THIELE CENTRE
FOR APPLIED MATHEMATICS IN NATURAL SCIENCE

# Statistics for Locally Scaled Point Processes

Michaela Prokešová, Ute Hahn
and Eva B. Vedel Jensen

Research Report

No. 06 | August 2004

# Statistics for Locally Scaled Point Processes

This Thiele Research Report is also Research Report number 444 in the Stochastics Series at Department of Mathematical Sciences, University of Aarhus, Denmark.

# STATISTICS FOR LOCALLY SCALED POINT PROCESSES

MICHAELA PROKEŠOVÁ*, UTE HAHN**, AND EVA B. VEDEL JENSEN***

Abstract. Recently, locally scaled point processes have been proposed as a new class of models for inhomogeneous spatial point processes. They are obtained as modifications of homogeneous template point processes and have the property that regions with different intensity differ only by a location dependent scale factor. The main emphasis of the present paper is on analysis of such models. Statistical methods are developed for estimation of scaling function and template parameters as well as for model validation. The proposed methods are assessed by simulation and used in the analysis of a vegetation pattern.

## 1. Introduction

The present paper deals with statistical analysis for inhomogeneous point processes that are obtained by local scaling. In these point processes, local geometry is constant, that is, subregions of the inhomogeneous process with different intensity appear to be scaled versions of the same homogeneous process. This property is characteristic of locally scaled point processes and not present in the other models for inhomogeneous point processes discussed in [Hahn et al., 2003]. Such patterns occur for example in vegetation of dry areas, as shown in Figure 1. Where water or other resources are short, plants grow sparsely and keep larger distances between individuals than in regions with better supply. Naturally there is no preference for a direction, and therefore the vegetation pattern is locally isotropic. Local scaling of an isotropic template process yields locally isotropic patterns in contrast to transformation of an isotropic template process ([Nielsen, 2001]).

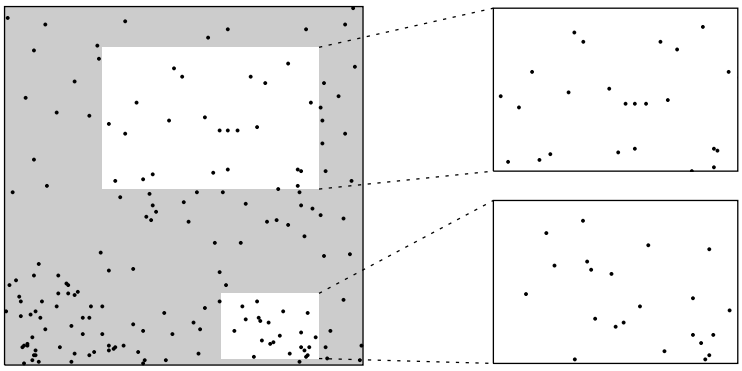

Figure 1. Left part: Map of 171 individuals of a *Scholtzia* aff. *involucrata* in Australian bush on a $220 \times 220$ m square. Right part: Two rectangular subregions with different intensity were rescaled such that they have the same number of individuals by unit area. Data from [Armstrong, 1991].

Similar locally scaled structures are found in arrangements of solid bodies with constant shape but location dependent size, such as the sinter filter discussed in [Hahn et al., 2003]

* Charles University, Department of Probability, Sokolovská 83, 18675 Praha 8, Czech Republic .
** University of Augsburg, Department of Applied Stochastics, 86135 Augsburg, Germany.
*** University of Aarhus, Department of Theoretical Statistics, Ny Munkegade, 8600 Aarhus C, Denmark
.

or in sponges with constant porosity but small pore size close to the surface and large pore size in the interior.

Locally scaled point processes are derived from a homogeneous template process which describes the interaction between points and is responsible for the local geometry of the resulting pattern. We will put the major focus on Markov template processes. Inhomogeneity is introduced through a location dependent function that gives the local scale, as explained in detail in Section 2 below.

Fitting a model to a given pattern thus consists of finding the parameters inherited from the template and choosing an appropriate scaling function. This can be achieved by simultaneous maximum (pseudo)likelihood estimation as discussed in Section 3. A less time consuming procedure is two step maximum likelihood estimation where the parameters of the template are estimated after having determined the scaling function. Section 4 is centred on two step estimation, which is assessed and demonstrated by a simulation study presented in Section 5. Section 6 addresses possibilities to estimate the scaling function by other methods.

A widely used popular method for model validation in the homogeneous case is to compare the empirical $K$-function with the theoretically known or simulated $K$-function of the fit. An inhomogeneous analogue of the $K$-function is proposed in Section 7. Furthermore we suggest an inhomogeneous version of the $Q^2$-statistic recently proposed by [Grabarnik and Chiu, 2002] for model validation of homogeneous point processes.

Finally, a statistical analysis of the point pattern in Figure 1 is presented in Section 8.

## 2. Locally scaled point processes

In this section, we introduce the locally scaled point processes and discuss some of their basic properties.

Let $X$ be a point process, defined on a full-dimensional bounded subset $\mathcal{X}$ of $\mathbb{R}^k$. We suppose that $X$ has a density $f_X$ with respect to the restriction of the unit rate Poisson point process $\Pi$ to $\mathcal{X}$. Let $\nu^* = (\nu^0, \ldots, \nu^k)$ be the set of $d$-dimensional volume measures $\nu^d$ in $\mathbb{R}^k$, $d = 0, 1, \ldots, k$. Let us suppose that $f_X$ is of the following form

$$f_X(\mathbf{x}) \propto g(\mathbf{x}; \nu^*), \quad \mathbf{x} \subset \mathcal{X} \text{ finite}, \tag{1}$$

where the function $g$ is scale-invariant, i.e.

$$g(c\mathbf{x}; \nu_c^*) = g(\mathbf{x}; \nu^*) \tag{2}$$

for all $\mathbf{x}$ and $c > 0$. Here, $\nu_c^* = (\nu_c^0, \ldots, \nu_c^k)$ and $\nu_c^d(A) = \nu_d(c^{-1}A)$, $A \in \mathbb{B}_k$. The classical homogeneous point processes have densities with this property.

The process $X$ will serve as a template process. In order to construct a locally scaled version of $X$ with scaling function $c : \mathbb{R}^k \to \mathbb{R}_+$, we replace the $d$-dimensional volume measure $\nu^d$ in $\mathbb{R}^k$ with a locally scaled version

$$\nu_c^d(A) = \int_A c(u)^{-d} \nu^d(\mathrm{d}u), \quad A \in \mathbb{B}_k,$$

$d = 0, 1, \ldots, k$. In what follows, we assume that the scaling function $c$ is bounded from below and from above, i.e. $0 < \underline{c} < c(u) < \overline{c}$, $u \in \mathbb{R}^k$. Furthermore, we will assume that $g(\cdot; \nu_c^*)$ is integrable with respect to the Poisson point process $\Pi_c$ with $\nu_c^k$ as intensity measure. A locally scaled point process $X_c$ on $\mathcal{X}$ with template process $X$ is then defined by the following density with respect to $\Pi_c$

$$f_{X_c}^{(c)}(\mathbf{x}) \propto g(\mathbf{x}; \nu_c^*). \tag{3}$$

Note that the density of $X_c$ with respect to $\Pi$ is

$$f_{X_c}(\mathbf{x}) = \exp\left( - \int_{\mathcal{X}} [c(u)^{-k} - 1] \nu^k(\mathrm{d}u) \right) \prod_{x \in \mathbf{x}} c(x)^{-k} \cdot f_{X_c}^{(c)}(\mathbf{x}). \tag{4}$$

**Example 1.** The Strauss process $X$ with intensity parameter $\beta > 0$, interaction parameter $\gamma \in [0, 1]$ and interaction distance $r > 0$ is given by the density

$$f_X(\mathbf{x}) \propto \beta^{n(\mathbf{x})}\gamma^{s(\mathbf{x})}, \quad \mathbf{x} \subset \mathcal{X} \text{ finite},$$

where $n(\mathbf{x})$ is the number of points in $\mathbf{x}$ and $s(\mathbf{x})$ is the number of $r$-close pairs, cf. [Strauss, 1975]. The density is of the form (1) with

$$g(\mathbf{x}; \nu^*) = \beta^{\nu^0(\mathbf{x})}\gamma^{\sum^{\neq}_{\{u,v\}\subseteq\mathbf{x}} \mathbf{1}\{\nu^1([u,v])\le r\}},$$

where the superscript $\neq$ in the summation indicates that $u$ and $v$ are different. It is easy to check that this function is indeed scale-invariant. The locally scaled Strauss process $X_c$ has density with respect to $\Pi_c$ of the form

$$f^{(c)}_{X_c}(\mathbf{x}) \propto \beta^{n(\mathbf{x})}\gamma^{s_c(\mathbf{x})},$$

where

$$s_c(\mathbf{x}) = \sum_{\{u,v\}\subseteq\mathbf{x}}^{\neq} \mathbf{1}\{\nu^1_c([u, v]) \le r\}.$$

Figure 2 shows locally scaled Strauss processes on $\mathcal{X} = [0, 1]^2$ with scaling function of the exponential form

$$c_\theta(u) = \sqrt{\frac{1 - \mathrm{e}^{-2\theta}}{2\theta}}\, \mathrm{e}^{\theta u_1}, \quad u = (u_1, u_2) \in \mathcal{X}, \tag{5}$$

for 4 different values of the inhomogeneity parameter $\theta \in \{0.25, 0.5, 1, 1.5\}$. The normalisation $\sqrt{\frac{1-\mathrm{e}^{-2\theta}}{2\theta}}$ ensures that the 4 point patterns have approximately the same number of points (see Section 4.2 for details).

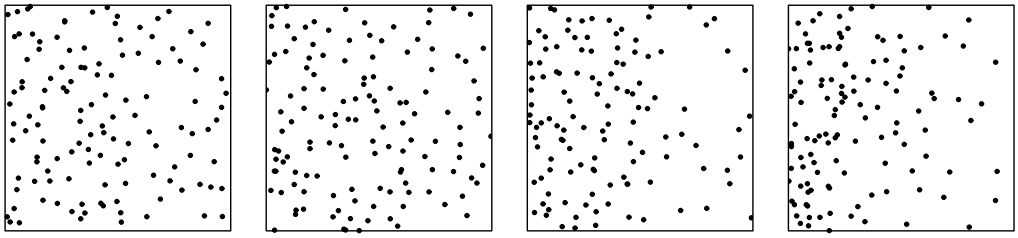

FIGURE 2. Simulation of locally scaled Strauss processes on $[0, 1]^2$ with exponential scaling function (5) for $\theta \in \{0.25, 0.5, 1, 1.5\}$ (from left to right) and template parameters $\beta = 250, \gamma = 0.3$ and $r = 0.05$.

□

**Example 2.** The area-interaction point process with intensity parameter $\beta > 0$, interaction parameter $\gamma > 0$ and interaction distance $r > 0$ is given by the density

$$f(\mathbf{x}) \propto \beta^{n(\mathbf{x})}\gamma^{-\nu^2(U_r(\mathbf{x}))}, \quad \mathbf{x} \subset \mathcal{X} \text{ finite},$$

where $U_r(\mathbf{x}) = \bigcup_{x\in\mathbf{x}} b(x, r)$ is the union of balls with centers in $\mathbf{x}$ and radius $r$. For $\gamma > 1$ the point pattern appears clustered, for $\gamma < 1$ regular, cf. [Baddeley and van Lieshout, 1995]. The density is again of the form (1) with scale invariant

$$g(\mathbf{x}; \nu^*) = \beta^{\nu^0(\mathbf{x})}\gamma^{-\nu^2(\cup_{x\in\mathbf{x}}\{v\in\mathcal{X}\,:\,\nu^1([v,x])\le r\})}.$$

The locally scaled area-interaction process has density with respect to $\Pi_c$ of the form

$$f^{(c)}_{X_c}(\mathbf{x}) \propto \beta^{n(\mathbf{x})}\gamma^{-\nu^2_c(U_{c,r}(\mathbf{x}))},$$

where $U_{c,r} = \bigcup_{x\in\mathbf{x}} b_c(x, r)$ and $b_c(x, r) = \{v \in \mathcal{X} : \nu^1_c([v, x]) \le r\}$ is the scaled ball. Figure 3 shows locally scaled area-interaction processes with the same scaling function (5) as in Example 2. The value of the interaction parameter $\gamma$ was chosen so that $\gamma^{-\pi r^2} \approx 0.1$ and the point patterns are visibly clustered.

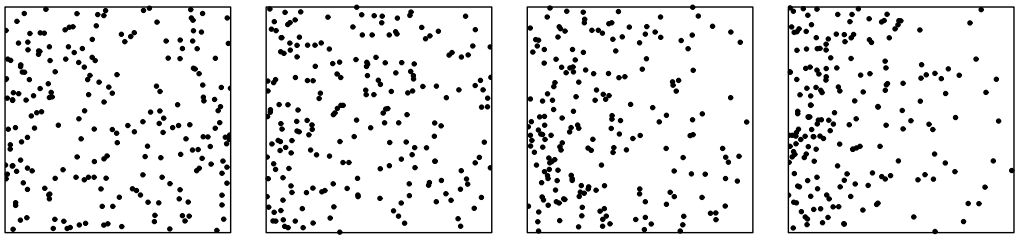

FIGURE 3. Simulation of locally scaled area-interaction processes on $[0,1]^2$ with exponential scaling function (5) for $\theta \in \{0.25, 0.5, 1, 1.5\}$ (from left to right) and template parameters $\beta = 180, \gamma = 6.7 \cdot 10^{31}$ and $r = 0.1$.

□

The Strauss process and the area-interaction process are examples of point processes from two large classes of homogeneous template processes, viz. the distance-interaction processes and the shot noise weighted processes. For these two classes, it has been shown in [Hahn et al., 2003] that the Papangelou conditional intensities $\lambda_{X_c}$ of the locally scaled process $X_c$ and $\lambda_X$ of the template process $X$, respectively, satisfy

$$\lambda_{X_c}(x \mid \mathbf{x}) = c_0^{-k} \lambda_X(\frac{x}{c_0} \mid \frac{\mathbf{x}}{c_0}), \tag{6}$$

if the scaling function $c$ is constant and equal to $c_0$ in a scaled neighbourhood of $x$. The right-hand side of (6) is the conditional intensity of a globally scaled template process with scaling factor $c_0$. It is therefore expected that the locally scaled process appears as a scaled version of the template process if the scaling function is slowly varying compared to the interaction radius. The development of further formal reasoning, supporting this statement, seems very hard.

It is also of interest to study the unconditional intensity function $\lambda_c(x), x \in \mathcal{X}$, of the locally scaled process. Let us suppose that the template process $X$ is homogeneous with intensity $\lambda_0$ ($X$ may, for instance, be defined on a torus with periodic boundary conditions). Then,

$$\lambda_c(x) = c(x)^{-k} \lambda_0, \quad x \in \mathcal{X}, \tag{7}$$

holds if the template process is Poisson or the scaling function is constant. Also, (7) holds for any locally scaled distance-interaction process in $\mathbb{R}^1$, see the Appendix. The equality (7) is expected to hold approximately if the scaling function is slowly varying, compared to the interaction radius.

For statistical inference of locally scaled models, we will distinguish two cases. In fully parametric models, both the scaling function $c$ and the homogeneous template process $X$ are specified by a set of parameters. In semiparametric models only the template process is parametrically specified.

In the following, the parameters of the template process are denoted by $\psi$, and $\theta$ is the parameter of the scaling function in fully parametric models. The parameter space of a fully parametric model is $\Theta \times \Psi$, while, in semiparametric models, the scaling function can be any function in the space $\mathcal{C}^+$ of measurable positive functions, satisfying the regularity conditions mentioned above.

A particularly attractive parametric form of the scaling function is the exponential form

$$c_\theta(u) = \alpha(\theta)\, \mathrm{e}^{\theta \cdot \tau(u)}, \quad u \in \mathbb{R}^k, \tag{8}$$

where $\theta \in \Theta \subseteq \mathbb{R}^l$, $\alpha(\theta) \in \mathbb{R}_+$, $\cdot$ indicates the inner product and $\tau(u) \in \mathbb{R}^l$. A locally scaled model with an exponential scaling function is called an exponentially scaled model. Note that if $\tau(u) = u$, then scaled distances can be calculated explicitly. Using the coarea formula

we get

$$\begin{aligned}\nu_c^1([u,v]) &= \int_{[u,v]} \alpha(\theta)^{-1}\mathrm{e}^{-\theta\cdot x}\ \nu^1(\mathrm{d}x) \ = \\ &= \int_0^1 \nu^1([u,v])\ \alpha(\theta)^{-1}\mathrm{e}^{-\theta\cdot(u+t(v-u))}\,\mathrm{d}\,t \ = \\ &= \nu^1([u,v])\,\frac{[c_\theta(u)^{-1}-c_\theta(v)^{-1}]}{\theta\cdot(v-u)}, \qquad u,v\in\mathbb{R}^k.\end{aligned}$$

## 3. Simultaneous maximum (pseudo)likelihood estimation of scaling function and template parameters

In a fully parametric model, the likelihood factorizes as, cf. (4),

$$L(\theta,\psi;\mathbf{x}) = L_0(\theta;\mathbf{x})\cdot L_1(\theta,\psi;\mathbf{x}), \tag{9}$$

where $L_0$ is the likelihood of an inhomogeneous Poisson point process $\Pi_c$ with intensity measure $\nu_c^k$, and $L_1(\theta,\psi;\mathbf{x}) = f_{X_c}^{(c)}(\mathbf{x};\psi)$ is the density of the scaled process $X_c$ with respect to $\Pi_c$. Recall that the scaling function is parametrized by $\theta$, i.e. $c = c_\theta$.

Maximum likelihood estimation is most feasible in exponential families, since it amounts to moment estimation there. Most popular homogeneous Markov point process models are partially exponential, and the set $\psi$ splits into two components – the nuisance parameters and the remaining parameters, that form exponential family parameters given the nuisance parameters. Since the likelihood in Markov point processes is known only up to the normalizing constant, one has to resort to MCMC methods for MLE, cf. e.g. [Møller and Waagepetersen, 2003]. Whilst moment estimation in these models can be done relatively precisely with affordable effort, estimation of the normalizing constant entails numerical pitfalls and should be avoided as much as possible. This suggests that MLE should be done on a grid of nuisance parameters, since given this component, the remaining parameters are exponential family parameters. In locally scaled processes, the inhomogeneity parameter also acts as a nuisance parameter.

Usually, the point process $X_c$ is observed in a sampling window $W\subseteq\mathcal{X}$. In such cases, a conditional likelihood may be used, based on the conditional density of $X_c\cap W$ given $X_c\cap W^c = \mathbf{x}_{W^c}$ where $\mathbf{x}_{W^c}$ is a finite subset of $W^c$. Since

$$f_{X_c}(\cdot\mid\mathbf{x}_{W^c})\propto f_{X_c}(\cdot\cup\mathbf{x}_{W^c}),$$

it follows from (4) that (9) still holds for the conditional likelihoods. This result is mainly of interest for locally scaled Markov point processes.

A less computational demanding procedure is based on the pseudolikelihood function, see [Baddeley and Turner, 2000] and references therein. The pseudolikelihood function of a point process density $f$ with respect to a Poisson point process with intensity measure $\mu$, based on observation in $W$, is defined by

$$\exp\left(-\int_W[\lambda(u\mid\mathbf{x})-1]\,\mu(\mathrm{d}u)\right)\prod_{x\in\mathbf{x}\cap W}\lambda(x\mid\mathbf{x}\backslash\{x\}),\quad W\subseteq\mathcal{X}$$

where $\mathbf{x}$ is the realized point pattern in $\mathcal{X}$ and

$$\lambda(u\mid\mathbf{x}) = \frac{f(\mathbf{x}\cup\{u\})}{f(\mathbf{x})},\quad u\notin\mathbf{x},$$

is the Papangelou conditional intensity associated with $f$.

Based on observation in $W$, let $PL_W(\theta,\psi;\mathbf{x})$ be the pseudolikelihood function for the density $f_{X_{c_\theta}}(\,\cdot\,;\psi)$ with respect to the unit rate Poisson point process and let $PL_{W,1}(\theta,\psi;\mathbf{x})$ be the pseudolikelihood for the density $f_{X_{c_\theta}}^{(c_\theta)}(\,\cdot\,;\psi)$ with respect to the Poisson point process with intensity measure $\nu_{c_\theta}^k$. Then, using (4), we find

$$PL_W(\theta,\psi;\mathbf{x}) = L_0(\theta;\mathbf{x}\cap W)\cdot PL_{W,1}(\theta,\psi;\mathbf{x}). \tag{10}$$

Note that

$$\lambda_{\theta,\psi}(u \mid \mathbf{x}) = c_\theta(u)^{-k}\, \lambda^{(c_\theta)}_{\theta,\psi}(u \mid \mathbf{x}), \quad u \notin \mathbf{x},$$

where $\lambda_{\theta,\psi}$ and $\lambda^{(c_\theta)}_{\theta,\psi}$ are the conditional intensities associated with $f_{X_{c_\theta}}$ and $f^{(c_\theta)}_{X_{c_\theta}}$, respectively.

A proof of (10) can be constructed as follows. From (4), we get

$$\begin{aligned}
&PL_W(\theta,\psi;\mathbf{x}) \\
&= \exp(-\int_W [\lambda_{\theta,\psi}(u \mid \mathbf{x}) - 1]\nu^k(\mathrm{d}u)) \prod_{x\in\mathbf{x}\cap W} \lambda_{\theta,\psi}(x \mid \mathbf{x}\backslash\{x\}) \\
&= \exp(-\int_W [c_\theta(u)^{-k} - 1]\nu^k(\mathrm{d}u)) \prod_{x\in\mathbf{x}\cap W} c_\theta(x)^{-k} \\
&\quad \times \exp(-\int_W [\lambda^{(c_\theta)}_{\theta,\psi}(u \mid \mathbf{x}) - 1] c_\theta(u)^{-k}\nu^k(\mathrm{d}u)) \prod_{x\in\mathbf{x}\cap W} \lambda^{(c_\theta)}_{\theta,\psi}(x \mid \mathbf{x}\backslash\{x\}) \\
&= L_0(\theta;\mathbf{x}) \cdot PL_{W,1}(\theta,\psi;\mathbf{x}).
\end{aligned}$$

As the values of the scaled interaction statistics (e.g. $s_{c_\theta}(\mathbf{x})$ in the Strauss model) and subsequently the values of $\lambda^{(c_\theta)}_{\theta,\psi}(u \mid \mathbf{x})$ depend on the inhomogeneity parameter $\theta$, the latter is a nuisance parameter also in the pseudolikelihood estimation. This means we have to evaluate the profile pseudolikelihood on a grid of nuisance parameters similarly to the maximum likelihood approach. However, this is much less computational intensive in maximum pseudolikelihood estimation than in maximum likelihood estimation, since $PL_1$ can be calculated directly without having to estimate an unknown normalizing constant by simulation as it is the case with $L_1$.

## 4. Two step maximum likelihood estimation of scaling parameters prior to template parameters

The structural similarity of the full likelihood in locally scaled models and the full likelihood in transformation models for point processes suggests that partial likelihood inference as in the paper by [Nielsen and Jensen, 2004] will be successful also for locally scaled models. [Nielsen and Jensen, 2004] estimated the inhomogeneity parameters by maximizing the Poisson part $L_0$ of the likelihood only, assuming no interaction. They chose an exponential model for the inhomogeneity function, since this largely simplifies calculations.

Below, this approach is followed for the locally scaled models. In Section 4.1, we find the maximum likelihood estimate $\widehat{\theta}_0$ of $\theta$ on the basis of $L_0$ and, in Section 4.2, it is shown that $\widehat{\theta}_0$ can be regarded as an approximate moment estimator. In Section 4.3, estimation of the template parameters is considered.

### 4.1. **Estimation of scaling parameters, using the Poisson likelihood**

We suppose that the scaling function is of the form

$$c(u) = \alpha\, \mathrm{e}^{\theta\cdot\tau(u)}, \qquad u \in \mathbb{R}^k, \tag{11}$$

where $\theta \in \Theta \subseteq \mathbb{R}^l$ and $\alpha \in \mathbb{R}_+$. In addition to the inhomogeneity parameter $\theta$, the scaling function contains a global scaling parameter $\alpha$. For the moment, these two parameters vary in a product set $\Theta \times \mathbb{R}_+$.

Then, the Poisson part of the likelihood of the process $X_c$, observed in a set $W$, is

$$L_0(\theta,\alpha;\mathbf{x}\cap W) = \mathrm{e}^{-\int_W (\alpha^{-k}\mathrm{e}^{-k\,\theta\cdot\tau(u)} - 1)\,\nu^k(\mathrm{d}u)} \prod_{x\in\mathbf{x}\cap W} (\alpha^{-k}\mathrm{e}^{-k\,\theta\cdot\tau(x)}).$$

The log-likelihood becomes

$$l_0(\theta,\alpha;\mathbf{x}\cap W) = \int_W 1\,\nu^k(\mathrm{d}u) - \int_W \alpha^{-k}\mathrm{e}^{-k\,\theta\cdot\tau(u)}\,\nu^k(\mathrm{d}u) - k\,n(\mathbf{x}\cap W)\ln\alpha + \sum_{x\in\mathbf{x}\cap W} (-k\,\theta\cdot\tau(x)).$$

Assume that $n(\mathbf{x} \cap W) > 0$ and $\|\tau(u)\| \, \mathrm{e}^{\theta \cdot \tau(u)}$ is uniformly bounded in $u \in W$ and $\theta \in \Theta$. Then by differentiating we get $l+1$ equations

$$\begin{aligned} k\alpha^{-k-1} \int_W \mathrm{e}^{-k\,\theta \cdot \tau(u)} \, \nu^k(\mathrm{d}u) &= k\, n(\mathbf{x} \cap W) \alpha^{-1} \\ \alpha^{-k} \int_W k\tau_i(u) \mathrm{e}^{-k\,\theta \cdot \tau(u)} \, \nu^k(\mathrm{d}u) &= k \sum_{x \in \mathbf{x} \cap W} \tau_i(x), \quad i = 1, \ldots, l. \end{aligned}$$

Dividing the last $l$ equations by the first equation we get the vector equation

$$\frac{t(\mathbf{x} \cap W)}{n(\mathbf{x} \cap W)} = m(\theta), \tag{12}$$

where $t(\mathbf{x} \cap W) = \sum_{x \in \mathbf{x} \cap W} \tau(x)$ and

$$m(\theta) = \frac{\int_W \tau(u) \mathrm{e}^{-k\,\theta \cdot \tau(u)} \, \nu^k(\mathrm{d}u)}{\int_W \mathrm{e}^{-k\,\theta \cdot \tau(u)} \, \nu^k(\mathrm{d}u)}. \tag{13}$$

Thus the estimate of $\theta$ does not depend on the estimate of the constant $\alpha$ and furthermore the estimate depends only on the statistic $t(\mathbf{x} \cap W)/n(\mathbf{x} \cap W)$.

It turns out that we get exactly the same estimate of $\theta$ if we impose the following normalizing condition on $c_\theta$

$$\int_W c_\theta(u)^{-k} \, \nu^k(\mathrm{d}u) = \nu^k(W), \tag{14}$$

implying that

$$\alpha = \alpha(\theta) = \left[ \int_W \mathrm{e}^{-k\,\theta \cdot \tau(u)} \nu^k(\mathrm{d}u) \,/\, \nu^k(W) \right]^{\frac{1}{k}}. \tag{15}$$

To see this, note that under (15), the Poisson likelihood takes the form

$$\begin{aligned} L_0(\theta; \mathbf{x} \cap W) &= \mathrm{e}^{-\int_W (\alpha(\theta)^{-k} \mathrm{e}^{-k\theta \cdot \tau(u)} - 1) \, \nu^k(\mathrm{d}u)} \prod_{x \in \mathbf{x} \cap W} (\alpha(\theta)^{-k} \mathrm{e}^{-k\,\theta \cdot \tau(x)}) \\ &= \left( \alpha(\theta)^{-k} \mathrm{e}^{-k\theta \cdot \frac{t(\mathbf{x} \cap W)}{n(\mathbf{x} \cap W)}} \right)^{n(\mathbf{x} \cap W)}. \end{aligned}$$

Taking the logarithm and differentiating with respect to $\theta$, we again get the vector equation (12). As we shall see in Section 4.2, (14) appears to be a very natural condition.

The existence and uniqueness of a solution $\widehat{\theta}_0$ to (12) have been studied in [Nielsen and Jensen, 2004] in a closely related set–up where the parameter of interest was $\tilde{\theta} = -k\,\theta$. The same type of arguments applies here. Using (15), it is seen that

$$\{ \frac{\alpha(\theta)^{-k}}{\nu^k(W)} \, \mathrm{e}^{-k\,\theta \cdot \tau(u)} \, : \theta \in \Theta \}$$

is an exponential family of densities on $W$, with respect to $\nu^k$. If the family is regular, then the function $m$ in (13) is a bijection of $\Theta$ on $\mathrm{int}S$ where $S$ is the convex support of the family, cf. e.g. [Barndorff-Nielsen, 1978]. Thus, under these conditions, there is a unique solution to (12) if $n(\mathbf{x} \cap W) > 0$ and $t(\mathbf{x} \cap W)/n(\mathbf{x} \cap W) \in \mathrm{int}S$.

**Example 3.** Let $\tau(u) = u$ and $W = [0,1]^k$. Then, $\Theta = \mathbb{R}^k$

$$\begin{aligned} c_\theta(u) &= \alpha(\theta) \mathrm{e}^{\theta \cdot u}, \\ \alpha(\theta) &= \left( \prod_{i=1}^k \frac{1 - \mathrm{e}^{-k\,\theta_i}}{k\,\theta_i} \right)^{\frac{1}{k}}, \end{aligned}$$

and $m(\theta) = (m_1(\theta), \ldots, m_k(\theta))$ where

$$m_i(\theta) = \frac{1 - \mathrm{e}^{-k\,\theta_i} - k\,\theta_i \mathrm{e}^{-k\,\theta_i}}{k\,\theta_i (1 - \mathrm{e}^{-k\,\theta_i})}, \quad i = 1, \ldots, k.$$

□

### 4.2. **Statistical properties of $\widehat{\theta}_0$**

The estimator $\widehat{\theta}_0$ is the maximum likelihood estimator of $\theta$ if the template process is Poisson. It is also possible to give theoretical support to the use of $\widehat{\theta}_0$ for general template processes, as shown below.

**Proposition 1.** *Suppose that the intensity of the locally scaled process $X_{c_\theta}$ satisfies*

$$\lambda_{\theta,\psi}(u) = c_\theta(u)^{-k}\lambda_{0\psi}. \tag{16}$$

*Then,*

$$\frac{\mathbb{E}_{\theta,\psi}[t(X_{c_\theta} \cap W)]}{\mathbb{E}_{\theta,\psi}[n(X_{c_\theta} \cap W)]} = m(\theta).$$

*Proof.* We use the following version of the Nguyen-Zessin formula for $X_{c_\theta}$, cf. [Nguyen and Zessin, 1976],

$$\mathbb{E}_{\theta,\psi} \sum_{x \in X_{c_\theta}} h(x) = \int_{\mathbb{R}^k} h(x)\lambda_{\theta,\psi}(x)\nu^k(\mathrm{d}x).$$

We get

$$\begin{aligned}
\mathbb{E}_{\theta,\psi}[t(X_{c_\theta} \cap W)] &= \mathbb{E}_{\theta,\psi} \sum_{x \in X_{c_\theta} \cap W} \tau(x) \\
&= \int_W \tau(x)\lambda_{\theta,\psi}(x)\nu^k(\mathrm{d}x) \\
&= \lambda_{0\psi} \int_W \tau(x)c_\theta(x)^{-k}\nu^k(\mathrm{d}x).
\end{aligned}$$

In particular,

$$\mathbb{E}_{\theta,\psi}[n(X_{c_\theta} \cap W)] = \lambda_{0\psi} \int_W c_\theta(x)^{-k}\nu^k(\mathrm{d}x). \tag{17}$$

The result now follows directly. □

If (16) holds, $\widehat{\theta}_0$ can thus be regarded as a moment estimator.

As mentioned in Section 2, the equation (16) holds if the template process is homogeneous and the scaling function is constant. More interestingly, (16) holds for a not necessarily constant scaling function for distance-interaction processes in $\mathbb{R}^1$, see the Appendix. Generally, equation (16) is expected to hold approximately if the scaling function varies slowly compared to the interaction radius.

### 4.3. **Estimation of the template parameters**

Having estimated the scaling parameter $\theta$ we can proceed by the estimation of the template process parameters. We will here concentrate on the case where the pseudolikelihood $PL_{W,1}(\widehat{\theta}_0, \psi; \mathbf{x})$ from the decomposition (10) is used. In the following we discuss the practical implementation of this method for the locally scaled models. We consider general parametric scaling functions.

Recall that the pseudolikelihood $PL_{W,1}(\theta, \psi; \mathbf{x})$ for the density $f^{(c_\theta)}_{X_{c_\theta}}(\,\cdot\,; \psi)$ with respect to the Poisson point process with intensity measure $\nu^k_{c_\theta}$, based on observation in a window $W \subset \mathbb{R}^k$, is defined as follows

$$PL_{W,1}(\theta, \psi; \mathbf{x}) = \exp\left(-\int_W [\lambda^{(c_\theta)}_{\theta,\psi}(u \mid \mathbf{x}) - 1]\, \nu^k_{c_\theta}(\mathrm{d}u)\right) \prod_{x \in \mathbf{x} \cap W} \lambda^{(c_\theta)}_{\theta,\psi}(x \mid \mathbf{x}\backslash\{x\}). \tag{18}$$

In the second step of the two-step estimation procedure we fix the scaling parameter $\theta$ to $\widehat{\theta}_0$ and maximize $PL_{W,1}(\widehat{\theta}_0, \psi; \mathbf{x})\, \mathrm{e}^{-\nu^k_{c_{\widehat{\theta}_0}}(W)}$ as a function of $\psi$. This can be done in a way similar to the procedure used in the homogeneous case, cf. [Baddeley and Turner, 2000].

We partition $W$ into a finite number of cells $C_i$, each containing one dummy point $u_i, i = 1, \ldots, l$. The union of the dummy points and the points of the observed pattern is denoted $\{u_j : j = 1, \ldots, m\}$. Furthermore let $C_{i(j)}$ be the unique cell containing $u_j, j = 1, \ldots, m$, with dummy point $u_{i(j)}$. Then we approximate the integral in the pseudolikelihood by

$$\int_W \lambda^{c_{\widehat{\theta}_0}}_{\widehat{\theta}_0,\psi}(u \mid \mathbf{x}) \nu^k_{c_{\widehat{\theta}_0}}(\mathrm{d}u) \approx \sum_{j=1}^{m} \lambda^{c_{\widehat{\theta}_0}}_{\widehat{\theta}_0,\psi}(u_j \mid \mathbf{x}\backslash\{u_j\})\, w_j,$$

where

$$w_j = \frac{\nu^k(C_{i(j)})}{c_{\widehat{\theta}_0}(u_{i(j)})^k} \frac{1}{(1 + n(\mathbf{x} \cap C_{i(j)}))} \approx \nu^k_{c_{\widehat{\theta}_0}}(C_{i(j)}) \frac{1}{(1 + n(\mathbf{x} \cap C_{i(j)}))}. \tag{19}$$

Here, $n(\mathbf{x} \cap C_{i(j)})$ is the total number of observed points in the cell $C_{i(j)}$. $\frac{\nu^k(C_{i(j)})}{c_{\widehat{\theta}_0}(u_{i(j)})^k}$ approximates $\nu^k_{c_{\widehat{\theta}_0}}(C_{i(j)})$ if the cells $C_{i(j)}$ are sufficiently small, such that the scaling function $c$ is approximately constant in $C_{i(j)}$. Let us denote $\lambda^{c_{\widehat{\theta}_0}}_{\widehat{\theta}_0,\psi}(u_j \mid \mathbf{x}\backslash\{u_j\})$ by $\lambda_j$, $j = 1, \ldots, m$. The pseudolikelihood can then be approximated as a weighted likelihood of independent Poisson variables $y_j$ with means $\lambda_j$ and weights $w_j$

$$\log(PL_{W,1}(\widehat{\theta}_0, \psi; \mathbf{x}) \mathrm{e}^{-\nu^k_{c_{\widehat{\theta}_0}}(W)}) \approx \sum_{j=1}^{m} (y_j \log \lambda_j - \lambda_j)\, w_j, \tag{20}$$

$$y_j = \frac{1}{w_j} \mathbf{1}\{u_j \in \mathbf{x}\}, \quad j = 1, \ldots, m. \tag{21}$$

When the conditional intensity $\lambda^{c_{\widehat{\theta}_0}}_{\widehat{\theta}_0,\psi}$ is of exponential family form, (20) can easily be maximized, using standard software for generalized linear models.

## 5. Simulation study

In order to further study the properties of the estimation procedure proposed in Section 4, a simulation study was carried out. The simulation experiment concerns the exponentially scaled Strauss point process with scaling function

$$c_\theta(u) = \sqrt{\frac{1 - \mathrm{e}^{-2\theta}}{2\theta}}\, \mathrm{e}^{\theta u_1}, \quad u = (u_1, u_2) \in \mathbb{R}^2, \tag{22}$$

observed on the unit square $W = [0, 1]^2$. We used four different values of the inhomogeneity parameter $\theta \in \{0.25, 0.5, 1, 1.5\}$. For the template Strauss process we fixed the interaction radius $r$ to 0.05 and used a dense set of $\gamma$-values in $\{0.01, 0.02, \ldots, 1.00\}$. For $\beta$, we used the two values of 250 and 100 to investigate the influence of the total intensity. Note that $\theta = 1.5$ represents quite strong inhomogeneity, compare with Figure 2.

For each combination of the parameters, 1000 point patterns were generated using MCMC and the distribution of $\widehat{\theta}_0$ was approximated by the empirical distribution from the 1000 realisations. To reduce the edge effects in the simulation the process was generated on a bigger window $[-0.2, 1.5] \times [-0.5, 1.5]$ so that $\bigcup_{x \in [0,1]^2} b_c(x, 2r)$ was included in this bigger window.

Figure 4 shows the empirical mean values for the estimator $\widehat{\theta}_0$. Since the function $m$ defined by (13) is concave and $t(\mathbf{x} \cap W)/n(\mathbf{x} \cap W)$ was found to be approximately unbiased for $m(\theta)$, $\widehat{\theta}_0$ tends to overestimate $\theta$. This can be seen in Figure 4 for $\theta = 1.5$ and 1, however the relative bias is not larger than 1% and it does not depend on the interaction parameter $\gamma$. The 95% envelopes for $\widehat{\theta}_0$ are also shown in Figure 4 and for reasonably high number of observed points (i.e. $\beta = 250$) the inhomogeneity is reliably detected by $\widehat{\theta}_0$. Notice, for example that for $\theta = 1$, 95% of the estimates $\widehat{\theta}_0$ falls into the interval $[0.75, 1.25]$ and even for $\theta = 0.25$ – an inhomogeneity often hardly recognizable from the realizations, 97.5% of the $\widehat{\theta}_0$ estimates are larger that zero.

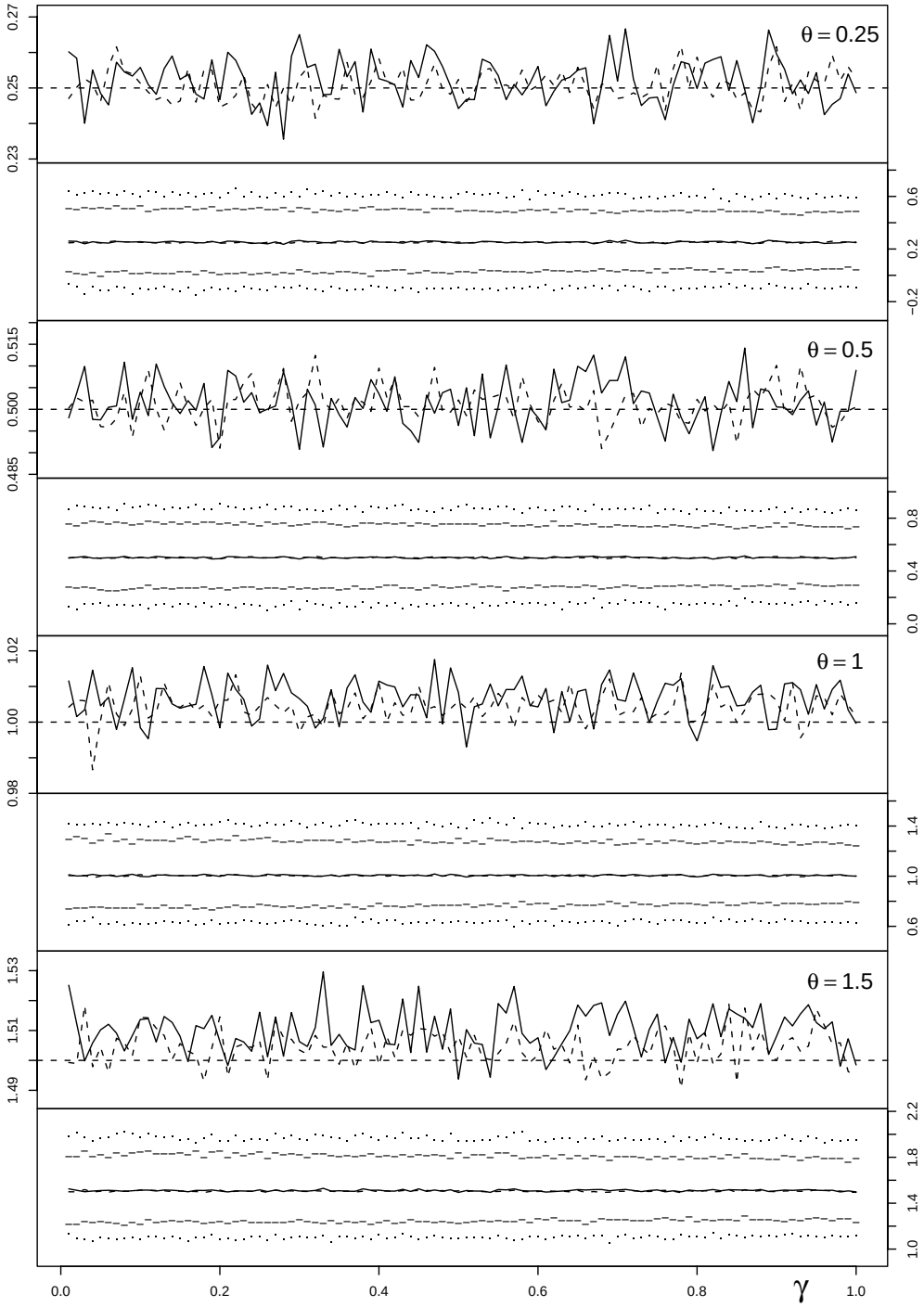


FIGURE 4. Empirical mean values and 95% envelopes for the estimator $\widehat{\theta_0}$ for four different values of the inhomogeneity parameter $\theta$ (values are indicated in the plots) and for template parameter $\beta = 250$ (full drawn lines, resp. dashed lines for envelopes) and $\beta = 100$ (dashed lines, resp. dotted lines for envelopes), as a function of the template parameter $\gamma$. The central lines in the envelope plots are the empirical means again.

Note that since the scaling function has been normalized as in (14), (16) implies that

$$\mathbb{E}_{\theta,\psi}\, n(X_{c_\theta} \cap W) = \lambda_{0\psi} \nu^k(W), \tag{23}$$

i.e. the mean number of points in $W$ does not depend on the inhomogeneity parameter $\theta$. Since (16) does not hold exactly for the Strauss process, we investigated whether (23) holds approximately, using the simulated data. The approximation is excellent in this example, cf. Figure 5.

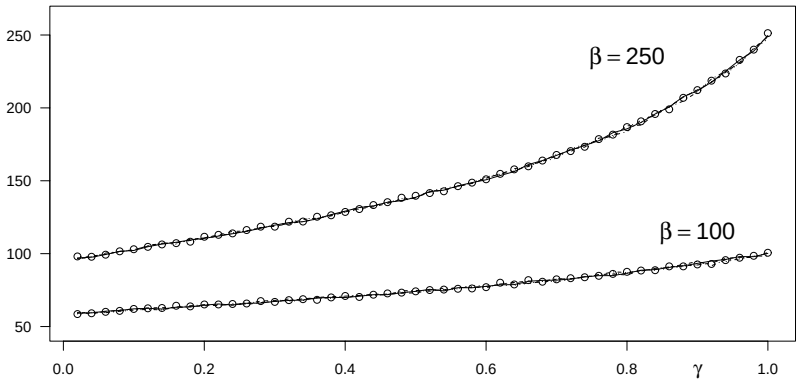


FIGURE 5. Comparison of intensities (empirical mean values) on the unit square for the exponentially scaled Strauss point process with different $\theta \in \{0.25, 0.5, 1, 1.5\}$ (dashed, chain-dotted, full and dotted curves) and the intensity of the template process (circles). As the maximal vertical difference between the curves is less than 3 these are hardly distinguishable.

Let us next study the estimation of the template parameters. The density of the Strauss process is of exponential family form with one nuisance parameter $r$ – the interaction radius (see Example 1). Thus $\psi = (\beta, \gamma, r)$ and

$$\log \lambda_j = \log \beta + s_c(u_j; \mathbf{x}) \log \gamma,$$

where

$$s_c(u_j; \mathbf{x}) = \sum_{x \in \mathbf{x} \setminus \{u_j\}} \mathbf{1}\{\nu_c^1([u_j, x]) \le r\}.$$

To find the estimate of $\psi$ we have to compute and compare the profile pseudolikelihood

$$\overline{PL}_{W,1}(r) = \max_{\beta,\gamma} PL_{W,1}(\widehat{\theta}_0, \beta, \gamma, r; \mathbf{x})$$

on a grid of values of $r$. We let $\beta_r$ and $\gamma_r$ be the values of $\beta$ and $\gamma$ at which

$$PL_{W,1}(\widehat{\theta}_0, \cdot, \cdot, r; \mathbf{x})$$

is maximal (the subscript $r$ indicates the dependence on $r$). In Figures 6 and 7 we illustrate the procedure on a simulated exponentially scaled Strauss point pattern with the scaling function (22) and parameters $\theta = 1$, $\beta = 250$, $\gamma = 0.25$, $r = 0.05$, $W = [0,1]^2$. The parameter $\theta$ has been fixed to the correct value and a regular grid of $100 \times 100$ dummy points was used. In the plots presented in Figures 6 and 7, the profile pseudolikelihood and the estimates $\hat{\beta}_r$ and $\hat{\gamma}_r$ are plotted as functions of the nuisance parameter $r$. The jaggedness of the plots is due to the discontinuity of the interpoint distance function $s_c$ as a function of $r$. In Figure 6 we used no border correction (the pseudolikelihood (18) with $\mathbf{x}$ replaced by $\mathbf{x} \cap W$) while in Figure 7 we used a border correction of $\nu_c^1 = 0.05$ (the psedolikelihood (18) with $W$ replaced by an irregular observation window $\widetilde{W} = \{u \in W : \nu_c^1(u, \partial W) > 0.05\}$, where $\partial W$ denotes the boundary of $W$).

The obtained estimates of $\psi$ are in good agreement with the true values, especially the estimate of the interaction radius is very precise. It is also important that the estimates with and without border correction do not differ substantially (which is probably caused by the sufficiently large number of observed points in $W$).

The results concerning pseudolikelihood estimation were confirmed in repeated simulation experiments.

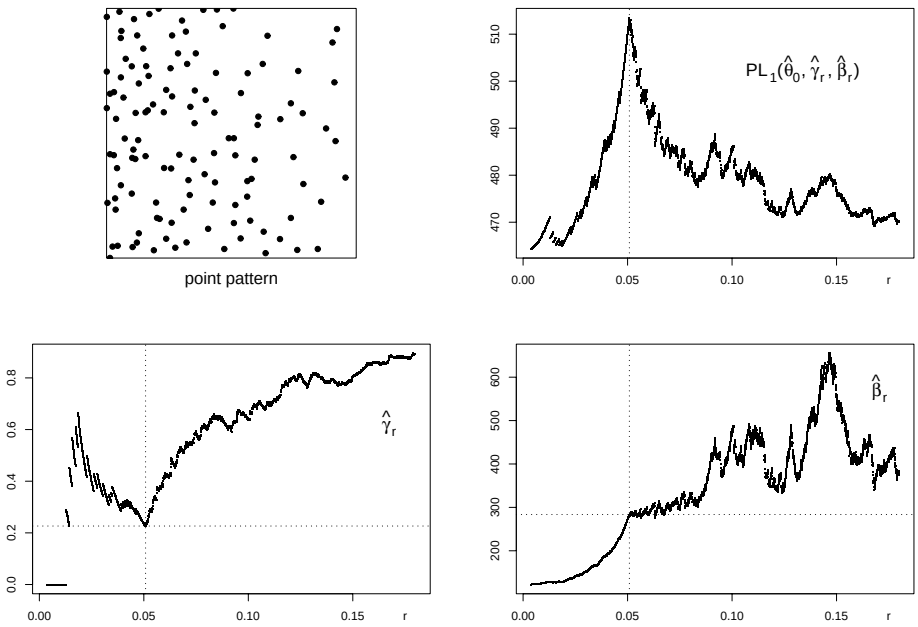


FIGURE 6. Profile pseudolikelihood estimation of the template parameters $\beta, \gamma, r$ of a simulated exponentially scaled Strauss process on $[0,1]^2$ without any border correction. The first picture shows the data $\mathbf{x}$. In the 3 graphs, the profile pseudolikelihood and the corresponding estimates $\hat{\beta}_r$ and $\hat{\gamma}_r$ are plotted as functions of $r$. The final estimates $\hat{r} = 0.0508$, $\hat{\beta} = 283$, $\hat{\gamma} = 0.23$ are indicated by the dotted lines. The true values are $r = 0.05$, $\beta = 250$, $\gamma = 0.25$.

## 6. Two step inference where scaling function is estimated using other (non ML) methods

Going one step further, one could also estimate $c$ in some other way from the local intensity

$$\lambda_c(u) = \mathbb{E}\, \lambda_c(u \mid X_c),$$

$u \in \mathcal{X}$, of the locally scaled process $X_c$, using the approximate relation

$$\lambda_c(u) \approx c(u)^{-k} \lambda_0, \tag{24}$$

where $\lambda_0$ is the intensity of the template process. In order to estimate the scaling function, we could use an estimate $\hat{\lambda}_c(u)$ of the local intensity, and set

$$\hat{c}(u) = \left[K \hat{\lambda}_c(u)\right]^{-1/k},$$

where $K = 1/\lambda_0$ is some constant that can be arbitrarily fixed. For convenience, one may choose $K = 1$.

If, in a parametric setting, $\hat{\lambda}_c(u)$ is the maximum likelihood estimator of the intensity of an inhomogeneous Poisson process, then $\hat{c}(u)$ is the same partial MLE as the one based on $L_0$. On the other hand, $\lambda_c(u)$ can also be estimated non parametrically, for example, using kernel methods or Voronoi tessellations. Or parametrically by other methods than maximum likelihood, e.g. regression methods.

## 7. Model validation

Since the two-step estimation procedure, suggested in the present paper, can only be justified theoretically in special cases, it is particularly important to develop effective procedures for model validation. This is the topic of the present section. For non-Poisson point processes only little is known about the theoretical distribution of characteristics that can be used for model validation. Therefore tests are usually simulation based.

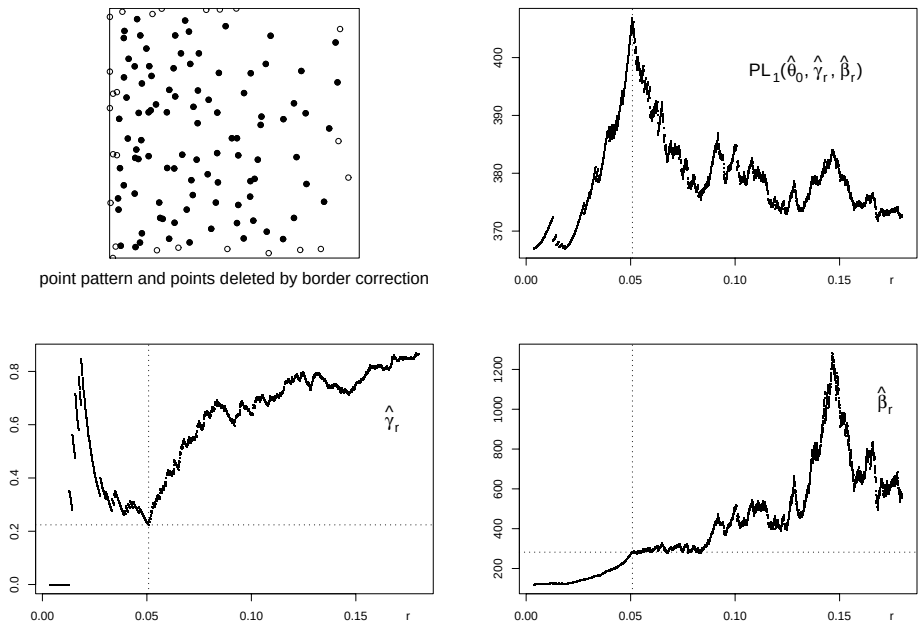


FIGURE 7. Profile pseudolikelihood estimation of the template parameters $\beta, \gamma, r$ of a simulated exponentially scaled Strauss process on $W = [0,1]^2$ with border correction $\nu_c^1 = 0.05$. The first picture shows the same data used in Figure 6. The full circles are the data points used for the estimation. The 3 graphs are constructed as in Figure 6. The obtained estimates are $\hat{r} = 0.0508$, $\hat{\beta} = 282$, $\hat{\gamma} = 0.22$. The true values are $r = 0.05$, $\beta = 250$, $\gamma = 0.25$.

In the case of homogeneous point processes, the probably most popular tests are based on second order summary statistics such as Ripley's $K$-function, see e.g. [Diggle, 1983] and [Møller and Waagepetersen, 2003]. [Baddeley et al., 2000] extend the definition of the $K$-function and other second order characteristics to a certain type of inhomogeneous point processes, so called second order intensity reweighted stationary processes. This class of processes comprises inhomogeneous processes obtained by independent thinning, but does not include locally scaled processes (apart from the Poisson process).

[Grabarnik and Chiu, 2002] consider another so-called $Q^2$ statistic for tests against Poisson processes, which simply spoken amounts to a goodness-of-fit test for the frequency distribution of number of neighbours in an $r$-neighbourhood. In this section we investigate how $K-$functions and $Q^2$-statistics can be adapted to locally scaled inhomogeneous point processes.

### 7.1. The $K$-function

The $K$-function for a stationary point process $X$ is defined as the expected number of points in $X \setminus \{0\}$, given $0 \in X$, in a ball of radius $r$ around 0, divided by the intensity $\lambda_0$ of $X$

$$K_0(r) = \frac{\mathbb{E}_0^! n(X \cap b(0,r) \setminus \{0\})}{\lambda_0}.$$

(In the case of finite processes with interaction, we assume that the domain of $X$ is large enough that $X \cap b(0,r)$ is virtually free of boundary effects.) By the Campbell-Mecke theorem, a ratio-unbiased estimator of $K_0(r)$ is given by

$$\widehat{K}_0(W,r) = \frac{1}{\lambda_0 n(\mathbf{x} \cap W)} \sum_{x \in \mathbf{x} \cap W} \sum_{y \in \mathbf{x} \setminus \{x\}} \mathbf{1}\{\nu^1([x,y]) \le r\},$$

where $\mathbf{x}$ is an observed point pattern from the stationary point process $X$. If instead a locally scaled point pattern $\mathbf{x}$ is observed, we suggest to use a locally scaled analogue of

$\widehat{K}_0(W,r)$, viz.

$$\widetilde{K}_0(W,r) = \frac{1}{\lambda_0 n(\mathbf{x}\cap W)} \sum_{x\in\mathbf{x}\cap W} \sum_{y\in\mathbf{x}\setminus\{x\}} \mathbf{1}\{\nu_c^1([x,y]) \le r\}. \tag{25}$$

Note that $\widetilde{K}_0(W,r)$ is ratio-unbiased for $K_0(r)$ if $c$ is constant. Furthermore, $\widetilde{K}_0(W,r)$ is ratio-unbiased for general scaling functions and distance-interaction point processes defined on an interval $I$ of $\mathbb{R}^1$. To see this, we use that in $\mathbb{R}^1$ a locally scaled distance-interaction process $X_c$ has the same distribution as $h(X)$ where $h$ is a 1-1 differentiable transformation of $I$ onto $I$ with $(h^{-1})' = c^{-1}$. (A proof of this result can be found in the Appendix.) Therefore, we have

$$\begin{aligned}
&\mathbb{E}\Big(\sum_{x\in X_c\cap W} \sum_{y\in X_c\setminus\{x\}} \mathbf{1}\{\nu_c^1([x,y]) \le r\}\Big)\\
&= \mathbb{E}\Big(\sum_{x\in h(X)\cap W} \sum_{y\in h(X)\setminus\{x\}} \mathbf{1}\{\nu^1([h^{-1}(x),h^{-1}(y)] \le r\}\Big)\\
&= \mathbb{E}\Big(\sum_{x\in X\cap h^{-1}(W)} \sum_{y\in X\setminus\{x\}} \mathbf{1}\{\nu^1([x,y]) \le r\}\Big).
\end{aligned}$$

Accordingly, the ratio-unbiasedness of $\widetilde{K}_0$ follows from the ratio-unbiasedness of $\widehat{K}_0$. Generally, $\widetilde{K}_0$ is expected to be (approximately) ratio-unbiased if $r$ is small such that $c$ varies little in a scaled neighbourhood. In any case, one should use simulations of the scaled null hypothesis model, not only of the template, for model validation.

A further simplification is accomplished by applying $\nu_c^1([x,y]) \approx \frac{2}{c(x)+c(y)}\nu^1([x,y])$, which was introduced for distance-interaction processes in [Hahn et al., 2003]. The corresponding statistic

$$\check{K}_0(W,r) = \frac{1}{\lambda_0 n(\mathbf{x}\cap W)} \sum_{x\in\mathbf{x}\cap W} \sum_{y\in\mathbf{x}\setminus\{x\}} \mathbf{1}\{\nu^1([x,y]) \le \tfrac{1}{2}(c(x)+c(y))r\} \tag{26}$$

is particularly useful if $c$ is estimated nonparametrically, because it requires evaluation of $c$ only in the data points.

In practical situations, both $\lambda_0$ and $c$ have to be estimated from the data. As discussed in the preceeding sections, the estimation of $c$ cannot be separated from the estimation of $\lambda_0$. Since the template is unique only up to a constant scale factor which determines $\lambda_0$, the scaling function $c$ is unique only up to a constant as well. We suggest to normalize $c$ such that $\nu_c^k(W) = \nu^k(W)$, see (14). Thus, we set $\hat{\lambda}_0 := n(\mathbf{x}\cap W)/\nu^k(W)$ since $\mathbb{E}n(X_c\cap W) = \int_W \lambda_c(x)\,\nu^k(\mathrm{d}x) \approx \int_W \lambda_0 c(x)^{-k}\,\nu^k(\mathrm{d}x) = \lambda_0\,\nu^k(W)$.

### 7.2. The $Q^2$ statistic

The $Q^2$-statistic proposed by [Grabarnik and Chiu, 2002] is (in the simplest case) based on the numbers $M_\ell(W,r)$ of points in $W$ with $\ell$ $r$-close neighbours, $\ell = 0,1,\ldots,q$. For a homogeneous Poisson point process, the expectation $\mu$ and the covariance matrix $\Sigma$ of the vector $M = (M_0, M_1, \ldots, M_q)^\top$ can easily be calculated. A finite range dependency argument is used to show that the statistic

$$Q^2 = (M-\mu)^\top \Sigma^{-1} (M-\mu)$$

is asymptotically $\chi^2$-distributed for increasing size of the observation window $W$. By simulation experiments, [Grabarnik and Chiu, 2002] showed that $Q^2$ discriminates well between patterns from a mixed cluster and regular point process and the Poisson process.

Since $\mu$ and $\Sigma$ can also be calculated for an inhomogeneous Poisson point process, it would be possible to use the same $Q^2$-statistic also for tests of inhomogeneous Poisson processes. However, the expected number of neighbours in a ball of radius $r$ around a point $x$ would depend on the local intensity $\lambda(x)$. Thence, inhomogeneity introduces much extra variation to $M$ which would largely cut down the diagnostic value of $Q^2$.

This effect can be avoided by adjusting $r$ to the local intensity. We propose to replace the Euclidean neighbour distance by the locally scaled neighbour distance. In an inhomogeneous Poisson point process with intensity $\lambda_c(x) = c(x)^{-k}\lambda_0$, the number of $r$-scaled-close neighbours of a point $x$ is Poisson distributed with parameter $\lambda_0\nu_c^k(b_c(x,r))$. Since

$$\frac{\nu_c^k(b_c(x,r))}{\nu^k(b(x,r))} \to 1 \text{ as } r \to 0,$$

the distribution of $r$-scaled-close neighbours does hardly depend on the location for small $r$, and is close to the distribution of $r$-close neighbour number in the homogeneous case.

The local scaling analogue of $M_\ell$ is

$$M_{\ell\text{inhom}}(W,r) = \sum_{x\in\mathbf{x}\cap W} \mathbf{1}\{n(b_c(x,r)\cap\mathbf{x}\setminus\{x\}) = \ell\}. \tag{27}$$

Since calculation of $\mu$ and $\Sigma$ is feasible only for the Poisson point process with slowly varying scaling function, we suggest to do simulation tests. This would allow to test any hypothesis. While any distance between observed and expected neighbour number distribution can be used, we still recommend to use the squared Mahalanobis distance $Q^2$, however to replace $\mu$ and $\Sigma$ with estimates obtained by simulation. Note that the simulations for estimating $\mu$ and $\Sigma$ are not to be reused for the test.

## 8. Data analysis

The plant point pattern in Figure 1 appears slightly clustered, so we need to model attractive interaction between the plants. The exponentially scaled area-interaction model appears to be a good candidate because the area of a location dependent neighbourhood around each plant enters explicitly into the model density.

We used the two-step fitting procedure. For convenience we rescaled the data to the unit square $W = [0,1]^2$. As the pattern exhibits obvious inhomogeneity in the vertical direction but appears quite homogeneous in the horizontal direction we used an exponential scaling function of the form

$$c(u) = \sqrt{\frac{1-e^{-2\theta}}{2\theta}}\,\mathrm{e}^{\theta u_2}, \quad u = (u_1,u_2) \in \mathbb{R}^2. \tag{28}$$

Based on $L_0(\theta;\mathbf{x}\cap W)$ we obtained the following estimate of $\theta$

$$\widehat{\theta}_0 = 1.0839,$$

with $\alpha(\widehat{\theta}_0) = 0.6391$, see (15).

Secondly, we maximized the pseudolikelihood $PL_{W,1}(\widehat{\theta}_0,\psi;\mathbf{x})$ with $\widehat{\theta}_0$ fixed. The density of the area-interaction process is of an exponential family form with one nuisance parameter $r$ - the interaction radius. As for the Strauss process, $\psi = (\beta,\gamma,r)$ and for the estimation we use the same weights as in (19) and

$$\begin{aligned} \log\lambda_j &= \log\beta - \nu_c^2(U_{c,r}(u_j;\mathbf{x}))\log\gamma, \\ U_{c,r}(u_j;\mathbf{x}) &= \{y\in W : \nu_c^1([y,u_j]) \le r, \nu_c^1([y,\mathbf{x}\setminus\{u_j\}]) > r\}. \end{aligned}$$

We used a grid of $100\times100$ dummy points which were equidistant in the horizontal direction and $\nu^1_{c_{\widehat{\theta}_0}}$ –equidistant in the vertical direction (actually this means that the dummy points were $\nu^1_{c_{\widehat{\theta}_0}}$ –equidistant in both directions – compare with (28) ).

We maximized the profile pseudolikelihood on a grid of $r$-values. The main problem is the computation of the scaled volumes $\nu_c^2(U_{c,r}(u_j;\mathbf{x}))$ for all the points $u_j$, $j = 1,\dots,m$. This can be done only approximately. To approximate these scaled volumes with a reasonable precision it is necessary to compute the scaled distances from the points $\{u_j, j = 1,\dots,m\}$ to each point in a very fine grid of points in $W$. This job is computationally quite demanding.

The approximate profile pseudolikelihood $PL_{W,1}(\widehat{\theta}_0,\psi;\mathbf{x})$ was computed with border correction $\nu_c^1 = 0.05$. This degree of border correction was chosen as a compromise between

minimizing the bias caused by missing unobserved points and not excluding too many observed points from the estimation (with the chosen border correction one forth of the points was not used in the estimation). The profile pseudolikelihood and estimates of the parameters as functions of $r$ are plotted in Figure 8. Note that the curves are smoother than in the case of the Strauss process because now the interaction function is continuous as a function of $r$. We obtained the following values

$$\widehat{r} = 0.085,\ \widehat{\beta} = 184,\ \widehat{\gamma} = 3.99 \cdot 10^{26},\ \widehat{\gamma}^{-\pi\widehat{r}^2} = 0.25. \tag{29}$$

The value of $\widehat{\gamma}^{-\pi\widehat{r}^2}$ is included because it gives a better impression of the strength of the interaction, as this is actually the term which appears in the template density. The fit indicates a slightly clustered point pattern as we expected.

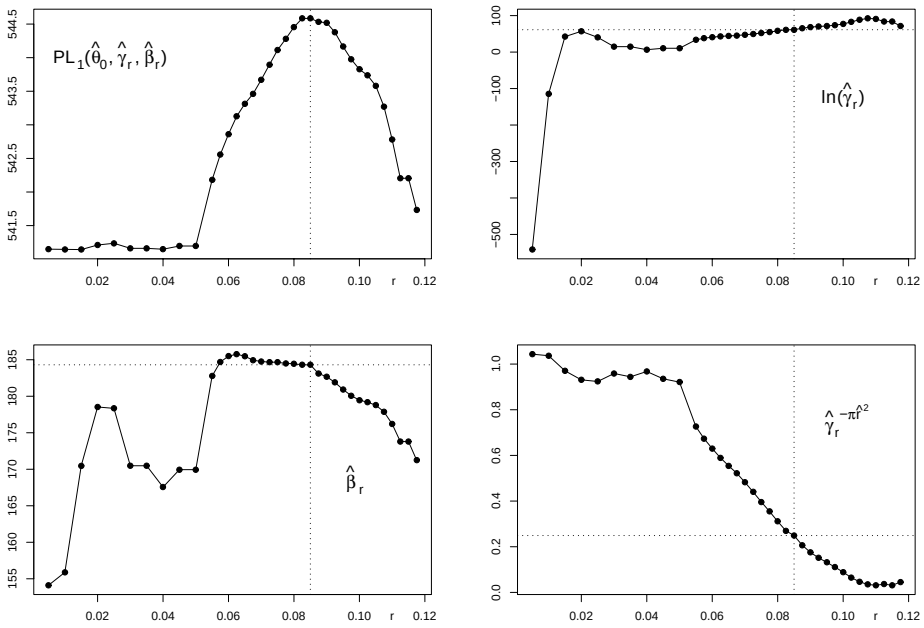


FIGURE 8. Pseudolikelihood estimation of the template parameters $\beta, \gamma, r$ of the exponentially scaled area-interaction process for the plant data from Figure 1 rescaled to $[0,1]^2$ with border correction $\nu_c^1 = 0.05$. The graphs show the profile pseudolikelihood and the corresponding estimates $\hat{\beta}_r$ and $\hat{\gamma}_r$ as functions of $r$. The last graph of $\widehat{\gamma}^{-\pi\widehat{r}^2}$ shows the strength of the attractive interaction. The resulting estimates are $\hat{r} = 0.085$, $\hat{\beta} = 184$, $\hat{\gamma} = 3.99 \cdot 10^{26}$.

For model validation we used the $\widetilde{K}_0(\widetilde{W}, r)$ and $Q^2$ statistics from Section 7. Figure 9 shows the locally scaled estimate $\widetilde{K}_0(\widetilde{W}, r)$ with $\widetilde{W} = \{u \in W : \nu_c^1(u, \partial W) > 0.05\}$ (full-drawn line) together with the empirical mean and 95% envelopes for $\widetilde{K}_0(\widetilde{W}, r)$ calculated from 399 simulations under the fitted exponentially scaled area-interaction model (dashed lines). The locally scaled estimate $\widetilde{K}_0(\widetilde{W}, r)$ for the plant data lies inside the envelopes of the fitted area-interaction model.

Next we tested the locally scaled Poisson hypothesis on the plant data. We used the

$$Q_{\mathrm{P}}^2 = (M_{\mathrm{inhom}} - \mu_P)^\top \Sigma_P^{-1} (M_{\mathrm{inhom}} - \mu_P)$$

statistic with $r = 0.05$ and $M_{\mathrm{inhom}} = (M_{0\mathrm{inhom}}, \ldots, M_{6\mathrm{inhom}})$ defined by (27). The subscript P indicates that in the formula for $Q^2$ we use as $\mu$ and $\Sigma$ the mean $\mu_{\mathrm{P}}$ and the covariance matrix $\Sigma_{\mathrm{P}}$ of $M_{\mathrm{inhom}}$ for the fitted locally scaled Poisson model with $\theta = \widehat{\theta}_0$. The values of $\mu_{\mathrm{P}}$ and $\Sigma_{\mathrm{P}}$ were estimated from 8000 simulated realizations of the fitted locally scaled Poisson model.

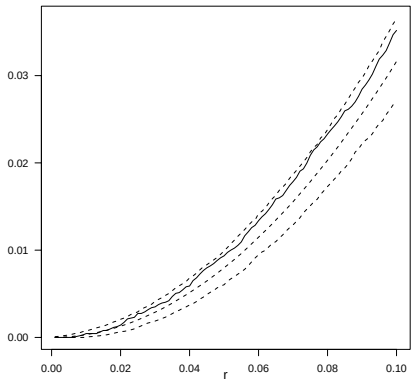


FIGURE 9. The estimate $\widetilde{K}_0$ for the plant data (full drawn line) and mean and 95% envelopes for $\widetilde{K}_0$ for the exponentially scaled area-interaction model (dashed lines).

The simulation test (using 499 realizations of the hypothesis locally scaled Poisson model with $\theta = \widehat{\theta}_0$) gives the p-value of 0.05. Thus the plant data is not very well described by the Poisson model.

Then we used the $Q_{\mathrm{A}}^2$ statistic (i.e. the mean value $\mu_{\mathrm{A}}$ and covariance matrix $\Sigma_{\mathrm{A}}$ of $M_{\mathrm{inhom}}$ are computed for the fitted exponentially scaled area-interaction model) for testing of the fitted locally scaled area-interaction model. The test gave the p-value of 0.106.

## References

[Armstrong, 1991] Armstrong, P. (1991). *Species patterning in heath vegetation of the Northern Sandplain.* Honours Thesis, Murdoch University, Western Australia.

[Baddeley et al., 2003] Baddeley, A. J., Hazelton, M., Møller, J., and Turner, R. (2003). Residuals and diagnostics for spatial point processes. Abstract, 54 ISI Session Berlin, 13-20 August 2003.

[Baddeley et al., 2000] Baddeley, A. J., Møller, J., and Waagepetersen, R. (2000). Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350.

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

[Baddeley and van Lieshout, 1995] Baddeley, A. J. and van Lieshout, M. N. M. (1995). Area-interaction point processes. *Ann. Inst. Statist. Math.*, 47:601–619.

[Barndorff-Nielsen, 1978] Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester.

[Diggle, 1983] Diggle, P. J. (1983). *Statistical Analysis of Point Patterns*. Chapman and Hall.

[Grabarnik and Chiu, 2002] Grabarnik, P. and Chiu, S. N. (2002). Goodness-of-fit test for complete spatial randomness against mixtures of regular and clustered spatial point processes. *Biometrika*, 89(2):411–421.

[Hahn et al., 2003] Hahn, U., Jensen, E. B. V., van Lieshout, M. N. M., and Nielsen, L. S. (2003). Inhomogeneous spatial point processes by location-dependent scaling. *Adv. Appl. Prob.*, 35:319–336.

[Møller and Waagepetersen, 2003] Møller, J., and Waagepetersen, R. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton.

[Nguyen and Zessin, 1976] Nguyen, X. X. and Zessin, H. (1976). Punktprozesse mit Wechselwirkung. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 37:91–126.

[Nielsen, 2001] Nielsen, L. S. (2001) *Point process models allowing for interaction and inhomogeneity*. PhD Thesis, Department of Mathematical Sciences, University od Aarhus.

[Nielsen and Jensen, 2004] Nielsen, L. S. and Jensen, E. B. V. (2003). Statistical inference for transformation inhomogeneous point processes. *Scand. J. Statist.*, 31:131–142.

[Strauss, 1975] Strauss, D. J. (1975). A model for clustering. *Biometrika*, 63:467–475.

## Appendix A. Proof of relation

Let us suppose that $X$ is a distance-interaction process on an interval $I = [a, b]$ of $\mathbb{R}^1$ with density

$$f_X(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \prod_{\mathbf{y} \subseteq_2 \mathbf{x}} \varphi(\{\nu^1([u,v]) : \{u,v\} \subseteq \mathbf{y}, u \neq v\}),$$

where $\subseteq_2$ indicates that $\mathbf{y}$ should have at least two elements. The density of $X_c$ is then

$$f_{X_c}(\mathbf{x}) \propto \prod_{x \in \mathbf{x}} c(x)^{-1} \times \beta^{n(\mathbf{x})} \prod_{\mathbf{y} \subseteq_2 \mathbf{x}} \varphi(\{\nu_c^1([u,v]) : \{u,v\} \subseteq \mathbf{y}, u \neq v\}).$$

Let us consider the 1-1 differentiable transformation $h$ of $I$ onto $I$ defined by

$$h^{-1}(x) = \int_a^x c(u)^{-1} \mathrm{d}u.$$

Then, the density of $X_c$ can be rewritten as

$$f_{X_c}(\mathbf{x}) \propto \prod_{x \in \mathbf{x}} Jh^{-1}(\mathbf{x}) \cdot f_X(h^{-1}(\mathbf{x})).$$

It follows that $X_c$ is distributed as $h(X)$. In particular, for $A \in \mathcal{B}(I)$,

$$\begin{aligned} \mathbb{E}n(X_c \cap A) &= \mathbb{E}n(X \cap h^{-1}(A)) \\ &= \int_{h^{-1}(A)} \lambda_0 \mathrm{d}x \\ &= \int_A c(u)^{-1} \lambda_0 \mathrm{d}u, \end{aligned}$$

or

$$\lambda_c(u) = c(u)^{-1} \lambda_0.$$