

Jiří Dvořák; Michaela Prokešová

Moment estimation methods for stationary spatial Cox processes - A comparison

*Kybernetika*, Vol. 48 (2012), No. 5, 1007--1026

Persistent URL: <http://dml.cz/dmlcz/143096>

## Terms of use:

© Institute of Information Theory and Automation AS CR, 2012

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://project.dml.cz>

# MOMENT ESTIMATION METHODS FOR STATIONARY SPATIAL COX PROCESSES – A COMPARISON

JIŘÍ DVOŘÁK AND MICHAELA PROKEŠOVÁ

In the present paper we consider the problem of fitting parametric spatial Cox point process models. We concentrate on the moment estimation methods based on the second order characteristics of the point process in question. These methods represent a simulation-free faster-to-compute alternative to the computationally intense maximum likelihood estimation. We give an overview of the available methods, discuss their properties and applicability. Further we present results of a simulation study in which performance of these estimating methods was compared for planar point processes with different types and strength of clustering and inter-point interactions.

*Keywords:* moment estimation methods, spatial Cox point process, parametric inference

*Classification:* 62M30, 60G55

## 1. INTRODUCTION

Estimation of parametric models for spatial point processes has been a very active research area in the last few years. Motivated by the need of analyzing always larger and more complicated data sets in reasonably short time, several simulation-free estimation methods based on composite likelihood and/or estimating equations have been developed as alternatives to the computationally more demanding maximum likelihood and Bayesian methods, see [21] for a recent overview.

In the present paper we will focus on Cox point processes (sometimes also called doubly stochastic point processes in the literature) ([4, 17]). Cox process is the preferred point process model when we need to analyze clustered point patterns. It is able to model clustering of different strength on different scales as well as inhomogeneity dependent on spatial covariates. As such it is used in a large spectrum of applications e. g. in biology, ecology and epidemiology (see [14] for examples). An important property of the Cox point process (from the statistical point of view) is the fact that its likelihood is not available in a closed or computationally tractable form whereas the product densities (i. e. densities of the moment measures) may be tractable. The consequence is that for performing the (approximate) maximum likelihood estimation computationally demanding Markov chain Monte Carlo simulation methods must be developed for each particular application (see [20, Chapter 10] for further details). Therefore several alternative simulation-free estimation methods for the stationary spatial Cox processes

were proposed in the past few years ([8, Chapter 6], [10, 18, 23]). All of them can be generalized to two-step estimation procedures applicable also to estimation of nonstationary Cox processes but in this paper we will consider only the original versions of the estimators for the stationary case.

All of the introduced estimation methods are based on the second order product density of the point process (or equivalent characteristics derived from it) and do not use any information about the moment measures of order higher than two. It follows that the resulting estimators are consistent but inefficient. Since they were introduced independently there exists no comparison of their performance. Therefore we have conducted a simulation study to compare the performance of the introduced methods in estimation of a parametric stationary Cox process observed on a medium sized observation window in  $\mathbb{R}^2$ . The comparison by means of asymptotic variance of the estimators is not feasible since the formulas for asymptotic variance contain integrals with respect to the third and fourth order moment measures of the point process we want to estimate. These higher order moment measures of course differ for different Cox process models we estimate, but more importantly they are usually too complicated to enable a direct comparison.

The paper is organized as follows. We give the necessary notation and background information in Section 2 and a discussion of the Cox process model in Section 3. In Section 4 we give an overview of the available moment estimation methods for stationary spatial Cox processes and their theoretical properties. Empirical performance of the estimators is assessed in Section 5 by means of a simulation study. Main results of the study are summarized in Section 6. We conclude with a discussion in Section 7.

## 2. BACKGROUND

In this section we first introduce some basic notation from the theory of spatial point processes. For a detailed introduction into the topic [5] or [8] can be consulted.

Let  $X$  denote a simple (i. e. without multiple points) point process in  $\mathbb{R}^d$ . Let  $\mathcal{B}^d$  be the Borel  $\sigma$ -algebra on  $\mathbb{R}^d$ . For a set  $B \in \mathcal{B}^d$  we denote by  $|B|$  its volume and by  $X(B)$  the number of points from  $X$  in  $B$ . The origin in  $\mathbb{R}^d$  will be denoted by  $o$ ,  $B(x, R)$  is the closed ball centered at  $x \in \mathbb{R}^d$  with radius  $R > 0$ . The Euclidean norm of the vector  $x$  is denoted by  $\|x\|$  and  $I$  is the indicator function.

The  $k$ th order factorial moment measure  $\alpha^{(k)}$  of the point process  $X$  is defined as follows:

$$\alpha^{(k)}(A) = \mathbb{E} \left( \sum_{u_1, \dots, u_k \in X}^{\neq} I[(u_1, \dots, u_k) \in A] \right), A \in \mathcal{B}^{dk},$$

where  $\neq$  denotes that the sum runs over  $k$ -tuples of distinct points of  $X$ . Particularly,  $\alpha^{(1)}(A) = \mathbb{E}X(A)$  is the mean number of points observed in the set  $A$ .

If  $X$  is stationary, i. e. its distribution is invariant with respect to translations, we have  $\alpha^{(1)}(A) = \lambda|A|$  for any  $A \in \mathcal{B}^d$  and the constant  $\lambda \geq 0$  is called intensity of  $X$ .

We will always assume that factorial moment measures  $\alpha^{(k)}$  have densities with respect to Lebesgue measure on  $\mathbb{R}^{dk}$  which are denoted by  $\lambda^{(k)}$  and called the  $k$ th order product densities of  $X$  or  $k$ th order intensity functions of  $X$ . In the sequel we will use mainly the first two intensity functions  $\lambda^{(1)}$  and  $\lambda^{(2)}$ .

Under the assumption of stationarity of  $X$  we get  $\lambda(x) = \text{const} = \lambda$  and  $\lambda^{(2)}(x, y) = \lambda^{(2)}(0, y - x) = \lambda^{(2)}(y - x)$  is effectively a function of one argument only. Moreover,

$$\lambda^{(2)}(x, y) = \lambda\lambda_0(y - x), \quad x, y \in \mathbb{R}^d,$$

where  $\lambda_0$  is the density of the second order reduced moment measure of the process  $X$ .

The function  $\lambda_0$  is also called conditional intensity or Palm intensity in the literature since  $\lambda_0$  is in fact the intensity function of the Palm distribution  $\mathcal{P}_0$  of the original point process  $X$  (see [6, Proposition 13.2.VI] for details).

Two popular characteristics of stationary point processes can be defined from  $\lambda_0$ , namely the pair correlation function

$$g(x, y) = g(y - x) = \lambda^{(2)}(x, y)/\lambda^2 = \lambda_0(y - x)/\lambda, \quad x, y \in \mathbb{R}^d,$$

and the  $K$ -function defined by

$$\lambda K(R) = \int_{B(o,R)} \lambda_0(u) \, du = \mathbb{E} [X(B(o, R) \setminus \{o\}) | X(\{o\}) = 1], \quad R > 0.$$

Note that  $\lambda K(R)$  can be interpreted as the mean number of further points from  $X$  in  $B(x, R)$  centered at a typical point  $x$  of the point process  $X$ .

### 3. COX PROCESSES

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

$$\lambda^{(k)}(u_1, \dots, u_k) = \mathbb{E} \left[ \prod_{i=1}^k \Psi(u_i) \right].$$

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

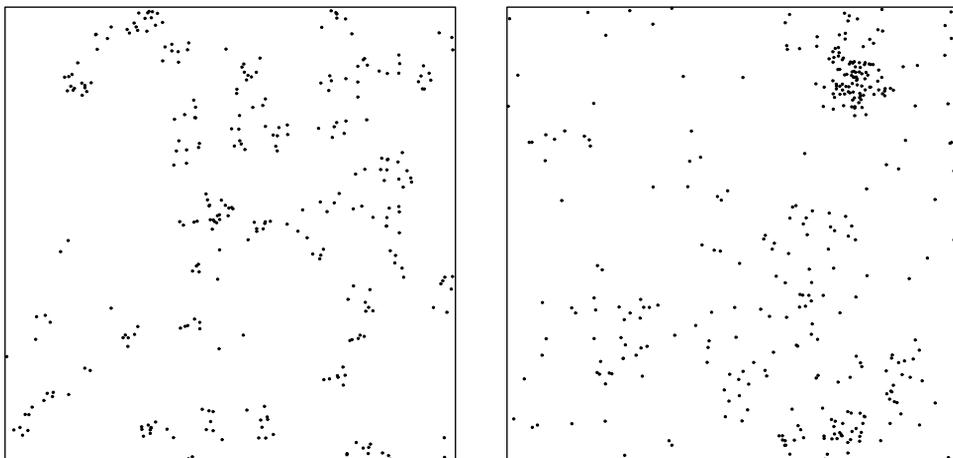
The two most often used classes of spatial Cox point processes are the log-Gaussian Cox processes ([18]) and the shot-noise Cox processes ([19]). Even though these can be generalized and unified in the framework of Lévy-based Cox processes ([13]), for ease of exposition we stick to the two classical models.

A shot noise Cox process  $X$  is characterized by the random intensity function

$$\Psi(u) = \sum_{(c,\gamma) \in \Phi} \gamma k(c, u),$$

where  $c \in \mathbb{R}^d, \gamma > 0, \Phi$  is a Poisson process on  $\mathbb{R}^d \times (0, \infty)$  and  $k$  is a kernel, i. e. a measurable, nonnegative function on  $\mathbb{R}^d \times \mathbb{R}^d$  such that  $k(c, \cdot)$  is a probability density for each  $c$ . Note that  $X$  is distributed as the superposition of independent Poisson processes  $X_{(c,\gamma)}$  with intensity functions  $\lambda^{(1)}(\cdot) = \gamma k(c, \cdot), (c, \gamma) \in \Phi$ , where we interpret  $X_{(c,\gamma)}$  as a cluster with center  $c$  and mean number of points  $\gamma$ . Thus the shot-noise Cox processes include several important cluster point process models, e. g. the Poisson Neyman–Scott processes.

**Example 3.1.** (Thomas process in  $\mathbb{R}^2$ ) A Thomas process in  $\mathbb{R}^2$  can be constructed as follows. Let  $C$  be a stationary planar Poisson point process of mother points with intensity  $\kappa > 0$ . Each mother point  $c \in C$  produces a random number of daughter points. Their number has Poisson distribution with mean  $\mu > 0$ . The daughters around  $c$  are i.i.d. with isotropic Gaussian displacement (with standard deviation  $\sigma > 0$ ) from the mother point  $c$ . A realization of the Thomas process is shown in Figure 1.



**Fig. 1.** Sample realizations of Cox point processes in observation window  $W = [0, 1]^2$  showing different types of clustering.  
 Left: Thomas process, right: log-Gaussian Cox process.

The intensity of the Thomas process is  $\lambda = \kappa \mu$  and the Palm intensity is

$$\lambda_0(u) = \kappa \mu + \frac{\mu}{4\pi\sigma^2} \exp\left\{-\frac{\|u\|^2}{4\sigma^2}\right\}.$$

Log-Gaussian Cox process is defined by the conditional intensity function  $\Psi(u) = \exp\{Z(u)\}$ , where  $Z$  is a Gaussian random field on  $\mathbb{R}^d$ . Since the distribution of the Gaussian field  $Z$  is completely determined by the mean value field  $m(u) = \mathbb{E}Z(u)$  and the covariance function  $c(u, v) = \text{cov}(Z(u), Z(v))$ , so is the distribution of  $X$ . Moreover, simple relationships hold for the first two intensity functions of  $X$ :

$$\lambda^{(1)}(u) = \exp\{m(u) + c(u, u)/2\}, \quad \lambda^{(2)}(u, v) = \lambda^{(1)}(u)\lambda^{(1)}(v) \exp\{c(u, v)\}.$$

**Example 3.2.** (log-Gaussian Cox process in  $\mathbb{R}^2$  with exponential covariance function) Consider a stationary Gaussian random field  $Z$  in  $\mathbb{R}^2$  with constant mean  $\mu$  and exponential covariance function  $c(x, y) = \sigma^2 \exp\{-\beta\|x - y\|\}$ , where  $\sigma^2 > 0$  and  $\beta > 0$  are parameters. Then the log-Gaussian Cox process  $X$  with conditional intensity function

$\Psi(u) = \exp \{Z(u)\}$  will have the first order intensity function  $\lambda = \exp \left\{ \mu + \frac{\sigma^2}{2} \right\}$  and the Palm intensity function will be

$$\lambda_0(u) = \exp \left\{ \mu + \frac{\sigma^2}{2} \right\} \cdot \exp \left\{ \sigma^2 \cdot \exp \{-\beta \|u\|\} \right\}.$$

A realization of such log-Gaussian Cox process is shown in Figure 1.

#### 4. MODEL FITTING FOR COX PROCESSES

Let  $\mathbf{x} = (x_1, \dots, x_n)$  denote a point pattern observed in a compact observation window  $W$ , i.e. the realization of the process  $X \cap W$ , where  $x_i, i = 1, \dots, n$ , denote locations of the observed points. We will assume a parametric model for  $X$  and the vector of unknown parameters will be denoted by  $\theta$ .

Let us first discuss the classical maximum likelihood estimation. To use this method we will consider the density  $f_\theta(\mathbf{x})$  of  $X$  with respect to the distribution of stationary unit-rate Poisson process (i.e.  $\lambda^{(1)}(u) = 1$ ). The maximum likelihood estimate  $\hat{\theta}$  is then obtained as the value of  $\theta$  maximizing  $f_\theta(\mathbf{x})$ .

For a Poisson process with the intensity function  $\lambda_\theta(x)$  the density is

$$f_\theta(\mathbf{x}) = \exp \left\{ |W| - \int_W \lambda_\theta(u) \, du \right\} \prod_{i=1}^n \lambda_\theta(x_i). \tag{1}$$

As long as a suitable parametrization of  $\lambda_\theta$  is available the density  $f_\theta$  has a tractable form and the estimate  $\hat{\theta}$  can be easily obtained.

For a Cox process driven by random conditional intensity function  $\Psi_\theta(u), u \in W$ , the density is

$$f_\theta(\mathbf{x}) = \mathbb{E}_\theta \left[ \exp \left\{ |W| - \int_W \Psi_\theta(u) \, du \right\} \prod_{i=1}^n \Psi_\theta(x_i) \right].$$

In order to obtain the maximum likelihood estimate of  $\theta$  it is needed to repetitively evaluate the expectation including a complicated integral term with respect to possible values of the random conditional intensity function  $\Psi_\theta$ . One can of course take advantage of MCMC or other techniques and use approximations of the likelihood function  $f_\theta(\mathbf{x})$ , (see e.g. [20, Chapter 10]). Nevertheless this approach is usually computationally very demanding and thus faster, simulation-free alternatives were sought.

Three such methods will be reviewed in the following subsections. They are in fact second order moment estimation methods because all of them are based on the second-order moment characteristics of the process  $X$  like  $K, g, \lambda_0$  or  $\lambda^{(2)}$ . In the following we will assume that  $X$  is a stationary Cox point process characterized by its second-order intensity function  $\lambda^{(2)}(\cdot; \theta)$  or by some of the other above listed characteristics. As was explained in Section 1 for many Cox process models these characteristics are available in a reasonably tractable form as functions of the parameter  $\theta$  and thus the maximization of the respective estimation criteria is numerically feasible.

### 4.1. Minimum contrast

This estimation method was in the context of spatial statistics first introduced in [7, Chapter 5]. It can be based either on the  $K$ -function or the pair correlation function  $g$ , see e.g. [8, Chapter 6]. In the version using the  $g$ -function this method requires isotropy of the process  $X$  in addition to its stationarity. In this case the  $g$ -function is a function of a scalar argument.

The vector of parameters  $\theta$  is estimated by minimizing the discrepancy measure

$$\int_r^R [\hat{K}^c(u) - K^c(u; \theta)]^2 du \quad \text{or} \quad \int_r^R [\hat{g}^c(u) - g^c(u; \theta)]^2 du \quad (2)$$

between the nonparametric estimate  $\hat{K}$  or  $\hat{g}$  and its theoretical value  $K(\cdot; \theta)$  or  $g(\cdot; \theta)$ , respectively.

Constants  $c$ ,  $r$  and  $R$  are used to control the sampling fluctuations in the estimates of  $K$  and  $g$ . In [8, Section 6.1.1] it is recommended that for fitting aggregated point patterns using the  $K$ -function the constant  $c = 0.25$  is used and that for data on a unit square  $R$  should be no larger than 0.25. The remaining constant  $r$  can be set to 0 or a small positive value, e.g. the minimum observed interpoint distance. For the pair correlation function  $g$  there is no standard recommendation available.

Minimum contrast estimators are popular partly due to the fact that they are implemented in the `spatstat` package for R (see [2]) for three important cluster point process models: Thomas process, isotropic log-Gaussian Cox process with exponential covariance function and Matern cluster point process.

The asymptotic properties of the minimum contrast estimators are discussed in [11] and [12]. In [12] strong consistency and asymptotic normality for minimum contrast estimates using the  $K$ -function was proved for Poisson cluster processes. In [11] asymptotic normality for minimum contrast estimates using the  $K$ -function was showed under a strong mixing assumption which is fulfilled also for a broad range of log-Gaussian Cox processes.

### 4.2. Composite likelihood

Composite likelihood approach is a general statistical methodology (see [16]). In the context of point processes it is based on adding together individual log-likelihoods for single points or pairs of points of the process  $X$  to form a composite log-likelihood. Several versions of composite likelihood have been suggested for estimation of different types of spatial point processes (see [1, 10, 21]). Composite likelihood suitable for estimation of Cox processes was introduced in [10]. It uses the second-order intensity function  $\lambda^{(2)}(\cdot, \theta)$  to obtain the probability density for two points of  $X$  occurring at locations  $x$  and  $y$ :

$$f(x, y; \theta) = \frac{\lambda^{(2)}(x - y; \theta)}{\int_W \int_W \lambda^{(2)}(u - v; \theta) \, dudv}.$$

After adding the individual log-likelihoods the composite log-likelihood is obtained:

$$\log CL(\theta) = \sum_{x \neq y \in X \cap W, \|x - y\| < R} \log \frac{\lambda^{(2)}(x - y; \theta)}{\int_W \int_W I(\|u - v\| < R) \lambda^{(2)}(u - v; \theta) \, dudv}.$$

Here only pairs of points within distance  $R$  are considered. Disregarding the pairs of points separated by distance larger than  $R$  is motivated by the fact that distant pairs of points are often nearly independent. They do not carry much information about the parameters but increase variability of the estimator.

Numerical maximization of  $\log CL(\theta)$  in the form presented above can be computationally demanding. This issue can be solved by application of the inner region correction for the edge effects. It leads to a simplified formula

$$\log CL(\theta) = \sum_{x \in X \cap (W \ominus R), y \in X \cap W, 0 < \|x - y\| < R} \log \frac{\lambda^{(2)}(x - y; \theta)}{\lambda^2 |W \ominus R| K(R)}, \tag{3}$$

where  $W \ominus R = \{w \in W : B(w, R) \subset W\}$ , i.e.  $W \ominus R$  is the window  $W$  eroded by distance  $R$ . Vector  $\hat{\theta}$  maximizing  $\log CL(\theta)$  is then taken for the estimate of  $\theta$ . Consistency and asymptotic normality of the composite likelihood estimator are proved in [10] under suitable mixing assumptions.

### 4.3. Palm likelihood

Palm likelihood estimator was introduced in [23] and uses a very “geometrical” approach. It is based on the process of differences among the points of the observed point process  $X$ . Let us start with considering the processes

$$Y_x = \{y - x, x \neq y \in X \cap W\}, \quad x \in X \cap W.$$

Each  $Y_x$  is again a point process (now inhomogeneous) with intensity function  $\lambda_0(\cdot; \theta)$  – the Palm intensity of the original process  $X$  (see [22] and [23] for more detailed explanation).

Now ignoring the interactions in the process  $Y_x$ , i.e. approximating  $Y_x$  by a Poisson process, the log-likelihood of  $Y_x \cap B(o, R)$  is the following (up to a constant – cf. (1)):

$$\sum_{y \in X \cap W, \|x - y\| < R} \log \lambda_0(x - y; \theta) - \int_{\mathbb{R}^2} I(\|u\| < R) \lambda_0(u; \theta) du.$$

Here we consider only pairs of points within distance  $R$  following the same reasoning as in the case of the composite likelihood estimator.

Treating all the  $Y_x, x \in X \cap W$ , as independent, identically distributed replications, we sum the individual log-likelihoods over  $x \in X \cap W$  and get the so-called Palm log-likelihood:

$$\log PL(\theta) = \sum_{x \neq y \in X \cap W, \|x - y\| < R} \log \lambda_0(x - y; \theta) - X(W) \int_{\mathbb{R}^2} I(\|u\| < R) \lambda_0(u; \theta) du.$$

Applying the inner region correction we get the following expression for the Palm log-likelihood which is more useful for the estimation in practice:

$$\log PL(\theta) = \sum_{x \in X \cap (W \ominus R), y \in X \cap W, 0 < \|x - y\| < R} \log \lambda_0(x - y; \theta) - X(W \ominus R) \lambda K(R). \tag{4}$$

Again, the maximizer of  $\log PL(\theta)$  is taken for the estimate of  $\theta$ .

Note that even though the Palm likelihood estimation was derived by using the process of differences it is a second-order moment method as well because it is based on the second-order characteristic  $\lambda_0$  of the observed point process  $X$ . Let us also remark that its use is not restricted to simple models such as the Thomas process. If the Palm intensity function  $\lambda_0$  is in such form that direct maximization of  $\log PL(\theta)$  is not possible (e.g. for general Neyman–Scott processes), numerical algorithms are available for finding  $\operatorname{argmax} \log PL(\theta)$  (see [23, Section 3]).

Strong consistency and asymptotic normality of the Palm likelihood method are proved in [22] under suitable mixing assumptions. These are fulfilled for a broad range of Poisson Neyman–Scott processes and log-Gaussian Cox processes.

Moreover we would like to stress here that since the Palm intensity is defined as a function of the parameter  $u \in \mathbb{R}^d$  and not just its length, the Palm likelihood estimation is not restricted to isotropic point processes only. It can be used for estimation of any suitably parametrized stationary point process Cox model in  $\mathbb{R}^d$  (the same holds for the composite likelihood estimation). A worked out example of an anisotropic Thomas process in  $\mathbb{R}^2$  can be found in [22].

**Example 3.1. (continued)** For the stationary Poisson Neyman–Scott processes with mother intensity  $\kappa > 0$ , mean number of daughter points in the cluster  $\mu > 0$  and the daughter displacement kernel  $k$  parametrized by  $\theta_k$ , we have  $\lambda_0(u) = \kappa\mu + \mu k(u)$ . Thus the score equation we get by differentiation of the log Palm likelihood (4) with respect to  $\mu$  will yield

$$\hat{\mu} = \frac{N}{\kappa K(R) |X \cap (W \ominus R)|}, \tag{5}$$

where  $N$  denotes the number of terms in the sum in (4) and the  $K$ -function does not depend on  $\mu$ . For the composite likelihood the score equation we get by differentiation of (3) with respect to  $\mu$  is identity  $\frac{2N}{\mu} = \frac{2N}{\mu}$  and the parameter  $\mu$  is non-identifiable. The same applies to the minimum contrast estimation using both the  $K$  and  $g$  function because none of them depends on  $\mu$ . Thus the parameter  $\mu$  must be estimated from the known relation between  $\mu$ , intensity  $\lambda$  of the process and the other estimated parameters, where we take  $\hat{\lambda} = X(W)/|W|$ .

When inserting  $\hat{\mu}$  into  $\log PL(\theta)$  we find that  $\log PL(\hat{\mu}, \kappa, \theta_k)$  treats the observed process of differences as i.i.d. observations with density  $f(u) = \lambda_0(u)/(\kappa K(R))$ ,  $u \in B(o, R)$  and the respective score equations are unbiased estimating equations. Also  $\log CL(\kappa, \theta_k)$  treats the process of differences between pairs of observed points from  $X$  as i.i.d. observations with the same density  $f$  and leads to a set of unbiased estimating equations although the estimating functions are different.

**Example 3.2. (continued)** For the log-Gaussian Cox process the situation is analogous. The parameter  $\mu$  – the mean of the random field  $Z$  – cannot be estimated by the minimum contrast method and the composite likelihood method and has to be calculated from the estimate of intensity by  $\log \left( \frac{X(W)}{|W|} \right) - \frac{\sigma^2}{2}$ . For the Palm likelihood

method we get the estimate

$$\hat{\mu} = \log \frac{N}{K(R)|X \cap (W \ominus R)|} - \frac{\sigma^2}{2}, \tag{6}$$

and after inserting  $\hat{\mu}$  into  $\log PL(\theta)$  we get again a sum over the observed process of differences of likelihoods corresponding to the density  $\lambda_0(u)/K(R)$ ,  $u \in B(o, R)$  and leading to a set of unbiased estimating equations. The same holds for  $\log CL(\theta_{-\mu})$ , where  $\theta_{-\mu}$  denotes the parameter vector  $\theta$  without the parameter  $\mu$ .

### 5. SIMULATION STUDY

To compare the empirical performance of the estimators we chose two different types of cluster processes – the Thomas process and the log-Gaussian Cox process, see Examples 3.1 and 3.2.

To assess the performance of the estimators on middle-sized to large point patterns exhibiting different degree of clustering we chose eight combinations of parameter values for the Thomas process ( $\kappa = 25$  or  $50$ ,  $\mu = 4$  or  $6$  and  $\sigma = 0.02$  or  $0.04$ , representing relatively strong and weak clustering, respectively) and for the log-Gaussian Cox process ( $\beta = 10$  or  $20$  representing relatively strong and weak dependence, respectively,  $\sigma^2 = 1$  and  $\mu$  calculated so that the intensity of the process is  $100, 150, 200$  and  $300$  as in the case of the Thomas process).

For each process and each combination of parameters we generated 500 independent realizations and re-estimated the parameters using the three moment estimation methods (i. e. using formulas (2), (3) and (4)).

#### 5.1. Parameter estimation – computational details

Realizations of the Thomas process and the log-Gaussian Cox process with appropriate parameters in a unit square window  $W$  were simulated using the package `spatstat` for R (for reference see [2]).

When simulating realizations of the log-Gaussian Cox process we need to use a finite representation of the random field  $Z$  in the observation window  $W$ . We chose a regular grid of  $25 \times 25$  points and instead of the Gaussian random field  $Z$  we simulated a Gaussian random vector representing values of  $Z$  in the given lattice points. Then we approximated values of the random field  $Z$  in  $W$  by the value in the nearest lattice point.

We used the minimum contrast method (MCE) with both the  $K$ -function (MCEK) and the pair correlation function  $g$  (MCEg). The tuning parameters in the case of the  $K$ -function were chosen according to the recommendation in [8, Section 6.1.1]. The same values of parameters were used for the  $g$ -function, i. e.  $c = 0.25$ ,  $R = 0.25$  and  $r$  equal to the minimum interpoint distance observed in the given realization. The estimation was conducted by the functions provided by `spatstat`.

The composite likelihood (CLE) and Palm likelihood (PLE) estimation methods have only one tuning parameter,  $R$ , determining the maximum distance of pairs of points that will be taken into account. The distance  $R$  has to be chosen carefully so that the variance of the estimators is reduced but not much information about the interaction in the point

pattern is lost. Since the observation window  $W$  is a unit square we chose the values of  $R$  to be 0.1, 0.2 and 0.3, respectively.

To take into account the interpoint interactions when choosing the value of  $R$  we also consider  $R = r_{data}$ , where  $r_{data}$  corresponds to a range of correlation. It is determined as the smallest  $r$  for which  $\hat{g}(r) < 1$  holds, where  $\hat{g}$  is a nonparametric estimate of the pair correlation function  $g$ .

For CLE and PLE we applied the inner region correction for the edge effects, see (3) and (4). It is straightforward and can be used for data observed in irregular windows. Other edge corrections could be used as well, for example the torus correction could perform better in the case of a rectangular window but it is difficult to use in practice for irregularly shaped windows.

The estimation for CLE and PLE was conducted in software `Mathematica 7`. Maximization of the appropriate log-likelihood functions was performed using a combination of two methods.

First, derivatives of the log-likelihood function with respect to the unknown parameters were calculated and a vector of parameter values was found for which all the derivatives were equal to zero using a default Newton's method.

Then, if the numerical method diverged (i. e. the estimate lies out of a generously long interval containing the true value of the parameter), the parameters are estimated again by direct maximization of the log-likelihood function using simulated annealing.

This procedure is motivated by the fact that the estimation using derivatives is fast but somewhat numerically unstable, while direct maximization is more computationally demanding but in general less likely to diverge. The combination of these two methods was a good compromise between the computational time and numerical stability.

For the minimum contrast methods and the composite likelihood method the value of the non-identifiable parameter  $\mu$  was estimated from the observed intensity of the point process as described in the examples at the end of Section 4.

## 6. RESULTS

Main results are summarized in Tables 1 and 2 for the Thomas process and the log-Gaussian Cox process, respectively. They show results for two variants of MCE (based on  $K$ -function and pair correlation function) and for CLE and PLE with different choices of the tuning parameter  $R$ . Complete results are available in [9].

The tables show relative mean biases of the estimators and relative mean squared errors (MSEs) (by relative we mean divided by the true value of the estimated parameter or by its square for the MSE). All the statistics were obtained from the middle 95% of the estimates from 500 replications.

Neglecting the 5% of the most extreme estimates was motivated by the fact that in certain situations the estimation methods can be numerically unstable, see Section 5.1. If such a situation was encountered in practice the estimates would easily be identified as unrealistic and one would alter the parameters of the underlying optimization methods or opt for an alternative estimation method. However, due to the extent of the computations involved, this was not possible in this simulation study. Note that if all the estimates were used to calculate the statistics these would be severely distorted by the

numerical instability and hence completely uninformative for a prospective researchers willing to find a method of choice for their particular dataset.

## 6.1. Thomas process

### 6.1.1. Parameter $\sigma$

Among the three parameters of the Thomas process model  $\sigma$  is the one estimated with most accuracy by any of the compared methods. Particularly MCE $g$  and CLE with  $R = 0.1$  give very precise estimates. From the two MCE methods MCEK is always worse than MCE $g$ . PLE with  $R = 0.1$  (as well as CLE with  $R = 0.2$ ) is comparable with MCEK for strong clustering. For weak clustering PLE is inferior to the other methods.

The reason for this is the strong positive bias (around 25% for  $R = 0.1$ ) of PLE in case of weak clustering. PLE generally overestimates the parameter  $\sigma$  while MCE always underestimates it. CLE with  $R = 0.1$  is virtually unbiased for strong clustering and positively biased for weak clustering. For strong clustering CLE with  $R = 0.1$  generally showed the smallest variance followed by MCE $g$ , PLE and MCEK. For weak clustering variance of CLE and PLE with  $R = 0.1$  becomes larger and MCE $g$  becomes the most stable estimator.

An important observation is the rapid deterioration of the quality of the CLE and PLE estimates with increasing value of the tuning parameter  $R$ . Increase in the variability of the estimators with growing  $R$  is to be expected, since with the employed inner region edge correction and with growing  $R$  we lose growing percentage of the data. However, for CLE and  $R = 0.3$  the estimator of  $\sigma$  is not just bad, it is even numerically unstable, and in case of weak clustering and higher intensity of the point process the same happens also for  $R = 0.2$ .

An explanation for this could be found in the exact form of the score equations for CLE - the estimating functions as described in Example 3.1 in Section 4 have a steep step around the correct value of the parameter  $\sigma$  and then become virtually constant (and nonzero). Thus for a small number of observed pairs of points from  $X$  and a clustered point pattern (with few distant pairs observed) the estimating equation can lead to a severe overestimation of the parameter  $\sigma$ . Therefore it is extremely important when estimating the interaction parameters for processes with weaker clustering by CLE or PLE to choose the tuning constant  $R$  not too large. Definitely not larger than the range of correlation and reasonably small with respect to the size of the observation window so that not too much data is lost by the edge correction.

### 6.1.2. Parameter $\kappa$

For the other interaction parameter  $\kappa$  estimated directly by all three methods the best estimates are provided by PLE (with  $R = 0.1$ , for weak clustering also with  $R = 0.2$ ) in the majority of cases. The exception are processes with low intensity and strong clustering, i. e. with a few number of tight well defined clusters in the observed point pattern – here MCE $g$  gives better results. Like with the other parameters the MCEK method is consistently worse than MCE $g$  and the worst results are generally provided by CLE (with  $R = 0.1$ ).

Note that the good results of PLE ( $R = 0.1$ ) are mainly caused by the comparatively small variability of the estimates because all the PLE estimates with  $R = 0.1$  have considerable negative bias. The bias becomes lower for  $R = 0.2$  and thus the best estimates for processes with low intensity and weak interaction (i. e. a few loose clusters in the observation window) are obtained by PLE with  $R = 0.2$ . This behaviour is again implied by the exact form of the estimating functions used. Note that for CLE the situation is somewhat similar like for the parameter  $\sigma$  – CLE with  $R = 0.1$  provides nearly unbiased estimates, whereas with  $R = 0.2$  we have considerable positive bias and with  $R = 0.3$  the estimates become useless. Nevertheless the variability of CLE is generally higher than variability of the other estimators.

When comparing the MCEg and MCEK methods the main factor is also variability of the estimators - the MCEg method is consistently less variable than MCEK (the same situation like with the parameter  $\sigma$ ).

### 6.1.3. Parameter $\mu$

The last parameter  $\mu$  is by all the methods determined by means of the observed number of points of  $X$  and the values of the already estimated parameters. It is true not just for the MCE and CLE where  $\hat{\mu} = X(W)/(|W|\hat{\kappa})$  but also for the PLE method. Namely (5) is derived just from the comparison of the expected and observed number  $N$  of pairs of points from  $X$ .

Thus it is natural that the quality of the estimates of  $\mu$  for MCE and CLE follows the pattern from estimation of  $\kappa$  with MCEg being better than MCEK and CLE providing even worse estimates. The PLE estimator uses mean number of pairs of points from  $X$  which could provide more exact estimates than just  $X(W)$ . On the other hand, the formula (5) includes both of the parameters  $\kappa$  and  $\sigma$  and thus can be influenced more by the bad quality of those estimates. The simulation study shows that in reality for weak clustering PLE (with  $R = 0.1$ ) provides the best estimates of  $\mu$  – again mainly due to low variability of the estimates since they are negatively biased. For strong clustering the estimates provided by MCE are better.

### 6.1.4. Procedure using $r_{data}$

The estimated values of  $r_{data}$  are for strong clustering concentrated around 0.1 as expected, with a fair proportion being smaller than 0.1 (see [9] for the histograms of  $r_{data}$ ). For CLE the estimates of  $\sigma$  using  $R = r_{data}$  are comparable with those obtained with  $R = 0.1$ ; they have only slightly larger bias and variance. The bias is nevertheless still very small (smaller than for MCE) and the variance is only a bit larger than for MCEg. Overall, the procedure with  $r_{data}$  for CLE estimation of  $\sigma$  provides very good estimates – better than for MCEK. Estimates of the other interaction parameter  $\kappa$  by CLE with  $r_{data}$  are worse but comparable with estimates produced by CLE with  $R = 0.1$ .

For PLE the estimates of the interaction parameters with  $R = r_{data}$  are also worse than for fixed value  $R = 0.1$ . By closer inspection of the simulated data we see that particularly for the parameter  $\kappa$  the cases with  $r_{data} < 0.1$  produce significantly biased estimates.

				MCE		CLE				PLE					
	$\kappa$	$\mu$	$\sigma$	K	$g$	.1	.2	.3	$r_{data}$	.1	.2	.3	$r_{data}$		
Bias	$\hat{\sigma}$	25	4	.02	-.044	-.029	.009	-.034	-.019	.028	.068	.023	.086	.118	
				.04	-.064	-.067	.077	.208	34.7	.188	.287	.494	1.27	1.22	
		50	6	.02	-.028	-.027	.001	-.045	-.086	.021	.096	.014	.052	.106	
				.04	-.044	-.047	.093	25.2	160	.238	.233	.849	3.55	2.41	
		$\hat{\kappa}$	25	4	.02	.103	.070	-.013	.271	.917	-.097	-.143	.083	.161	-.260
					.04	.148	.159	.098	.196	.987	-.087	-.331	-.178	.042	-.107
	50	6	.02	.060	.022	.071	.318	.818	.004	-.122	.115	.139	-.157		
			.04	.109	.102	.064	.405	1.02	.075	-.293	-.133	.024	-.087		
	$\hat{\mu}$	25	4	.02	-.074	-.067	.101	-.145	-.303	.386	-.416	-.431	-.375	-.365	
				.04	-.094	-.117	.098	.022	-.187	.896	-.252	-.272	-.289	-.272	
	50	6	.02	-.054	-.038	-.057	-.199	-.338	-.002	-.405	-.434	-.344	-.378		
			.04	-.068	-.076	-.035	-.230	-.260	-.011	-.257	-.274	-.254	-.266		
MSE	$\hat{\sigma}$	25	4	.02	.035	.014	.016	.032	.094	.023	.031	.045	.135	.051	
				.04	.050	.034	.061	.317	1.10 <sup>4</sup>	.168	.211	.764	6.85	5.47	
		50	6	.02	.022	.009	.006	.018	.036	.011	.026	.026	.055	.027	
				.04	.035	.023	.036	2.10 <sup>4</sup>	1.10 <sup>5</sup>	.332	.077	2.45	32.6	16.8	
		$\hat{\kappa}$	25	4	.02	.125	.073	.207	.364	3.08	.244	.108	.186	.225	.233
					.04	.270	.217	.429	.576	5.67	.321	.170	.085	.192	.186
	50	6	.02	.091	.052	.097	.304	1.65	.111	.039	.100	.099	.109		
			.04	.189	.148	.184	.562	4.48	.268	.093	.039	.145	.150		
	$\hat{\mu}$	25	4	.02	.075	.044	.266	.142	.250	1.38	.192	.225	.185	.166	
				.04	.144	.116	.582	.435	.410	42.0	.076	.112	.159	.133	
	50	6	.02	.055	.031	.062	.101	.200	.108	.173	.211	.140	.160		
			.04	.130	.103	.133	.151	.219	.179	.071	.088	.133	.125		

Tab. 1. Summary of simulation results – Thomas process.

For weak clustering the estimated value of  $r_{data}$  is scattered between 0.1 and 0.2 with a few cases taking the value of 0.25 which was the upper bound for the estimated range of interaction (since the estimate of  $g$  function is not very stable for values larger than 0.25). For estimation of  $\sigma$  CLE with  $R = r_{data}$  has significantly larger bias and variance than with  $R = 0.1$  and PLE produces impractical estimates. For estimation of  $\kappa$  CLE with  $R = r_{data}$  produces sometimes better estimates than with  $R = 0.1$  but they are still worse than those produced by any other method. PLE with  $R = r_{data}$  produces generally worse estimates of  $\kappa$  than with  $R = 0.1$  but these are typically still better than those produced by the other methods.

In conclusion we can say that the method using  $R = r_{data}$  is not superior to the fixed value of the tuning parameter  $R = 0.1$  for CLE and PLE. Partially it can be explained by the larger variability caused by the changing value of  $R$  but very often also the bias of the estimators is larger than with the fixed value of  $R = 0.1$ . The question of the ideal choice of  $R$  for a given point process model and a given observation window is not a simple one. Even the amount of information expressed by the mean number of observed pairs of points with distance  $\leq R$  for a clustered point process  $X$  need not be a monotone function of  $R$  (see [22, Section 3]). More sophisticated methods of the adaptive choice of  $R$  must be used to produce a considerable improvement in the quality

of the estimates than just a simple choice  $R = \text{range of interaction}$ .

### 6.1.5. Overall performance

In conclusion we can say that quality of all the compared estimators improves with higher intensity of the process  $X$  and stronger interactions in the point process (i. e. tighter clusters). Concerning minimum contrast estimation the version using pair correlation function always yields better estimates than the version using the  $K$ -function. When using the CLE and PLE methods for estimation of the interaction parameters it is important to use reasonably small values of the tuning parameter  $R$  which provide reasonably good estimates. High values of  $R$  lead to numerical instability of the methods and impractical estimates – this is especially important for CLE, to a lesser extent also for PLE.

To address the overall performance, the minimum contrast estimation using the pair correlation function provides the best (in the sense of the relative MSE) estimates of the interaction parameters although for point patterns with strong clustering when estimating  $\sigma$  CLE (with  $R = 0.1$ ) yields fully comparable and sometimes better quality estimates.

For point patterns with weak clustering PLE (with  $R = 0.2$  and  $R = 0.1$ ) yields the best estimates of  $\kappa$ . However, this is due to the minimal variability of the PLE estimator which has a serious negative bias in this case. The second best (according to MSE) estimator is MCE $g$  which has half-size bias (i. e. acceptable 10%).

Parameter  $\mu$  is calculated from the intensity (of points or pairs of points) of  $X$  and as such it depends on the quality of the other parameter estimates. The best values of MSE were achieved for strong clustering by MCE $g$  and for weak clustering by (again significantly biased) PLE.

## 6.2. Log-Gaussian Cox process

### 6.2.1. Parameter $\beta$

For the log-Gaussian Cox process from Example 3.2  $\beta$  is the scale parameter of the covariance function of the Gaussian driving field and the hardest one to estimate. Unlike in the case of Thomas process for minimum contrast estimation MCEK showed always better performance than MCE $g$ . Nevertheless for stronger interaction ( $\beta = 10$ ) the best results are obtained by CLE with  $R = 0.1$  which has both minimal bias and minimal variability. Also here (as in the case of Thomas process) it is important to have reasonably small value of the tuning parameter  $R$ , i. e.  $R = 0.1$ . For larger values of  $R$  the CLE (and also PLE) deteriorates fast and the estimates become useless.

Quality of the estimates of course improves with growing intensity  $\lambda$ ; the growing amount of information seems to be best used by CLE since for high intensity ( $\lambda = 300$ ) CLE (with  $R = 0.1$ ) outperforms MCEK even in the case of weak dependence. CLE (with  $R = 0.1$ ) generally slightly underestimates  $\beta$  but the absolute value of bias is much smaller than for any other compared estimator and this holds uniformly for any log-Gaussian Cox process considered in the study. When the variability of this estimator becomes small enough (thanks to the sufficient amount of data in the large  $\lambda$  case) it becomes superior to MCE estimation even in the case of weak dependence.

To understand this results better it is good to note that the case  $\beta = 10$  which we call stronger interaction/dependence means that values of  $g(u) - 1$  (where  $g$  is the pair correlation function) are significantly positive for a larger range of values  $u$  than for the case of  $\beta = 20$ . Thus the observed point pattern of  $X$  for  $\beta = 10$  is much more variable and may contain a few clusters (or sometimes none if  $\lambda$  is low) with large scale and large number of points (like the one in the upper right corner of the right panel in Figure 1) whereas for  $\beta = 20$  the observed point pattern is more homogeneous with large number of smaller (both in terms of scale and number of points) clusters. The large clusters from  $X$  with  $\beta = 10$  make the estimation of the functional characteristics  $K$  and  $g$  less stable than in the case of  $\beta = 20$  or the Thomas process, which influences negatively the quality of the MCE estimation. Obviously the CLE method is less influenced by occurrence of these large clusters.

The same fact may explain higher efficiency of MCEK when compared with MCEg. Even in the case of weaker dependence ( $\beta = 20$ ) the clusters observed in  $X$  are highly variable in terms of the number of points and there are always some with a fairly high number of points. In such a situation the estimate of the  $K$ -function as a cumulative function is more stable than the estimate of the  $g$  function (which corresponds to the density of  $K$ -function).

The performance of PLE is inferior to the other methods and for weak dependence ( $\beta = 20$ ) the estimator is unusable. For stronger dependence in the point pattern the performance for PLE with  $R = 0.1$  is comparable with the MCE methods, for larger values of  $R$  the estimator becomes unusable again.

### 6.2.2. Parameter $\sigma^2$

The other interaction parameter  $\sigma^2$  is best estimated by the MCE methods. The estimates are only slightly biased and the variance is lower than for the other estimators. Note that the sign of the bias depends on the value of  $\beta$  – for  $\beta = 10$  (i.e. point patterns with a few large and heavy clusters)  $\sigma^2$  is underestimated while for  $\beta = 20$  it is overestimated by both MCEK and MCEg. The performance of the two MCE methods is very similar.

Worse but still reasonable results are obtained for PLE with  $R = 0.1$  and strong dependence in the point pattern. In the case of weak dependence the PLE method needs higher intensity  $\lambda$  of the point process to provide usable estimates. The worst results are provided by CLE. Both CLE and PLE have considerable negative bias (larger for the case of strong dependence) but PLE shows consistently lower variability than CLE which makes it superior.

As in the case of estimation of  $\beta$  both PLE and CLE perform reasonably well only with the tuning parameter  $R = 0.1$ , for larger values both the bias and variability increase to impractical values.

### 6.2.3. Parameter $\mu$

The intensity parameter  $\mu$  is the easiest one to estimate and it is estimated very well by any of the compared methods. Remember that MCE and CLE estimate  $\mu$  by the same formula (see Example 3.2 in Section 4) which is influenced by the value of the estimated

parameter  $\sigma^2$ . Thus it follows from the properties of the estimates of  $\sigma^2$  that the MCE estimates of  $\mu$  has to be superior to the CLE estimates and that MCEK and MCE $g$  perform equally well. Note moreover that the CLE estimates with  $R = 0.2, 0.3$  are still very good even though the quality of the estimates of  $\sigma^2$  was not good at all. Since  $\hat{\mu}$  is a linear function of  $\hat{\sigma}^2$  the influence of the quality of the estimates of interaction parameters on the estimate of the intensity parameter  $\mu$  is much smaller than for the Thomas process case.

Another interesting observation is that the efficiency of the PLE estimate which is determined by formula (6) is worse than for the other methods and improves with larger value of the tuning parameter  $R$ . The formula (6) uses the number of observed pairs of points with distance smaller than  $R$  and provides negatively biased estimates – the smaller the value of  $R$  the larger the bias of the estimates. Although the PLE estimates of  $\mu$  are still fairly exact the conclusion is that the intensity parameter  $\mu$  is better estimated by the simpler formula using just the observed number of points of the point process  $X$ .

#### 6.2.4. Procedure using $r_{data}$

The procedure using  $R = r_{data}$  was even less successful for the log-Gaussian Cox process than for the Thomas process. This could be explained partially by a more uniform distribution of the estimated values of  $r_{data}$  over the whole range between 0 and 0.25 (see [9] for the histograms of  $r_{data}$ ). Moreover, when  $\beta = 10$  we get quite often the estimate  $r_{data} = 0.25$  and for such  $R$  both CLE and PLE become impractical.

For PLE the use of data dependent  $R$  produces impractical estimates of  $\beta$  and estimates of  $\sigma^2$  with larger variance and for strong interaction also with larger bias than the procedure with fixed  $R = 0.1$ .

For CLE the estimate of the interaction parameter  $\beta$  using  $R = r_{data}$  has in case of strong dependence generally smaller bias than with fixed  $R = 0.1$  but the variance is larger and in total the estimate is worse than with fixed  $R = 0.1$ . Still it is superior to MCE. In case of weak dependence the small bias is the same like with fixed  $R = 0.1$  and in total larger variance makes the estimate inferior to MCE. The estimate of  $\sigma^2$  by CLE with  $R = r_{data}$  is worse than the other estimates and for small values of the intensity it is even impractical.

#### 6.2.5. Overall performance

The quality of all the compared estimators improves with higher intensity of the point process  $X$  and in the majority of cases also with weaker dependence (i. e. larger value of the parameter  $\beta$ ). However, there is one important exception – the CLE with  $R = 0.1$  provides more precise estimates of  $\beta$  for processes with stronger dependence. And in this case CLE is much better than the other estimation methods. A plausible explanation is that the score equations of CLE are less influenced by the high variability of the observed point pattern (caused by the variability of a few large and heavy clusters) than are the estimates of the functions  $K$  and  $g$  used for MCE. For point processes with  $\beta = 20$  and small enough intensity ( $\lambda < 300$ ) the observed clusters in the point pattern are

				MCE		CLE				PLE			
		$\lambda$	$\beta$	K	$g$	.1	.2	.3	$r_{data}$	.1	.2	.3	$r_{data}$
Bias	$\hat{\beta}$	100	10	.281	.498	-.070	.190	.645	.087	-.221	1.76	2.44	.747
			20	.146	.216	.090	.380	.401	.095	.527	1.80	4.41	.359
		300	10	.259	.360	-.143	-.036	.258	-.062	-.224	1.51	2.25	.847
			20	.140	.180	-.051	.437	.272	.092	.417	.798	.629	.484
	$\hat{\sigma}^2$	100	10	-.065	-.016	-.311	-.559	-.688	.032	-.218	-.467	-.570	-.357
			20	.050	.059	-.140	-.357	-.603	.536	-.295	-.269	-.525	-.122
		300	10	-.046	-.023	-.241	-.535	-.722	-.509	-.160	-.384	-.484	-.375
			20	.043	.072	-.131	-.336	-.668	-.196	-.131	-.212	-.422	-.129
	$\hat{\mu}$	100	10	-.001	-.006	.026	.047	.064	-.018	-.146	-.052	-.016	-.111
			20	-.005	-.006	.017	.042	.068	-.067	-.094	-.079	-.005	-.147
		300	10	.000	-.003	.019	.041	.061	.037	-.115	-.043	-.013	-.050
			20	-.006	-.008	.012	.027	.059	.014	-.086	-.062	-.011	-.082
MSE	$\hat{\beta}$	100	10	.538	.825	.112	.756	2.16	.649	.513	7.59	11.0	3.35
			20	.309	.357	.428	.983	.798	.576	2.14	8.67	51.5	2.21
		300	10	.235	.299	.085	.206	.566	.089	.268	6.12	14.0	3.57
			20	.124	.132	.091	.631	.394	.220	1.13	2.10	2.00	1.24
	$\hat{\sigma}^2$	100	10	.146	.151	.380	.548	.774	2.90	.184	.436	.623	.430
			20	.196	.198	.476	.561	.756	4.16	.294	.519	.714	.529
		300	10	.060	.058	.184	.485	.736	.480	.132	.267	.450	.285
			20	.064	.063	.136	.354	.688	.233	.121	.226	.402	.138
	$\hat{\mu}$	100	10	.005	.005	.007	.008	.012	.054	.038	.012	.011	.041
			20	.004	.004	.008	.010	.012	.066	.016	.016	.009	.055
		300	10	.002	.002	.003	.006	.007	.006	.021	.006	.005	.009
			20	.001	.001	.002	.003	.006	.003	.011	.006	.003	.011

**Tab. 2.** Summary of simulation results – log-Gaussian Cox process,  $\sigma^2 = 1$  in all cases.

more homogeneous and the MCEK becomes the best (in the sense of the relative MSE) estimator of  $\beta$  followed by MCEg.

The parameter  $\sigma^2$  is best estimated by the minimum contrast methods, which show very similar performance, followed by PLE with  $R = 0.1$ . Like in the other cases small value of the tuning parameter  $R$  is crucial for estimation of the interaction parameters by both PLE and CLE; for larger  $R = 0.2, 0.3$  both the bias and variability of the PLE and CLE become too large.

Estimators of the intensity parameter  $\mu$  are very precise for any of the compared methods. The important conclusion from the simulation study is that for log-Gaussian Cox processes a simple estimate by means of the observed (first order) intensity of the point process is superior to the more complicated estimate provided by PLE which uses the observed intensity of pairs of points from  $X$ .

### 7. DISCUSSION

We have considered the problem of model fitting for Cox point processes which can be used to model aggregated point patterns. For Cox processes the classical maximum

likelihood estimation must be implemented by the computationally demanding MCMC simulation methods and an individual numerical algorithm must be developed and tuned for each particular application. For cluster processes particularly the profile likelihood for the scale parameter of the cluster distribution must be used, which makes the procedure even more computationally demanding [21, Section 7.3]. Bayesian estimation with a suitable choice of priors can make use of the demarginalization strategy by treating the unobserved driving field of the Cox processes as missing data [21, Section 7.4]. But still the computational load of the MCMC simulation is larger by at least an order of magnitude when compared with the simulation-free moment estimation methods reviewed in this paper.

Asymptotic results for the maximum likelihood estimation were established in [15] for certain Cox processes on a real line. However, we are not aware of any such results for spatial Cox processes.

All of the moment estimation procedures reviewed in this paper besides the MCE $g$  were proved to be consistent and asymptotically normal under suitable mixing conditions (see the references in Section 4). However, the theoretical expressions for the asymptotic variance of the estimators are too complicated to be used for direct comparison of the efficiency of the estimators. Moreover the convergence may be too slow and the bias can be a more important issue than the variance for certain point processes observed on medium sized windows (like in the case of PLE of  $\kappa$  – see Section 6.1.2). Therefore we assessed the empirical performance of the discussed moment estimation procedures by a simulation study. To cover different types of clustering in the observed point pattern both a shot-noise Cox model and a log-Gaussian Cox model with several combinations of parameter values were used.

A short conclusion of the simulation study is that the minimum contrast method using the pair correlation function  $g$  provides the best estimates in the majority of cases (see the next paragraph for an exceptional situation). It seems that the functional criterion (2) is able to get more information from the point pattern than the composite and Palm likelihood. On the other hand, when it comes to more complicated models like those considered in [23] where the pair correlation function does not have a closed form anymore but is expressed by means of an integral the maximization of (2) may be numerically more demanding. Then the more feasible Palm or composite likelihood estimation can be a good alternative.

There is one important situation where the estimators' performance was different. For the estimation of the interaction range  $\beta$  for log-Gaussian Cox model the best results were provided by composite likelihood and MCE $K$  was superior to MCE $g$ . A plausible explanation is that the population of clusters observed in the log-Gaussian Cox process is much more variable (both in terms of scale and number of points in the clusters) than the one observed in the Poisson Neyman–Scott processes. A small (and highly variable) number of large and heavy clusters observed in the point pattern makes the estimates of the function  $K$  and even more so  $g$  unstable. In such a case the simpler estimating equations implied by CLE are probably more stable and provide better estimates of the interaction range parameter  $\beta$ .

In biological applications which study e. g. the interactions among members of a plant community or seed dispersal curves the interaction range can be a parameter of primary

interest. Our results suggest that especially in the case of a nonhomogenous population of observed clusters composite likelihood should be the preferred estimation method.

Another interesting issue we observed in the simulation study is a rather extreme version of the bias-variance trade-off when estimating the number of clusters (i.e. parameter  $\kappa$ ) for Thomas process with weak clustering (i.e. loose clusters). Observed point patterns are very homogeneous in this case and the individual clusters are completely unrecognisable. Thus it is understandable that for  $\kappa$  we got estimates with the worst performance (compared to other parameters) and as the results show the amount of data we used is still not enough to estimate this parameter safely. The smallest MSE was exhibited by PLE in spite of the heavy bias it had, followed by only slightly biased MCEg. Thus if one is not prepared to accept the granted serious bias provided by PLE it is better to prefer MCEg also for the estimation of  $\kappa$ .

The results we obtained about the better efficiency of the MCEg method in comparison with the MCEK method for Poisson Neyman–Scott processes are in agreement with the results published in [3] for the so called G-shot noise Cox processes. Even though these processes are not distinguishable from the Poisson Neyman–Scott processes by just the  $g$  or  $K$  function the MCE method can be used for the parameter estimation of these models. [3, Section 4.2.1] states that MCEK can be unstable in some cases and that MCEg should be the preferable method.

Further we experimented with a simple data-based procedure for the choice of the tuning parameter  $R$  for CLE and PLE methods. The conclusion is that such a choice does not improve the quality of the estimates in comparison with the a priori chosen fixed value  $R = 0.1$ . More sophisticated methods of the adaptive choice of  $R$  must be used to produce a considerable improvement in the quality of the estimates, such as e.g. the crossvalidation procedure suggested in [10, Section 2.4].

## ACKNOWLEDGEMENT

The work was supported by Czech Science Foundation, project no. P201/10/0472, and the grant SVV 265 315.

(Received September 7, 2011)

## REFERENCES

---

- [1] A. J. Baddeley and R. Turner: Practical maximum pseudolikelihood for spatial point processes. *Aust. N. Z. J. Statist.* *42* (2000), 283–322.
- [2] A. J. Baddeley and R. Turner: Spatstat: an R package for analyzing spatial point patterns. *J. Statist. Softw.* *12* (2005), 1–42.
- [3] A. Brix: Generalized gamma measures and shot-noise Cox processes. *Adv. in Appl. Probab.* *31* (1999), 929–953.
- [4] D. R. Cox: Some statistical models related with series of events. *J. Roy. Statist. Soc. Ser. B* *17* (1955), 129–164.
- [5] D. J. Daley and D. Vere-Jones: *An Introduction to the Theory of Point Processes. Volume 1: Elementary Theory and Methods.* Second edition. Springer Verlag, New York 2003.

- [6] D. J. Daley and D. Vere-Jones: An Introduction to the Theory of Point Processes. Volume 2: General Theory and Structure. Second edition. Springer Verlag, New York 2008.
- [7] P. J. Diggle: Statistical Analysis of Spatial Point Patterns. Academic Press, London 1983.
- [8] P. J. Diggle: Statistical Analysis of Spatial Point Patterns. Second edition. Oxford University Press, New York 2003.
- [9] J. Dvořák and M. Prokešová: Moment estimation methods for stationary spatial Cox processes – a simulation study. Preprint (2011), <http://www.karlin.mff.cuni.cz/~dvorak/CoxEstimation/CoxEstimation.SimulationStudy.pdf>.
- [10] Y. Guan: A composite likelihood approach in fitting spatial point process models. J. Amer. Statist. Assoc. *101* (2006), 1502–1512.
- [11] Y. Guan and M. Sherman: On least squares fitting for stationary spatial point processes. J. Roy. Statist. Soc. Ser. B *69* (2007), 31–49.
- [12] L. Heinrich: Minimum contrast estimates for parameters of spatial ergodic point processes. In: Trans. 11th Prague Conference on Random Processes, Information Theory and Statistical Decision Functions. Academic Publishing House, Prague 1992.
- [13] G. Hellmund, M. Prokešová, and E. B. Vedel Jensen: Lévy-based Cox point processes. Adv. in Appl. Probab. *40* (2008), 603–629.
- [14] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan: Statistical Analysis and Modelling of Spatial Point Patterns. John Wiley and Sons, Chichester 2008.
- [15] A. T. Jensen: Statistical Inference for Doubly Stochastic Poisson Processes. Ph.D. Thesis, Department of Applied Mathematics and Statistics, University of Copenhagen 2005.
- [16] B. G. Lindsay: Composite likelihood methods. Contemp. Math. *80* (1988), 221–239.
- [17] B. Matérn: Doubly stochastic Poisson processes in the plane. In: Statistical Ecology. Volume 1. Pennsylvania State University Press, University Park 1971.
- [18] J. Møller, A. R. Syversveen, and R. P. Waagepetersen: Log-Gaussian Cox processes. Scand. J. Statist. *25* (1998), 451–482.
- [19] J. Møller: Shot noise Cox processes. Adv. in Appl. Probab. (SGSA) *35* (2003), 614–640.
- [20] J. Møller and R. P. Waagepetersen: Statistical Inference and Simulation for Spatial Point Processes. Chapman and Hall/CRC, Florida 2003.
- [21] J. Møller and R. P. Waagepetersen: Modern statistics for spatial point processes. Scand. J. Statist. *34* (2007), 643–684.
- [22] M. Prokešová and E. B. Vedel Jensen: Asymptotic Palm likelihood theory for stationary spatial point processes. Accepted to Ann. Inst. Statist. Math. (2012).
- [23] U. Tanaka, Y. Ogata, and D. Stoyan: Parameter estimation and model selection for Neyman–Scott point processes. Biometrical J. *49* (2007), 1–15.

*Jiří Dvořák, Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Praha 8. Czech Republic.*

*e-mail: [dvorak@karlin.mff.cuni.cz](mailto:dvorak@karlin.mff.cuni.cz)*

*Michaela Prokešová, Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Praha 8. Czech Republic.*

*e-mail: [prokesov@karlin.mff.cuni.cz](mailto:prokesov@karlin.mff.cuni.cz)*