Towards optimal Takacs–Fiksel estimation

Jean-François Coeurjolly, Yongtao Guan, Mahdieh Khanmohammadi, Rasmus Waagepetersen

Towards optimal Takacs–Fiksel estimation

Jean-François Coeurjolly[1], Yongtao Guan[2], Mahdieh Khanmohammadi[3,4] and Rasmus Waagepetersen[5]

[1]Laboratory Jean Kuntzmann, Grenoble Alpes University, France, Jean-Francois.Coeurjolly@upmf-grenoble.fr
[2]Department of Management Science, University of Miami, USA, yguan@bus.miami.edu
[3]Department of Computer Science, University of Copenhagen, Denmark
[4]Department of Electrical Engineering and Computer Science, University of Stavanger, Norway, mh.khanmohammadi@gmail.com
[5]Department of Mathematical Sciences, Aalborg University, Denmark, rw@math.aau.dk

December 21, 2015

Abstract

The Takacs–Fiksel method is a general approach to estimate the parameters of a spatial Gibbs point process. This method embraces standard procedures such as the pseudolikelihood and is defined via weight functions. In this paper we propose a general procedure to find weight functions which reduce the Godambe information and thus outperform pseudolikelihood in certain situations. The performance of the new procedure is investigated in a simulation study and it is applied to a standard dataset. Finally, we extend the procedure to handle replicated point patterns and apply it to a recent neuroscience dataset.

Keywords: Gibbs point processes; Godambe information; optimal estimation; pseudolikelihood; spatial point processes.

1 Introduction

Spatial Gibbs point processes are important models for spatial dependence in point patterns (van Lieshout, 2000) with a broad range of applications (e.g. Stoyan and Penttinen, 2000; Illian *et al.*, 2008). Such processes are specified by a density with respect to a Poisson point process or, equivalently, by the Papangelou conditional intensity. When the density or Papangelou conditional intensity has a parametric form, popular options for parameter estimation include maximum likelihood (e.g. Ogata and Tanemura, 1984; Møller and Waagepetersen, 2004), maximum pseudolikelihood (e.g. Besag, 1977; Jensen and Møller, 1991; Baddeley and Turner, 2000; Billiot *et al.*, 2008), maximum logistic regression likelihood (Baddeley *et al.*, 2014) and Takacs-Fiksel estimation (e.g. Fiksel, 1984; Takacs, 1986; Coeurjolly *et al.*, 2012).

Maximum likelihood estimation for a Gibbs point process requires computationally intensive estimation of an unknown normalizing constant in the density function. This explains why alternative estimation methods have been studied. Takacs-Fiksel estimation is an estimating function method based on the general Georgii-Nguyen-Zessin integral equation involving the Papangelou conditional intensity and a user-specified weight function. A particular choice of the weight function allows us to recover the score of the pseudolikelihood. Pseudolikelihood estimation has an intuitively appealing motivation and is by far the most popular estimation method in practical applications of Gibbs point processes with a user-friendly implementation in the `spatstat` package (Baddeley and Turner, 2005). Logistic regression likelihood estimation for Gibbs point processes was recently introduced to eliminate a bias problem coming from the Berman-Turner approximation of the pseudolikelihood (Berman and Turner, 1992; Baddeley and Turner, 2000). The logistic regression can be viewed as a computationally efficient approximation of the pseudolikelihood. Hence in the following, we may not differentiate between the pseudolikelihood and the logistic regression methods. It is not known whether pseudolikelihood is the optimal Takacs-Fiksel method in terms of minimizing parameter estimation variance. The various alternative weight functions for Takacs-Fiksel estimation considered in Diggle *et al.* (1994) and Coeurjolly *et al.* (2012) did for example not outperform pseudolikelihood.

In this paper our aim is to develop a systematic approach to construct a weight function that can lead to a more efficient estimation approach than existing methods. Our approach is motivated by the one considered in Guan *et al.* (2015) who considered estimation of the intensity function of a spatial point process and identified the optimal estimating function within a class of estimating functions based on the Campbell formula (e.g. Møller and Waagepetersen, 2004). Their optimal estimating function was derived from a sufficient condition equating the sensitivity matrix for the optimal estimating function and the covariance between the optimal estimating function and an arbitrary estimating function.

Extending the ideas in Guan *et al.* (2015) to Gibbs point processes is not straightforward. One problem is that covariances of Takacs-Fiksel estimating functions are not available in closed forms. For this reason our derived weight function only approximately satisfies the aforementioned sufficient condition. Nevertheless, we show in a simulation study that the new weight function may yield better estimation accuracy and is closer to fulfilling the sufficient condition than pseudolikelihood. Another issue is that the practical implementation of the new method is more computationally demanding than the pseudolikelihood, especially for point patterns of high cardinality. On the other hand, fine-tuning the estimation method is more a concern for small datasets.

The rest of the paper is organized as follows. Section 2 gives background on Gibbs point processes and Takacs-Fiksel estimation and presents our new methodology. We discuss implementation issues in Section 3. Section 4 contains a simulation study and applications to the standard Spanish towns dataset as well as a recent replicated point pattern dataset from neuroscience.

2 Background and methodology

2.1 Gibbs point processes

A point process $\mathbf{X}$ on $\Lambda \subseteq \mathbb{R}^d$ is a random subset of Λ that is locally finite, meaning that $\mathbf{X} \cap W$ is almost surely finite for every bounded $W \subseteq \Lambda$. When Λ is bounded, then $\mathbf{X}$ is a finite point process almost surely, taking values in Ω, the set of finite point configurations in Λ.

In this paper, we consider parametric inference for a finite Gibbs point process $\mathbf{X}$. The distribution of $\mathbf{X}$ is specified by a density $f(\cdot;\theta) : \Omega \to [0,\infty)$ with respect to the Poisson process of unit intensity. The density is of the form

$$f(\mathbf{y};\theta) \propto H(\mathbf{y})e^{V(\mathbf{y};\theta)}$$

where $\theta \in \Theta \subseteq \mathbb{R}^p$ is a p-dimensional parameter vector, $H : \Omega \to [0,\infty)$ serves as a baseline or reference factor, and $V : \Omega \to \mathbb{R}$ is often called the potential. Assuming that H is hereditary, i.e. $H(\mathbf{y} \cup u) > 0$ implies $H(\mathbf{y}) > 0$ for any $u \in \Lambda$ and $\mathbf{y} \in \Omega$, the Papangelou conditional intensity of $\mathbf{X}$ exists and is defined by

$$\lambda(u,\mathbf{y};\theta) = \frac{f(\mathbf{y} \cup u;\theta)}{f(\mathbf{y};\theta)} = H(u,\mathbf{y})e^{V(u,\mathbf{y};\theta)}$$

where $H(u,\mathbf{y}) = \mathbf{1}[H(\mathbf{y}) > 0]H(\mathbf{y} \cup u)/H(\mathbf{y})$ and $V(u,\mathbf{y};\theta) = V(\mathbf{y} \cup u;\theta) - V(\mathbf{y};\theta)$. Note that λ and f are in one-to-one correspondence. Hence the distribution of $\mathbf{X}$ can equivalently be specified in terms of the Papangelou conditional intensity. Intuitively, $\lambda(u,\mathbf{x};\theta)\,\mathrm{d}u$ is the conditional probability that a point of $\mathbf{X}$ occurs in a small neighbourhood B_u of volume $\mathrm{d}u$ around the location u, given $\mathbf{X}$ outside B_u is equal to $\mathbf{x}$; see Georgii (1976) for a general presentation and Coeurjolly *et al.* (2015) for links with Palm distributions. Gibbs point processes can also be characterized through the Georgii-Nguyen-Zessin formula (see Georgii, 1976; Nguyen and Zessin, 1979), which states that for any $h : \Lambda \times \Omega \to \mathbb{R}$ (such that the following expectations are finite)

$$\mathrm{E}\sum_{u \in \mathbf{X}} h(u, \mathbf{X} \setminus u) = \mathrm{E}\int_\Lambda h(u,\mathbf{X})\lambda(u,\mathbf{X};\theta)\,\mathrm{d}u. \tag{2.1}$$

Conditions ensuring the existence of Gibbs point processes, including the unbounded case $\Lambda = \mathbb{R}^d$, constitute a full research topic (see e.g. Dereudre *et al.*, 2012, and the references therein). We here restrict our attention to two specific examples. Let, for $R \geq 0$, $s_R(u,\mathbf{y})$ denote the number of R-close neighbours of u in $\mathbf{y}$. The Strauss model is then defined by $H(u,\mathbf{y}) = 1$ and $V(u,\mathbf{y};\theta) = \theta_1 + \theta_2 s_R(u,\mathbf{y})$ for $\theta_1 \in \mathbb{R}$ and $\theta_2 \leq 0$. For the Strauss hard core model, $V(u,\mathbf{y};\theta)$ is defined as for the Strauss model while $H(u,\mathbf{y})$ is 1 if all the points of $\mathbf{y} \cup u$ are separated by some hard core distance $\delta > 0$ and zero otherwise. The Strauss hard core model exists for all $\theta_1, \theta_2 \in \mathbb{R}$. Other examples can be found in Møller and Waagepetersen (2004); see also Section 4.3 for an example of an inhomogeneous model.

In this paper we mainly focus on the case of a bounded Λ and assume $W = \Lambda$. We further assume finite range, i.e. for any $u \in \Lambda$ and $\mathbf{y} \in \Omega$

$$\lambda(u,\mathbf{y};\theta) = \lambda(u,\mathbf{y} \cap B(u,R);\theta) \tag{2.2}$$

where $B(u, R)$ is the ball with center u and radius $0 < R < \infty$. Thus the conditional intensity of a point u given $\mathbf{y}$ only depends on the R-close neighbors in $\mathbf{y}$. Extensions to marked Gibbs point processes and/or Gibbs point processes of unbounded range are possible at the expense of more notation and technicalities. We avoid this to focus on the new statistical methodology that we propose.

2.2 Takacs-Fiksel estimation

The class of Takacs-Fiksel estimating functions (Fiksel, 1984; Takacs, 1986) consists of functions of the form

$$e_h(\theta) = \sum_{u \in \mathbf{X} \cap W} h(u, \mathbf{X} \setminus u; \theta) - \int_W h(u, \mathbf{X}; \theta) \lambda(u, \mathbf{X}; \theta) \, \mathrm{d}u \tag{2.3}$$

for weight functions $h : \Lambda \times \Omega \to \mathbb{R}^p$ parameterized by θ. By the Georgii-Nguyen-Zessin formula (2.1) these estimating functions are unbiased, i.e., $\mathrm{E}\, e_h(\theta) = 0$. An estimate obtained by solving $e_h(\theta) = 0$ with respect to θ is called a Takacs-Fiksel estimate.

When $h(u, \mathbf{y}; \theta) = \mathrm{d} \log \lambda(u, \mathbf{y}; \theta) / \mathrm{d}\theta$, (2.3) is the score function for the log pseudolikelihood function. The corresponding estimate can be obtained using standard statistical software and its statistical properties have been deeply studied in the literature (e.g. Jensen and Møller, 1991; Mase, 1999; Jensen and Künsch, 1994; Billiot *et al.*, 2008; Baddeley *et al.*, 2014). However, the pseudolikelihood score has not been shown to be optimal within the class of Takacs-Fiksel estimating functions. In the following section our aim is to construct a competitor to the pseudolikelihood in terms of statistical efficiency.

2.3 Towards optimality

For an estimating function e_h, two important quantities are the sensitivity matrix $S_h = -\mathrm{E}\left(\frac{\mathrm{d}}{\mathrm{d}\theta^\top} e_h(\theta)\right)$ and the covariance matrix $\Sigma_h = \mathrm{Var}(e_h(\theta))$ of the estimating function. From these the Godambe information matrix is obtained as

$$G_h = S_h^\top \Sigma_h^{-1} S_h.$$

In applications of estimating functions, the inverse Godambe matrix provides the approximate covariance matrix of the associated parameter estimate. An estimating function e_ϕ is said to be Godambe optimal in a class of estimating functions e_h indexed by a set C of functions h, if the difference $G_\phi - G_h$ is non-negative definite for all $h \in C$. Following Guan *et al.* (2015), a sufficient condition for e_ϕ to be optimal is that for every estimating function e_h, $h \in C$,

$$\mathrm{Cov}(e_h(\theta), e_\phi(\theta)) = S_h. \tag{2.4}$$

In the context of Takacs-Fiksel estimating functions (2.3),

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\theta^\top} e_h(\theta) = \sum_{u \in \mathbf{X} \cap W} \frac{\mathrm{d}}{\mathrm{d}\theta^\top} h(u, \mathbf{X} \setminus u; \theta) - \int_W \frac{\mathrm{d}}{\mathrm{d}\theta^\top} h(u, \mathbf{X}; \theta) \lambda(u, \mathbf{X}; \theta) \, \mathrm{d}u \\ - \int_W h(u, \mathbf{X}; \theta) \frac{\mathrm{d}}{\mathrm{d}\theta^\top} \lambda(u, \mathbf{X}; \theta) \, \mathrm{d}u. \end{aligned}$$

So by the Georgii-Nguyen-Zessin formula (2.1),

$$S_h = \mathrm{E} \int_W h(u, \mathbf{X}; \theta) \frac{\mathrm{d}}{\mathrm{d}\theta^\top} \lambda(u, \mathbf{X}; \theta) \, \mathrm{d}u. \tag{2.5}$$

The definition of the estimating function e_h actually corresponds to the concept of innovations for spatial point processes (Baddeley *et al.*, 2005). Coeurjolly and Rubak (2013) investigated the problem of estimating the covariance between two innovations which here corresponds to the covariance between two estimating functions. Assuming the right hand side below is finite, Coeurjolly and Rubak (2013, Lemma 3.1) established that

$$\begin{aligned}\mathrm{Cov}(e_h(\theta), e_g(\theta)) = \mathrm{E} \Big[& \int_W h(u, \mathbf{X}; \theta) g(u, \mathbf{X}; \theta)^\top \lambda(u, \mathbf{X}; \theta) \, \mathrm{d}u \\ & + \int_W \int_W h(u, \mathbf{X}; \theta) g(u, \mathbf{X}; \theta)^\top \left(\lambda(u, \mathbf{X}; \theta) \lambda(v, \mathbf{X}; \theta) - \lambda(\{u, v\}, \mathbf{X}; \theta) \right) \mathrm{d}u \, \mathrm{d}v \\ & + \int_W \int_W \Delta_v h(u, \mathbf{X}; \theta) \Delta_u g(v, \mathbf{X}; \theta)^\top \lambda(\{u, v\}, \mathbf{X}; \theta) \, \mathrm{d}u \, \mathrm{d}v \Big]\end{aligned} \tag{2.6}$$

where for any $u, v \in W$ and any $\mathbf{y} \in \Omega$, the second order Papangelou conditional intensity $\lambda(\{u, v\}, \mathbf{y}; \theta)$ and the difference operator $\Delta_u h(v, \mathbf{y}; \theta)$ are given by

$$\begin{aligned}\lambda(\{u, v\}, \mathbf{y}; \theta) &= \lambda(u, \mathbf{y}) \lambda(v, \mathbf{y} \cup u; \theta) = \lambda(v, \mathbf{y}; \theta) \lambda(u, \mathbf{y} \cup v; \theta) \\ \Delta_u h(v, \mathbf{y}; \theta) &= h(v, \mathbf{y} \cup u; \theta) - h(v, \mathbf{y}; \theta).\end{aligned}$$

Returning to the condition (2.4), we introduce for any $\mathbf{y} \in \Omega$ the operator $T_{\mathbf{y}}$ acting on $\mathbb{R}^p$ valued functions g,

$$T_{\mathbf{y}} g(u) = \int_W g(v) t(u, v, \mathbf{y}; \theta) \, \mathrm{d}v, \tag{2.7}$$

where

$$t(u, v, \mathbf{y}; \theta) = \lambda(v, \mathbf{y}; \theta) \left(1 - \frac{\lambda(v, \mathbf{y} \cup u; \theta)}{\lambda(v, \mathbf{y}; \theta)} \right).$$

The finite range property of the Papangelou conditional intensity implies that for any $v \notin B(u, R)$, $t(u, v, \mathbf{y}; \theta) = 0$. So the domain of integration in (2.7) is actually just $W \cap B(u, R)$. From (2.5)-(2.6), (2.4) is equivalent to $\mathrm{E}(A) + \mathrm{E}(B) = 0$ where

$$A = \int_W h(u, \mathbf{X}; \theta) \lambda(u, \mathbf{X}; \theta) \left\{ \phi(u, \mathbf{X}; \theta) - \frac{\lambda^{(1)}(u, \mathbf{X}; \theta)}{\lambda(u, \mathbf{X}; \theta)} + T_{\mathbf{X}} \phi(u, \mathbf{X}; \theta) \right\}^\top \mathrm{d}u \tag{2.8}$$

$$B = \int_W \int_W \Delta_v h(u, \mathbf{X}; \theta) \Delta_u \phi(v, \mathbf{X}; \theta)^\top \lambda(\{u, v\}, \mathbf{X}; \theta) \, \mathrm{d}u \, \mathrm{d}v \tag{2.9}$$

and $\lambda^{(1)}(u, \mathbf{X}; \theta) = \mathrm{d}\lambda(u, \mathbf{X}; \theta) / \mathrm{d}\theta$. The expectation $\mathrm{E}(B)$ is very difficult to evaluate. Moreover, in the context of asymptotic covariance matrix estimation for the pseudolikelihood, Coeurjolly and Rubak (2013) remarked that the contribution of the term (2.9) to the covariance $\mathrm{Cov}(e_h(\theta), e_\phi(\theta))$ was negligible. In the following

we will neglect the term (2.9) and call 'semi-optimal' a function $\phi : W \times \Omega \to \mathbb{R}^p$ (parameterized by θ) such that for any $h : W \times \Omega \to \mathbb{R}^p$, $\mathrm{E}(A) = 0$. This holds if for any $\mathbf{y} \in \Omega$, $\phi(\cdot, \mathbf{y}; \theta)$ is the solution to the Fredholm integral equation (e.g. chapter 3 in Hackbusch, 1995)

$$\phi(\cdot, \mathbf{y}; \theta) + T_{\mathbf{y}}\phi(\cdot, \mathbf{y}; \theta) = \frac{\lambda^{(1)}(\cdot, \mathbf{y}; \theta)}{\lambda(\cdot, \mathbf{y}; \theta)} \tag{2.10}$$

Compared to Guan *et al.* (2015) the problem of optimal Takacs-Fiksel estimation poses additional challenges. One is the term B which does not appear in the context of optimal estimation of the intensity function. Another is that ϕ is a function of both $u \in W$ and $\mathbf{y} \in \Omega$ which implies additional computational complexity as detailed in the following section.

Having solved (2.10), the covariance and sensitivity matrices for the resulting estimating function

$$e_\phi(\theta) = \sum_{u \in \mathbf{X} \cap W} \phi(u, \mathbf{X} \setminus u; \theta) - \int_W \phi(u, \mathbf{X}; \theta)\lambda(u, \mathbf{X}; \theta)\,\mathrm{d}u \tag{2.11}$$

are given by

$$\begin{aligned} S &= \mathrm{E}\int_W \phi(u, \mathbf{X}; \theta)\lambda^{(1)}(u, \mathbf{X}; \theta)^\top\,\mathrm{d}u \\ \Sigma &= S + \mathrm{E}\int_W \int_W \Delta_u\phi(v, \mathbf{X}; \theta)\Delta_v\phi(u, \mathbf{X}; \theta)^\top \lambda(\{u, v\}, \mathbf{X}; \theta)\,\mathrm{d}u\,\mathrm{d}v. \end{aligned}$$

Note that for a truly optimal ϕ, we would have $S = \Sigma$. In the simulation studies in Section 4.1 we assess how close S and Σ are for our semi-optimal ϕ.

3 Implementation

Let $\mathbf{x} = \{x_1, \ldots, x_n\}$ denote a realization of $\mathbf{X}$. To solve $e_\phi(\theta) = 0$, we use Newton-Raphson iterations starting at the pseudolikelihood estimate with the Hessian matrix estimated by the empirical sensitivity $\hat{S} = \int_W \phi(u, \mathbf{x}; \theta)\lambda^{(1)}(u, \mathbf{x}; \theta)^\top\,\mathrm{d}u$. To evaluate e_ϕ and $\hat{S}$ we need to solve (2.10) with respect to $\phi(\cdot; \mathbf{y}; \theta)$ for all $\mathbf{y} = \mathbf{x}, \mathbf{x} \setminus x_1, \ldots, \mathbf{x} \setminus x_n$.

3.1 Symmetrization

To ease the implementation and in particular the use of Cholesky decompositions, we symmetrize the operator $T_{\mathbf{y}}$. This is possible if we assume for any $u, v \in W$ and $\mathbf{y} \in \Omega$, the ratio $\lambda(v, \mathbf{y} \cup u; \theta)/\lambda(v, \mathbf{y}; \theta)$ is symmetric in u and v. This assumption is valid for instance for all pairwise interaction point processes. Indeed, the Papangelou conditional intensity of such processes is given by $\lambda(u, \mathbf{y}; \theta) = e^{\sum_{w \in \mathbf{y}} \psi(\{w,u\}; \theta)}$ where ψ is a real valued function, whereby $\lambda(v, \mathbf{y} \cup u; \theta)/\lambda(v, \mathbf{y}; \theta) = e^{\psi(\{v,u\}; \theta)}$.

We now multiply each term of (2.10) by $\sqrt{\lambda(\cdot, \mathbf{y}; \theta)}$ and reformulate the problem to solve

$$\widetilde{\phi}(\cdot, \mathbf{y}; \theta) + \widetilde{T}_{\mathbf{y}}\widetilde{\phi}(\cdot, \mathbf{y}; \theta) = \frac{\lambda^{(1)}(\cdot, \mathbf{y}; \theta)}{\sqrt{\lambda(\cdot, \mathbf{y}; \theta)}} \tag{3.1}$$

with respect to the function $\widetilde{\phi}(\cdot, \mathbf{y}; \theta) = \sqrt{\lambda(\cdot, \mathbf{y}; \theta)}\phi(\cdot, \mathbf{y}; \theta)$ where $\widetilde{T}_{\mathbf{y}}$ is the operator with kernel

$$\widetilde{t}(u, v, \mathbf{y}; \theta) = \sqrt{\lambda(u, \mathbf{y}; \theta)\lambda(v, \mathbf{y}; \theta)} \left(1 - \frac{\lambda(v, \mathbf{y} \cup u; \theta)}{\lambda(v, \mathbf{y}; \theta)}\right).$$

Once we have obtained the function $\widetilde{\phi}$, we obtain the semi-optimal function ϕ by $\phi(u, \mathbf{y}; \theta) = \widetilde{\phi}(u, \mathbf{y}; \theta)/\sqrt{\lambda(u, \mathbf{y}; \theta)}$.

3.2 Numerical solution using Nyström approximation

The equation (3.1) is solved numerically using the Nyström approximation (Nyström, 1930). We introduce a quadrature scheme with m quadrature points $u_1, \ldots, u_m \in W$ and associated weights w_j, $j = 1, \ldots, m$, and approximate the operator $T_{\mathbf{y}}$ for any $\mathbb{R}^p$ valued function g by

$$T_{\mathbf{y}} g(u) \approx \sum_{j=1}^{m} g(u_j) t(u, u_j, \mathbf{y}; \theta) w_j.$$

Introducing the quadrature approximation in (3.1) and multiplying each term by $\sqrt{w_i}$ we obtain $\sqrt{w_i}\widetilde{\phi}(u_i, \mathbf{y}; \theta)$, $i = 1, \ldots, m$, as solutions of the linear equations

$$\sqrt{w_i}\widetilde{\phi}(u_i, \mathbf{y}; \theta) + \sum_{j=1}^{m} \sqrt{w_i w_j}\widetilde{t}(u_i, u_j, \mathbf{y}; \theta)\sqrt{w_j}\widetilde{\phi}(u_j, \mathbf{y}; \theta) = \sqrt{w_i}\left(\frac{\lambda^{(1)}(u_i, \mathbf{y}; \theta)}{\sqrt{\lambda(u_i, \mathbf{y}; \theta)}}\right),$$

for $i = 1, \ldots, m$. These equations can be reformulated as the matrix equation

$$\left(\mathbf{I}_m + \widetilde{\mathbf{T}}(\mathbf{y}; \theta)\right) \sqrt{w}\widetilde{\phi}(\mathbf{y}; \theta) = \ell(\mathbf{y}; \theta) \tag{3.2}$$

where $\mathbf{I}_m$ is the $m \times m$ identity matrix, $\widetilde{\mathbf{T}}(\mathbf{y}; \theta) = \left[\sqrt{w_i w_j}\widetilde{t}(u_i, u_j, \mathbf{y}; \theta)\right]_{ij}$, $i, j = 1, \ldots, m$, $\ell(\mathbf{y}; \theta)$ is the $m \times p$ matrix with rows $\sqrt{w_i}\lambda^{(1)}(u_i, \mathbf{y}; \theta)^\top/\sqrt{\lambda(u_i, \mathbf{y}; \theta)}$, $i = 1, \ldots, m$, and $\sqrt{w}\widetilde{\phi}(\mathbf{y}; \theta)$ is the $m \times p$ matrix with rows $\sqrt{w_i}\widetilde{\phi}(u_i, \mathbf{y}; \theta)^\top$. The symmetric matrix $\widetilde{\mathbf{T}}(\mathbf{y}; \theta)$ is sparse due to the finite range property. Thus, provided that $\mathbf{I}_m + \widetilde{\mathbf{T}}(\mathbf{y}; \theta)$ is positive definite, the matrix equation can be solved with respect to $\sqrt{w}\widetilde{\phi}(\mathbf{y}; \theta)$ using sparse Cholesky factorization (see Davis, 2006, and the `R` package `Matrix`).

Having solved (3.2) with respect to $\sqrt{w}\widetilde{\phi}(\mathbf{y}; \theta)$, and thus obtaining estimates of $\sqrt{w_i}\widetilde{\phi}(u_i, \mathbf{y}; \theta)$, we obtain estimates $\hat{\phi}(u_i, \mathbf{y}; \theta)$ of $\phi(u_i, \mathbf{y}; \theta)$ via the relation $\hat{\phi}(u_i, \mathbf{y}; \theta) = \widetilde{\phi}(u_i, \mathbf{y}; \theta)/\sqrt{w_i\lambda(u_i, \mathbf{y}; \theta)}$. Letting $\hat{\phi}(\mathbf{y}; \theta)$ be the $m \times p$ matrix with rows $\hat{\phi}(u_j, \mathbf{y}; \theta)^\top$ the Nyström approximation of $\phi(u, \mathbf{y}; \theta)$ for any $u \in W$ is

$$\hat{\phi}(u, \mathbf{y}; \theta) \approx \frac{\lambda^{(1)}(u, \mathbf{y}; \theta)}{\lambda(u, \mathbf{y}; \theta)} - \hat{\phi}(\mathbf{y}; \theta)^\top (w_j t(u, u_j, \mathbf{y}; \theta))_{j=1}^{m}.$$

In particular we obtain the approximations $\hat{\phi}(u, \mathbf{x} \setminus u; \theta)$ of $\phi(u, \mathbf{x} \setminus u; \theta)$, $u \in \mathbf{x}$, which are needed to evaluate the first term in (2.11). Finally, the integral term in (2.11) and the empirical sensitivity are approximated by

$$\hat{\phi}(\mathbf{x}; \theta)^\top (w_j \lambda(u_j, \mathbf{x}; \theta))_{j=1}^{m} \text{ and } \hat{\phi}(\mathbf{x}; \theta)^\top w\lambda^{(1)}(\mathbf{x}; \theta)$$

where $w\lambda^{(1)}(\mathbf{x}; \theta)$ is the $m \times p$ matrix with rows $w_j\lambda^{(1)}(u_j, \mathbf{x}; \theta)^\top$.

3.3 Some computational considerations

The matrix $\mathbf{I}_m + \widetilde{\mathbf{T}}(\mathbf{y};\theta)$ is not guaranteed to be positive definite. In case of purely repulsive point processes (Papangelou conditional intensity always decreasing when neighbouring points are added) all entries in $\widetilde{\mathbf{T}}(\mathbf{y};\theta)$ are positive and we did not experience negative definite $\mathbf{I}_m + \widetilde{\mathbf{T}}(\mathbf{y};\theta)$. However, with models allowing for positive interaction we occasionally experienced negative definiteness in which case a solution for $\phi(\cdot,\mathbf{y};\theta)$ cannot be obtained. In such case we simply returned the pseudolikelihood estimate.

In case of a quadrature scheme corresponding to a subdivision of W into square cells of sidelength s the computational complexity of one Newton-Raphson update is roughly of the order $(n+1)m(R/s)^2$. Thus, the semi-optimal approach is less feasible for data with a high number n of points. On the other hand, fine-tuning the estimation method is probably less a concern in data rich situations.

In case of e.g. the Strauss hard core process we may encounter $\lambda(u_j,\mathbf{y};\theta) = 0$. In this case we use the conventions $\lambda^{(1)}(u_j,\mathbf{y};\theta)/\sqrt{\lambda(u_j,\mathbf{y};\theta)} = 0$ and $\lambda(u_j,\mathbf{y}\cup u_j;\theta)/\lambda(u_j,\mathbf{y};\theta) = 0$.

4 Numerical experiments and applications

In this section we consider first a simulation study for the Strauss process followed by two data examples: the classical Spanish towns data and a recent dataset from neuroscience.

4.1 Simulation study

The performance of our semi-optimal estimating function relative to the pseudolikelihood score is studied by applying both estimating functions to simulations of a Strauss process (Section 2.1) on the unit square. We use the `spatstat` (Baddeley and Turner, 2005) procedure `rStrauss()` to generate exact simulations of the Strauss process for $\beta = \exp(\theta_1) = 100$ and all combinations of $R = 0.04, 0.08, 0.12$ and $\gamma = \exp(\theta_2) = 0.1, 0.2, 0.4, 0.8$.

The semi-optimal estimating function is implemented using a 50×50 or a 75×75 grid. For the pseudolikelihood we use the unbiased logistic likelihood implementation introduced in Baddeley *et al.* (2014) with a stratified quadrature point process on the same grids. For both estimation methods, R is assumed known and equal to the value used to generate the simulations. For the parameters θ_1 and θ_2, Table 1 shows for each parameter setting, the root mean square error (RMSE) of the pseudolikelihood estimates minus the RMSE of the semi-optimal estimates relative to the RMSE of the pseudolikelihood estimates. For each parameter setting, the RMSEs are estimated from a sample of 1000 parameter estimates obtained from 1000 simulations. Since the parameter estimates are close to being unbiased we obtain essentially the same results by replacing RMSE with the standard error of the simulated estimates. We omit simulations where all interpoint distances are larger than R and thus neither the pseudolikelihood nor the semi-optimal estimates exist. The reported relative differences in RMSE are subject to Monte Carlo error. Table 1 therefore also shows

estimated standard errors for these obtained by applying a bootstrap to each Monte Carlo sample.

	Interaction parameter $\gamma = \exp(\theta_2)$							
Range	0.1		0.2		0.4		0.8	
and grid	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
$R = 0.04$								
(50,50)	-1 (0.5)	1 (0.3)	-1 (0.4)	0 (0.2)	-2 (0.3)	0 (0.3)	-2 (0.2)	0 (0.3)
(75,75)	-1 (0.4)	0 (0.2)	-1 (0.3)	0 (0.2)	-1 (0.3)	0 (0.2)	-1 (0.2)	0 (0.2)
$R = 0.08$								
(50,50)	3 (1)	1 (0.6)	0 (0.9)	0 (0.6)	5 (0.8)	2 (0.7)	4 (0.5)	2 (0.7)
(75,75)	2 (0.8)	2 (0.5)	3 (0.9)	2 (0.6)	2 (0.8)	1 (0.7)	1 (0.5)	1 (0.6)
$R = 0.12$								
(50,50)	3 (1.2)	5 (0.8)	4 (1.1)	3 (1)	6 (1.3)	4 (1.3)	5 (1)	4 (1.2)
(75,75)	3 (1.3)	5 (0.9)	7 (1.2)	4 (1)	7 (1.3)	5 (1.2)	7 (1.1)	6 (1.4)

Table 1: RMSE for pseudolikelihood minus RMSE for semi-optimal relative to RMSE for pseudolikelihood (in percent) in case of estimation of θ_1 and θ_2. Grids of 50×50 or 75×75 quadrature points are considered. Numbers between brackets are bootstrap standard errors (in percent).

In case of $R = 0.04$, there is no efficiency improvement by using the semi-optimal estimating function. In fact the semi-optimal approach appears to be sometimes slightly worse than pseudolikelihood, even when taking into account the Monte Carlo error of the estimated relative differences in RMSE. However, for $R = 0.08$ and $R = 0.12$ the semi-optimal approach is always better with decreases up to 5 and 7% in RMSE for semi-optimal relative to RMSE for pseudolikelihood. The results are fairly similar for the two choices of grids with consistently slightly better results for the 75×75 grid in case of $R = 0.12$.

For an optimal estimating function and at the true parameter value, the covariance Σ of the estimating function coincides with the sensitivity matrix S. Table 2 shows the estimated relative differences $(S_{ij} - \Sigma_{ij})/\Sigma_{ij}$ in percent for the pseudolikelihood and semi-optimal estimating functions. For neither of the estimating functions,

	Interaction parameter $\gamma = \exp(\theta_2)$											
Interaction range	0.1			0.2			0.4			0.8		
$R = 0.04$	3	-17	-48	5	-5	-49	2	15	-44	3	-4	-46
0.08	-5	-31	-50	-2	-16	-44	3	-7	-40	1	1	-33
012	-10	-38	-45	-10	-36	-49	-7	-23	-41	-5	-10	-26
$R = 0.04$	41	-4	-46	40	11	-47	27	34	-41	12	4	-43
0.08	87	-4	-46	80	20	-35	71	35	-27	30	27	-20
0.12	119	-3	-37	102	6	-35	87	31	-18	41	32	0

Table 2: Relative difference (in percent) between sensitivity and variance of estimating function, $(S_{ij} - \Sigma_{ij})/\Sigma_{ij}$, $ij = 11, 12, 22$. Upper three rows: semi-optimal. Lower three rows: pseudolikelihood. A grid of 75×75 quadrature points is used.

the covariance and sensitivity agree. However, in general the relative deviations are larger for pseudolikelihood than for semi-optimal.

4.2 Application to the Spanish towns dataset

In this section, we consider the Spanish towns dataset (see Figure 1) which Ripley (1988) and then Illian *et al.* (2008) proposed to model using a Strauss hard core model. We compare results regarding estimation of θ_1 and θ_2 using respectively the

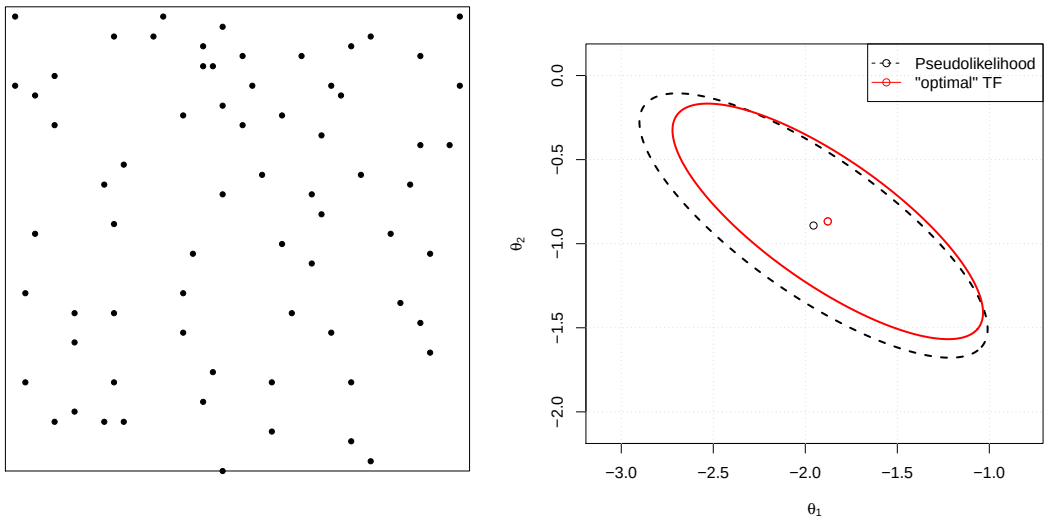

Figure 1: Left: locations of 69 Spanish towns in a 40 mile by 40 mile region. Right: 95% confidence ellipses for the parameters (θ_1, θ_2) for the pseudolikelihood and the semi-optimal Takacs-Fiksel methods.

new method and the pseudolikelihood. For the hard core distance and the interaction range we use the values $\widehat{\delta} = 0.83$ and $\widehat{R} = 3.5$ obtained by Illian *et al.* (2008). As in the previous section, for the pseudolikelihood we use the unbiased logistic likelihood implementation with a stratified quadrature point process on a 50×50 grid. The same grid is used for the semi-optimal Takacs-Fiksel estimator.

Pseudolikelihood and semi-optimal estimates are presented in Table 3. Standard

	θ_1	θ_2
SO	-1.88 (0.12)	-0.87 (0.08)
PL	-1.96 (0.15)	-0.89 (0.10)
Ratio se	0.79	0.79

Table 3: Results for Strauss model applied to the Spanish towns dataset. First and second row: semi-optimal (SO) and pseudolikelihood (PL) estimates with estimated standard errors in parantheses. Last row: ratio of semi-optimal standard errors to pseudolikelihood standard errors.

errors of the pseudolikelihood estimates (respectively the semi-optimal Takacs-Fiksel estimates) are estimated from 500 simulations of the model fitted using the pseudolikelihood (respectively the semi-optimal procedure). The standard errors are clearly reduced with the semi-optimal Takacs-Fiksel method. In addition to these results, we compute the Frobenius norm of the estimated covariance matrices for the pseudolikelihood and optimal Takacs-Fiksel estimates (which actually corresponds to estimate of the inverse Godambe matrices). We obtained approximately the values 0.25 and 0.20 for the pseudolikelihood and the semi-optimal Takacs-Fiksel methods respectively. Based on the asymptotic normality results established by Baddeley *et al.* (2014) and Coeurjolly *et al.* (2012), we construct 95% confidence ellipses for the parameter vector (θ_1, θ_2). The ellipses are depicted in the right plot of Figure 1. The area of the confidence region for the semi-optimal Takacs-Fiksel method is 81% of the one for the pseudolikelihood. In line with the simulation study, these empirical findings show that more precise parameter estimates can be obtained with the semi-optimal Takacs-Fiksel method compared to pseudolikelihood.

4.3 Application to synaptic vesicles

Synapses are regions in the brains where nerve impulses are transmitted or received. Inside the synapses, neurotransmitters are carried by small membrane-bound compartments called synaptic vesicles. Recently Khanmohammadi *et al.* (2014) studied whether stress affects the spatial distribution of vesicles within the synapse. The data used for the study originated from microscopial images of slices of brains from respectively a group of 6 control rats and a group of 6 stressed rats. The images were annotated to identify the boundaries of the synapses and possible mitochondria in the synapses, the location of the vesicles, and the extents of the so-called active zones where the vesicles release their contents of neurotransmitters. Figure 2 shows the annotations of two synapses. In total 7 synapses from the control rats and 6 synapses from the stressed rats were annotated. For each synapse several images corresponding to several slices of the synapse were annotated. We restrict here attention to the middle slice for each synapse. Thus, our data consist of 7 and 6 annotated images for respectively the control and the stressed rats. The side lengths of enclosing rectangles for the synapses range between 395.6 and 1009.0nm with a mean of 663.0nm. Further details on the dataset can be found in Khanmohammadi *et al.* (2014).

4.3.1 Point process models for locations of vesicles

For the ith image of type $t = C, S$ (control or stressed) we consider the centers of the vesicles as a realization of a finite spatial point process $\mathbf{X}_{ti}$ with observation window W_{ti} defined by the boundary of the synapse excluding areas occupied by mitochondria. We further assume that the pairs $(W_{ti}, \mathbf{X}_{ti})$ of the observation windows and the spatial point processes for different images are independent, that the W_{ti} are all identically distributed and that the $\mathbf{X}_{ti}$ of the same type t are identically distributed. Khanmohammadi *et al.* (2014) modelled the locations of the vesicles as an inhomogeneous Strauss hard core process, but noted some evidence of aggregation of vesicles at a larger scale not accommodated by this model. After considering

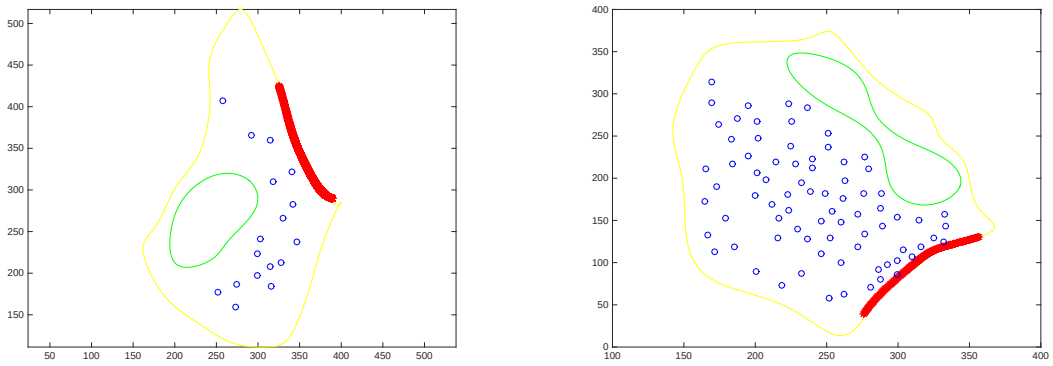

Figure 2: Plots of vesicle locations for a control (left) and a stressed (right) synapse. The active zone is shown by the red curve while the green curves show the boundary of mitochondria.

the Strauss hard core model for reference, we therefore extend it to a multiscale model with an additional interaction term. More precisely, for a location u and a configuration $\mathbf{x}$ of vesicle locations in a synapse of type t, the conditional intensity is of the form

$$\lambda_t(u, \mathbf{x}; \theta) = \exp[\theta_{0t} + \theta_{1t} d(u) + \theta_{2t} s_r(u, \mathbf{x}) + \theta_{3t} s_{r,R}(u, \mathbf{x})] H_\delta(u, \mathbf{x}) \tag{4.1}$$

where $t = C, S$, $0 \leq \delta < r < R$, $\theta_{0t}, \theta_{1t}, \theta_{2t}, \theta_{3t} \in \mathbb{R}$, $d(u)$ is distance to the active zone, $s_r(u, \mathbf{x})$ is the number of points in $\mathbf{x}$ with distance to u smaller than r, $s_{r,R}(u, \mathbf{x})$ denotes the number of points in $\mathbf{x}$ with distance to u in the interval $[r, R]$, and the hard core term $H_\delta(u, \mathbf{x})$ is 1 if all points in $\mathbf{x} \cup u$ are separated by a distance greater than δ and zero otherwise. The inhomogeneous Strauss hard core model is the special case with $\theta_{3t} = 0$. As in Khanmohammadi *et al.* (2014) the hard core distance h is set to 17.5nm corresponding to the average diameter of a vesicle and the interaction distance r is set to 32.5nm. We further choose the value of $R = 107.5$nm to maximize a profile pseudolikelihood based on the data for all 13 images. We finally scale the distances $d(\cdot)$ by a factor 10^{-3} in order to obtain θ_{lt} estimates of the same order of magnitude.

4.3.2 Inference for replicated point patterns

Let e^{ti} denote an estimating function (either the pseudolikelihood score or the semi-optimal) for the tith image and denote by θ_t the vector of parameters to be inferred, $t = C, S$. Then for each group we optimally form a pooled estimating function e^t as the sum of the e^{ti}'s. Following standard asymptotic arguments for pooled independent estimating functions, the variance of the corresponding parameter estimate is approximated as

$$\operatorname{Var} \sqrt{n_t}(\hat{\theta}_t - \theta_t) \approx (S^{t1})^{-1} \operatorname{Var} e^{t1} (S^{t1})^{-1}$$

(the so-called sandwich estimator, Song, 2007) where n_t is the number of replicates of type t and S^{t1} is the sensitivy matrix associated with e^{t1}. In practice we replace

	θ_{0C}	θ_{1C}	θ_{2C}
SO	-6.74 (0.21)	-0.68 (0.20)	-2.79 (1.00)
PL	-6.80 (0.21)	-0.66 (0.22)	-2.61 (1.00)
Ratio se	1.01	0.80	1.02
	θ_{0S}	θ_{1S}	θ_{2S}
SO	-5.51 (0.23)	-1.14 (0.47)	-0.61 (0.21)
PL	-5.63 (0.24)	-1.12 (0.49)	-0.45 (0.24)
Ratio se	0.93	0.91	0.78

Table 4: Results for Strauss hard core process. Semi-optimal (SO) and pseudolikelihood (PL) estimates with estimated standard errors in parantheses and ratios of semi-optimal standard errors to pseudolikelihood standard errors. First three rows: control. Last three rows: stressed.

S^{t1} and $\operatorname{Var} e^{t1}$ by their empirical estimates replacing the unknown θ_t with its estimate. We conduct the pseudolikelihood estimation for the replicated data using the user-friendly `mppm` procedure in the `R` package `spatstat` while using our own code to evaluate the approximate variances of the pseudolikelihood or semi-optimal estimates.

4.3.3 Results for synapse data

Table 4 shows parameter estimates and associated standard errors obtained for the Strauss hard core model with either the pseudolikelihood or the semi-optimal approach. Except for θ_{2S}, the semi-optimal and pseudolikelihood estimates are fairly similar. The qualitative conclusions based on the two types of estimates are identical: negative dependence of conditional intensity on distance and strong repulsion between vesicles both in the control and the stressed group. The estimated semi-optimal standard errors are smallest for all parameters in the stressed group and θ_{1C} in the control groups. For θ_{0C} and θ_{2C} the estimated semi-optimal and pseudolikelihood standard errors are very similar.

Due to the aforementioned evidence of large scale aggregation we turn to the multiscale model for a more detailed comparison of the control and the stressed group. Parameter estimates and associated standard errors for the multiscale model are shown in Table 5. The qualitative conclusions based on the pseudolikelihood and the semi-optimal estimates coincide. All parameters are significantly different from zero (assuming estimate divided by its standard error is approximately $N(0,1)$). In particular the positive estimates of θ_{3C} and θ_{3S} confirm that there is aggregation at a larger scale. The negative estimates of θ_{1t}, $t = C, S$ indicate that the conditional intensity is decreasing as a function of distance to the active zone while there appears to be a strong repelling interaction between vesicles according to the estimates of θ_{2t}.

To test the hypotheses: H_l: $\theta_{lS} = \theta_{lC}$, $l = 0, 1, 2, 3$, we consider statistics of the form $(\hat{\theta}_{lS} - \hat{\theta}_{lC})/\sqrt{\mathrm{se}_{lS}^2 + \mathrm{se}_{lC}^2}$ where se_{1t}^2 denotes the estimated standard error of the estimate $\hat{\theta}_{lt}$ $t = S, C$ and the estimates are obtained either using the semi-optimal approach or pseudolikelihood. Under the hypothesis this statistic is approximately $N(0,1)$. According to these tests, H_0 and H_2 and H_3 are rejected while H_1 is not

	θ_{0C}	θ_{1C}	θ_{2C}	θ_{3C}
SO	-8.87 (0.12)	-0.47 (0.11)	-3.40 (0.78)	0.45 (0.02)
PL	-8.71 (0.13)	-0.46 (0.12)	-3.35 (0.80)	0.40 (0.02)
Ratio se	0.94	0.93	0.98	1.01
	θ_{0S}	θ_{1S}	θ_{2S}	θ_{3S}
SO	-7.87 (0.21)	-0.48 (0.26)	-1.17 (0.18)	0.26 (0.02)
PL	-7.84 (0.19)	-0.69 (0.27)	-1.12 (0.18)	0.24 (0.02)
Ratio se	1.12	0.96	0.97	1.07

Table 5: Results for multiscale process. Semi-optimal (SO) and pseudolikelihood (PL) estimates with estimated standard errors in parantheses and ratios of semi-optimal standard errors to pseudolikelihood standard errors. First three rows: control. Last three rows: stressed.

irrespective of the estimation method (p-values < 0.0001, 0.96, 0.005, < 0.0001 and < 0.0001, 0.44, 0.006 and < 0.0001 for semi-optimal and pseudolikelihood, respectively). There is thus evidence that the negative interaction between vesicles is stronger for the control rats than for the stressed rats which means that the vesicles tend to form more regular patterns for the control rats.

For the control data, the estimated standard errors are smallest with the semi-optimal approach except for θ_{3C}. For the stressed rats the semi-optimal standard errors are smallest for θ_{1S} and θ_{2S} but not for θ_{0S} and θ_{3S}. Overall, a clear pattern is not visible. Note also that the ratios of estimated standard errors should be interpreted with care as they are obviously subject to sampling error.

5 Discussion

In this paper, we have investigated the scope for outperforming the pseudolikelihood by tuning weight functions for the Takacs-Fiksel estimator. Due to the complicated nature of moments for Gibbs point processes, the method is less straightforward than the one proposed by Guan *et al.* (2015) for estimating the intensity function of a spatial point process. Therefore our new Takacs-Fiksel method is not guaranteed to be optimal. It is also computationally more expensive than the approach in Guan *et al.* (2015) because the weight function needs to be evaluated for both the observed point pattern and all patterns obtained by omitting one point at a time.

In the simulation study and the applications we have demonstrated that for purely repulsive point processes, the new semi-optimal approach can yield better statistical efficiency than the pseudolikelihood while the picture is less clear for more complex situations involving both repulsive and attractive interactions. When comparing the methods the higher computational complexity of the semi-optimal approach should also be taken into account. Thus, while we have made a significant step towards optimal Takacs-Fiksel estimation there is still room for further improvement.

References

Baddeley, A. & Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics* **42**, 283–322.

Baddeley, A. & Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* **12**, 1–42.

Baddeley, A., Turner, R., Møller, J. & Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(5), 617–666.

Baddeley, A., Coeurjolly, J.-F., Rubak, E. & Waagepetersen, R. (2014). Logistic regression for spatial Gibbs point processes. *Biometrika* **101**, 377–392.

Berman, M. & Turner, R. (1992). Approximating point process likelihoods with GLIM. *Applied Statistics* **41**, 31–38.

Besag, J. (1977). Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute* **47**, 77–92.

Billiot, J.-M., Coeurjolly, J.-F. & Drouilhet, R. (2008). Maximum pseudolikelihood estimator for exponential family models of marked Gibbs point processes. *Electronic Journal of Statistics* **2**, 234–264.

Coeurjolly, J.-F. & Rubak, E. (2013). Fast covariance estimation for innovations computed from a spatial Gibbs point process. *Scandinavian Journal of Statistics* **40**(4), 669–684.

Coeurjolly, J.-F., Dereudre, D., Drouilhet, R. & Lavancier, F. (2012). Takacs–Fiksel method for stationary marked Gibbs point processes. *Scandinavian Journal of Statistics* **39**(3), 416–443.

Coeurjolly, J.-F., Møller, J. & Waagepetersen, R. (2015). Conditioning in spatial point processes. Submitted for publication. Available at arXiv:1512.05871.

Davis, T. (2006). *Direct methods for sparse linear systems*. Part of the SIAM Book Series on the Fundamentals of Algorithms, Philadelphia.

Dereudre, D., Drouilhet, R. & Georgii, H.-O. (2012). Existence of Gibbsian point processes with geometry-dependent interactions. *Probability Theory and Related Fields* **153**(3-4), 643–670.

Diggle, P., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. & Tanemura, M. (1994). On parameter estimation for pairwise interaction point processes. *International Statistical Review* **62**(1), 99–117.

Fiksel, T. (1984). Estimation of parameterized pair potentials of marked and non-marked Gibbsian point processes. *Elektronische Informationsverarbeitung und Kybernetik* **20**, 270–278.

Georgii, H. O. (1976). Canonical and grand canonical Gibbs states for continuum systems. *Communications in Mathematical Physics* **48**(1), 31–51.

Guan, Y., Jalilian, A. & Waagepetersen, R. (2015). Quasi-likelihood for spatial point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77**(3), 677–697.

Hackbusch, W. (1995). *Integral Equations: Theory and Numerical Treatment*. Birkhäuser, Basel, Switzerland.

Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Statistics in Practice, Wiley, Chichester.

Jensen, J. L. & Künsch, H. R. (1994). On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Annals of the Institute of Statistical Mathematics* **46**, 475–486.

Jensen, J. L. & Møller, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Annals of Applied Probability* **1**, 445–461.

Khanmohammadi, M., Waagepetersen, R., Nava, N., Nyengaard, J. R. & Sporring, J. (2014). Analysing the distribution of synaptic vesicles using a spatial point process model. In: *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB '14, Newport Beach, California, USA, September 20-23, 2014*, 73–78.

van Lieshout, M. (2000). *Markov Point Processes and their Applications*. Imperial College Press, London.

Mase, S. (1999). Marked Gibbs processes and asymptotic normality of maximum pseudo-likelihood estimators. *Mathematische Nachrichten* **209**, 151–169.

Møller, J. & Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton.

Nguyen, X. X. & Zessin, H. (1979). Ergodic theorems for spatial processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **48**, 133–158.

Nyström, E. J. (1930). Über die praktische auflösung von integralgleichungen mit anwendungen auf randwertaufgaben. *Acta Mathematica* **54**(1), 185–204.

Ogata, Y. & Tanemura, M. (1984). Likelihood analysis of spatial point patterns. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **46**, 496–518.

Ripley, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press, Cambridge.

Song, P. X.-K. (2007). *Correlated data analysis: modeling, analytics, and applications*. Springer Series in Statistics, Springer, New York, NY.

Stoyan, D. & Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science* **15**(1), 61–78.

Takacs, R. (1986). Estimator for the pair-potential of a Gibbsian point process. *Statistics* **17**, 429–433.