parameter  $\hat{\beta}$  from the first order moment properties *i.e.*, by maximization of the Bernoulli composite likelihood  $L_B$  derived from the first order intensity function

$$L_B(\beta) = \exp\left(-\int_{\mathcal{X}} \rho_\beta(u) du\right) \prod_{x \in Y \cap \mathcal{X}} \rho_\beta(x). \quad (24)$$

If there is a unique  $\beta$  which maximizes  $L_B$  we can get it as the solution of the estimating equation

$$\sum_{x \in Y \cap \mathcal{X}} \frac{d(\log \rho_\beta(x))}{d\beta} - \int_{\mathcal{X}} \frac{d(\log \rho_\beta(x))}{d\beta} \rho_\beta(u) du = 0. \quad (25)$$

This is an unbiased estimating equation for processes of location dependent thinning type. Note that  $L_B$  is actually equal to the likelihood of an inhomogeneous Poisson process with intensity function  $\rho_\beta(u)$ . Thus the estimate of the first-order parameter  $\beta$  is obtained by ignoring the interactions in the point process  $Y$ . In the sequel we will denote  $\hat{\rho} = \rho_{\hat{\beta}}$ . Asymptotic properties of the estimator  $\hat{\beta}$  were investigated in Waagepetersen (2007) and Waagepetersen and Guan (2009).

In the second step we estimate the interaction parameter  $\theta$  conditionally on  $\beta = \hat{\beta}$ . There are several options, how to do this.

The oldest one is the minimum contrast estimation for the  $g_i$  or  $K_i$  function (Diggle, 2003; Waagepetersen and Guan, 2009). Or we can use the composite likelihood approach (Guan, 2006) and find  $\theta$  as the argument of maxima for the composite (log) likelihood function

$$\log CL(\theta) = \sum_{x \neq y \in Y \cap \mathcal{X}, \|x - y\| < R} \left[ \log(\hat{\rho}(x)\hat{\rho}(y)g_\theta(y - x)) - \log\left(\int_{\mathcal{X}} \int_{\mathcal{X}} \hat{\rho}(u)\hat{\rho}(v)g_\theta(u - v)I(\|u - v\| < R) du dv\right) \right],$$

where  $I(\cdot)$  denotes the indicator function and  $R > 0$  is a user specified tuning constant.

Another version of the composite log likelihood function may be defined by

$$\log CL'(\theta) = \sum_{x \neq y \in Y \cap \mathcal{X}, \|x - y\| < R} \log(\hat{\rho}(x)\hat{\rho}(y)g_\theta(y - x)) - \int_{B(0, R)} \int_{\mathcal{X} \cap (\mathcal{X} - u)} \hat{\rho}(x)\hat{\rho}(x + u)g_\theta(u) dx du.$$

This was introduced under the name of second order object function in Waagepetersen (2007).

Or we can apply a so called Palm likelihood function  $L_P$ . This was recently introduced for the stationary case by Tanaka *et al.* (2007) and the method is based on the Poisson approximation of the likelihood of the process of differences  $(X \cap W) - (X \cap W)$  between the observed points of  $X$ . Because the intensity function of  $(X \cap W) - (X \cap W)$  is expressed by means of the intensity  $\lambda_0$  of the Palm measure of  $X$  the function

$$\log L_P(\theta) = \sum_{x \neq y \in X \cap \mathcal{X}, \|x-y\| < R} \log(|X \cap W| \lambda_0(x-y; \theta)) - |X \cap W| \int_{\mathbb{R}^d} I(\|u\| < R) \lambda_0(u; \theta) du,$$

is called the (log)-Palm likelihood. For details see Tanaka *et al.* (2007). This method can be directly generalized to the inhomogeneous case of location dependent thinned processes if we define

$$\log L_P = \sum_{x \neq y \in Y \cap \mathcal{X}, \|x-y\| < R} \log(\hat{\rho}(y) g_\theta(y-x)) - \sum_{x \in Y \cap \mathcal{X}} \int_{\mathcal{X} \cap B(x,R)} \hat{\rho}(u) g_\theta(u-x) du,$$

where  $R > 0$  is a tuning constant. The maximum Palm likelihood estimate of  $\theta$  is the value which maximizes  $\log L_P = \log L_P(\theta)$ . Further details and a simulation study comparing the methods introduced above can be found in (Prokešová and Jensen, in preparation).

*Example.* For the inhomogeneous Thomas process with log-linear intensity function the (vector) first order estimation equation (Eq. 25) becomes simply

$$\sum_{x \in Y \cap \mathcal{X}} (1, z(x)) = \int_{\mathcal{X}} (1, z(u)) \exp(\beta_0 + z(u)\beta^T) du. \quad (26)$$

The implementation of the composite or Palm likelihood estimation of  $\theta = (\mu, \sigma)$  is then straightforward when using the closed form of the  $g$ -function (Eq. 22). The last parameter  $\nu$  is obtained from  $\beta_0 = \log(\mu\nu/M)$  and the value  $M = \max_{u \in \mathcal{X}} (\exp(z(u)\hat{\beta}^T))$ .

## INTRODUCING INHOMOGENEITY INTO COX PROCESSES

Let us start again with the log Gaussian Cox processes. We said already that their distribution is completely determined by the mean value field and

the covariance function of the corresponding Gaussian field  $\Psi$  or equivalently it is completely determined by the first and second order intensity function (since we have the 1-1 correspondence given by the Eqs. 8 and 9). This means that as far as the covariance function  $c(u, v)$  is translation invariant, the (inhomogeneous) log Gaussian Cox process  $X$  is of the location dependent thinning type. For the case of the log-linear intensity function dependent on the covariates  $z(u)$  we can reparametrize

$$\Lambda(u) = \exp(z(u)\beta^T + \Psi(u)), \quad (27)$$

with  $E\Psi(u) = -c(0,0)/2$  which means  $E \exp(\Psi(u)) = 1$  ( $\Psi$  is now a stationary Gaussian field) and having the covariance function parametrized by  $\theta$  we can apply the estimation procedures from the preceding chapter. Since in the literature there are not any cases with covariance function which is not translation invariant we can conclude that all practically usable log Gaussian Cox processes are of the location dependent thinning type.

We would like to remark here, that the main advantage of the log Gaussian Cox processes, *i.e.*, the simplicity of the specification of the model, changes to a disadvantage from the viewpoint of the modelling of point patterns with diverse geometrical structure. Namely simple specification by only  $\mu$  and  $c$  corresponds to the full determination of the model by only  $\rho^{(1)}$  and  $\rho^{(2)}$ . It follows that all the higher ( $\geq 3$ ) order product intensity functions are determined by the first and second order intensity functions and consequently the possible geometrical structure of the resulting point patterns is highly restricted. Thus, *e.g.*, for modelling the first and third point pattern from Fig. 2 log Gaussian Cox processes are not a good choice. Here one should use the shot noise Cox processes which are more flexible.

We can introduce inhomogeneity in the shot noise Cox process in several ways. Let us suppose that  $f_\beta : \mathcal{X} \rightarrow \mathbb{R}^+$  is a function bounded away from zero and infinity on  $\mathcal{X}$ , parametrized by the inhomogeneity parameter  $\beta$  and  $k$  and  $\zeta$  are a translation invariant kernel and an intensity measure on  $(0, \infty) \times \mathcal{X}$  invariant with respect to shifts in the spatial coordinate  $x \in \mathcal{X}$ .

We can define a model with a new kernel

$$k'_\beta(w, u) = f_\beta(u) k(w, u), \quad (28)$$

(thus the kernel  $k'$  no longer integrates to 1, nevertheless it is possible to reparametrize this model in such a way, that we still have  $k'$  a probability kernel, but the formulas are more complicated (Hellmund *et al.*, 2008). Then we obtain a point process of the location dependent thinning type with

$\rho(u) \propto f_\beta(u)$  and a well defined inhomogeneous  $g$ -function equal to the  $g$ -function of the homogeneous model and we can use the estimation procedures from the preceding chapter.

An example of such a point process can be seen in the middle panel of Fig. 2. This is a realization of the inhomogeneous Thomas process model from the preceding chapter with  $\beta = (\beta_0, \beta_1)$ ,  $z(u) = u_2$  and  $f_\beta = \exp(\beta_1 u_2)$ . We see clearly how the strength of the clustering disappears in the areas with low intensity since the clusters are gradually thinned out even to individual points in the bottom part of the panel. Such a point pattern would correspond, *e.g.*, to a situation when we want to model a biological population of seedlings, produced by a homogeneous population of older plants, where the seedlings were thinned by different ecological conditions at their respective location.

A completely different point pattern can be seen in the first panel of Fig. 2 – here we have the same clusters (same size of clusters and number of points in the clusters) but their number (=intensity) in different parts of the observation window is different. Keeping to our plant community example this would correspond to a situation, where we have homogeneous outer ecological conditions but the sources of seedlings (older plants) are unevenly distributed in space. This point process was derived from the homogeneous one by setting

$$\zeta'(dr, dw) = f_\beta(w) \zeta(dr, dw), \quad (29)$$

and keeping the same kernel  $k$ .

Concerning the question of parameter estimation of the model (Eq. 29) by moment methods we can plug in into Eqs. 11 and 12

$$\rho(u) = \iint r k(w, u) f_\beta(w) \zeta(dr, dw), \quad (30)$$

$$\begin{aligned} \rho^{(2)}(u, v) &= \rho(u) \rho(v) \\ &+ \iint r^2 k(w, u) k(w, v) f_\beta(w) \zeta(dr, dw), \end{aligned} \quad (31)$$

and see that the separation of the inhomogeneity parameters is no longer possible. Nevertheless we can still accept the pragmatic attitude and under the assumption that the scale of change of the inhomogeneity function is larger than the size of the clusters (*i.e.*, the scale/practical range of the kernel function  $k$ ) we can use the approximation

$$\rho_{\beta, \theta}(u) \sim f_\beta(u) \iint r k_\theta(w, u) \zeta_\theta(dr, dw), \quad (32)$$

to obtain from the estimation (Eq. 25) a (biased, but under the above mentioned assumption reasonable)

estimate  $\hat{\beta}$  of the inhomogeneity parameter  $\beta$  and subsequently with the approximation

$$\begin{aligned} \rho_{\hat{\beta}, \theta}^{(2)}(u, v) &\sim \rho_{\hat{\beta}, \theta}(u) \rho_{\hat{\beta}, \theta}(v) \\ &+ f_{\hat{\beta}}\left(\frac{u+v}{2}\right) \iint r^2 k_\theta(w, u) k_\theta(w, v) \zeta_\theta(dr, dw), \end{aligned}$$

we can use the composite or Palm likelihood to obtain the estimate of the interaction parameter  $\theta$ .

*Example.* Let us consider the inhomogeneous Thomas process with  $\zeta(dr, dw) = f_\beta(w) \mu \delta_v(dr) dw$ , *i.e.*, the mothers are distributed inhomogeneously according to the intensity function  $f_\beta \mu$ . Then by the approximation (Eq. 32) we get

$$\rho_{\beta, \theta}(u) \sim f_\beta(u) \cdot \text{const}, \quad (33)$$

and by choosing, *e.g.*,  $f_\beta(u) = \exp(z(u) \beta^T) / M$  we see that we obtain the same estimate of  $\beta$  like for the log-linear inhomogeneous Thomas process of location dependent thinning type from the previous chapter. Nevertheless for the second order intensity function we get the approximation

$$\begin{aligned} \rho_{\hat{\beta}, \theta}^{(2)}(u, v) &\sim \rho_{\hat{\beta}}(u) \rho_{\hat{\beta}}(v) \cdot \\ &\left( 1 + \frac{f_{\hat{\beta}}\left(\frac{u+v}{2}\right) \exp\left(-\frac{\|u-v\|^2}{4\sigma^2}\right)}{f_{\hat{\beta}}(u) f_{\hat{\beta}}(v) 4\pi\sigma^2\mu} \right), \end{aligned} \quad (34)$$

thus the obtained estimate of  $\theta$  by either the composite or Palm likelihood method will be influenced by the weighting factor of  $f_{\hat{\beta}}\left(\frac{u+v}{2}\right) / (f_{\hat{\beta}}(u) f_{\hat{\beta}}(v))$ .

When constructing the inhomogeneous models it is also possible to include the inhomogeneity function  $f_\beta$  into several ingredients of the shot noise Cox process specification, thus obtaining point patterns with more complicated geometrical structure. Our last example is the point pattern in the right hand panel of Fig. 2. We can notice a certain resemblance with the right hand panel in Fig. 1 – the locally scaled point pattern. Here also the different parts of the figure look like a scaled version of the same clustered point pattern, where the intensity is high, the clusters are more compact and where the intensity is low, they are more loose.

This type of inhomogeneous shot noise Cox process was derived from the homogeneous model by defining the kernel  $k'$

$$k'(w, u) = k_\theta\left(\frac{w}{f_\beta(w)}, \frac{u}{f_\beta(w)}\right) \frac{1}{f(w)^d}, \quad (35)$$

(note that  $k'$  is again a probability density in  $u$ ) and the intensity measure  $\zeta'$  of  $\Phi'$  by

$$\zeta'(dr, dw) = \zeta_\theta(dr, dw) / f_\beta(w)^d. \quad (36)$$

Here  $d$  is the dimension of the space  $\mathcal{X}$ , *i.e.*,  $d = 2$  for the planar case and  $f_\beta$  has the meaning of the scaling function – *i.e.*, the point pattern looks at different places like a scaled version of some homogeneous template point process scaled with the factor  $f_\beta(u)$ .

Concerning the parameter estimation we would suggest to proceed analogically to the Markov point process case (Prokešová *et al.*, 2006), *i.e.*, with a two step estimation procedure, first estimating the function  $f_\beta$  from the Poisson likelihood (Eq. 24) using the approximation

$$\rho_\beta(u) \sim \frac{1}{f_\beta^d} \iint r k_\theta(w, u) \zeta_\theta(dr, dw), \quad (37)$$

and then instead of the pseudolikelihood (used in the Markov point process case) use the composite likelihood or Palm likelihood with the approximation for the second order intensity function

$$\begin{aligned} \rho_{\beta, \theta}^{(2)}(u, v) &\sim \\ &\frac{1}{f_{\hat{\beta}}(u)^d f_{\hat{\beta}}(v)^d} \left[ \left( \iint r k_\theta(w, u) \zeta_\theta(dr, dw) \right)^2 \right. \\ &+ \left. \iint r^2 k_\theta\left(w, \frac{u}{f_{\hat{\beta}}(\frac{u+v}{2})}\right) k_\theta\left(w, \frac{v}{f_{\hat{\beta}}(\frac{u+v}{2})}\right) \zeta_\theta(dr, dw) \right] \\ &\sim \rho_{\hat{\beta}}(u) \rho_{\hat{\beta}}(v) g_X\left(\frac{u}{f_{\hat{\beta}}(\frac{u+v}{2})}, \frac{v}{f_{\hat{\beta}}(\frac{u+v}{2})}\right), \end{aligned}$$

where  $g_X$  denotes the homogeneous version of the  $g$ -function for the homogeneous shot noise Cox process with kernel  $k_\theta$  and intensity measure  $\zeta_\theta$ .

## CONCLUSIONS

In this paper we discussed the problem of defining an inhomogeneous model for clustered point process data, which possess a specific geometrical structure which makes the modelling by the locally thinned point process models inappropriate. The disadvantage of such models is a complex structure of the intensity functions which makes the parameter estimation by moment methods (be it minimum contrast of  $K$ -function, Poisson likelihood, composite likelihood or Palm likelihood estimation) intractable without accepting certain approximations which lead to biased estimates. But parameter estimation in these models is possible, which was illustrated by the methods in this paper. Nevertheless the suggested methods and their properties deserve further investigation.

## ACKNOWLEDGEMENTS

This work was supported by projects GAČR 201/08/P100 from the Czech Science Foundation and MSM0021620839 financed by the Ministry of Education of the Czech Republic.

## REFERENCES

- Baddeley AJ, Møller J, Waagepetersen R (2000). Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Stat Neerl* **54**:329–50.
- Daley DJ, Vere-Jones D (1988). An introduction to the theory of point processes. New York: Springer Verlag.
- Diggle P (2003). Statistical analysis of spatial point patterns. New York: Oxford University Press.
- Guan Y (2006). A composite likelihood approach in fitting spatial point process models. *J Am Stat Assoc* **101**:1502–12..
- Hahn U, Jensen EBV, van Lieshout MNM, Nielsen LS (2003). Inhomogeneous spatial point processes by location-dependent scaling. *Adv Appl Probab* **35**:319–36.
- Hellmund G, Prokešová M, Jensen EBV (2008). Lévy-based Cox point processes. *Adv Appl Probab* **40**:603–29.
- Jensen EBV, Nielsen LSN (2001). A review on inhomogeneous point processes. In: Basawa I, Heyde CC, Taylor RL, eds. *Select Proc Symp Infer Stoch Proces*. Beachwood, Ohio: Institute of Mathematical Statistics, 297–318.
- Møller J (2003). Shot noise Cox processes. *Adv Appl Probab* **35**:614–40.
- Møller J, Syversveen AR, Waagepetersen RP (1998). Log Gaussian Cox processes. *Scand J Statist* **25**:451–82.
- Nielsen LS, Jensen EBV (2004). Statistical inference for transformation inhomogeneous point processes. *Scand J Statist* **31**:131–42.
- Ogata Y, Tanemura M (1986). Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In: Francis IS, Manly BFJ, Lam FC, eds. *Proc Pacific Statist Congr – 1985*. Amsterdam: Elsevier, 150–4.
- Prokešová M, Hahn U, Jensen EBV (2006). Statistics for locally scaled point processes. In: Baddeley A, Gregori P, Mateu J, Stoica R, Stoyan D, eds. *Case Studies in Spatial Point Process Modelling*. New York: Springer Verlag, 99–123.
- Stoyan D, Kendall WS, Mecke J (1995). *Stochastic geometry and its applications*. 2nd ed. Chichester: Wiley.
- Tanaka U, Ogata Y, Stoyan D. (2007). Parameter estimation and model selection for Neyman-Scott point processes. *Biometrical J* **49**:1–15.



Thomas M (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika* [36:18–25](#).

Waagepetersen RP (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott

processes. *Biometrics* [63:252–8](#).

Waagepetersen RP, Guan Y (2009). Two-step estimation for inhomogeneous spatial point processes. *J Roy Stat Soc B* [71:685-702](#).