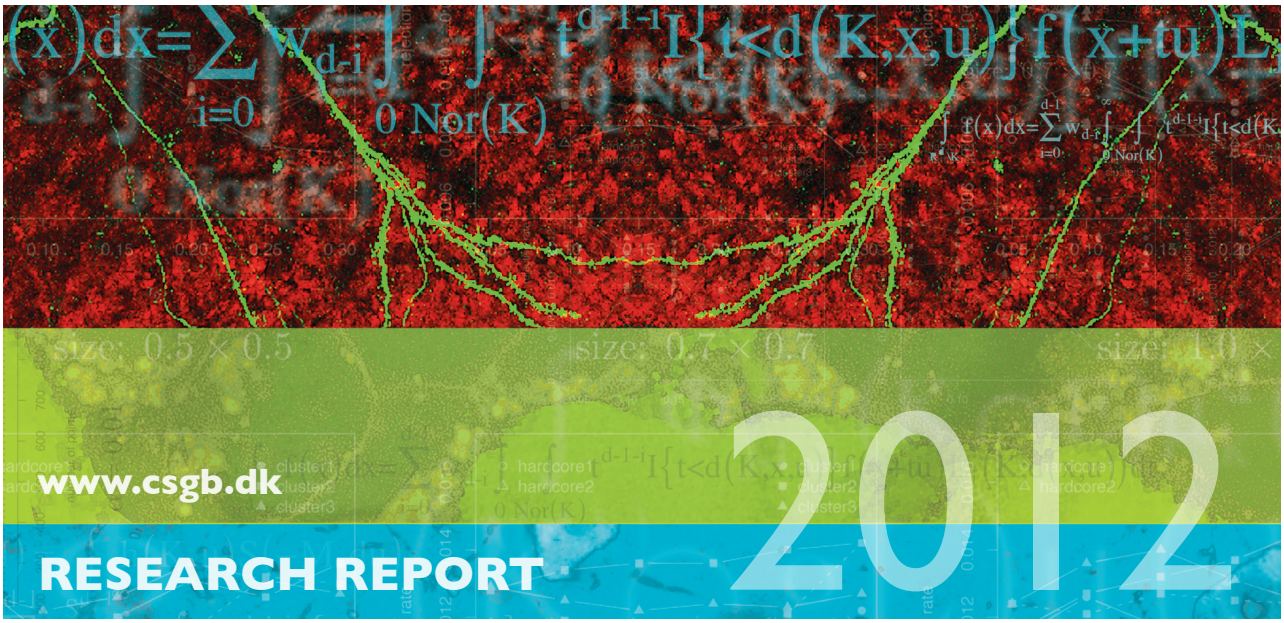
CSGB CENTRE FOR **STOCHASTIC GEOMETRY** AND ADVANCED **BIOIMAGING**



Jean-François Coeurjolly and Ege Rubak

Fast covariance estimation for innovations computed from a spatial Gibbs point process

No. 03, April 2012

Fast covariance estimation for innovations computed from a spatial Gibbs point process

Jean-François Coeurjolly[1] and Ege Rubak[2]

[1]Laboratory Jean Kuntzmann, Grenoble University, France ,
Jean-Francois.Coeurjolly@upmf-grenoble.fr.
[2]Department of Mathematical Sciences, Aalborg University, rubak@math.aau.dk.

Abstract

In this paper, we derive an exact formula for the covariance of two innovations computed from a spatial Gibbs point process and suggest a fast method for estimating this covariance. We show how this methodology can be used to estimate the asymptotic covariance matrix of the maximum pseudo-likelihood estimate of the parameters of a spatial Gibbs point process model. This allows us to construct asymptotic confidence intervals for the parameters. We illustrate the efficiency of our procedure in a simulation study for several classical parametric models. The procedure is implemented in the statistical software R and it is included in `spatstat`, which is an R package for analyzing spatial point patterns.

Keywords: innovation process, maximum pseudo-likelihood, confidence intervals, exponential family models, Georgii-Nguyen-Zessin formula.

1 Introduction

Spatial point patterns are datasets containing the random locations of some event of interest. Such datasets appear in many scientific fields such as biology, epidemiology, geography, astrophysics, physics, and economics. The stochastic mechanism generating such a dataset is modeled as a spatial point process and general references covering as well theoretical as practical aspects of this topic are e.g. Møller and Waagepetersen (2004); Stoyan et al. (1995); Illian et al. (2008). The basic spatial point process model is the Poisson process, which models complete spatial randomness in the sense that points appear independently of each other. In many applications there is dependence (or interaction) between the points, and the Poisson point process model cannot be applied. In this case Gibbs (or Markov) point processes constitute one of the main alternatives to the Poisson process, and they allow for both repulsive and attractive interaction between points. Gibbs point processes are typically defined through the so-called Papangelou conditional intensity, and a parametric class of Gibbs point process models is obtained by defining a parametric

class of Papangelou conditional intensities. For the sake of simplicity, this paper deals with exponential family models meaning that the Papangelou conditional intensity is log-linear in terms of the parameters. However, extensions to nonlinear models may be undertaken on the basis of this paper.

In the literature several methods for estimating parameters of Gibbs point process models have been suggested, and we refer to Møller and Waagepetersen (2007) for a recent overview of this problem. One of the most widely used methods is to use the maximum pseudo-likelihood estimate (MPLE) originally suggested by Besag (1975). Theoretical aspects of the MPLE for stationary Gibbs point processes have been considered in *e.g.* Jensen and Møller (1991); Jensen and Künsch (1994); Billiot et al. (2008) while practical aspects were tackled in Baddeley and Turner (2000). The popularity of this procedure is mainly due to its computational simplicity compared to the classical maximum likelihood method, and it is the default method for estimating parameters of spatial Gibbs point processes in the R package `spatstat` (Baddeley and Turner, 2005).

Typically the uncertainty of the MPLE is assessed by parametric bootstrap methods. This is computationally expensive since it requires both Monte-Carlo simulations of the fitted model and computation of the MPLE for each realization. As an alternative Billiot et al. (2008) proved the asymptotic normality of the MPLE and derived a formula for the asymptotic covariance matrix as well as an estimator of this matrix. However, this estimator is also computationally expensive due to numerical approximation of several integrals. In this paper, we express the entries of the covariance matrix as covariances between certain spatial point process innovations as defined by Baddeley et al. (2005). We prove an exact formula for the covariance between two innovations and derive a consistent estimate of this covariance. The proposed estimate does not involve any integration making it very fast compared to the alternative methods.

The rest of the paper is organized as follows. Section 2 introduces relevant notation and background material on spatial point processes including some known asymptotic results for the MPLE. Section 3 contains the main results of the paper. Here we studythe covariance between two innovations and suggest an estimator of the asymptotic covariance matrix for the MPLE. Section 4 illustrates the performance and efficiency of the developed methodology through a simulation study and applies the method to a real dataset. Finally, auxiliary results and proofs are deferred to Appendix A.

2 Gibbs point processes and pseudo-likelihood

2.1 Definition of (Gibbs) point processes

A point process X in $\mathbb{R}^d$ is a locally finite random subset of $\mathbb{R}^d$ meaning that the restriction of X to any bounded Borel set is finite. The elements of X are referred to as points and we think of them as locations of some objects or events of interest. In applications this may be locations of trees, mineral deposits, disease cases, galaxies, etc.

In this paper we keep measure theoretical details to a minimum and we will only introduce some necessary notation and terminology. The point process X takes values in the set Ω consisting of all locally finite subsets of $\mathbb{R}^d$. Thus the distribution of X is a probability measure P on an appropriate σ-algebra consisting of subsets of Ω. If the distribution of X is translation invariant we say that X is stationary.

Often the points of a point process have extra information attached to them such as e.g. the size of the tree or the type of the disease. Such information is called a mark taking values in a mark space $\mathbb{M}$. It is a random variable on $\mathbb{M}$ with distribution λ^{m}. In this case X is called a marked point process with state space $\mathbb{S} = \mathbb{R}^d \times \mathbb{M}$, and a typical element of $\mathbb{S}$ is denoted $u^\mu := (u, \mu)$. The mark space $\mathbb{M}$ may be quite general, but the reader will miss no fundamental concepts by letting $\mathbb{M}$ be $\mathbb{R}^m$ or a countable set. The state space $\mathbb{S}$ is equipped with the product measure $\lambda^d \otimes \lambda^{\mathrm{m}}$, where λ^d is the Lebesgue measure on $\mathbb{R}^d$, and with a slight abuse of notation we let $\mathrm{d}u^\mu := \lambda^d \otimes \lambda^{\mathrm{m}}(u, \mu) = \lambda^d(u)\lambda^{\mathrm{m}}(\mu)$. We call a marked point process stationary if the point process on $\mathbb{R}^d$ induced by discarding the marks is stationary. For marked point processes, we let Ω denote the set of locally finite subsets of $\mathbb{S}$.

Throughout the paper Λ is exclusively used to denote bounded Borel sets of $\mathbb{R}^d$ and $|\cdot|$ denotes the volume of such a set. For $x \in \Omega$, we let $x_\Lambda := x \cap (\Lambda \times \mathbb{M})$ and $n(x)$ denotes the number of points in x. For brevity, we say that "X is observed in Λ" for some Λ when the locations of X are in Λ and the marks are in $\mathbb{M}$.

In this paper we work with stationary (marked) Gibbs point processes models, which may be defined through a parametric family of Papangelou conditional intensities $\lambda_\theta : \mathbb{S} \times \Omega \to \mathbb{R}^+$, $\theta \in \Theta$, where θ is the parameter vector and Θ is the parameter space. Heuristically, the Papangelou conditional intensity has the interpretation that $\lambda_\theta(u^\mu, X)\mathrm{d}u^\mu$ is the conditional probability of observing a marked point in a ball of volume $\mathrm{d}u^\mu$ around u^μ given the rest of the point process is X (see *e.g.* Møller and Waagepetersen (2004)). We will not discuss how to consistently specify the Papangelou conditional intensity to ensure the existence of a Gibbs point process on $\mathbb{S}$, but rather we simply assume we are given a well-defined Gibbs point process. The reader interested in a deeper presentation of Gibbs point processes and the existence problem is referred to Ruelle (1969); Preston (1976) or Dereudre et al. (2011). In Section 2.2 we give several examples of Gibbs point processes.

Throughout the paper we will often use the following two concepts for a function $f : \mathbb{S} \times \Omega \to \mathbb{R}$,

(i) f has finite interaction range $R \geq 0$, i.e.

$$f(u^\mu, x) = f(u^\mu, x_{\mathcal{B}(u,R)}) \tag{2.1}$$

where $\mathcal{B}(u, R)$ is the euclidean ball centered at u with radius R.

(ii) f is translation invariant, i.e.

$$f(u^\mu, x) = f(0^\mu, \tau_u x) \tag{2.2}$$

where $\tau_u x$ is the translation of the locations of x by the vector $-u$.

In the remainder of the paper we will assume the following general model assumption:

[Model] For any $u^\mu \in \mathbb{S}$ and $x \in \Omega$, let $\mathbf{v}(u^\mu, x) = (v_1(u^\mu, x), \ldots, v_p(u^\mu, x))^T$, where $v_i : \mathbb{S} \times \Omega \to \mathbb{R}$ for $i = 1, \ldots, p$. For $\theta \in \Theta \subseteq \mathbb{R}^p$, let $\lambda_\theta : \mathbb{S} \times \Omega \to \mathbb{R}$ be a function of the form

$$\lambda_\theta(u^\mu, x) \;=\; \exp(\theta^T \mathbf{v}(u^\mu, x)) \tag{2.3}$$

satisfying (2.1) and (2.2). Let P_θ denote the distribution of a (well-defined) stationary hereditary marked Gibbs point process with Papangelou conditional intensity λ_θ, and let $X \sim P_{\theta^\star}$.

Under this assumption, the Papangelou conditional intensity completely characterizes the Gibbs point process in terms of the Georgii-Nguyen-Zessin (GNZ) Formula (see Papangelou (2009) and Zessin (2009) for historical comments and Georgii (1976) or Nguyen and Zessin (1979a) for a general presentation).

Lemma 2.1 (Georgii-Nguyen-Zessin Formula)**.** *For any measurable function $h : \mathbb{S} \times \Omega \to \mathbb{R}$ such that the following quantities are defined and finite, then*

$$\mathbf{E}\Big(\sum_{u^\mu \in X} h(u^\mu, X \setminus u^\mu)\Big) = \mathbf{E}\Big(\int_{\mathbb{R}^d \times \mathbb{M}} h(u^\mu, X)\lambda_{\theta^\star}(u^\mu, X)\,du^\mu\Big) \tag{2.4}$$

where $\mathbf{E}$ denotes the expectation with respect to $P_{\theta^\star}$.

Based on this formula, Baddeley et al. (2005) defined the concept of h-innovations of a spatial point process (for a function $h : \mathbb{S}\times\Omega \to \mathbb{R}$). The h-innovations computed in a bounded domain Λ is the centered random variable defined by

$$I_\Lambda(X, h) := \sum_{u^\mu \in X_\Lambda} h(u^\mu, X \setminus u^\mu) - \int_{\Lambda\times\mathbb{M}} h(u^\mu, X)\lambda_{\theta^\star}(u^\mu, X)\mathrm{d}u^\mu. \tag{2.5}$$

Baddeley et al. (2005) proposed to replace $\theta^\star$ in (2.5) by a consistent estimate to obtain residuals for spatial point processes. Such residuals can be used as a diagnostic tool of goodness-of-fit and they have also been considered by Coeurjolly and Lavancier (2012) and Baddeley et al. (2011) both from a theoretical and practical point of view.

2.2 Examples of Gibbs point processes

In this section we present some classical examples of parametric point process models (see e.g. Møller and Waagepetersen (2004) for more details). In particular, these examples will be used in the simulation study in Section 4 to assess the methodology proposed in this paper. Let $u^\mu \in \mathbb{S}$ and $x \in \Omega$. Most of the examples presented hereafter are not marked, and in these cases we omit the mark notation.

(i) *Poisson point process.* Basic example for which the log-Papangelou conditional intensity is a constant, i.e. $\log \lambda_\theta(u, x) = \theta$. The assumption **[Model]** is satisfied for any value of θ.

(ii) *Strauss point process.* Defined by

$$\log \lambda_\theta(u, x) = \theta_1 + \theta_2 n_{[0,R]}(u, x)$$

where $n_{[0,R]}(u, x) = \sum_{v\in x} \mathbf{1}(\|v - u\| \le R)$ is the number of R-close neighbours of u in x. This process has range of interaction R, and assumption **[Model]** is satisfied if $R < \infty$ and $\theta_2 \le 0$.

(iii) *Piecewise Strauss point process.* Generalization of the Strauss point process obtained by substituting the indicator function with a step function. It is defined by

$$\log \lambda_\theta(u,x) = \theta_1 + \sum_{j=1}^{p} \theta_j n_{(R_{j-1},R_j]}(u,x)$$

where $n_{(R_{j-1},R_j]}(u,x) = \sum_{v\in x} \mathbf{1}(R_{j-1} < \|v-u\| \le R_j)$ for $R_0 = 0 < R_1 < \cdots < R_p$. This process has range of interaction R_p, and assumption **[Model]** is satisfied if $R_p < \infty$ and $\theta_2, \ldots, \theta_p \le 0$.

(iv) *Geyer saturation point process* (with saturation threshold 1). Defined by

$$\log \lambda_\theta(u,x) = \theta_1 + \theta_2\Big(\sum_{v\in x\cup u} \mathbf{1}(d(v, x\cup u \setminus v) \le R) - \sum_{v\in x} \mathbf{1}(d(v, x\setminus v) \le R)\Big)$$

where $d(u,x) = \min_{w\in x} \|w-u\|$ is the distance from u to the nearest point of x. This process has range of interaction $2R$, and assumption **[Model]** is satisfied if $R < \infty$.

(v) *Multi-type Strauss point process.* This is a marked point process with m discrete marks ($\mathbb{M} = \{1,\ldots,m\}$). It is defined by

$$\log \lambda_\theta(u^j, x) = \theta_j + \sum_{k=1}^{m} \theta_{jk} n_{[0,R_{jk})}(u^j, x^k), \quad j = 1,\ldots,m$$

where $\theta_{jk} = \theta_{kj}$ and $R_{jk} = R_{kj}$. Here $n_{[0,R_{jk})}(u^j, x^k)$ denotes the number of points in x of type k which are R_{jk}-close neighbours to the point u^j of type j. The process has range of interaction $R = \max R_{jk}$, and assumption **[Model]** is satisfied when $R < \infty$ and $\theta_{jk} \le 0$, for all $j,k \in \{1,\ldots,m\}$.

2.3 Maximum pseudo-likelihood estimate

Assume we observe X_{Λ^+}, where $\Lambda^+ \subset \mathbb{R}^d$ is bounded, and let $\Lambda = \Lambda^+ \ominus R$ be the erosion of Λ^+ by R, i.e.

$$\Lambda = \Lambda^+ \ominus R = \{u \in \Lambda^+ | \mathcal{B}(u,R) \subseteq \Lambda^+\}. \tag{2.6}$$

The maximum pseudo-likelihood estimate (MPLE) is the value $\theta = \widehat{\theta}$ which maximizes the pseudo-likelihood

$$PL_\Lambda(X;\theta) = \prod_{u^\mu \in X_\Lambda} \lambda_\theta(u^\mu, X\setminus u^\mu) \exp\Big(-\int_{\Lambda\times\mathbb{M}} \lambda_\theta(u^\mu, X)\mathrm{d}u^\mu\Big).$$

This maximum is attained at the root of the score function with jth component

$$\frac{\partial}{\partial\theta_j} \log PL_\Lambda(X;\theta) = \sum_{u^\mu\in X_\Lambda} v_j(u^\mu, X\setminus u^\mu) - \int_{\Lambda\times\mathbb{M}} v_j(u^\mu, X)\lambda_\theta(u^\mu, X)\mathrm{d}u^\mu$$

for $j = 1,\ldots,p$.

To detail the asymptotic properties of the MPLE, we now let $\Lambda = \Lambda_n$ depend on an index n. For each n we assume Λ_n is a cube with volume $|\Lambda_n| \to \infty$ for $n \to \infty$. Furthermore, we need the following technical assumption:

[MPLE] The parameter space $\Theta \subset \mathbb{R}^p$ is compact, $\theta^\star \in \mathring{\Theta}$ and for any $\theta \neq \theta^\star$, the following identifiability condition holds

$$P_{\theta^\star}\left((\theta - \theta^\star)^T \mathbf{v}(0^M, X) \neq 0\right) > 0.$$

Furthermore, for all $u^\mu \in \mathbb{S}$ and $x \in \Omega$ there exists a constant $\kappa \geq 0$ such that one of the following two assumptions is satisfied:

$$\theta_i \leq 0 \quad \text{and} \quad -\kappa \leq v_i(0^\mu, x) \leq \kappa\, n(x_{\mathcal{B}(0,R)}) \tag{2.7}$$

or

$$-\kappa \leq v_i(0^\mu, x) \leq \kappa \tag{2.8}$$

where R is the range of interaction defined in (2.1).

Billiot et al. (2008) extended the results in Jensen and Møller (1991) and Jensen and Künsch (1994) and obtained consistency and asymptotic normality of the MPLE for a large class of models including the examples presented in Section 2.2. We now state the central limit theorem for the MPLE.

Proposition 2.2 (Billiot et al. (2008))**.** *Assume that the distribution of X is ergodic and that **[MPLE]** is satisfied. Then, for $n \to \infty$, the MPLE is strongly consistent and satisfies the following central limit theorem*

$$|\Lambda_n|^{1/2}(\widehat{\theta}_n - \theta^\star) \xrightarrow{d} \mathcal{N}(0, U^{-1}\Sigma U^{-1}),$$

where U and Σ are (p,p) matrices with entries

$$U_{jk} = \mathbf{E}[v_j(0^M, X) v_k(0^M, X) \lambda_{\theta^\star}(0^M, X)] \tag{2.9}$$

$$\Sigma_{jk} = \lim_{n\to\infty} |\Lambda_n|^{-1} \operatorname{Cov}\left(\frac{\partial}{\partial\theta_j} \log PL_\Lambda(X;\theta^\star), \frac{\partial}{\partial\theta_k} \log PL_\Lambda(X;\theta^\star)\right) \tag{2.10}$$

where M is a random variable with distribution λ^{m}.

To propose a computationally efficient way of estimating the asymptotic covariance matrix for the MPLE, the key point is to note that

$$\frac{\partial}{\partial\theta_j} \log PL_\Lambda(X;\theta^\star) = I_\Lambda(X, v_j). \tag{2.11}$$

Thus, from (2.10) we need to be able to estimate the covariance between innovations, which we detail in the following section.

3 Covariance of innovations

Several properties of the innovations are established in Baddeley et al. (2005) and Baddeley et al. (2008). In particular, Proposition 4 in Baddeley et al. (2005) presents a formula for the variance of $I_\Lambda(X,h)$. Our Lemma A.1 in Appendix A.1 extends this result by providing a formula for the covariance between two innovations $I_\Lambda(X,g)$ and $I_\Lambda(X,h)$. In this Section, we use Lemma A.1 to study the asymptotic covariance between innovations. In particular, we propose a consistent estimate of this covariance which requires no numerical integration. Finally, we use the results to estimate the asymptotic covariance matrix of the MPLE, which allows us to quantify the uncertainty of the MPLE much faster than previously possible.

To obtain the asymptotic results in this section, we need the second order Papangelou conditional intensity

$$\lambda_\theta(\{u^\mu, v^\nu\}, X) = \lambda_\theta(u^\mu, X \cup v^\nu)\lambda_\theta(v^\nu, X) = \lambda_\theta(v^\nu, X \cup u^\mu)\lambda_\theta(u^\mu, X), \quad u^\mu, v^\nu \in \mathbb{S}. \tag{3.1}$$

Also, for any $v^\nu \in \mathbb{S}$, we define the difference operator Δ_{v^ν} applied to a function $h : \mathbb{S} \times \Omega \to \mathbb{R}$ as

$$\Delta_{v^\nu} h(u^\mu, X) := h(u^\mu, X \cup v^\nu) - h(u^\mu, X). \tag{3.2}$$

Furthermore, certain conditions on the functions g and h are needed, as detailed in the following assumption:

[H(g,h)] The functions $g, h : \mathbb{S} \times \Omega \to \mathbb{R}$ satisfy (2.1) and (2.2). Furthermore, there exists an open neighbourhood $\mathcal{V}$ of $\theta^\star$ such that for any $\theta \in \mathcal{V}$, the random variables I_1, I_2, I_3 given by

$$I_1(g,h) := \left| g(0^M, X) h(0^M, X) \lambda_{\theta^\star}(0^M, X) \right| \tag{3.3}$$

$$I_2(g,h) := \int_{\mathcal{B}(0,R)\times\mathbb{M}} \left| g(0^M, X) h(v^\nu, X) \lambda_{\theta^\star}(\{0^M, v^\nu\}, X) \times \left(\frac{\lambda_\theta(0^M, X)\lambda_\theta(v^\nu, X)}{\lambda_\theta(\{0^M, v^\nu\}, X)} - 1 \right) \right| \mathrm{d}v^\nu \tag{3.4}$$

$$I_3(g,h) := \int_{\mathcal{B}(0,R)\times\mathbb{M}} \left| \Delta_{v^\nu} g(0^M, X) \Delta_{0^M} h(v^\nu, X) \lambda_{\theta^\star}(\{0^M, v^\nu\}, X) \right| \mathrm{d}v^\nu \tag{3.5}$$

have finite expectation.

Note that **[Model]** implies that $\lambda_\theta(\{u^\mu, v^\nu\}, X)$ is almost surely positive for any $u^\mu, v^\nu \in \mathbb{S}$ and any $\theta \in \Theta$. In particular the ratio in (3.4) is therefore well-defined. Now we study the normalized covariance of innovations

$$C_{\Lambda_n}(g,h) := |\Lambda_n|^{-1} \ \mathrm{Cov}\left(I_{\Lambda_n}(X,g), I_{\Lambda_n}(X,h) \right),$$

for cubes Λ_n with $|\Lambda_n| \to \infty$ as $n \to \infty$. For the result below, the neighbourhood $\mathcal{V}$ appearing in **[H(g,h)]** could be replaced by $\{\theta^\star\}$.

Proposition 3.1. *Assume* ***[H(g,h)]*** *and let M be a random variable with distribution λ^{m}. Then, as $n \to \infty$,*

$$C_{\Lambda_n}(g,h) \to C(g,h) = A_1(g,h) + A_2(g,h) + A_3(g,h)$$

where

$$\begin{aligned}
A_1(g,h) &= \mathbf{E}\Big[g(0^M,X)h(0^M,X)\lambda_{\theta^\star}(0^M,X)\Big] \\
A_2(g,h) &= \int_{\mathcal{B}(0,R)\times\mathbb{M}} \mathbf{E}\Big[g(0^M,X)h(v^\nu,X)\big(\lambda_{\theta^\star}(0^M,X)\lambda_{\theta^\star}(v^\nu,X) \\
&\qquad\qquad - \lambda_{\theta^\star}(\{0^M,v^\nu\},X)\big)\Big]\,dv^\nu \\
A_3(g,h) &= \int_{B(0,R)\times\mathbb{M}} \mathbf{E}\Big[\Delta_{v^\nu} g(0^M,X)\Delta_{0^M} h(v^\nu,X)\lambda_{\theta^\star}(\{0^M,v^\nu\},X)\Big]\,dv^\nu.
\end{aligned}$$

The following main result of this paper establishes a strongly consistent and computationally fast estimate of $C(g,h)$. The idea behind our result is to combine a consistent estimate of $\theta^\star$ with estimates of the matrices $A_i(g,h)$, $i = 1,2,3$ of Proposition 3.1.

Theorem 3.2. *Let $g_\theta, h_\theta : \mathbb{S}\times\Omega \to \mathbb{R}$ be parametric families of functions, which are (almost surely) continuous in θ. Assume there exists an open neighbourhood $\mathcal{V}$ of $\theta^\star$ such that for all $\theta \in \mathcal{V}$ the assumption $[\mathbf{H}(\mathbf{g}_\theta, \mathbf{h}_\theta)]$ holds, and let $\widehat{\theta} = \widehat{\theta}_n(X)$ be a strongly consistent estimate of $\theta^\star$. Then, as $n \to \infty$, we have the following almost sure convergence*

$$\widehat{C}(g_{\widehat{\theta}}, h_{\widehat{\theta}}) := \widehat{A}_1(g_{\widehat{\theta}}, h_{\widehat{\theta}}) + \widehat{A}_2(g_{\widehat{\theta}}, h_{\widehat{\theta}}) + \widehat{A}_3(g_{\widehat{\theta}}, h_{\widehat{\theta}}) \to C(g_{\theta^\star}, h_{\theta^\star})$$

where

$$\begin{aligned}
\widehat{A}_1(g_{\widehat{\theta}}, h_{\widehat{\theta}}) &= \frac{1}{|\Lambda_n|}\sum_{u^\mu \in X_{\Lambda_n}} g_{\widehat{\theta}}(u^\mu, X\setminus u^\mu) h_{\widehat{\theta}}(u^\mu, X\setminus u^\mu) \\
\widehat{A}_2(g_{\widehat{\theta}}, h_{\widehat{\theta}}) &= \frac{1}{|\Lambda_n|}\sum_{\substack{u^\mu, v^\nu \in X_{\Lambda_n} \\ u\neq v, \|u-v\|\le R}} g_{\widehat{\theta}}(u^\mu, X\setminus\{u^\mu,v^\nu\}) h_{\widehat{\theta}}(v^\nu, X\setminus\{u^\mu,v^\nu\}) \\
&\qquad\qquad \times\left(\frac{\lambda_{\widehat{\theta}}(u^\mu, X\setminus\{u^\mu,v^\nu\})\lambda_{\widehat{\theta}}(v^\nu, X\setminus\{u^\mu,v^\nu\})}{\lambda_{\widehat{\theta}}(\{u^\mu,v^\nu\}, X\setminus\{u^\mu,v^\nu\})} - 1\right) \\
\widehat{A}_3(g_{\widehat{\theta}}, h_{\widehat{\theta}}) &= \frac{1}{|\Lambda_n|}\sum_{\substack{u^\mu, v^\nu \in X_{\Lambda_n} \\ u\neq v, \|u-v\|\le R}} \Delta_{v^\nu} g_{\widehat{\theta}}(u^\mu, X\setminus\{u^\mu,v^\nu\}) \Delta_{u^\mu} h_{\widehat{\theta}}(v^\nu, X\setminus\{u^\mu,v^\nu\}).
\end{aligned}$$

From Proposition 3.1 and (2.9)-(2.11) we have $U_{jk} = A_1(v_j, v_k)$ and $\Sigma_{jk} = C(v_j, v_k)$. Then the corollary below follows by combining Proposition 2.2 with Theorem 3.2.

Corollary 3.3. Let the matrices $\widehat{A}_i(v_j, v_k)$, $i = 1,2,3$, be as in Theorem 3.2 with $\widehat{\theta}$ given by the MPLE. Under the assumption **[MPLE]**, the (p,p) matrices $\widehat{U}$ and $\widehat{\Sigma}$

with entries $\widehat{U}_{jk} = \widehat{A}_1(v_j, v_k)$ and $\widehat{\Sigma}_{jk} = \widehat{C}(v_j, v_k) = \widehat{A}_1(v_j, v_k)+\widehat{A}_2(v_j, v_k)+\widehat{A}_3(v_j, v_k)$ are strongly consistent estimates of U and Σ. Moreover, if Σ is positive definite, we have the following convergence in distribution as $n \to \infty$

$$|\Lambda_n|^{1/2}\widehat{\Sigma}^{-1/2}\widehat{U}(\widehat{\theta}_n - \theta^\star) \xrightarrow{d} \mathcal{N}(0, I_p). \tag{3.6}$$

We point out that (3.6) does not require the ergodicity of $P_{\theta^\star}$ and it therefore applies even if a phase transition occurs (see Jensen and Künsch (1994) for a proof of this). Furthermore, we refer to Billiot et al. (2008) for a proof of the positive definiteness of the matrix Σ for a large class of models (including the ones presented in this paper).

4 Applications

In this section we describe how the theory of Sections 2-3 is applied in practice (for $d = 2$). In Section 4.1 we detail the methodology for a Strauss point process. Section 4.2 describes a simulation study involving the models presented in Section 2.2. In Section 4.3 we exemplify the methodology using a dataset of marked points.

We assume we are given a realization x^+ of X_{Λ^+}, and we let $x = x^+_\Lambda$ denote the realization of X_Λ, where Λ is given by (2.6). Let $\widehat{\theta}$ denote the MPLE based on X_{Λ^+}. From Corollary 3.3 we use the approximation $\widehat{\theta} \sim \mathcal{N}(\theta^\star, \widehat{\Sigma}_{\text{MPLE}})$, where $\widehat{\Sigma}_{\text{MPLE}} = |\Lambda|^{-1}\widehat{U}^{-1}\widehat{\Sigma}\widehat{U}^{-1}$. If $\hat{s}_i^2$ denotes the ith diagonal element of $\widehat{\Sigma}_{\text{MPLE}}$, then the approximate 95 % confidence interval for $\theta_i^\star$ is $[\widehat{\theta}_i - 1.96\hat{s}_i, \widehat{\theta}_i + 1.96\hat{s}_i]$, $i = 1, \ldots, p$. The approximate 95 % confidence region for $\theta^\star$ is $\{\theta : (\widehat{\theta}-\theta)^T\widehat{\Sigma}^{-1}_{\text{MPLE}}(\widehat{\theta}-\theta) \le q_{95\,\%}\}$, where $q_{95\,\%}$ is the 95 % quantile of a χ^2_p distribution.

4.1 Strauss point process

When X is a Strauss point process the formulas for $\widehat{A}_1$, $\widehat{A}_2$ and $\widehat{A}_3$ defining $\widehat{\Sigma}_{\text{MPLE}}$ simplify considerably and we detail these in the following to underline the computational simplicity of $\widehat{\Sigma}_{\text{MPLE}}$. Let $n = n(x)$ be the number of points in $x = (x_1, \ldots, x_n)$. We denote by T (resp. T^+) the vector of length n with ith component given by the number of R-close neighbours of x_i in $x \setminus x_i$ (resp. R-close neighbours of x_i in $x^+ \setminus x_i$). Then

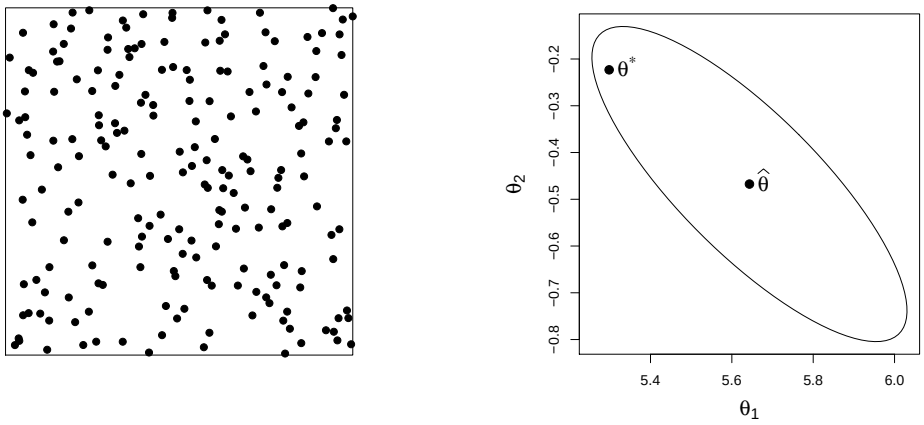
$$\widehat{A}_1 = |\Lambda|^{-1} \begin{pmatrix} n & \sum_i T_i^+ \\ \sum_i T_i^+ & \sum_i T_i^{+2} \end{pmatrix}$$

$$\widehat{A}_2 = |\Lambda|^{-1}(e^{-\widehat{\theta}_2} - 1) \begin{pmatrix} \sum_i T_i & \sum_i T_i(T_i^+ - 1) \\ \sum_i T_i(T_i^+ - 1) & \sum_{I_R}(T_i^+ - 1)(T_j^+ - 1) \end{pmatrix}$$

$$\widehat{A}_3 = |\Lambda|^{-1} \begin{pmatrix} 0 & 0 \\ 0 & \sum_i T_i \end{pmatrix}$$

where $I_R = \{i, j = 1, \ldots, n : \|x_i - x_j\| \le R, x_i \ne x_j\}$.

As an example consider a realization in the unit square of a Strauss point process with interaction range $R = 0.05$ and parameters $\theta_1^\star = \log(\beta) = \log(200) \approx 5.3$ and $\theta_2^\star = \log(\gamma) = \log(0.5) \approx -0.69$. Such a realization is generated via a perfect simulation algorithm in `spatstat` as follows:



(a) **(b)**

Figure 1: (a) Realization of a Strauss point process. (b) Approximate 95 % confidence region for the MPLE.

```
> X <- rStrauss(beta=200, gamma=0.5, R=0.05)
```

In this case the point pattern with 204 points, shown in Figure 1(a), was generated. Then the MPLE of the parameters of a Strauss point process model with interaction range $R = 0.05$ is calculated via:

```
> fit <- ppm(X, interaction=Strauss(0.05))
```

The result `fit` contains relevant information about the fitted model and the MPLE, which was $(\widehat{\theta}_1, \widehat{\theta}_2) = (5.64, -0.47)$ in this case. The approximate covariance matrix of the MPLE is estimated using the formulas above via:

```
> sigmaMPLE <- vcov(fit)
```

The result is simply the estimated covariance matrix of the MPLE. From this we can calculate the approximate 95 % confidence region, shown in Figure 1(b), and the individual confidence intervals, which in this case were $[5.33, 5.95]$ and $[-0.74, -0.20]$ for $\theta_1^\star$ and $\theta_2^\star$ respectively.

Note that the procedure `vcov` is not specific to the Strauss model, but works for any point process model implemented in `spatstat` satisfying **[Model]**.

4.2 Simulation study

In this section we present a simulation study using the following models:

- Strauss point processes with $R = 0.05$ and $\theta_1^\star = \log(200)$, where models S1, S2 and S3 respectively have $\theta_2^\star = \log(0.8)$, $\theta_2^\star = \log(0.5)$, and $\theta_2^\star = \log(0.2)$.
- Piecewise Strauss point processes with $R_1 = 0.05$, $R_2 = 0.1$, and $\theta_1^\star = \log(200)$, where models P1 and P2 respectively have $(\theta_2^\star, \theta_3^\star) = (\log(0.8), \log(0.2))$, and $(\theta_2^\star, \theta_3^\star) = (\log(0.2), \log(0.8))$.

Table 1: Results for different simulated Gibbs point process models based on 500 replications with $\Lambda = [0, \ell]^2$, $\ell = 1, 2$. Left column: empirical coverage rates (i.e. the fraction of confidence ellipsoids covering the true parameter value). Right: minimum and maximum of empirical 1-dimensional coverage rates (i.e. the fraction of confidence intervals covering the true parameter value for each parameter).

	Coverage (%)		**1-dim. coverage (%)**	
	ℓ=1	ℓ=2	ℓ=1	ℓ=2
S1	95.0	93.8	[95.2,96.0]	[94.2,94.6]
S2	94.4	95.2	[95.4,96.6]	[95.2,97.0]
S3	95.0	97.0	[96.2,96.8]	[97.0,97.2]
P1	88.0	94.2	[93.4,97.6]	[93.0,95.8]
P2	92.2	94.2	[94.8,95.0]	[95.2,95.8]
G1	96.4	95.4	[96.4,97.4]	[95.6,95.8]
G2	95.6	94.2	[96.4,96.4]	[94.2,94.6]
M1	94.6	94.8	[94.0,95.8]	[95.2,96.2]
M2	92.2	96.0	[93.6,95.8]	[92.3,95.2]

- Geyer point processes with $R = 0.05$ and $\theta_1^\star = \log(100)$, where models G1 and G2 respectively have $\theta_2^\star = \log(1.2)$ and $\theta_2^\star = \log(0.8)$.
- Multi-type Strauss point processes with two types, $R_{11} = R_{22} = R_{12} = 0.05$ and $\theta_1^\star = \theta_2^\star = \log(200)$, where models M1 and M2 respectively have $\theta_{11}^\star = \theta_{22}^\star = \theta_{12}^\star = \log(0.5)$, and $\theta_{11}^\star = \theta_{22}^\star = \log(0.8)$, $\theta_{12}^\star = \log(0.2)$.

For each model 500 realizations were generated using the Metropolis-Hastings algorithm with birth, death and shift proposals as detailed in Geyer and Møller (1994) (except for the Strauss point processes which were generated using the perfect simulation algorithm of Berthelsen and Møller (2002); Berthelsen and Møller (2003)). For all the models $\Lambda^+ = [-R, \ell + R]^2$, $\ell = 1, 2$, where R is the interaction range of each model. Based on these simulations we calculated the approximate 95 % confidence region (respectively confidence intervals for each parameter) and checked whether it covered $\theta^\star$ (respectively $\theta_j^\star$). The results given in Table 1 show that the coverage rates are close to the expected 95 % for all the models.

4.3 Illustration on the dataset `amacrine`

In this section we use a real dataset to illustrate how the methodology developed in this paper is applied. We consider the dataset `amacrine` available in the `spatstat` package. This dataset is a spatial marked point pattern of displaced amacrine cells in the retina of a rabbit. The marks have two discrete values *off* and *on* classifying the type of the cell, and the locations of 142 *off* cells and 152 *on* cells are given in a 1060 µm by 662 µm sampling frame. This dataset was first analyzed by Diggle (1986) and later revisited by Baddeley (2010, Chapter 21, p. 400) to illustrate the use of the approximate maximum likelihood estimate (AMLE) suggested by Huang and Ogata (1999). The data was modeled as a multi-type Strauss point process with interaction

radii $R_{11} = R_{12} = R_{12} = 60\,\mu\text{m}$. The parameter estimates and their standard errors (obtained using parametric Bootstrap techniques) presented in Baddeley (2010) are recalled in Table 2. This table also contains the corresponding maximum pseudo-likelihood estimates (MPLE) and their standard errors calculated by the method developed in this paper. The MPLE and standard errors were obtained in a few seconds. In contrast, calculation of the AMLE and its standard errors may take a few hours depending on the number of replications used in the Huang-Ogata approximation and on the number of replications of the fitted model used to estimate the standard error. We have parametrized the Multi-type Strauss point process as in Section 2.2 with the mark 1 (respectively 2) representing *off* (respectively *on*) cells.

We observe that using AMLE or MPLE yields qualitatively similar results. In particular, we note that the estimates of θ_{11} and θ_{22} are large, suggesting a strong repulsion between cells of the same type, whereas the small estimate of θ_{12} suggests a weaker repulsion between cells of different types. As it was done in Baddeley (2010), we can propose an asymptotic hypothesis testing procedure to test the hypothesis $H_0 : \theta_{12} = 0$ against $H_1 : \theta_{12} \neq 0$. If $\hat{\theta}_{12}$ is the MPLE of θ_{12} and $\hat{\sigma}_{12}$ denotes the standard error, then our asymptotic result implies that $\hat{\theta}_{12}/\hat{\sigma}_{12}$ approximately follows a standard Gaussian distribution under the null hypothesis. Based on this we cannot reject the null hypothesis ($p \simeq 46\,\%$), so a model with independence between the two types of cells might be appropriate.

Table 2: Parameter estimates and standard errors for the Huang-Ogata approximate maximum likelihood estimate (AMLE) and the maximum pseudo-likelihood estimate (MPLE) for the amacrine dataset. The index 1 (resp. 2) corresponds to the cell type *off* (resp. *on*). The cross $\times$ indicates that this standard error was not computed in Baddeley (2010).

Method	**Parameter**	θ_1	θ_2	$\theta_2 - \theta_1$	θ_{11}	θ_{12}	θ_{22}
AMLE	estimate	-6.045	-5.798	0.247	-1.346	-0.100	-1.335
	se	0.325	$\times$	0.323	0.160	0.085	0.170
MPLE	estimate	-4.424	-4.541	-0.116	-2.140	-0.164	-1.978
	se	0.732	0.794	0.669	0.190	0.224	0.259

Acknowledgements

This research was supported by Joseph Fourier University of Grenoble (project "SpaComp"), by Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation, by the Danish Natural Science Research Council, grant 09-072331, "Point process modelling and statistical inference" and by l'Institut Français du Danemark. A part of this research was done while the first author was visiting the Department of Mathematical Sciences at Aalborg University. He would like to thank the members of the department for their kind hospitality.

A Appendix

A.1 Auxiliary Lemma

The following result provides an exact formula for the covariance of two spatial point process innovations.

Lemma A.1. *Assume $g, h : \mathbb{S} \times \Omega \to \mathbb{R}$ are such that the following quantities are defined and finite, then*

$$\mathrm{Cov}\left(I_\Lambda(X,g), I_\Lambda(X,h)\right) = \widetilde{A}_{1,\Lambda}(g,h) + \widetilde{A}_{2,\Lambda}(g,h) + \widetilde{A}_{3,\Lambda}(g,h)$$

with

$$\begin{aligned}
\widetilde{A}_{1,\Lambda}(g,h) &= \mathbf{E}\left[\int_{\Lambda\times\mathbb{M}} g(u^\mu, X)h(u^\mu, X)\lambda_{\theta^\star}(u^\mu, X)\,du^\mu\right] \\
\widetilde{A}_{2,\Lambda}(g,h) &= \mathbf{E}\Big[\int_{(\Lambda\times\mathbb{M})^2} g(u^\mu, X)h(v^\nu, X)\big(\lambda_{\theta^\star}(u^\mu, X)\lambda_{\theta^\star}(v^\nu, X) \\
&\qquad\qquad - \lambda_{\theta^\star}(\{u^\mu, v^\nu\}, X)\big)\,du^\mu\,dv^\nu\Big] \\
\widetilde{A}_{3,\Lambda}(g,h) &= \mathbf{E}\left[\int_{(\Lambda\times\mathbb{M})^2} \Delta_{v^\nu} g(u^\mu, X)\Delta_{u^\mu} h(v^\nu, X)\lambda_{\theta^\star}(\{u^\mu, v^\nu\}, X)\,du^\mu\,dv^\nu\right].
\end{aligned}$$

Proof. From the GNZ Formula (2.4), $\mathbf{E}[I_\Lambda(X,g)] = \mathbf{E}[I_\Lambda(X,h)] = 0$. Now, we decompose the covariance into four terms

$$\mathbf{E}[I_\Lambda(X,g)I_\Lambda(X,h)] = T_1 + T_2 + T_3 + T_4.$$

These different terms are defined and simplified using again the GNZ formula as follows

$$T_1 = \mathbf{E}\Big[\int_{(\Lambda\times\mathbb{M})^2} g(u^\mu, X)\lambda_{\theta^\star}(u^\mu, X)h(v^\nu, X)\lambda_{\theta^\star}(v^\nu, X)\mathrm{d}u^\mu\mathrm{d}v^\nu\Big] \tag{A.1}$$

$$\begin{aligned}
T_2 &= -\mathbf{E}\Big[\int_{\Lambda\times\mathbb{M}} g(u^\mu, X)\lambda_{\theta^\star}(u^\mu, X)\mathrm{d}u^\mu \sum_{v^\nu\in X_\Lambda} h(v^\nu, X\setminus v^\nu)\Big] \\
&= -\mathbf{E}\Big[\sum_{v^\nu X_\Lambda}\left(h(v^\nu, X\setminus v^\nu)\int_{\Lambda\times\mathbb{M}} g(u^\mu, X)\lambda_{\theta^\star}(u^\mu, X)\mathrm{d}u^\mu\right)\Big] \\
&= -\mathbf{E}\Big[\int_{(\Lambda\times\mathbb{M})^2} h(v^\nu, X)g(u^\mu, X\cup v^\nu)\lambda_{\theta^\star}(u^\mu, X\cup v^\nu)\lambda_{\theta^\star}(v^\nu, X)\mathrm{d}u^\mu\mathrm{d}v^\nu\Big]
\end{aligned} \tag{A.2}$$

$$\begin{aligned}
T_3 &= -\mathbf{E}\Big[\int_{\Lambda\times\mathbb{M}} h(v^\nu, X)\lambda_{\theta^\star}(v^\nu, X)\mathrm{d}v^\nu \sum_{u^\mu\in X_\Lambda} g(u^\mu, X\setminus u^\mu)\Big] \\
&= -\mathbf{E}\Big[\int_{(\Lambda\times\mathbb{M})^2} g(u^\mu, X)h(v^\nu, X\cup u^\mu)\lambda_{\theta^\star}(\{u^\mu, v^\nu\}, X)\mathrm{d}u^\mu\mathrm{d}v^\nu\Big]
\end{aligned} \tag{A.3}$$

and

$$
\begin{aligned}
T_4 &= \mathbf{E}\Big[\sum_{u^\mu, v^\nu \in X_\Lambda} g(u^\mu, X \setminus u^\mu) h(v^\nu, X \setminus v^\nu)\Big] \\
&= \mathbf{E}\Big[\sum_{\substack{u^\mu, v^\nu \in X_\Lambda \\ u^\mu \neq v^\nu}} g(u^\mu, X \setminus u^\mu) h(v^\nu, X \setminus v^\nu)\Big] \\
&\quad + \mathbf{E}\Big[\sum_{u^\mu \in X_\Lambda} g(u^\mu, X \setminus u^\mu) h(u^\mu, X \setminus u^\mu)\Big] \\
&= \mathbf{E}\Big[\int_{(\Lambda \times \mathbb{M})^2} g(u^\mu, X \cup v^\nu) h(v^\nu, X \cup u^\mu) \lambda_{\theta^\star}(\{u^\mu, v^\nu\}, X) \mathrm{d}u^\mu \mathrm{d}v^\nu\Big] \\
&\quad + \mathbf{E}\Big[\int_{\Lambda \times \mathbb{M}} g(u^\mu, X) h(u^\mu, X) \lambda_{\theta^\star}(u^\mu, X) \mathrm{d}u^\mu\Big]. \qquad \text{(A.4)}
\end{aligned}
$$

Rearranging (A.1)–(A.4) leads to the result. □

A.2 Proof of Proposition 3.1

Proof. From Lemma A.1, we just have to prove that $|\Lambda_n|^{-1}\widetilde{A}_{i,\Lambda_n}(g,h) \to A_i(g,h)$, for $i = 1, 2, 3$. The stationarity of the point process is sufficient for $i = 1$ since $|\Lambda_n|^{-1}\widetilde{A}_{1,\Lambda_n} = A_1(g,h)$. For the other terms, let $u^\mu, v^\nu \in \mathbb{S}$ such that $\|u - v\| \geq R$. Then for any function $f : \mathbb{S} \times \Omega \to \mathbb{R}$ satisfying (2.1), we have $f(u^\mu, x \cup v^\nu) = f(u^\mu, x)$, which implies

$$
\lambda_{\theta^\star}(\{u^\mu, v^\nu\}, X) = \lambda_{\theta^\star}(u^\mu, X \cup v^\nu)\lambda_{\theta^\star}(v^\nu, X) = \lambda_{\theta^\star}(u^\mu, X)\lambda_{\theta^\star}(v^\nu, X)
$$

and

$$
\Delta_{v^\nu} g(u^\mu, X) = g(u^\mu, X \cup v^\nu) - g(u^\mu, X) = 0.
$$

Then we focus on the convergence of the second term (the third one follows similar arguments). Let us decompose $\widetilde{A}_{2,\Lambda_n}(g,h) = \widetilde{A}^1_{2,\Lambda_n}(g,h) + \widetilde{A}^2_{2,\Lambda_n}(g,h)$ where

$$
\begin{aligned}
\widetilde{A}^1_{2,\Lambda_n}(g,h) &:= \mathbf{E}\Big[\int_{(\Lambda_n \ominus R) \times \mathbb{M}} \int_{(\mathcal{B}(u,R) \cap \Lambda_n) \times \mathbb{M}} f(u^\mu, v^\nu, X) \mathrm{d}v^\nu \mathrm{d}u^\mu\Big] \\
\widetilde{A}^2_{2,\Lambda_n}(g,h) &:= \mathbf{E}\Big[\int_{(\Lambda_n \setminus (\Lambda_n \ominus R)) \times \mathbb{M}} \int_{(\mathcal{B}(u,R) \cap \Lambda_n) \times \mathbb{M}} f(u^\mu, v^\nu, X) \mathrm{d}v^\nu \mathrm{d}u^\mu\Big],
\end{aligned}
$$

and $f(u^\mu, v^\nu, X) := g(u^\mu, X)h(v^\nu, X)\left(\lambda_{\theta^\star}(u^\mu, X)\lambda_{\theta^\star}(v^\nu, X) - \lambda_{\theta^\star}(\{u^\mu, v^\nu\}, X)\right)$. From the stationarity of X and since f satisfies (2.2), we get

$$
\begin{aligned}
|\Lambda_n|^{-1}\ \widetilde{A}^1_{2,\Lambda_n}(g,h) &= |\Lambda_n|^{-1}\mathbf{E}\Big[\int_{(\Lambda_n \ominus R) \times \mathbb{M}} \int_{\mathcal{B}(u,R) \times \mathbb{M}} f(u^\mu, v^\nu, X) \mathrm{d}v^\nu \mathrm{d}u^\mu\Big] \\
&= \frac{|\Lambda_n \ominus R|}{|\Lambda_n|} A_2(g,h) \\
&\to A_2(g,h)
\end{aligned}
$$

and

$$
\begin{aligned}
|\Lambda_n|^{-1}\,|\widetilde{A}^2_{2,\Lambda_n}(g,h)| &\leq |\Lambda_n|^{-1}\mathbf{E}\Big[\int_{(\Lambda_n\setminus(\Lambda_n\ominus R))\times\mathbb{M}}\int_{\mathcal{B}(u,R)\times\mathbb{M}}|f(u^\mu,v^\nu,X)|\mathrm{d}v^\nu\mathrm{d}u^\mu\Big]\\
&= \frac{|\Lambda_n\setminus(\Lambda_n\ominus R)|}{|\Lambda_n|}\mathbf{E}\Big[\int_{\mathcal{B}(0,R)\times\mathbb{M}}|f(0^M,v^\nu,X)|\mathrm{d}v^\nu\Big]\\
&\to 0
\end{aligned}
$$

as $n\to\infty$. □

A.3 Proof of Theorem 3.2

Assumption **[Model]** asserts the existence of at least one stationary Gibbs measure. If this measure is unique, it is ergodic. Otherwise, it can be represented as a mixture of ergodic measures (see Georgii (1988), Theorem 14.10). Therefore, we can assume, for this proof, that $P_{\theta^\star}$ is ergodic.

Proof. For $j=1,2,3$, let us denote by $\widehat{A}_j(\theta)$ the quantity $\widehat{A}_j(g,h)$ where $\widehat{\theta}$ is replaced by θ, for $\theta\in\mathcal{V}$. In the following, the general ergodic theorem for spatial point processes obtained by Nguyen and Zessin (1979b) (see also Lemma 2 in Coeurjolly et al. (2012)) combined with the GNZ Formula (2.4) will be widely used (as $n\to\infty$). These uses are justified by the assumptions **[Model]** and **[H(g_θ,h_θ)]**. Using the arguments above, we immediately obtain the following almost sure convergence.

$$
\widehat{A}_1(\theta)\to A_1(\theta):=E\left[g_\theta(0^M,X)h_\theta(0^M,X)\lambda_{\theta^\star}(0^M,X)\right]. \tag{A.5}
$$

As in the proof of Proposition 3.1, we focus on the convergence of the second term $\widehat{A}_2(\theta)$ (the third one follows similar arguments). Let us decompose $\widehat{A}_2(\theta)=\widehat{A}^1_2(\theta)+\widehat{A}^2_2(\theta)$ where

$$
\begin{aligned}
\widehat{A}^1_2(\theta) = \frac{1}{|\Lambda_n|}\sum_{u^\mu\in X_{\Lambda_n\ominus R}}\ \sum_{v^\nu\in X_{\mathcal{B}(u,R)\setminus u^\mu}} & g_\theta(u^\mu,X\setminus\{u^\mu,v^\nu\})h_\theta(v^\nu,X\setminus\{u^\mu,v^\nu\})\\
&\times\left(\frac{\lambda_\theta(u^\mu,X\setminus\{u^\mu,v^\nu\})\lambda_\theta(v^\nu,X\setminus\{u^\mu,v^\nu\})}{\lambda_\theta(\{u^\mu,v^\nu\},X\setminus\{u^\mu,v^\nu\})}-1\right)\\
\widehat{A}^2_2(\theta) = \frac{1}{|\Lambda_n|}\sum_{u^\mu\in X_{\Lambda_n\setminus\Lambda_n\ominus R}}\ \sum_{v^\nu\in X_{\Lambda_n\cap\mathcal{B}(u,R)\setminus u^\mu}} & g_\theta(u^\mu,X\setminus\{u^\mu,v^\nu\})h_\theta(v^\nu,X\setminus\{u^\mu,v^\nu\})\\
&\times\left(\frac{\lambda_\theta(u^\mu,X\setminus\{u^\mu,v^\nu\})\lambda_\theta(v^\nu,X\setminus\{u^\mu,v^\nu\})}{\lambda_\theta(\{u^\mu,v^\nu\},X\setminus\{u^\mu,v^\nu\})}-1\right).
\end{aligned}
$$

Since $|\Lambda_n \ominus R|/|\Lambda_n| \sim 1$ as $n \to \infty$, we get the following almost sure convergence

$$\begin{aligned}
\widehat{A}_2^1(\theta) \to \mathbf{E}\Big[\sum_{v^\nu \in X_{\mathcal{B}(0,R)}} & g_\theta(0^M, X \setminus v^\nu) h_\theta(v^\nu, X \setminus v^\nu) \times \\
& \times \left(\frac{\lambda_\theta(0^M, X \setminus v^\nu)\lambda_\theta(v^\nu, X \setminus v^\nu)}{\lambda_\theta(\{0^M, v^\nu\}, X \setminus v^\nu)} - 1 \right) \lambda_{\theta^\star}(0^M, X) \Big] \\
= \mathbf{E}\Big[\int_{\mathcal{B}(0,R)\times\mathbb{M}} & g_\theta(0^M, X) h_\theta(v^\nu, X) \\
& \times \left(\frac{\lambda_\theta(0^M, X)\lambda_\theta(v^\nu, X)}{\lambda_\theta(\{0^M, v^\nu\}, X)} - 1 \right) \underbrace{\lambda_{\theta^\star}(0^M, X \cup v^\nu)\lambda_{\theta^\star}(v^\nu, X)}_{=\lambda_{\theta^\star}(\{0^M, v^\nu\}, X)} \mathrm{d}v^\nu \Big] \\
=: A_2(\theta).
\end{aligned}$$

Now, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ the following holds almost surely

$$\begin{aligned}
|\widehat{A}_2^2(\theta)| \leq \frac{1}{|\Lambda_n|} \sum_{u^\mu \in X_{\Lambda_n \setminus (\Lambda_n \ominus R)}} \sum_{v^\nu \in X_{\mathcal{B}(u,R)} \setminus u^\mu} & \Big| g_\theta(u^\mu, X \setminus \{u^\mu, v^\nu\}) h_\theta(v^\nu, X \setminus \{u^\mu, v^\nu\}) \\
& \times \left(\frac{\lambda_\theta(u^\mu, X \setminus \{u^\mu, v^\nu\})\lambda_\theta(v^\nu, X \setminus \{u^\mu, v^\nu\})}{\lambda_\theta(\{u^\mu, v^\nu\}, X \setminus \{u^\mu, v^\nu\})} - 1 \right) \Big| \\
\leq 2 \, \frac{|\Lambda_n \setminus (\Lambda_n \ominus R)|}{|\Lambda_n|} & I_2(g_\theta, h_\theta) \\
\to 0.
\end{aligned}$$

In the previous equations $I_2(g_\theta, h_\theta)$ given by (3.4). With similar arguments, we may prove that $\widehat{A}_3(\theta) \to A_3(\theta)$ where

$$A_3(\theta) := \mathbf{E}\Big[\int_{\mathcal{B}(0,R)\times\mathbb{M}} \Delta_{v^\nu} g_\theta(0^M, X) \Delta_{0^M} h_\theta(v^\nu, X) \lambda_{\theta^\star}(\{0^M, v^\nu\}, X) \mathrm{d}v^\nu \Big].$$

For any $\theta \in \mathcal{V}$, $\widehat{C}(\theta) := \sum_{j=1}^3 \widehat{A}_j(\theta)$ converges $P_{\theta^\star}$-almost surely towards $C(\theta) := \sum_{j=1}^3 A_j(\theta)$ as $n \to \infty$. Under the assumption **[H(g_θ,h_θ)]**, $\widehat{C}(\theta)$ and $C(\theta)$ are continuous functions in θ which implies $\widehat{C}(g_{\widehat{\theta}}, h_{\widehat{\theta}}) \to C(\theta^\star)$. The proof is therefore finished since $C(\theta^\star) = C(g_{\theta^\star}, h_{\theta^\star})$. □

A.4 Proof of Corollary 3.3

Proof. Since the MPLE is a strongly consistent estimate of $\theta^\star$ (Proposition 2.2), the only thing to prove is that for all $j, k = 1, \ldots, p$ the assumption **[H(v_j, v_k)]** is fulfilled. In particular we have to verify that the variables $I_i(v_j, v_k)$, $i = 1, 2, 3$ defined by (3.3)-(3.5) have finite expectation. We note that **[MPLE]** implies the local stability property, i.e. there exists $\widetilde{\lambda} < \infty$, such that for any $u^\mu, v^\nu \in \mathbb{S}$, $x \in \Omega$ and $\theta \in \Theta$ we have $\lambda_\theta(u^\mu, x) \leq \widetilde{\lambda}$ and $\lambda_\theta(\{u^\mu, v^\nu\}, X) = \lambda_\theta(u^\mu, x)\lambda_\theta(v^\nu, x \cup u^\mu) \leq \widetilde{\lambda}^2$. For ease of presentation we assume in the following that $v_i(u^\mu, x)$ satisfies (2.7) for $i = 1, \ldots, p$. Similar arguments can be used when some of $v_i(u^\mu, x)$, $i = 1, \ldots, p$

satisfy (2.8). Then for any $u^\mu, v^\nu \in \mathbb{S}$ such that $\|u - v\| \leq R$ we have $1/\lambda_\theta(u^\mu, x) \leq \exp(\tilde{\kappa} n(x_{\mathcal{B}(u,R)}))$, where $\tilde{\kappa} = \sup_\theta(-\kappa \sum_{i=1}^p \theta_i) > 0$ and

$$\frac{1}{\lambda_\theta(u^\mu, v^\nu, x)} \leq \mathrm{e}^{\tilde{\kappa}(n(x_{\mathcal{B}(u,R)})+n(x_{\mathcal{B}(v,R)}))} \leq \mathrm{e}^{\tilde{\kappa}\, n(x_{\mathcal{B}(u,2R)})}.$$

Then we derive

$$\begin{aligned}
I_1(v_j, v_k) &\leq \widetilde{\lambda}\kappa^2\, n(X_{\mathcal{B}(0,R)})^2, \\
I_2(v_j, v_k) &\leq 2\widetilde{\lambda}^3\kappa^2 \mathrm{e}^{\tilde{\kappa}\, n(x_{\mathcal{B}(0,2R)})} \int_{\mathcal{B}(0,R)} n(X_{\mathcal{B}(0,R)})\, n(X_{\mathcal{B}(v,R)}) \mathrm{d}v^\nu \\
&\leq 2\widetilde{\lambda}^3\kappa^2 |\mathcal{B}(0,R)|\, n(X_{\mathcal{B}(0,R)}) n(X_{\mathcal{B}(0,2R)}) \mathrm{e}^{\tilde{\kappa}\, n(x_{\mathcal{B}(0,2R)})} \\
I_3(v_j, v_k) &\leq 4\widetilde{\lambda}^2\kappa^2 |\mathcal{B}(0,R)|\, (1 + 2n(X_{\mathcal{B}(0,R)}))\, (1 + 2n(X_{\mathcal{B}(0,2R)})).
\end{aligned}$$

The result is therefore proved since we recall that for any spatial Gibbs point process satisfying a local stability property, we have in particular $\mathbf{E}[n(X_A)^k \mathrm{e}^{c\, n(X_A)}] < \infty$ for any integer k, constant c and bounded Borel set A (see e.g. Bertin et al., 2008, Proposition 11). □

References

A. Baddeley. *Multivariate and Marked Point Processes.* in Handbook of spatial statistics Edited by Gelfand, A.E. and Diggle, P.J. and Fuentes, M. and Guttorp, P., CRC Press, 2010.

A. Baddeley and R. Turner. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics*, 42(3):283–322, 2000.

A. Baddeley and R. Turner. Modelling spatial point patterns in `R`. *Journal of Statistical Software*, 12(6):1–42, 2005.

A. Baddeley, R. Turner, J. Møller, and M. Hazelton. Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society: Series B*, 67(5): 1–35, 2005.

A. Baddeley, J. Møller, and A.G. Pakes. Properties of residuals for spatial point processes. *Annals of the Institute of Statistical Mathematics*, 60(3):627–649, 2008.

A. Baddeley, E. Rubak, and J. Møller. Score, pseudo-score and residual diagnostics for spatial point process models. *Statistical Science*, 26(4):613–646, 2011.

K.K. Berthelsen and J. Møller. A primer on perfect simulation for spatial point processes. *Bulletin of the Brazilian Mathematical Society*, 33(3):351–367, 2002.

K.K. Berthelsen and J. Møller. Likelihood and non-parametric Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling. *Scandinavian Journal of Statistics*, 30(3):549–564, 2003.

E. Bertin, J.-M. Billiot, and R. Drouilhet. R-local Delaunay inhibition model. *Journal of Statistical Physics*, 132(4):649–667, 2008.

Julian Besag. Statistical analysis of Non-Lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 24(3):179–195, 1975.

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

J.-F. Coeurjolly and F. Lavancier. Residuals for stationary marked Gibbs point processes. *in revision*, 2012. `http://arxiv.org/abs/1002.0857`.

J.-F. Coeurjolly, D. Dereudre, R. Drouilhet, and F. Lavancier. Takacs-Fiksel method for stationary marked Gibbs point processes. *To appear in Scandinavian Journal of Statistics*, 2012. doi: 10.1111/j.1467-9469.2011.00738.x.

D. Dereudre, R. Drouilhet, and H.O. Georgii. Existence of Gibbsian point processes with geometry-dependent interactions. *Probability Theory and Related Fields*, 2011. doi 10.1007/s00440-011-0356-5.

P.J. Diggle. Displaced amacrine cells in the retina of a rabbit: analysis of a bivariate spatial point pattern. *Journal of neuroscience methods*, 18(1-2):115–125, 1986.

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

H.O. Georgii. *Gibbs measures and Phase Transitions*. de Gruyter, Berlin, 1988.

C.J. Geyer and J. Møller. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, 21(4):359–373, 1994.

F. Huang and Y. Ogata. Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *Journal of Computational and Graphical Statistics*, 8 (3):510–530, 1999.

J. Illian, A. Penttinen H., and Stoyan. *Statistical analysis and modelling of spatial point patterns*. Wiley-Interscience, 2008.

J.L. Jensen and H.R. Künsch. On asymptotic normality of pseudolikelihood estimates of pairwise interaction processes. *Annals of the Institute of Statistical Mathematics*, 46(3): 475–486, 1994.

J.L. Jensen and J. Møller. Pseudolikelihood for exponential family models of spatial point processes. *Annals of Applied Probability*, 1(3):445–461, 1991.

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

J. Møller and R.P. Waagepetersen. Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, 34(4):643–684, 2007.

X. Nguyen and H. Zessin. Integral and differential characterizations of Gibbs processes. *Mathematische Nachrichten*, 88(1):105–115, 1979a.

X.X. Nguyen and H. Zessin. Ergodic theorems for spatial processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 48:133–158, 1979b.

F. Papangelou. The Armenian Connection: Reminiscences from a time of interactions. *Journal of contemporary mathematical analysis*, 44(1):14–19, 2009.

C.J. Preston. *Random fields*. Springer Verlag, 1976.

D. Ruelle. *Statistical Mechanics*. Benjamin, New York-Amsterdam, 1969.

D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic geometry and its applications*. John Wiley and Sons, Chichester, 1995.

H. Zessin. Der papangelou prozess. *Journal of contemporary mathematical analysis*, 44(1): 36–44, 2009.