HAL
open science

Poisson intensity parameter estimation for stationary Gibbs point processes of finite interaction range

Jean-François Coeurjolly, Nadia Morsli

To cite this version:

Jean-François Coeurjolly, Nadia Morsli. Poisson intensity parameter estimation for stationary Gibbs point processes of finite interaction range. Spatial Statistics, 2013, 4, pp.45-56. 10.1016/j.spasta.2013.02.003 . hal-00742374v2

HAL Id: hal-00742374

https://hal.science/hal-00742374v2

Submitted on 22 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Poisson intensity parameter estimation for stationary Gibbs point processes of finite interaction range

Jean-François Coeurjolly and Nadia Morsli
Jean-Francois.Coeurjolly@upmf-grenoble.fr; nadia.morsli@imag.fr
Laboratory Jean Kuntzmann, Grenoble University, France.

Abstract

We introduce a semi-parametric estimator of the Poisson intensity parameter of a spatial stationary Gibbs point process. Under very mild assumptions satisfied by a large class of Gibbs models, we establish its strong consistency and asymptotic normality. We also consider its finite-sample properties in a simulation study.

Keywords: central limit theorem; Georgii-Nguyen-Zessin formula; Papangelou conditional intensity; semi-parametric estimation.

1 Introduction

Spatial Gibbs point processes form a major class of stochastic processes allowing the modelling of dependence of spatial point patterns. They have applications in many scientific fields such as biology, epidemiology, geography, astrophysics, physics, and economics. General references covering as well theoretical as practical aspects of these processes are e.g. [23], [16] or [14].

Gibbs point processes in $\mathbb{R}^d$ can be defined and characterized through the Papangelou conditional intensity (see [16]) which is a function $\lambda : \mathbb{R}^d \times \Omega \to \mathbb{R}_+$ where Ω is the space of locally finite configurations of points in $\mathbb{R}^d$. The Papangelou conditional intensity can be interpreted as follows: for any $u \in \mathbb{R}^d$ and $\mathbf{x} \in \Omega$, $\lambda(u, \mathbf{x})\mathrm{d}u$ corresponds to the conditional probability of observing a point in a ball of volume $\mathrm{d}u$ around u given the rest of the point process is $\mathbf{x}$. In the present paper, we assume that the Papangelou conditional intensity is decomposed as

$$\lambda(u, \mathbf{x}) = \beta\, \widetilde{\lambda}(u, \mathbf{x}), \tag{1.1}$$

where β is a positive real parameter and where $\widetilde{\lambda} : \mathbb{R}^d \times \Omega \to \mathbb{R}_+$. If the point process corresponds to a homogeneous Poisson point process $\widetilde{\lambda}(u, \mathbf{x}) = 1$. We further assume that $\widetilde{\lambda}(u, \emptyset) = 1$ which in case means that $\beta = \lambda(u, \emptyset)$ and that $\widetilde{\lambda}$ represents the higher order interaction term. Therefore, we suggest the name *Poisson intensity*

parameter for the parameter β. The goal of this paper is to develop an estimator of β without specifying anything on the function $\widetilde{\lambda}$. Our semi-parametric estimator is based on a smart application of the Georgii-Nguyen-Zessin formula (see [11] for a general presentation and Section 2.1) which is also a way to characterize a Gibbs point process. We propose a very simple ratio-estimator of β based on a single observation of a stationary spatial Gibbs point process observed in a bounded window and we prove its asymptotic properties (strong consistency and asymptotic normality) as the window expands to $\mathbb{R}^d$. The problem considered in this paper is related to two different works that we detail below.

First, due to its easy interpretability the most popular Gibbs model is without any doubt the (isotropic) pairwise interaction point process. Its Papangelou conditional intensity is given by

$$\lambda(u, \mathbf{x}) = \beta\widetilde{\lambda}(u, \mathbf{x}) \quad \text{with} \quad \widetilde{\lambda}(u, \mathbf{x}) = \prod_{v \in \mathbf{x} \setminus u} g(\|v - u\|). \tag{1.2}$$

The function g is called the (isotropic) pairwise interaction function. Estimating the model (1.2) consists in estimating β and the function g. To the best of our knowledge, this awkward problem has been considered only in [10]. The authors propose several methods to estimate non parametrically the function g. Our objective is to make some advances in that direction which justifies first to propose an estimator of β independently of the function g and to understand its properties.

Second, many popular methods exist to estimate a parametric Gibbs model. This includes methods based on the likelihood or the pseudo-likelihood and the Takacs-Fiksel method. We refer for instance to [16], [14] or [15] and the references therein for an overview of these methods. A recent contribution [1] investigates a new approach. The following exponential model is considered

$$\lambda(u, \mathbf{x}) = \beta\widetilde{\lambda}(u, \mathbf{x}) \quad \text{with} \quad \log \widetilde{\lambda}(u, \mathbf{x}) = \theta^\top v(u, \mathbf{x}) \tag{1.3}$$

where θ is a real p-dimensional parameter vector, $v(u, \mathbf{x}) = (v_1(u, \mathbf{x}), \ldots, v_p(u, \mathbf{x}))^\top$ for measurable functions $v_i : \mathbb{R}^d \times \Omega \to \mathbb{R}$, $i = 1, \ldots, p$. [1] develops a new estimator of the parameter θ where the parameter β is treated as a nuisance parameter. The main advantage of the estimator is that it is obtained via the resolution of a linear system of equations whereas all previous methods cited above require an optimization procedure. The ratio-estimator of β we propose is also very computationally cheap. Hence, it could serve as a complement of the estimate of θ proposed in [1] to completely identify the model (1.3).

The rest of the paper is organized as follows. Section 2 gives the background and notation, presents examples of Gibbs point processes and our main assumptions. Section 3 deals with the core of the paper. We present our estimator and derive its asymptotic properties as the window of observation expands to $\mathbb{R}^d$. A simulation study is conducted in Section 4 where we illustrate the efficiency of our estimator. Proofs of the results are postponed to Appendix.

2 Background and assumptions

2.1 Basic definitions and tools

A point process $\mathbf{X}$ in $\mathbb{R}^d$ is a locally finite random subset of $\mathbb{R}^d$, i.e. the number of points $N(\Lambda) = n(\mathbf{X}_\Lambda)$ of the restriction of $\mathbf{X}$ to Λ is a finite random variable whenever Λ is a bounded Borel set of $\mathbb{R}^d$ (see [8]). We let Ω be the space of locally finite point configurations. If the distribution of $\mathbf{X}$ is translation invariant, we say that $\mathbf{X}$ is stationary.

The Papangelou conditional intensity completely characterizes the Gibbs point process in terms of the Georgii-Nguyen-Zessin (GNZ) Formula (see [20] and [24] for historical comments and [11] or [18] for a general presentation). The GNZ formula states that for any measurable function $h : \mathbb{R}^d \times \Omega$ such that the left or right hand side exists

$$\mathrm{E} \sum_{u \in \mathbf{X}} h(u, \mathbf{X} \setminus u) = \mathrm{E} \int_{\mathbb{R}^d} h(u, \mathbf{X}) \lambda(u, \mathbf{X}) \mathrm{d}u. \tag{2.1}$$

We will not discuss how to consistently specify the Papangelou conditional intensity to ensure the existence of a Gibbs point process on $\mathbb{R}^d$, 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 [22, 21] or [9], see also Section 2.2 for a few examples.

The concept of innovation for spatial point processes is proposed in [4] and is a key-ingredient in this paper. Inspired by the GNZ formula, it is defined as follows: for some measurable function $h : \mathbb{R}^d \times \Omega \to \mathbb{R}$, the $h-$weighted innovation computed in a bounded domain Λ is the centered random variable defined by

$$I_\Lambda(\mathbf{X}, h) = \sum_{u \in \mathbf{X}_\Lambda} h(u, \mathbf{X} \setminus u) - \int_\Lambda h(u, \mathbf{X}) \lambda(u, \mathbf{X}) \mathrm{d}u. \tag{2.2}$$

We end this paragraph with a few notation. The volume of a bounded Borel set Λ of $\mathbb{R}^d$ is denoted by $|\Lambda|$. The norm $\|\cdot\|$ stands for the standard Euclidean norm. For two subsets Λ_1 and Λ_2 of $\mathbb{R}^d$, $d(\Lambda_1, \Lambda_2)$ corresponds to the minimal distance between these two subsets. Finally, $B(u, R)$ denotes the closed ball centered at $u \in \mathbb{R}^d$ with radius R.

2.2 Examples of Gibbs point processes

We present in this paragraph several classical examples through their Papangelou conditional intensity. Further details on these models can be found in [16] for examples (i)-(iv) and (vi)-(vii) and to [13] for the example (v). Let $u \in \mathbb{R}^d, \mathbf{x} \in \Omega$ and $R > 0$.

(i) Strauss point process.

$$\lambda(u, \mathbf{x}) = \beta \gamma^{n_{[0,R]}(u, \mathbf{x} \setminus u)}$$

where $\beta > 0, \gamma \in [0, 1]$ and $n_{[0,R]}(u, \mathbf{x}) = \sum_{v \in \mathbf{x}} \mathbb{1}(\|v - u\| \leq R)$ represents the number of R-closed neighbours of u in $\mathbf{x}$.

(ii) Strauss point process with hard-core. Let $0 < \delta < R$ the Papangelou conditional intensity is

$$\lambda(u, \mathbf{x}) = \begin{cases} 0 & \text{if } d(u, \mathbf{x} \setminus u) \leq \delta \\ \beta\gamma^{n_{]\delta,R]}(u,\mathbf{x}\setminus u)} & \text{otherwise} \end{cases}$$

where $\beta > 0, \gamma \in [0, 1]$.

(iii) Piecewise Strauss point process.

$$\lambda(u, \mathbf{x}) = \beta \prod_{j=1}^{p} \gamma_j^{n_{[R_{j-1},R_j]}(u,\mathbf{x}\setminus u)}$$

where $\beta > 0, \gamma_j \in [0, 1], n_{[R_{j-1},R_j]}(u, \mathbf{x}) = \sum_{v\in\mathbf{x}} \mathbb{1}(\|v - u\| \in [R_{j-1}, R_j])$ and $R_0 = 0 < R_1 < \ldots < R_p = R < \infty$.

(iv) Triplets point process.

$$\lambda(u, \mathbf{x}) = \beta\gamma^{s_{[0,R]}(\mathbf{x}\cup u)-s_{[0,R]}(\mathbf{x}\setminus u)}$$

where $\beta > 0, \gamma \in [0, 1]$ and $s_{[0,R]}(\mathbf{x})$ is the number of unordered triplets that are closer than R.

(v) Geyer saturation point process with saturation threshold s

$$\lambda(u, \mathbf{x}) = \beta\gamma^{t(\mathbf{x}\cup u)-t(\mathbf{x}\setminus u)}$$

where $\beta > 0, \gamma \in [0, 1], s \geq 1$ and $t(\mathbf{x}) = \sum_{v\in\mathbf{x}} \min(s, n_{[0,R/2]}(v, \mathbf{x} \setminus v))$.

(vi) Lennard-Jones model with finite range. This corresponds to the pairwise interaction point process defined by

$$\lambda(u, \mathbf{x}) = \beta \prod_{v\in\mathbf{x}\setminus u} g(\|v - u\|))$$

with for $r > 0$, $\log g(r) = \left(\theta^6 r^{-6} - \theta^{12} r^{-12}\right)\mathbb{1}_{]0,R]}(r)$, $\theta > 0$ and $\beta > 0$.

(vii) Area-interaction point process.

$$\lambda(u, \mathbf{x}) = \beta\gamma^{A(\mathbf{x}\cup u)-A(\mathbf{x}\setminus u)}$$

for $\beta > 0, \gamma > 0$ and where $A(\mathbf{x}) = |\cup_{v\in\mathbf{x}} B(v, R/2)|$.

Examples (i)-(iii) and (v)-(vi) are pairwise interaction point processes whereas example (iv) (resp. (vii)) is based on interactions on cliques of order 3 (resp. of any order). Example (vi) is the only example which is not locally stable but only Ruelle superstable, see [22]. We recall that the local stability property states that (for stationary Gibbs models) the Papangelou conditional intensity is uniformly bounded by a constant (see [16]).

2.3 Main assumptions and discussion

In this section, we consider the model (1.1) and we present the assumptions we require to propose in Section 3.1 an estimator of the Poisson intensity parameter.

To stress on the fact that the Papangelou conditional intensity satisfies (1.1) and that we aim at estimating β we denote, from now on, the corresponding Papangelou conditional intensity λ_β.

The two following assumptions will be considered and discussed throughout the paper.

(a) The Papangelou conditional intensity has a finite range R, i.e.

$$\lambda_\beta(u, \mathbf{x}) = \lambda_\beta(u, \mathbf{x}_{B(u,R)}), \tag{2.3}$$

for any $u \in \mathbb{R}^d$, $\mathbf{x} \in \Omega$, $\beta > 0$.

(b) The function $\widetilde{\lambda}$ satisfies for any $u \in \mathbb{R}^d$

$$\widetilde{\lambda}(u, \emptyset) = 1. \tag{2.4}$$

As this paper deals with stationary Gibbs point processes, we stress on the fact tha the Papangelou conditional intensity is invariant by translation for any $\beta > 0$. We emphasize that no more assumption will be required. In particular we do not require that the Gibbs point process is locally stable. This is noticeable since local stability is often an assumption made when dealing with such models, see e.g. [5]. We also underline that in the following, we do not assume that R is known. In place of this, we assume we are given an upper-bound of R.

Let us discuss these assumptions regarding the examples presented in Section 2.2. Examples (i)-(vii) exhibit Papangelou conditional intensities which satisfy (1.1) and (2.3) with finite range R. Example (i)-(vi) satisfy the assumption (2.4), whereas example (vii) is a counter-example since for the latter one $\widetilde{\lambda}(u, \emptyset) = \gamma^{\pi(R/2)^2} \neq 1$ for $\gamma > 0$. Condition (b) can thus be viewed as an identifiability condition.

3 Semi-parametric estimator of $\beta^\star$

3.1 Definition

In the following, we let $\beta^\star > 0$ denote the true parameter and we denote by $P_{\beta^\star}$ the probability measure of $\mathbf{X}$, i.e. $\mathbf{X} \sim P_{\beta^\star}$.

Our estimator is based on an application of the GNZ formula (2.1) for the following function h given for all $u \in \mathbb{R}^d$, $\mathbf{x} \in \Omega$ and $\widetilde{R} \geq R$ by

$$h(u, \mathbf{x}) = \mathbb{1}(d(u, \mathbf{x}) > \widetilde{R}) = \mathbb{1}(\mathbf{x} \cap B(u, \widetilde{R}) = \emptyset). \tag{3.1}$$

The idea is formulated by the following result. Let Λ be a bounded region of $\mathbb{R}^d$ and let $N_\Lambda(\mathbf{X}; \widetilde{R})$ and $V_\Lambda(\mathbf{X}; \widetilde{R})$ denote the following random variables

$$N_\Lambda(\mathbf{X}; \widetilde{R}) = \sum_{u \in \mathbf{X}_\Lambda} \mathbb{1}((\mathbf{X} \setminus u) \cap B(u, \widetilde{R}) = \emptyset)$$

$$V_\Lambda(\mathbf{X}; \widetilde{R}) = \int_\Lambda \mathbb{1}(\mathbf{X} \cap B(u, \widetilde{R}) = \emptyset) \mathrm{d}u.$$

The variable N_Λ corresponds to the number of points in X_Λ that are separated from other points in X by a distance greater than $\widetilde{R}$ while the variable V_Λ is the volume of Λ after digging a disk hole of radius $\widetilde{R}$ centered at each point of X.

Proposition 3.1. *Assume* (1.1), (2.3) *and* (2.4), *then we have for any* $\widetilde{R} \geq R$

$$\mathrm{E}\, N_\Lambda(\mathbf{X}; \widetilde{R}) = \beta^\star \,\mathrm{E}\, V_\Lambda(\mathbf{X}; \widetilde{R}) \tag{3.2}$$

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

Proof. Using the GNZ formula (2.1) applied to the function h defined by (3.1), we obtain

$$\begin{aligned}\mathrm{E}\, N_\Lambda(\mathbf{X}; \widetilde{R}) &= \mathrm{E} \sum_{u \in \mathbf{X}_\Lambda} \mathbb{1}((\mathbf{X} \setminus u) \cap B(u, \widetilde{R}) = \emptyset) \\ &= \beta^\star \,\mathrm{E} \int_\Lambda \mathbb{1}(\mathbf{X} \cap B(u, \widetilde{R}) = \emptyset) \widetilde{\lambda}(u, \mathbf{X}) \mathrm{d}u.\end{aligned}$$

Now, using the finite range property (2.3), the assumption (2.4) and the fact that $\widetilde{R} \geq R$ we continue with

$$\begin{aligned}\mathrm{E}\, N_\Lambda(\mathbf{X}; \widetilde{R}) &= \beta^\star \,\mathrm{E} \int_\Lambda \mathbb{1}(\mathbf{X} \cap B(u, \widetilde{R}) = \emptyset) \widetilde{\lambda}(u, \mathbf{X} \cap B(u, R)) \mathrm{d}u \\ &= \beta^\star \,\mathrm{E} \int_\Lambda \mathbb{1}(\mathbf{X} \cap B(u, \widetilde{R}) = \emptyset) \widetilde{\lambda}(u, \emptyset) \mathrm{d}u \\ &= \beta^\star \,\mathrm{E}\, V_\Lambda(\mathbf{X}; \widetilde{R}).\end{aligned}$$

□

Using a classical ergodic theorem for spatial point processes obtained in [19], we can expect that when the observation domain is large $N_\Lambda(\mathbf{X}; \widetilde{R}) \simeq \mathrm{E}\, N_\Lambda(\mathbf{X}; \widetilde{R})$ and $V_\Lambda(\mathbf{X}; \widetilde{R}) \simeq \mathrm{E}\, V_\Lambda(\mathbf{X}; \widetilde{R})$. Using this and Proposition 3.1, we are suggested a natural and computationally cheap ratio-estimator of the parameter $\beta^\star$. We define this estimator and present its asymptotic properties in the next section.

3.2 Asymptotic properties

To derive asymptotic properties of our procedure when the window expands to $\mathbb{R}^d$, we assume that $\mathbf{X}$ is observed in Λ_n where $(\Lambda_n)_{n \geq 1}$ is a sequence of cubes growing up to $\mathbb{R}^d$.

When u is a point close to the boundary of the window Λ_n, we can only evaluate $h(u, \mathbf{X} \cap \Lambda_n \cap B(u, \widetilde{R}))$. One simple strategy for eliminating the edge effect bias is the border method, i.e. by considering the erosion of Λ_n by $\widetilde{R}$ defined by

$$\Lambda_{n,\widetilde{R}} = \Lambda_{n \ominus \widetilde{R}} = \{u \in \Lambda_n : B(u, \widetilde{R}) \subseteq \Lambda_n\}.$$

Then we define our estimator by

$$\widehat{\beta}_n = \widehat{\beta}_n(\mathbf{X}; \widetilde{R}) = \frac{N_{\Lambda_{n,\widetilde{R}}}(\mathbf{X}; \widetilde{R})}{V_{\Lambda_{n,\widetilde{R}}}(\mathbf{X}; \widetilde{R})}. \tag{3.3}$$

We let F denote the empty space function (see [16]) given, in the stationary case, by

$$F(r) = P_{\beta^\star}(\mathbf{X} \cap B(0, r) \neq \emptyset) = 1 - P_{\beta^\star}(\mathbf{X} \cap B(0, r) = \emptyset)$$

for $r > 0$. Following this definition, by $F_{u,v} : \mathbb{R}_+ \to [0,1]$ for any $u, v \in \mathbb{R}^d$ we denote the function given by

$$F_{u,v}(r) = 1 - P_{\beta^\star}\big(\mathbf{X} \cap B(u,r) = \emptyset, \mathbf{X} \cap B(v,r) = \emptyset\big)$$

for $r > 0$. We now state our main result.

Proposition 3.2. *Assume* (1.1), (2.3) *and* (2.4), *and let* $\widetilde{R} \geq R$. *Then we have the following statements as* $n \to \infty$.
(i) $\widehat{\beta}_n$ *is a strongly consistent estimator of* $\beta^\star$.
(ii) If $P_{\beta^\star}$ *is ergodic, the following convergence in distribution holds*

$$|\Lambda_{n,\widetilde{R}}|^{1/2}(\widehat{\beta}_n - \beta^\star) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \tag{3.4}$$

where

$$\sigma^2 = \sigma^2(\beta^\star, \widetilde{R}) = \frac{\beta^\star}{1 - F(\widetilde{R})} + \frac{\beta^{\star 2}}{(1 - F(\widetilde{R}))^2} \int_{B(0,\widetilde{R})} (1 - F_{0,v}(\widetilde{R}))\mathrm{d}v. \tag{3.5}$$

(iii) The following convergence in distribution holds

$$|\Lambda_{n,\widetilde{R}}|^{1/2} \, \frac{(\widehat{\beta}_n - \beta^\star)}{\widehat{\sigma}_n} \xrightarrow{d} \mathcal{N}(0,1) \tag{3.6}$$

where

$$\widehat{\sigma}_n^2 = \widehat{\sigma}_n^2(\mathbf{X}; \widetilde{R}) = |\Lambda_{n,\widetilde{R}}| \left(\frac{\widehat{\beta}_n}{V_{\Lambda_{n,\widetilde{R}}}(\mathbf{X}; \widetilde{R})} + \frac{\widehat{\beta}_n^2 \, W_{\Lambda_{n,\widetilde{R}}}(\mathbf{X}; \widetilde{R})}{V^2_{\Lambda_{n,\widetilde{R}}}(\mathbf{X}; \widetilde{R})} \right) \tag{3.7}$$

is a strongly consistent estimator of σ^2 *where*

$$W_{\Lambda_{n,\widetilde{R}}}(\mathbf{X}; \widetilde{R}) = \int_{\Lambda_{n,\widetilde{R}}} \int_{B(u,\widetilde{R}) \cap \Lambda_{n,\widetilde{R}}} \mathbb{1}(\mathbf{X} \cap B(u, \widetilde{R}) = \emptyset) \mathbb{1}(\mathbf{X} \cap B(v, \widetilde{R}) = \emptyset) \mathrm{d}u \mathrm{d}v.$$

If the measure $P_{\beta^\star}$ is not ergodic, then it can be represented as a mixture of ergodic measures (see [12, Theorem 14.10]). In this case, the asymptotic distribution in (3.4) becomes a mixture of Gaussian distributions. Since a mixture of standard Gaussian distribution is a standard Gaussian distribution, this explains why $P_{\beta^\star}$ is assumed to be ergodic for the result (ii) and omitted for the result (iii).

We underline the interesting and unexpected fact that the asymptotic variance $\sigma^2(\beta^\star, \widetilde{R})$ (and its estimate) are independent on $\widetilde{\lambda}$ which is the unspecified part of the Papangelou conditional intensity.

4 Simulation study

In this section, we investigate the efficiency of the Poisson intensity estimate in a simulation study. We consider the models (i)-(v) described in Section 2.2. We set $R = 0.05$ and $\beta^\star = 200$ for all the examples except for the model `g2` where we set $\beta^\star = 50$. We consider

- Strauss point process where models `s1` and `s2` have respectively $\gamma = 0.2$ and $\gamma = 0.8$.
- Strauss point process with hard-core with $\delta = R/2$ and where models `shc1` and `shc2` have respectively $\gamma = 0.2$ and $\gamma = 0.8$.
- Piecewise Strauss point process with $p = 3$ and $(R_1, R_2, R_3) = (R/3, 2R/3, R)$ and where models `ps1` and `ps2` have respectively $(\gamma_1, \gamma_2, \gamma_3) = (0.8, 0.5, 0.2)$ and $(\gamma_1, \gamma_2, \gamma_3) = (0.2, 0.8, 0.2)$.
- Triplets point process where models `t1` and `t2` have respectively $\gamma = 0.2$ and $\gamma = 0.8$.
- Geyer saturation point process with saturation threshold $s = 1$ where models `g1` and `g2` have respectively $\gamma = 0.5$ and $\gamma = 1.5$.

The estimate of the Poisson intensity we propose depends on a parameter $\widetilde{R}$ and ought to work for any value larger than R. To illustrate this, $\beta^\star$ is estimated for different values of $\widetilde{R}$. We consider $\widetilde{R} = R \times p$, where p is a parameter varying from 0.9 to 1.2. Each replication has been generated on the domain $[0, L]^2$ and the estimates are computed on $[\widetilde{R}, L - \widetilde{R}]^2$. To illustrate the convergence as the window grows up, the simulations are done with $L = 1$ and $L = 2$. The simulations are done using the `R` package `spatstat` ([3]). The empirical results based on 500 replications are presented in Table 1.

The results are very satisfactory. We observe that for values of $\widetilde{R} \geq R$ (i.e. for values equal or larger than the finite range), the estimates are almost unbiased whereas a non negligible bias is observed when $\widetilde{R} < R$. The bias is negative when the point process is repulsive (i.e. when $\lambda(u, \mathbf{x}) \geq 1$) and positive when the point process is attractive (i.e. when $\lambda(u, \mathbf{x}) \leq 1$) in particular for the model `g2`. As underlined by one of the reviewers the sign of the bias when $\widetilde{R} \geq R$ is violated can be explained by the fact that V is underestimated for attractive point processes and overestimated for repulsive point processes. The closer $\widetilde{R}$ is to the finite range R, the lower the standard deviations are. The number of points with no $\widetilde{R}$-close neighbours is lower and lower as $\widetilde{R}$ grows up, which explains this variance increase. Finally, we can also highlight that the asymptotic results are partly checked. When the window varies from $[0, 1]^2$ to $[0, 2]^2$, the standard deviations of the estimates are divided by 2 which turns out to be the square root of the ratios of the domains areas.

Table 1 points out the dangers to choose a wrong value for the finite range parameter R. From a practical point of view, if this parameter is not set by the user, it has to be estimated. When the Papangelou conditional intensity is entirely modeled the parameters can be estimated using for instance the maximum likelihood or the pseudo-likelihood (e.g. [16]) and the irregular parameters (such as the finite range R) can be estimated using the profile pseudo-likelihood ([2]). Such an approach can not be used here since we do not want to model the higher order interaction terms. We noticed that for a given (repulsive) point pattern our estimate is increasing with the value of $\widetilde{R}$ until a certain value where the regime is changing. Therefore we propose to estimate the finite range parameter R as follows: a) Select a reasonable grid of values for $\widetilde{R}$ (we chose $[0.02, 0.08]$ in our example). b) Compute $\widehat{\beta}(\widetilde{R})$ for each value of the grid. c) We consider a piecewise linear regression of $\widehat{\beta}(\widetilde{R})$ in terms

of $\widetilde{R}$ and we define $\widehat{R}$ as the breakpoint on the slope (for this step we used the `R` package `segmented` [17]). Empirical results for the models `s1` and `s2` are presented in Table 2. Using the strategy described above, we manage to estimate R quite well. The resulting estimates of the parameter $\beta^\star$ seem to be still unbiased. We note that the standard deviation is greater than in the case where R is exaclty known (i.e. column $p = 1$ in Table 1). However the loss of efficiency is quite reasonable.

Finally, we investigate the finite sample properties of the estimator of the asymptotic variance of $\widehat{\beta}$ given by (3.7) and the asymptotic normality result (3.6). Table 3 focuses on the models `s1` and `s2` and displays the 95% empirical coverage rates (i.e. the fraction of asymptotic confidence intervals constructed from the asymptotic normality result (3.6) covering the true parameter value, here $\beta^\star = 200$). As expected, the theoretical coverage rate is respected whatever the value of $\widetilde{R} \geq R$ and far from the result when $\widetilde{R} = 90\%R$. Depite Proposition 3.2 does not cover this case, it is interesting to note that the empirical results are still quite satisfactory if we replace R by its estimated value.

			Mean (std dev.) for $\widetilde{R} = R \times p$			
Model		$\overline{n}$	$p = 0.9$	$p = 1$	$p = 1.1$	$p = 1.2$
`s1`	$L = 1$	99	174.0 (26.2)	203.6 (34.2)	204.3 (38.3)	205.5 (44.4)
	$L = 2$	393	172.6 (12.7)	200.5 (16.4)	200.3 (18.5)	200.6 (20.8)
`s2`	$L = 1$	156	192.0 (31.3)	203.3 (41.8)	203.9 (48.1)	204.7 (55.8)
	$L = 2$	622	190.7 (16.0)	200.7 (19.5)	200.3 (22.8)	200.4 (26.5)
`shc1`	$L = 1$	94	174.3 (23.2)	201.9 (30.3)	203.3 (36.8)	205.0 (41.9)
	$L = 2$	379	174.1 (12.1)	201.6 (15.7)	201.9 (17.8)	202.1 (21.1)
`shc2`	$L = 1$	130	194.3 (30.2)	205.4 (36.7)	207.0 (42.8)	209.1 (51.6)
	$L = 2$	514	190.2 (15.2)	198.5 (17.8)	199.5 (21.1)	200.6 (24.4)
`ps1`	$L = 1$	111	172.2 (25.1)	202.6 (33.2)	203.5 (38.7)	206.5 (47.2)
	$L = 2$	445	171.5 (12.4)	201.6 (16.3)	201.7 (18.6)	202.4 (21.6)
`ps2`	$L = 1$	134	191.1 (30.7)	201.7 (37.7)	206.2 (45.2)	208.8 (54.2)
	$L = 2$	535	192.8 (14.5)	201.7 (17.5)	202.1 (20.6)	202.0 (24.4)
`t1`	$L = 1$	159	181.6 (34.6)	204.5 (40.6)	205.2 (48.7)	205.8 (58.8)
	$L = 2$	634	180.5 (16.7)	201.6 (19.5)	202.1 (22.9)	202.8 (26.7)
`t2`	$L = 1$	178	188.7 (30.7)	200.7 (35.0)	202.4 (39.9)	205.1 (50.2)
	$L = 2$	727	189.3 (17.5)	200.0 (20.2)	199.8 (24.5)	199.8 (30.8)
`g1`	$L = 1$	110	174.8 (26.6)	201.4 (34.4)	201.5 (40.2)	203.6 (46.0)
	$L = 2$	440	175.0 (13.3)	200.7 (17.4)	201.3 (19.9)	201.6 (23.1)
`g2`	$L = 1$	70	54.5 (10.4)	50.0 (10.4)	50.2 (11.4)	50.5 (12.1)
	$L = 2$	276	54.4 (5.4)	50.2 (5.4)	50.2 (5.8)	50.3 (6.3)

Table 1: Mean and standard deviations of estimates of the Poisson intensity parameter $\beta^\star$ for different parameters $\widetilde{R}$ based on 500 replications of different models generated on the window $[0, L]^2$ and estimated on the window $[\widetilde{R}, L - \widetilde{R}]^2$ for $L = 1, 2$. The second column, $\overline{n}$, indicates the average number of points (over the 500 replications) falling into $[0, L]^2$, i.e. the Monte-Carlo estimation of $\mathrm{E}\, n(\mathbf{X} \cap [0, L]^2)$.

		$\widehat{R}$ mean (std dev.)	$\widehat{\beta}$ mean (std dev.)
s1	$L=1$	0.052 (0.004)	197.7 (38.8)
	$L=2$	0.051 (0.003)	198.9 (22.1)
s2	$L=1$	0.051 (0.004)	201.7 (38.9)
	$L=2$	0.051 (0.003)	198.4 (21.9)

Table 2: Mean and standard deviations of estimates of the finite range parameter and related Poisson intensity parameter estimates (i.e. $\widehat{\beta}(\widehat{R})$) based on 500 replications of Strauss models on the window $[0, L]^2$.

		$\widetilde{R} = p \times R$				$\widehat{R}$
		$p = 0.9$	$p = 1$	$p = 1.1$	$p = 1.2$	
s1	$L=1$	77.4%	94.8%	95.0%	94.2%	89.2%
	$L=2$	40.0%	94.8%	93.8%	95.4%	90.4%
s2	$L=1$	82.6%	93.2%	94.6%	93.8%	93%
	$L=2$	69.8%	95.0%	93.6%	92.8%	90.8%

Table 3: Empirical coverage rates (i.e. the fraction of 95% confidence intervals covering the true parameter value $\beta^\star = 200$) based on 500 replications of Strauss models on the window $[0, L]^2$.

A Proof of Proposition 3.2

We use the following additional notation in the proofs. The notation $\sum^{\neq}$ stands for summation over distinct pairs of points, Λ^c denotes the complementary set of Λ in $\mathbb{R}^d$. Let $\tau = (\tau_x)_{x \in \mathbb{R}^d}$ be the shift group, where τ_x is the translation by the vector $-x \in \mathbb{R}^d$. For brevity, when there is no ambiguity, we skip the dependence on $\mathbf{X}$ and on $\widetilde{R}$ of the variables $N_{\Lambda_{n,\widetilde{R}}}, V_{\Lambda_{n,\widetilde{R}}}, W_{\Lambda_{n,\widetilde{R}}}, \widehat{\sigma}_n^2$.

Proof. (i) There exists 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 [12, Theorem 14.10]). Therefore, we can assume, for this proof, that $P_{\beta^\star}$ is ergodic. We apply a general ergodic theorem for spatial point processes obtained by [18]. For this, we have to check that N_Λ and V_Λ are additive, invariant by translation and satisfy $\mathrm{E}\, N_{\Lambda_0} < \infty$ and $\mathrm{E}\, V_{\Lambda_0} < \infty$, where for instance Λ_0 is the unit cube $[0, 1]^d$. The additivity property is straightforward. Regarding the invariance by translation,

let $\mathbf{x} \in \Omega$, $\Lambda \subset \mathbb{R}^d$ and $y \in \mathbb{R}^d$, we have

$$\begin{aligned} N_{\tau_y\Lambda}(\tau_y\mathbf{x}; \widetilde{R}) &= \sum_{u\in\tau_y\mathbf{x}} \mathbb{1}_{\tau_y\Lambda}(u)\mathbb{1}(d(u, \tau_y\mathbf{x} \setminus u) > \widetilde{R}) \\ &= \sum_{u\in\tau_y\mathbf{x}} \mathbb{1}_{\Lambda}(u+y)\mathbb{1}(d(u, \tau_y\mathbf{x} \setminus u) > \widetilde{R}) \\ &= \sum_{v\in\mathbf{x}} \mathbb{1}_{\Lambda}(v)\mathbb{1}(d(v-y, \tau_y\mathbf{x} \setminus (v-y)) > \widetilde{R}) \\ &= \sum_{v\in\mathbf{x}} \mathbb{1}_{\Lambda}(v)\mathbb{1}(d(v, \mathbf{x} \setminus v) > \widetilde{R}) = N_\Lambda(\mathbf{x}; \widetilde{R}). \end{aligned}$$

Following similar arguments, we also have $V_{\tau_y\Lambda}(\tau_y\mathbf{x}; \widetilde{R}) = V_\Lambda(\mathbf{x}; \widetilde{R})$. Let us now focus on the integrability conditions. First, we have

$$\mathrm{E}\left|V_{\Lambda_0}\right| = \mathrm{E}\, V_{\Lambda_0} = \mathrm{E}\int_{\Lambda_0} \mathbb{1}(\mathbf{X} \cap B(u, \widetilde{R}) = \emptyset)\mathrm{d}u = |\Lambda_0|(1 - F(\widetilde{R})) = 1 - F(\widetilde{R}) < \infty.$$

Second, from Proposition 3.1

$$\mathrm{E}\left|N_{\Lambda_0}\right| = \mathrm{E}\, N_{\Lambda_0} = \beta^\star\, \mathrm{E}\, V_{\Lambda_0} = \beta^\star(1 - F(\widetilde{R})).$$

Therefore as $n \to \infty$, we derive the following almost sure convergences

$$\begin{aligned} |\Lambda_{n,\widetilde{R}}|^{-1} N_{\Lambda_{n,\widetilde{R}}} &\longrightarrow \mathrm{E}\, N_{\Lambda_0} = \beta^\star(1 - F(\widetilde{R})) \\ |\Lambda_{n,\widetilde{R}}|^{-1} V_{\Lambda_{n,\widetilde{R}}} &\longrightarrow \mathrm{E}\, V_{\Lambda_0} = 1 - F(\widetilde{R}). \end{aligned} \tag{A.1}$$

Since the function $(y, z) \mapsto \frac{y}{z}$ is continuous for any (y, z) in the open set $\{(y, z) \in \mathbb{R}^2 : z \neq 0\}$, we deduce the expected result for any fixed $\widetilde{R} \geq R$.

(ii) From the definition of the innovations process (2.2), we have

$$|\Lambda_{n,\widetilde{R}}|^{1/2}(\widehat{\beta}_n - \beta^\star) = |\Lambda_{n,\widetilde{R}}|^{1/2}\frac{N_{\Lambda_{n,\widetilde{R}}} - \beta^\star V_{\Lambda_{n,\widetilde{R}}}}{V_{\Lambda_{n,\widetilde{R}}}} = \frac{|\Lambda_{n,\widetilde{R}}|^{-1/2} I_{\Lambda_{n,\widetilde{R}}}(\mathbf{X}, h)}{|\Lambda_{n,\widetilde{R}}|^{-1} V_{\Lambda_{n,\widetilde{R}}}}. \tag{A.2}$$

Henceforth from Slutsky's Theorem and (A.1), the result will be deduced if we prove that $|\Lambda_{n,\widetilde{R}}|^{-1/2} I_{\Lambda_{n,\widetilde{R}}}(\mathbf{X}, h)$ tends to a zero mean Gaussian distribution. This will be done by combining two papers [6] and [7] dealing with asymptotic normality and variance calculation of innovations processes. For brevity, we write I_Λ in place of $I_\Lambda(\mathbf{X}, h)$.

Following [6, Lemma 3], we have to assume that $P_{\beta^\star}$ is ergodic and to check that for any bounded domain $\Lambda \subset \mathbb{R}^d$: (I) $\mathrm{E}\,|I_\Lambda|^3 < \infty$, (II) $\mathrm{E}\, I^2_{\Gamma_n} \to 0$ for any sequence of bounded domains $\Gamma_n \subset \mathbb{R}^d$ such that $\Gamma_n \to 0$ as $n \to \infty$ and (III) I_Λ depends on $X_{\Lambda\oplus\widetilde{R}}$. (III) follows from the finite range property (2.3) and the definition of h given by (3.1). We now focus on verifying (I). We have from Cauchy-Schwarz inequality

$$\mathrm{E}\,|I_\Lambda|^3 \leq 4\left(\mathrm{E}\, N_\Lambda^3 + \beta^{\star 3}\, \mathrm{E}\, V_\Lambda^3\right). \tag{A.3}$$

On the one hand, we have since for any $m \geq 1$, $h^m(u, \mathbf{x}) = h(u, \mathbf{x})$

$$\begin{aligned} \mathrm{E}\, N_\Lambda^3 = & \mathrm{E}\, N_\Lambda + 3\, \mathrm{E} \sum_{u,v \in X_\Lambda}^{\neq} h(u, \mathbf{X} \setminus u) h(v, \mathbf{X} \setminus v) \\ & + \mathrm{E} \sum_{u,v,w \in X_\Lambda}^{\neq} h(u, \mathbf{X} \setminus u) h(v, \mathbf{X} \setminus v) h(w, \mathbf{X} \setminus w). \end{aligned} \tag{A.4}$$

Let T_1, T_2 and T_3 denote the three terms of the right-hand side of (A.4). Following the proof of (i) in Proposition 3.2, $\mathrm{E}\, T_1 \leq \beta^\star |\Lambda|$. Now, for any $u, v \in \mathbb{R}^d$ and $\mathbf{x} \in \Omega$, we define the second order Papangelou conditional intensity by

$$\lambda_{\beta^\star}(u, v, \mathbf{x}) = \lambda_{\beta^\star}(u, \mathbf{x}) \lambda_{\beta^\star}(v, \mathbf{x} \cup u) = \lambda_{\beta^\star}(v, \mathbf{x}) \lambda_{\beta^\star}(u, \mathbf{x} \cup v).$$

By using iterated versions of the GNZ formula (2.1), we derive the following computations

$$\begin{aligned} T_2 &= 3\, \mathrm{E} \int_\Lambda \int_\Lambda h(u, \mathbf{X} \cup v) h(v, \mathbf{X} \cup u) \lambda_{\beta^\star}(u, v, \mathbf{X}) \mathrm{d}u \mathrm{d}v \\ &= 3\beta^{\star 2}\, \mathrm{E} \int_\Lambda \int_\Lambda \mathbb{1}(d(u, \mathbf{X} \cup v) > \widetilde{R}) \mathbb{1}(d(v, \mathbf{X} \cup u) > \widetilde{R}) \widetilde{\lambda}(u, \mathbf{X}) \widetilde{\lambda}(v, \mathbf{X} \cup u) \mathrm{d}u \mathrm{d}v \\ &= 3\beta^{\star 2}\, \mathrm{E} \int_\Lambda \int_\Lambda \mathbb{1}(d(u, \mathbf{X} \cup v) > \widetilde{R}) \mathbb{1}(d(v, \mathbf{X} \cup u) > \widetilde{R}) \widetilde{\lambda}^2(u, \emptyset) \mathrm{d}u \mathrm{d}v \\ &\leq 3\beta^{\star 2} |\Lambda|^2. \end{aligned}$$

Using similar ideas, we leave the reader to check that $T_3 \leq \beta^{\star 3} |\Lambda|^3$. On the other hand, we also check that

$$\begin{aligned} \mathrm{E}\, V_\Lambda^3 &= \mathrm{E} \int_\Lambda \int_\Lambda \int_\Lambda \mathbb{1}(d(u, \mathbf{X}) > \widetilde{R}) \mathbb{1}(d(v, \mathbf{X}) > \widetilde{R}) \mathbb{1}(d(w, \mathbf{X}) > \widetilde{R}) \mathrm{d}u \mathrm{d}v \mathrm{d}w \\ &\leq |\Lambda|^3. \end{aligned}$$

We finally obtain

$$\mathrm{E}\, |I_\Lambda|^3 \leq 4\beta^\star |\Lambda| + 12\beta^{\star 2} |\Lambda|^2 + 8\beta^{\star 3} |\Lambda|^3 < \infty$$

whereby (I) is checked. Following the proof of (I), we may prove that for any bounded domain Λ

$$\mathrm{E}\, I_\Lambda^2 \leq 2\beta^\star |\Lambda| + 4\beta^{\star 2} |\Lambda|^2$$

which proves (II). Therefore we can invoke [6, Lemma 3] to obtain the asymptotic normality of $\widehat{\beta}_n$.

The derivation of the asymptotic variance of $|\Lambda_{n,\widetilde{R}}|^{-1} \operatorname{Var} I_{\Lambda_{n,\widetilde{R}}}(\mathbf{X}, h)$ is done using Lemma B.1 in Appendix B with the function $g = h$ given by (3.1). We omit the details and leave the reader to check that h fulfills the integrability assumptions in Lemma B.1 and focus on the derivations of the quantities A_1, A_2 and A_3. First, it is clear that $A_1 = \beta^\star(1 - F(\widetilde{R}))$. Second, note that from (2.3) we have for any $v \in B^c(0, R)$ and $x \in \Omega$

$$\lambda_{\beta^\star}(0, \mathbf{x}) \lambda_{\beta^\star}(v, \mathbf{x}) - \lambda_{\beta^\star}(0, v, \mathbf{x}) = 0.$$

Since $B(0,R) \subseteq B(0,\widetilde{R})$

$$\mathrm{E}\int_{B(0,\widetilde{R})\setminus B(0,R)} \big(\lambda_{\beta^\star}(0,\mathbf{X})\lambda_{\beta^\star}(v,\mathbf{X}) - \lambda_{\beta^\star}(0,v,\mathbf{X})\big)dv = 0$$

which means that we can substitute R by $\widetilde{R}$ in the term A_2. Then, we derive

$$A_2 = \mathrm{E}\int_{B(0,\widetilde{R})} \mathbb{1}(d(0,\mathbf{X}) > \widetilde{R})\mathbb{1}(d(v,\mathbf{X}) > \widetilde{R})\big(\beta^{\star 2} - \lambda_{\beta^\star}(0,v,\mathbf{X})\big)\mathrm{d}v.$$

Third,

$$\begin{aligned} A_3 =& \mathrm{E}\int_{B(0,\widetilde{R})} \big(\mathbb{1}(d(0,\mathbf{X}\cup v) > \widetilde{R}) - \mathbb{1}(d(0,\mathbf{X}) > \widetilde{R})\big) \\ &\times \big(\mathbb{1}(d(v,\mathbf{X}\cup 0) > \widetilde{R}) - \mathbb{1}(d(v,\mathbf{X}) > \widetilde{R})\big)\lambda_{\beta^\star}(0,v,\mathbf{X})\mathrm{d}v \\ =& \mathrm{E}\int_{B(0,\widetilde{R})} \mathbb{1}(d(0,\mathbf{X}) > \widetilde{R})\mathbb{1}(d(v,\mathbf{X}) > \widetilde{R})\lambda_{\beta^\star}(0,v,\mathbf{X})\mathrm{d}v \end{aligned}$$

which leads to

$$\begin{aligned} A_2 + A_3 &= \beta^{\star 2}\,\mathrm{E}\int_{B(0,\widetilde{R})} \mathbb{1}(d(0,\mathbf{X}) > \widetilde{R})\mathbb{1}(d(v,\mathbf{X}) > \widetilde{R})\mathrm{d}v \\ &= \beta^{\star 2}\int_{B(0,\widetilde{R})} (1 - F_{0,v}(\widetilde{R}))\mathrm{d}v. \end{aligned}$$

In other words

$$|\Lambda_{n,\widetilde{R}}|^{-1}\,\mathrm{Var}\, I_{\Lambda_{n,\widetilde{R}}}(\mathbf{X},h) \to \beta^\star(1 - F(\widetilde{R})) + \beta^{\star 2}\int_{B(0,\widetilde{R})} (1 - F_{0,v}(\widetilde{R}))\mathrm{d}v$$

which leads to the result from (A.1) and (A.2).

(iii) We check that the proposed estimate $\widehat{\sigma}_n^2$ tends almost surely towards σ^2 as $n \to \infty$ (a convergence in probability would be sufficient). From (3.7), (A.1) and (3.5), we need to prove that

$$|\Lambda_{n,\widetilde{R}}|^{-1} W_{\Lambda_{n,\widetilde{R}}} \longrightarrow \int_{B(0,\widetilde{R})} (1 - F_{0,v}(\widetilde{R}))\mathrm{d}v.$$

We decompose $W_{\Lambda_{n,\widetilde{R}}}$ as $W^{(1)}_{\Lambda_{n,\widetilde{R}}} + W^{(2)}_{\Lambda_{n,\widetilde{R}}}$ where

$$\begin{aligned} W^{(1)}_{\Lambda_{n,\widetilde{R}}} &= \int_{\Lambda_{n,\widetilde{R}\ominus\widetilde{R}}}\int_{B(u,\widetilde{R})} \mathbb{1}(d(u,\mathbf{X}) > \widetilde{R})\mathbb{1}(d(v,\mathbf{X}) > \widetilde{R})\mathrm{d}u\mathrm{d}v \\ W^{(2)}_{\Lambda_{n,\widetilde{R}}} &= \int_{\Lambda_{n,\widetilde{R}}\setminus\Lambda_{n,\widetilde{R}\ominus\widetilde{R}}}\int_{B(u,\widetilde{R})\cap\Lambda_{n,\widetilde{R}}} \mathbb{1}(d(u,\mathbf{X}) > \widetilde{R})\mathbb{1}(d(v,\mathbf{X}) > \widetilde{R})\mathrm{d}u\mathrm{d}v \end{aligned}$$

where $\Lambda_{n,\widetilde{R}\ominus\widetilde{R}} = \{u \in \Lambda_{n,\widetilde{R}} : B(u,\widetilde{R}) \subseteq \Lambda_{n,\widetilde{R}}\}$. Regarding the first term, we can now apply the ergodic theorem to prove that almost surely

$$|\Lambda_{n,\widetilde{R}\ominus\widetilde{R}}|^{-1} W^{(1)}_{\Lambda_{n,\widetilde{R}}} \longrightarrow \mathrm{E}\int_{B(0,\widetilde{R})} \mathbb{1}(d(0,\mathbf{X}) > \widetilde{R})\mathbb{1}(d(v,\mathbf{X}) > \widetilde{R})\mathrm{d}v$$

where the latter comes from the finite range assumption (2.3) and the fact that $\widetilde{R} \geq R$. Since $|\Lambda_{n,\widetilde{R}}| \sim |\Lambda_{n,\widetilde{R}\ominus\widetilde{R}}|$ as $n \to \infty$, the proof is done by verifying that $|\Lambda_{n,\widetilde{R}}|^{-1} W^{(2)}_{\Lambda_{n,\widetilde{R}}} \to 0$ which is straightforward since

$$|\Lambda_{n,\widetilde{R}}|^{-1} W^{(2)}_{\Lambda_{n,\widetilde{R}}} \leq \frac{|\Lambda_{n,\widetilde{R}} \setminus \Lambda_{n,\widetilde{R}\ominus\widetilde{R}}|}{|\Lambda_{n,\widetilde{R}}|} |B(u,\widetilde{R})| = o(1).$$

□

B Auxiliary lemma

The following result is a particular case of [7, Proposition 3.2].

Lemma B.1. *Assume the Papangelou conditional intensity of the point process* $\mathbf{X}$ *satisfies* (1.1) *and* (2.3) *with finite range* R. *Let* $g : \mathbb{R}^d \times \Omega \to \mathbb{R}$ *be a nonnegative measurable function satisfying for any* $u \in \mathbb{R}^d$ *and* $\mathbf{x} \in \Omega$

$$g(u,\mathbf{x}) = g(0,\tau_u\mathbf{x}) \quad \text{and} \quad g(u,\mathbf{x}) = g(u,\mathbf{x}_{B(u,\widetilde{R})})$$

with $\widetilde{R} \geq R$. *We also assume that the following integrability conditions are fulfilled*

$$\begin{aligned}
&\mathrm{E}[g^2(0,\mathbf{X})\lambda_{\beta^\star}(0,\mathbf{X})] < \infty \\
&\mathrm{E}\int_{B(0,R)} g(0,\mathbf{X})g(v,\mathbf{X})\big|\lambda_{\beta^\star}(0,\mathbf{X})\lambda_{\beta^\star}(v,\mathbf{X}) - \lambda_{\beta^\star}(0,v,\mathbf{X})\big|\mathrm{d}v < \infty \\
&\mathrm{E}\int_{B(0,\widetilde{R})} \Big|\big(g(0,\mathbf{X}\cup v) - g(0,\mathbf{X})\big)\big(g(v,\mathbf{X}\cup 0) - g(v,\mathbf{X})\big)\Big|\lambda_{\beta^\star}(0,v,\mathbf{X})\mathrm{d}v < \infty.
\end{aligned}$$

Then, as $n \to \infty$

$$|\Lambda_n|^{-1}\operatorname{Var} I_{\Lambda_n}(\mathbf{X},g) \longrightarrow A_1 + A_2 + A_3$$

where

$$\begin{aligned}
A_1 &= \mathrm{E}[g^2(0,\mathbf{X})\lambda_{\beta^\star}(0,\mathbf{X})] \\
A_2 &= \mathrm{E}\int_{B(0,R)} g(0,\mathbf{X})g(v,\mathbf{X})\big(\lambda_{\beta^\star}(0,\mathbf{X})\lambda_{\beta^\star}(v,\mathbf{X}) - \lambda_{\beta^\star}(0,v,\mathbf{X})\big)\mathrm{d}v \\
A_3 &= \mathrm{E}\int_{B(0,\widetilde{R})} \big(g(0,\mathbf{X}\cup v) - g(0,\mathbf{X})\big)\big(g(v,\mathbf{X}\cup 0) - g(v,\mathbf{X})\big)\lambda_{\beta^\star}(0,v,\mathbf{X})\mathrm{d}v.
\end{aligned}$$

References

[1] A. Baddeley and D. Dereudre. Variational estimators for the parameters of Gibbs point process models. *To appear in Bernoulli*, 2013.

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

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

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

[5] 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.

[6] J.-F. Coeurjolly, D. Dereudre, R. Drouilhet, and F. Lavancier. Takacs-Fiksel method for stationary marked Gibbs point processes. *Scandinavian Journal of Statistics*, 49(3):416–443, 2012.

[7] J.-F. Coeurjolly and E. Rubak. Fast covariance estimation for innovations computed from a spatial Gibbs point process. *To appear in Scandinavian Journal of Statistics*, 2013.

[8] D. Daley and D. Vere-Jones. *An introduction to the Theory of Point Processes.* Springer Verlag, New York, 1988.

[9] D. Dereudre, R. Drouilhet, and H.O. Georgii. Existence of Gibbsian point processes with geometry-dependent interactions. *Probability Theory and Related Fields*, 153:647–670, 2012.

[10] P.J. Diggle, D.J. Gates, and A. Stibbard. A nonparametric estimator for pairwise-interaction point processes. *Biometrika*, 74(4):763–770, 1987.

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

[12] H.O. Georgii. Gibbs measures and Phase Transitions. *Walter de Gruyter, Berlin*, 1988.

[13] C.J. Geyer. Likelihood inference for spatial point processes. In W.S. Kendall O.E. Barndorff-Nielsen and M.N.M. Van Lieshout (eds), editors, *Stochastic Geometry: Likelihood and Computation*, number 80, chapter 3, pages 79–140. Chapman and Hall / CRC, Monographs on Statistics and Applied Probability, 1999.

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

[15] J. Møller. Parametric methods. In A.E. Gelfand, P. Diggle, M. Fuentes, and P. Guttorp, editors, *A Handbook of Spatial Statistics*, pages 317–337. Chapman and Hall/CRC Press, 2010.

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

[17] V.M.R. Muggeo. Segmented: an R package to fit regression models with broken-line relationships. *R news*, 8:20–25, 2008.

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

[19] X.X. Nguyen and H. Zessin. Ergodic theorems for Spatial Process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 48:133–158, 1979.

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

[21] C.J. Preston. *Random fields*. Springer Verlag, 1976.

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

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

[24] H. Zessin. Der papangelou prozess. *Contemporary mathematical analysis*, 44(1):36–44, 2009.