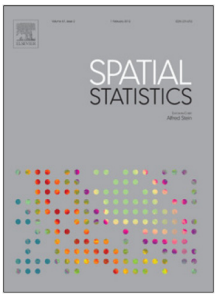
Estimating the intensity function of spatial point processes outside the observation window

E. Gabriel, Jérôme Coville, Joel Chadoeuf

Accepted Manuscript



Estimating the intensity function of spatial point processes outside the observation window

E. Gabriel, J. Coville, J. Chadœuf

PII: S2211-6753(17)30035-0
DOI: http://dx.doi.org/10.1016/j.spasta.2017.07.008
Reference: SPASTA 250

To appear in: *Spatial Statistics*

Received date : 1 February 2017
Accepted date : 25 July 2017

Please cite this article as: Gabriel, E., Coville, J., Chadœuf, J., Estimating the intensity function of spatial point processes outside the observation window. *Spatial Statistics* (2017), http://dx.doi.org/10.1016/j.spasta.2017.07.008

Estimating the intensity function of spatial point processes outside the observation window

E. Gabriel[1,2,*] and J. Coville[2] and J. Chadœuf[3]

[1] *Laboratory of Mathematics, Avignon University, F-84916 Avignon, France; edith.gabriel@univ-avignon.fr*
[2] *Biostatistic and Spatial Processes Unit, INRA, F-84911 Avignon, France; jerome.coville@inra.fr*
[3] *Statistics, UR1052, INRA, F-84911 Avignon, France; joel.chadoeuf@inra.fr*
[*] *Corresponding author*

Abstract. *Mapping the intensity of objects, as animal or plant species in ecological studies, is cumbersome as soon as these objects are not accessible by automated methods. The knowledge at large scale of the underlying process variability can then only be obtained through sampling and spatial prediction. Here, we aim to predict the intensity of a point process, at locations where it has not been observed, conditional to the observation using the best linear unbiased combination of the point process realization in the observation window. We show that the weight function associated to the predictor is the solution of a Fredholm equation of second kind. Both the kernel and the source term of the Fredholm equation are related to the second-order characteristics of the point process through the pair correlation function. We propose here several approximations to solve the Fredholm equation in order to obtain practical solutions and restrict the solution space to that generated by linear combinations of* (i) *step functions, which lead to a direct solution and* (ii) *elementary functions of a finite element basis, which provide a continuous approximation. Results are illustrated on simulations and to predict the intensity of Black Locust in a region of France.*

Keywords. *Fredholm equation; Finite element approach; Intensity estimation; Point process; Prediction; Spatial statistics.*

1 Introduction

In many applications the study window is too large to extensively map local intensity variations of the point process of interest since observation methods may be available at a much smaller scale only. That is for instance the case when studying the spatial repartition of a bird species at a regional scale, while the observations are made in windows of few hectares; or when detecting disease at the field scale, while observations correspond to spots of a few meter squares; or when mapping the presence of plant species at the catchment scale, while the observation scale is the meter square. The intensity must then be estimated from data issued out of samples spread in the study window, and hence, from a partial realization of the point process in this window.

We thus want to predict the intensity of a stationary point process conditional to its realization within the observation window W at any point $x_o \notin W$. In the sequel, this conditional intensity is called local intensity [10]. It allows us, through the conditioning, to take into account the second-order structure in

the prediction. As an example, let us consider the Thomas process which is a Poisson cluster process where the cluster centers (parents) are assumed to be Poisson and the offsprings are normally distributed around their parent points. This process is stationary. When the boundary of W splits a cluster, the local intensity across the boundary should be larger than the global intensity, i.e. the mean number of points per unit area. Another example is a soft core process, which is a pairwise interaction process. In this case, when an observed data point is located close to the boundary of W, the local intensity should be smaller than the global intensity as fewer points are expected around this point due to an interpoint interaction which decreases with distance.

Few ways exist to predict the local intensity. A first way consists of using [19]'s reconstruction method based on the first- and second-order characteristics of the point process. Once the empirical point pattern predicted within a given window, one can get the intensity by kernel smoothing. As it is a simulation-based method, it requires long computation times, especially when the prediction window is large and/or the point process is complex. In alternative methods, the intensity of the point process is driven by a stationary random field, and whilst based on different concepts, Bayesian [9, 8] or geostatistics [14, 1], they are constrained within the class of Cox processes. [20] developed for a wider class of parametric models a Bayesian approach for extrapolating and interpolating clustered point patterns. [16] compared geostatistical methods for mapping object counts collected from strip transects. Although no model is then required, the direct application of kriging methods does not allow to make good prediction for repulsive processes or processes with peaky local variations. Indeed, variograms are directly computed on count data and do not take explicitly into account the structure of the point process through the pair correlation function for instance. In what follows, we propose to predict the local intensity from the first- and second-order characteristics of the point process, which we can classically estimate in practice and which allow to handle a large scope of models. A first version of such an approach has been developed in [10], where the point process is regularized to get a count process over a grid and where the ordinary kriging is then adapted, with kriging weights defined from the structure of the point process (intensity and pair correlation function). This empirical approach thus mimics geostatistics through the link between the variogram and the pair correlation function. Whilst it can be applied in practice, it does not allow us to understand how the weights are built. The continuous approach developed in this paper generalizes the previous one and offers new (and better) approximations.

The predictor of the local intensity is defined in Section 2 as a linear combination of the point process realization. We get an exact, but not explicit, solution of the related weights, given by an integral equation (second kind Fredholm equation). Our approach clearly brings out how the pair correlation function operates and then why the proposition given in Gabriel *et al.* (2016) is a good approximation of our new theoretical solution. It also opens to other approximations, in particular we present here a finite element approach. Finally in Section 3, we illustrate the method on simulated and real data, including mixing processes with different interaction structures, at different scales.

2 Method

Let Φ be a stationary point process, with global intensity λ and pair correlation function g. We assume that Φ is absolutely continuous w.r.t the Poisson measure and we denote Φ_W its realization in the observation window W. We define the *local intensity* of Φ by its intensity conditional to Φ_W and we denote it $\lambda(x|\Phi_W)$. We denote W_{pred} the window in which we want to predict the local intensity, $W \cap W_{pred} = \emptyset$.

2.1 Definition of the predictor

We propose to predict the local intensity at locations where the process has not been observed, $x_o \in W_{pred}$, using the best linear unbiased combination of the point process realization in the observation window (which may be generally not a connected set due to sampling rules). Our predictor is similarly built as the kriging interpolator in geostatistics, except that we don't have a continuous random field in $\mathbb{R}$.

Proposition 2.1 *For $x_o \in W_{pred}$, the predictor*

$$\widehat{\lambda}(x_o|\Phi_W) = \int_{\mathbb{R}^2} w(x;x_o) \sum_{y\in\Phi_W} \delta(x-y)\,dx = \sum_{x\in\Phi_W} w(x;x_o)$$

is the Best Linear Unbiased Predictor of $\lambda(x_o|\Phi_W)$. Its variance is

$$\mathbb{V}\mathrm{ar}\left(\widehat{\lambda}(x_o|\Phi_W)\right) = \lambda \int_W w^2(x;x_o)\,dx + \lambda^2 \int_{W\times W} w(x;x_o)w(y;x_o)\,(g(x-y)-1)\,dx\,dy. \tag{1}$$

The weight function $w(x;x_o)$ satisfies $\int_W w(x;x_o)\,dx = 1$ and is solution of the Fredholm equation of the second kind:

$$w(x;x_o) + \int_W w(y;x_o)k(x,y)\,dy = f(x;x_o), \tag{2}$$

with kernel

$$k(x,y) = \lambda\left(g(x-y) - \frac{1}{\nu(W)}\int_W g(x-y)\,dx\right) \tag{3}$$

and source term

$$f(x;x_o) = \frac{1}{\nu(W)} + \lambda\left(g(x-x_o) - \frac{1}{\nu(W)}\int_W g(x-x_o)\,dx\right). \tag{4}$$

Note that the weights depend on the location x_o at which the prediction is made (see the proof below). However for ease of notation, we will use the notation $w(x)$ throughout the paper.

Similar results were found by [17, 4] on continuous random fields, where the kernel and the source term of the Fredholm equation were related to the variogram.

Sketch of proof:

We build the predictor as a kriging interpolator. Thus, we want it to be (i) a linear combination of the point process realization, (ii) unbiased and (iii) minimizing the variance prediction error.

(i) Using [6]'s notations, we set $\widehat{\lambda}(x_o|\Phi_W) = \int_{\mathbb{R}^2} w(x;x_o)\sum_{y\in\Phi_W}\delta(x-y)\,\mathrm{d}x = \sum_{x\in\Phi_W} w(x)$.

(ii) The local intensity at a location x_o given Φ_W is the limit $\lambda(x_o|\Phi_W) = \lim_{\nu(B)\to 0}\dfrac{\mathbb{E}\left[\Phi(B\oplus x_o)|\Phi_W\right]}{\nu(B)}$, where B is an elementary surface around x_o.

As we want the predictor to be unbiased, we get the constraint on the weight function:

$$\begin{aligned}\mathbb{E}\left[\widehat{\lambda}(x_o|\Phi_W)-\lambda(x_o|\Phi_W)\right]=0 &\iff \int_W \lambda w(x)\,\mathrm{d}x-\mathbb{E}\left[\lim_{\nu(B)\to 0}\frac{\mathbb{E}\left[\Phi(B\oplus x_o)|\Phi_W\right]}{\nu(B)}\right]=0\\ &\iff \lambda\left(\int_W w(x)\,\mathrm{d}x-1\right)=0\\ &\iff \int_W w(x)\,\mathrm{d}x=1.\end{aligned}$$

(*iii*) Finally, the predictor must minimize the error prediction variance. Reminding the variance of the number of points of a point process (see e.g. [6])

$$\mathbb{V}\mathrm{ar}\left(\Phi(B)\right)=\lambda\nu(B)+\lambda^2\int_{B\times B}\left(g(x-y)-1\right)\mathrm{d}x\mathrm{d}y$$

and the following limit for a continuous pair correlation function

$$\lim_{\nu(B)\to 0}\frac{1}{\nu(B)}\int_{B_o\times W}\left(g(x-y)-1\right)\mathrm{d}x\mathrm{d}y=\int_W\left(g(x_o-x)-1\right)\mathrm{d}x,$$

we have that minimizing the error prediction variance is equivalent to minimize

$$\lambda\int_W w^2(x)\,\mathrm{d}x+\lambda^2\int_{W\times W}w(x)w(y)\left(g(x-y)-1\right)\mathrm{d}x\mathrm{d}y-2\lambda^2\int_W w(x)\left(g(x_o-x)-1\right)\mathrm{d}x.$$

We use Lagrange multipliers under the constraint on the weight function and we set

$$\begin{aligned}T(w(x))=\lambda\int_W w^2(x)\,\mathrm{d}x+\lambda^2\int_{W\times W}w(x)w(y)\left(g(x-y)-1\right)\mathrm{d}x\mathrm{d}y\\ -2\lambda^2\int_W w(x)\left(g(x_o-x)-1\right)\mathrm{d}x+\mu\left(\int_W w(x)\,\mathrm{d}x-1\right),\end{aligned}$$

where μ is the Lagrange multiplier. Then for $\alpha(x)=w(x)+\varepsilon(x)$, we get

$$T(\alpha(x))\approx T(w(x))+2\lambda\int_W\varepsilon(x)\left[w(x)+\lambda w(y)\left(g(x-y)-1\right)\mathrm{d}y-\lambda\left(g(x_o-x)-1\right)+\frac{\mu}{2\lambda}\right]\mathrm{d}x.$$

Variational calculation and the Riesz representation theorem lead to

$$\begin{aligned}T(\alpha(x)) - T(w(x)) &= o(\varepsilon(x))\\ \Leftrightarrow \int_W\varepsilon(x)\left[w(x)+\lambda\int_W w(y)\left(g(x-y)-1\right)\mathrm{d}y-\lambda\left(g(x_o-x)-1\right)+\frac{\mu}{2\lambda}\right]\mathrm{d}x&=0\\ \Leftrightarrow\ w(x)+\lambda\int_W w(y)\left(g(x-y)-1\right)\mathrm{d}y-\lambda\left(g(x_o-x)-1\right)+\frac{\mu}{2\lambda}&=0 \qquad (5)\end{aligned}$$

The integration of Equation (5) leads to

$$1+\lambda\int_{W^2}w(y)\left(g(x-y)-1\right)\mathrm{d}y\mathrm{d}x-\lambda\int_W\left(g(x_o-x)-1\right)\mathrm{d}x+\frac{\nu(W)}{2\lambda}\mu=0,$$

from which we can obtain parameter μ. Then we replace μ in Equation (5) to get

$$\begin{aligned}w(x)+\lambda\int_W w(y)\left(g(x-y)-1\right)\mathrm{d}y-\frac{1}{\nu(W)}\left[1+\lambda\int_{W^2}w(y)\left(g(x-y)-1\right)\mathrm{d}x\mathrm{d}y\right]\\ =\lambda\left(g(x_o-x)-1\right)-\frac{\lambda}{\nu(W)}\int_W\left(g(x_o-x)-1\right)\mathrm{d}x.\end{aligned}$$

and then the Fredholm equation (2). From this integral equation, we can see that $w(x)$ is a weight function defined on W and which depends on x_o. □

2.2 Approximated solutions

Our objective is now to solve the Fredholm equation (2). Any existing solution already considered in the literature can be used. However, we don't aim to predict the local intensity at few number of points only, but rather to map the intensity in a given window. The problem is then to compute fast and accurately solutions of this Fredholm equation. We propose here to restrict the solution space to that generated by linear combinations of basis functions. We consider different bases: step functions, finite elements and splines.

Step functions

Let us consider that the observation window W is partitioned as an union of squared cells. We denote B an elementary square centered at 0, $B_i = x_i \oplus B$ the elementary square centered at x_i such that $B_i \cap B_j = \emptyset$, and n is the number of grid cell centers lying in W. For weight function defined as step function

$$w(x) = \sum_{j=1}^{n} w_j \frac{\mathbf{1}_{\{x \in B_j\}}}{\nu(B)}.$$

The solution of the Fredholm equation is direct as we thus get the classical kriging equations (see e.g. [5]). The predictor corresponds to the ordinary kriging interpolator of the number of points in each grid cell over the elementary area of a cell (see [10]):

$$\widehat{\lambda}(x_o|\Phi_W) = \sum_{j=1}^{n} w_j \frac{\Phi(B_j)}{\nu(B)},$$

where the weights w_j depend on the local structure of the point process through the covariance matrix C and the covariance vector C_o which are both expressed from the intensity and the pair correlation function of the point process Φ:

$$w = (w_1, \ldots, w_n) = \frac{1}{\nu(W)} C^{-1} C_o + \frac{1 - \mathbf{1}^T C^{-1} C_o}{\mathbf{1}^T C^{-1} \mathbf{1}} C^{-1} \mathbf{1},$$

where $C = \lambda\nu(B)\mathbb{I} + \lambda^2\nu^2(B)(G-1)$ with $G = \{g_{ij}\}_{i,j=1,\ldots,n}$, $g_{ij} = \frac{1}{\nu^2(B)} \int_{B\times B} g(x_i - x_j + u - v)\,\mathrm{d}u\,\mathrm{d}v$, $\mathbb{I}$ is the $n \times n$-identity matrix, and $C_o = \lambda^2\nu^2(B)(G_o - 1)$, with $G_o = \{g_{io}\}_{i=1,\ldots,n}$.

It is then possible to get some properties of this predictor. In particular, [10] used Neuman series to invert the covariance matrix, when $\lambda\nu(B)$ tends to 0,

$$C^{-1} = \frac{1}{\lambda\nu(B)} \left[\mathbb{I} + \lambda\nu(B) J_\lambda\right],$$

where a generic element of the matrix J_λ is given by

$$J_\lambda[i,j] = \sum_{k=1}^{\infty} (-1)^k \lambda^{k-1} \left(g(x_i, x_{l_1}) - 1\right) \left(g(x_{l_{k-1}}, x_j) - 1\right) \times \int_{W^{k-1}} \prod_{m=1}^{k-2} \left(g(x_{l_m}, x_{l_{m+1}}) - 1\right)\, \mathrm{d}x_{l_1} \ldots\, \mathrm{d}x_{l_{k-1}}.$$

This leads to the following variance of the predictor

$$\begin{aligned} \mathbb{V}\mathrm{ar}\left(\widehat{\lambda}(x_o|\Phi_W)\right) &= \lambda^3\nu^2(B)(G_o-1)^T(G_o-1) + \lambda^4\nu^3(B)(G_o-1)^T J_\lambda (G_o-1) \\ &\quad \frac{1 - \left[\lambda\nu(B)\mathbf{1}^T(G_o-1) + \lambda^2\nu^2(B)\mathbf{1}^T J_\lambda (G_o-1)\right]^2}{\frac{\nu(W)}{\lambda} + \nu^2(B)\mathbf{1}^T J_\lambda \mathbf{1}}. \end{aligned}$$

Finite Element Approach

In this section we use a Galerkin method (see for example [13] for a detailed presentation of the method) and as in [3, 7] we use a Finite Element Space for the approximation space. To be more precise, let $\mathcal{T}_h$ be a mesh of the observation window W, that is a collection of geometrically simple elements (we used triangles in the simulations) partitioning[1] W_{obs}. The positive real parameter h is related to the size of the elements in $\mathcal{T}_h$. We construct an approximation space V_h, which is a finite dimensional subspace of $L^2(W_{obs})$. As in a standard Galerkin scheme, setting $N = \dim V_h$ and considering a basis $\{\varphi_i\}_{i=1,\dots,N}$ of V_h, we can then approximate $w(x) \approx \sum_{j=1}^{N} \omega_j \varphi_j(x)$ which, plugged into the equation leads to the linear problem for all φ_i

$$\sum_{j=1}^{N} w_j \int_W \left(\varphi_i(x)\varphi_j(x) + \int_W k(x,y)\varphi_j(y)\varphi_i(x)\,\mathrm{d}y \right) \mathrm{d}x = \int_W f(x)\varphi_i(x)\,\mathrm{d}x, \tag{6}$$

where the kernel $k(x,y)$ and the source term $f(x)$ are defined by (3) and (4) respectively, By introducing a matrix formulation equation (6) reformulates as the so-called Galerkin equation

$$Mw + Kw = F, \tag{7}$$

where M is the FEM mass matrix, F is the vector $\left(\int_W f(x)\varphi_i(x)\,\mathrm{d}x \right)_{i=1,\dots,N}$ and K is a matrix with elements defined by

$$K_{ij} = \int_{W_{obs}} \int_{W_{obs}} k(x,y)\varphi_i(x)\varphi_j(y)\,\mathrm{d}x\,\mathrm{d}y.$$

To simplify the resolution of (7), we approximate the matrix K by using the projection on $V_h \times V_h$,

$$k(x,y) \approx \sum_{l,m} \mathcal{K}_{lm}\varphi_l(x)\varphi_m(y).$$

By setting $\mathcal{K}$ the matrix defined by this projection, we then get $K = M\mathcal{K}M$ and the system (7) leads to consider the problem:

$$(Id + \mathcal{K}M)w = M^{-1}F,$$

where Id denotes the identix matrix. When the mesh $\mathcal{T}_h$ is sufficiently fine, the above linear problem inherits the resolvability of the Fredholm equation ensuring the consistency of the approximation.

Linear splines

The weight function of $\widehat{\lambda}(x_o|\Phi_W) = \sum_{x\in\Phi_W} w(x)$ could be defined as a degree d spline curve. For sake of clarity, in the sequel we illustrate this approximation in $\mathbb{R}$ and we consider linear splines with equally spaced knots.

Let $W = [t_0, t_n] \cup [t'_0, t'_{n'}]$ be an observation window defined by two disjoint intervals as illustrated Figure 1 and Δ be the distance between two consecutive knots of each interval, $\Delta = t_{i+1} - t_i$, $t_i < t_{i+1}$.

[1] In practice, the mesh may produce only an *approximation* of the domain, which can possibly entice several technical difficulties when analysing the convergence of the method.

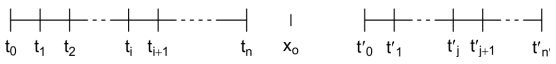


Figure 1: Observation window W defined as the union of two disjoint intervals

We define the weight function as a piecewise linear curve on W with real parameters a_i, b_i, a'_i, b'_i:

$$w(x) = \sum_{i=0}^{n} \left(a_i + b_i(x-t_i)\right)\mathbb{I}_{[t_i,t_{i+1}]}(x) + \sum_{j=0}^{n'} \left(a'_j + b'_j(x-t'_j)\right)\mathbb{I}_{[t'_j,t'_{j+1}]}(x).$$

The continuity property of the spline curve and the constraint $\int_W w(x)\,\mathrm{d}x = 1$ lead to

$$\begin{aligned} w(x) = a_0 + b_0(x-t_0)\mathbb{I}_{[t_0,t_1]}(x) + \sum_{i=1}^{n-1}\left(\Delta\sum_{j=0}^{i-1} b_j + b_i(x-t_i)\right)\mathbb{I}_{[t_i,t_{i+1}]}(x) \\ + a'_0 + b'_0(x-t'_0)\mathbb{I}_{[t'_0,t'_1]}(x) + \sum_{j=1}^{n'-1}\left(\Delta\sum_{k=0}^{j-1} b'_k + b'_j(x-t'_j)\right)\mathbb{I}_{[t'_j,t'_{j+1}]}(x). \end{aligned}$$

The Fredholm equation (2) becomes

$$\begin{aligned} \mathcal{F}(x) = a_0\left[1+\lambda A(x) - \frac{\lambda}{\nu(W)}\int_W A(x)\,\mathrm{d}x\right] + a'_0\left[1+\lambda A'(x) - \frac{\lambda}{\nu(W)}\int_W A'(x)\,\mathrm{d}x\right] \\ + b_0\left[(x-t_0)\mathbb{I}_{[t_0,t_1]}(x) + \Delta\mathbb{I}_{[t_1,t_n]}(x) + \lambda C_0(x) - \frac{\lambda}{\nu(W)}\int_W C_0(x)\,\mathrm{d}x\right] \\ + \sum_{i=1}^{n-2} b_i\left[(x-t_i)\mathbb{I}_{[t_i,t_{i+1}]}(x) + \Delta\mathbb{I}_{[t_{i+1},t_n]}(x) + \lambda C_i(x) - \frac{\lambda}{\nu(W)}\int_W C_i(x)\,\mathrm{d}x\right] \\ + b_{n-1}\left[(x-t_{n-1})\mathbb{I}_{[t_{n-1},t_n]}(x) + \lambda B(x) - \frac{\lambda}{\nu(W)}\int_W B(x)\,\mathrm{d}x\right] \\ + b'_0\left[(x-t'_0)\mathbb{I}_{[t'_0,t'_1]}(x) + \Delta\mathbb{I}_{[t'_1,t'_{n'}]}(x) + \lambda C'_0(x) - \frac{\lambda}{\nu(W)}\int_W C'_0(x)\,\mathrm{d}x\right] \\ + \sum_{j=1}^{n'-2} b'_j\left[(x-t'_j)\mathbb{I}_{[t'_j,t'_{j+1}]}(x) + \Delta\mathbb{I}_{[t'_{j+1},t_{n'}]}(x) + \lambda C'_j(x) - \frac{\lambda}{\nu(W)}\int_W C'_j(x)\,\mathrm{d}x\right] \\ + b'_{n'-1}\left[(x-t'_{n'-1})\mathbb{I}_{[t'_{n'-1},t_{n'}]}(x) + \lambda B'(x) - \frac{\lambda}{\nu(W)}\int_W B'(x)\,\mathrm{d}x\right] \\ - \lambda\left(g(x-x_o)-1\right) + \frac{\lambda}{\nu(W)}\int_W \left(g(x-x_o)-1\right)\mathrm{d}x = 0, \quad (8) \end{aligned}$$

where $A(x) = \sum_{i=0}^{n-1}\int_0^\Delta \left(g(x-t_i-u)-1\right)\mathrm{d}u$, $A'(x) = \sum_{j=0}^{n'-1}\int_0^\Delta \Big(g(x-t'_j-u)-1\Big)\,\mathrm{d}u$,

$B(x) = \int_0^\Delta u\left(g(x-t_{n-1}-u)-1\right)\mathrm{d}u$, $B'(x) = \int_0^\Delta u\left(g(x-t'_{n'-1}-u)-1\right)\mathrm{d}u$,

$C_i(x) = \Delta\sum_{j=i+1}^{n-1}\int_0^\Delta \left(g(x-t_j-u)-1\right)\mathrm{d}u + \int_0^\Delta u\left(g(x-t_i-u)-1\right)\mathrm{d}u$ and

$C'_j(x) = \Delta\sum_{k=j+1}^{n'-1}\int_0^\Delta \Big(g(x-t'_j-u)-1\Big)\,\mathrm{d}u + \int_0^\Delta u\Big(g(x-t'_j-u)-1\Big)\,\mathrm{d}u$.

We thus obtain a linear system w.r.t. a_0, a'_0, b_i and b'_j when considering a set of test points x inside W, whose resolution leads to an approximated solution of the Fredholm equation.

3 Illustrative examples

In this section, we provide illustrative examples of our method both on simulations and real data. A focus is put on the continuous solutions and we refer the reader to [10] for further simulation-based results in the case of weights defined by step functions. We also illustrate the method to map the local intensity of Black Locust in a region of France, whose pattern shows aggregation and inhibition at several spatial scales.

3.1 Simulated examples

3.1.1 Finite Element Approach

The method described in Section 2.2 was implemented using FreeFem++ [12]. The same mesh was used in all the simulations and composed of approximately 30000 triangles. We also choose a P1 finite element space as our interpolation space for the weights.

To illustrate the method on simulated data we considered two cluster point processes. The first is a Thomas process with parameters $\kappa = 10$ the intensity of the Poisson parent process, $\mu = 50$ the mean number of offsprings per parents and $\sigma = 0.05$ the standard deviation of Gaussian displacement of the offsprings around the parent. The top left panel of Figure 2 shows one realization of this process in the unit square. The second, which we refer to as mixing process, has a hard-core process with interaction radius $R = 0.5$ for the parent process, on average $\mu = 10$ offsprings per parent and the offsprings are normally distributed around the parent with standard deviation $\sigma = 0.1$. Thus, this process shows clustering at some scale and regularity at some other scale. One realization in $W = [0,10] \times [0,10]$ is illustrated on the top right panel of Figure 2.

Our predictor depends on the global intensity and the pair correlation function of the point process. For the Thomas process we considered the theoretical moments: $\lambda = \kappa\mu = 500$ and $g(r) = 1 + \frac{1}{4\pi\kappa\sigma^2}\exp\left(-\frac{r^2}{4\sigma^2}\right)$. For the mixing process, we estimated the first- and second-order moments as follows: $\hat{\lambda} = \frac{\Phi(W)}{\nu(W)} = 12.58$ and $g(r) = 1 + \alpha\left(\frac{\gamma}{r}\exp\left(-\left(\frac{r}{\gamma}\right)^{\beta}\right)\right)\sin\left(\frac{r}{\gamma}\right)$, with $\hat{\alpha} = 11.65$, $\hat{\beta} = 0.35$ and $\hat{\gamma} = 0.093$. The related pair correlation functions are plotted on the bottom panels of Figure 2. Note that [10] showed the little effect of the pair correlation function estimation in the prediction, as confirmed in the next section.

For the prediction purpose, we removed three horizontal bands (grey shades on the top panels of Figure 2) and hence the points of the process in it (red dots). Middle panels of Figure 2 depict the overlay of the prediction of the local intensity in the prediction window and a kernel smoothing of the data in the observation window (as it is the best estimation when we have all points). We could take the average of m kernel-smoothed empirical intensity functions to compare with our prediction. But it is not straight and requires conditional simulations as discussed in Section 4. We considered Gaussian kernels, with bandwidth selected by cross-validation [2].

These examples illustrate that our method is able to reproduce the second-order structure of the point process. In particular, it reproduces clusters as soon as there are points close enough to the boundary of the prediction window. However as $\|x_o - \Phi_W\|$ is larger than the interaction range, $\widehat{\lambda}(x_o|\Phi_W)$ tends to

the global intensity λ. These results are even more clear in the one-dimensional example.

Figure 3 illustrates the approximated weight function $w(x;x_o)$ associated to the point $x_o = (0.18, 0.57)$ (cyan dot on the top left panel of Figure 2) when predicting the local intensity of the Thomas process. It shows a radial function, with high positive values for locations close to x_o, negative values at intermediate range, and 0 values when $\|x_o - x\|$ is larger than the interaction range.

3.1.2 Linear splines in 1D

We simulate a Thomas process on the interval $[0,20]$, with a Poisson process of parents of intensity $\kappa = 5$, an average of $\mu = 10$ offsprings per parent, normally distributed around the parent point with standard deviation $\sigma = 1$. The intensity and the pair correlation function of the Thomas process are given by:

$$\lambda = \kappa\mu \quad \text{and} \quad g(r) = 1 + \frac{1}{2\sqrt{\pi\kappa}\sigma} \exp\left(-\frac{r^2}{4\sigma^2}\right).$$

To avoid border effects, we restrict the observation of the process to $W = [3, \delta_1] \cup [\delta_2, 17]$ and we predict the local intensity in $[\delta_1, \delta_2] \in \{[9,11], [8,12], [7,13], [6,14]\}$. In the spline approximation, knots are equally spaced, with distance $\Delta \in \{0.5, 1, 2\}$. Parameters a_0, a'_0, b_i, b'_j, $i = 1, \ldots n-1$, $j = 1, \ldots, n'-1$, of Equation (8) minimize $\sum_{x \in X} \mathcal{F}^2(x)$, where X is a set of test points regularly located in W. Figure 4 (left panel) shows one realization of the Thomas process in S (magenta vertical bars) and its restriction to W (black vertical bars).

At first, the method is applied assuming the pair correlation function to be known. The local intensity prediction in $[\delta_1, \delta_2] = [9, 11]$ obtained with $\Delta = 0.5$ (red curve), $\Delta = 1$ (orange curve) and $\Delta = 2$ (yellow curve) is superimposed on the empirical local intensity obtained by kernel smoothing on S (grey curve) and W (black curve) using the reflection method [18] to correct edge effects on the neighborhood of δ_1 and δ_2. Figure 4 illustrates the behavior of the prediction w.r.t. the spacing distance Δ between knots. Here, the prediction is good whatever Δ. We note however a better prediction when $\Delta = 0.5$, in the sense that it is closer to the empirical local intensity. This is particularly true at δ_2, but not at δ_1. Indeed, in the latter case the empirical local intensity is underestimated as there are only few points close to δ_1. More generally, the smaller Δ, the better the prediction as the weight function $w(x)$ is then better approximated. The right panel of Figure 4 shows the piecewise approximation of the weight function obtained for $x_o = 10$ and with $\Delta = 0.5$ (red curve), $\Delta = 1$ (orange curve) and $\Delta = 2$ (yellow curve). Note that in this example, the observation window is symmetric around $x_o = 10$. Thus, the weight function for $x_o = 10$ should be symmetric around this point, but this requires both a sufficient number of knots (in Figure 4 the weight function tends to be symmetric as Δ decreases) and a sufficient number of test points used to estimate parameters a_0, b_i, a'_0 and b'_j. Here, only forty test points have been used. Further simulations with $\Delta = 0.25$ and $\Delta = 0.1$ and eighty test points confirm these results but are not showed her for conciseness.

Figure 5 illustrates the prediction of the local intensity when the size of the prediction window $[\delta_1, \delta_2]$ varies: the smaller the prediction window, the more the point locations inside W influence the predictor and better the prediction w.r.t. the empirical local intensity of the point process (grey curve). In fact, the quality of the prediction in $[\delta_1, \delta_2]$ is related to the range of the pair correlation function. We get good predictions when $\delta_2 - \delta_1$ is less than 4σ and the prediction of the local intensity tends to the intensity of the point process when the distance between x_o and the observation window is much larger than the

interaction range. It corresponds to the fact that point locations in W do not bring any information on the intensity. This is particularly clear in the fourth case where $[\delta_1, \delta_2] = [6, 14]$ (Figure 5 right panel).

So far the pair correlation function was assumed to be know, but in practice it must be estimated. Figure 6 (left panel) compares the prediction of the local intensity when using the theoretical pair correlation function (red curve) or an estimate (orange curve) fitted from a realization of the process on $[0, 300]$. Figure 6 (right panel) illustrates the theoretical pair correlation function (black curve) and its estimate modelled as $g(r) = 1 + \beta \exp(-\alpha r^\gamma)$. This model has been empirically chosen and fitted by least squares in order to mimic practical situations. Here, we get $\hat{\alpha} = 0.213$, $\hat{\beta} = 0.048$ and $\hat{\gamma} = 2.33$ (red curve). As the estimated pair correlation function is close to the theoretical one, the related weight function, and thus the prediction, are rather similar.

We also computed an empirical confidence band defined as twice the standard deviation (green curve on Figure 6). From Equation (1), the standard deviation is proportional to the global intensity of the point process. Thus, in this example because the ratio between the standard deviation and the local intensity is close to one, the confidence band seems to be large. Its relative width decreases as the local intensity increases.

3.2 An example with real data

To illustrate how the method can be used in practice, we applied it to the spatial distribution of 582 Black locust (*Robinia p.pseudoacacia L.*) of a 25 year old coppice of 8*ha* located in the Val-de-Loire region (France) [15]. Coppicing consists in cutting the tree trunk off close to the ground in order to stimulate the growth of new shoots on the stump. The realization of shoots and the prediction windows (grey shades) are plotted on the top left panel of Figure 7. We remove the points in these areas and use the remaining Black Locust locations to predict the local intensity.

The pair correlation function globally shows an aggregative pattern (bottom left panel of Figure 7) attached to a given stump at the origin of the coppice. Looking more closely (bottom right panel of Figure 7) we can notice a repulsive pattern at small scale ($\approx 10\ m$) reflecting the woodland management. To describe this complex structure, we modelled the pair correlation function as

$$g(r) = \begin{cases} 1 + \alpha \left(\frac{\gamma}{r}\right)^\beta \sin\left(\frac{r}{\gamma}\right), & \text{if } r \geq 0.01\ m \\ 0, & \text{otherwise.} \end{cases}$$

In this model, we imposed a hard-core radius of $0.01\ m$ corresponding to the fact that tree trunk bases cannot overlay. We get $\hat{\alpha} = 3.5$, $\hat{\beta} = 1.89$ and $\hat{\gamma} = 2.34$.

We consider a finite element approach, with a P1 finite element basis constructed on a mesh composed of about 45000 triangles, to approximate the weight function and predict the local intensity. The result (top right panel of Figure 7) shows that the predictor is able to reproduce the structures revealed by the pair correlation function: high aggregation at small scale ($< 5\ m$), inhibition at $10\ m$, and aggregation at $20\ m$. Clearly, the prediction continuously match the clusters split by the boundary of the prediction window, i.e. $\widehat{\lambda}(x_o|\Phi_W) >> \lambda$ as $\min_{x \in \Phi_W} \|x_o - x\| < 5\ m$. Then, values of the predicted local intensity tend to be smaller (close to 0) at distances close to the inhibition range. Finally, $\lambda(x_o|\Phi_W)$ tends to λ as $\min_{x \in \Phi_W} \|x_o - x\| > 20\ m$. This behavior is straightly related to the weight function. Figure 8 illustrates $w(x; x_o)$ at $x_o = (106, 251)$ and $x_o = (106, 43.75)$ (blue and green dots respectively, on the left panel of Figure 8). It clearly shows how much the weight function, and thus the prediction, depends on the pair

correlation function.

4 Discussion

Our approach aims to predict the local intensity of a point process outside the observation window, i.e. its intensity conditional to the realization of the point process. It uses the concept of geostatistics of defining an unbiased linear predictor of the local intensity, which minimizes the error prediction variance. However, there is a fundamental difference between the two approaches: in geostatistics data locations are issued out of samples and do not contain information *a priori*; it is the variable of interest evaluated at these locations which is informative. Conversely, in our context of point processes, point locations are informative. Consequently, we have to take into account all the observed area, meaning both points and empty space. Then, the kriging equations given by geostatistics become an integral equation and we do not have an explicit solution. We consider solution spaces generated by linear combinations of step functions, finite elements and splines. In the latter, the rough fitting of the spline approximation could be improved by refining the number of knots, the number of test points and using splines of higher order. In all cases, we could explore the sensitivity of the method w.r.t the parameters involved in the approximations, but also its performances.

Several approximations have been proposed in this paper to predict the local intensity and their respective performances need to be compared. To achieve it, we need to know the true local intensity. This knowledge becomes then the crucial issue since it is not known for many processes. Here, we used a kernel smoothing of the realization to approximate it. It would be better to take the average of m kernel-smoothed empirical intensity functions. This requires conditional simulations and thus raises the issue to know how to perform it. The two main methods of conditional simulations assume either that the point process distribution is known, or that its first- and second-order moments are known. For many point processes, the conditional point process distribution is also not known so that the method proposed by [19] remains the only available one. However, relying only on first- and second-order moments, this conditional simulation procedure is only an approximated one. As an example let us consider a Neyman-Scott model where offsprings are spread on a segment centered on the parent point, segment orientations being randomly uniformly distributed. This leads to an isotropic point process with visually observable random alignments which cannot be described with the first two moments. Thus, there remains to clarify when such procedures based on the two first moments are valid.

Our method can be used to improve or to optimize sampling schemes. Indeed, as in classical geostatistics, this can be achieved from the variance of the predictor. This could particularly be useful when mapping the intensity of cryptic species in large areas in an adaptative scheme.

Our approach depends heavily on the pair correlation function and therefore on its estimation at any distances up to the interaction range. This leads to a trade-off between having adapted samples catching structures at different scales and spending a minimal time for data collection. The sample spatial unit must be large enough to get an accurate estimation of the pair correlation function at the origin, but must also allow its estimation at distances covered by the prediction and observation windows. A solution could be to consider a combination of thin transects to get estimates at moderate to large distances and large quadrats to get estimates at small distances.

This paper offers a conceptual toolbox which could be extended to more general cases covered in geostatistics, provided that features of point processes are taken into account. In particular we are show-

ing that the extension to non-stationary point processes is rather straightforward when assuming the point process to be second-order intensity reweighted stationary, with an intensity driven by non-stationary covariates. However for spatio-temporal point processes, predicting beyond the last observation date raises some issues about how to integrate the structure of the spatio-temporal pair correlation function. If the dependence structure varies with time, one could stationarize the process (as for inhomogeneous point processes, see e.g. [11]). But this issue is still open for processes with a more complex dependence structure.

References

[1] E Bellier, P Monestiez, G Certain, J Chadœuf, and V Bretagnolle. Reducing the uncertainty of wildlife population abundance: model-based versus design-based estimates. *Environmetrics*, 24(7):476–488, 2013.

[2] M Berman and PJ Diggle. Estimating weighted integrals of the second-order intensity of a spatial point process. *J Roy Stat Soc B*, 51:81–92, 1989.

[3] O Bonnefon, J Coville, and G Legendre. Concentration phenomenon in somme nonlocal equation. *Discrete and Continuous Dynamical Systems - Series B*, 22(3):763–781, 2017.

[4] C-C. Cheng. Investigations into green's function as inversion-free solution of the kriging equation, with geodetic applications. Technical Report Report No. 472, The Ohio State University, 2004.

[5] J-P Chilès and P Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, New York, second edition, 2012.

[6] S Chiu, D Stoyan, W Kendall, and J Mecke. *Stochastic Geometry and Its Applications*. John Wiley & Sons, New York, third edition, 2013.

[7] J Coville. Contribution à l'étude d'équations non locales en dynmaique de spopulations. Technical report, HDR de l'Université Aix-MArseille, CNRS, I2M UMR 7373, 2015.

[8] PJ Diggle, P Moraga, B Rowlingson, and B Taylor. Spatial and spatio-temporal log-gaussian cox processes: Extending the geostatistical paradigm. *Statistical Science*, 28(4):542–563, 2013.

[9] PJ Diggle and P Ribeiro. *Model-Based Geostatistics*. Springer, New York, 2007.

[10] E Gabriel, F. Bonneu, P. Monestiez, and J. Chadœuf. Adapted kriging to predict the intensity of partially observed point process data. *Spatial Statistics*, 18:54–71, 2016.

[11] U Hahn, E Vedel-Jensen, M-C Van Lieshout, and L Stougaard Nielsen. Inhomogeneous spatial point processes by location-dependent scaling. *Advances in Applied Probability*, 35:319–336, 2003.

[12] F Hecht. New development in freefem++. *Journal of Numerical Mathematics*, 20:251–265, 2012.

[13] R Kress. *Integral Equations*. Applied Mathematical Sciences Series, Springer, New York, 2013.

[14] P Monestiez, L Dubroca, E Bonnin, JP Durbec, and C Guinet. Geostatistical modelling of spatial distribution of balaenoptera physalus in the northwestern mediterranean sea from sparse count data and heterogeneous observation efforts. *Ecological Modelling*, 193:615–628, 2006.

[15] L Pages. Lois de croissance en biomasse du taillis : le robinier dans le val-de-loire. *Annales des sciences forestières*, 43(4):533–550.

[16] H Saito, S McKenna, D Zimmerman, and T Coburn. Geostatistical interpolation of object counts collected from multiple strip transects: Ordinary kriging versus finite domain kriging. *Stochastic Environmental Research and Risk Assessment*, 19:71–85, 2005.

[17] P. D. Sampson, D. Damian, and P. Guttorp. *geoENV III — Geostatistics for Environmental Applications: Proceedings of the Third European Conference on Geostatistics for Environmental Applications held in Avignon, France, November 22–24, 2000*, chapter Advances in Modeling and Inference for Environmental Processes with Nonstationary Spatial Covariance, pages 17–32. Springer Netherlands, Dordrecht, 2001.

[18] BW Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, London, 1986.

[19] A Tscheschel and D Stoyan. Statistical reconstruction of random point patterns. *Computational Statistics and Data Analysis*, 51:859–871, 2006.

[20] M-C. van Lieshout and A. Baddeley. Extrapolating and interpolating spatial patterns. In *In Spatial Cluster Modelling, A.B. Lawson and D.G.T. Denison (EDS.) Boca Raton: Chapman And Hall/CRC*, pages 61–86. Press, 2001.

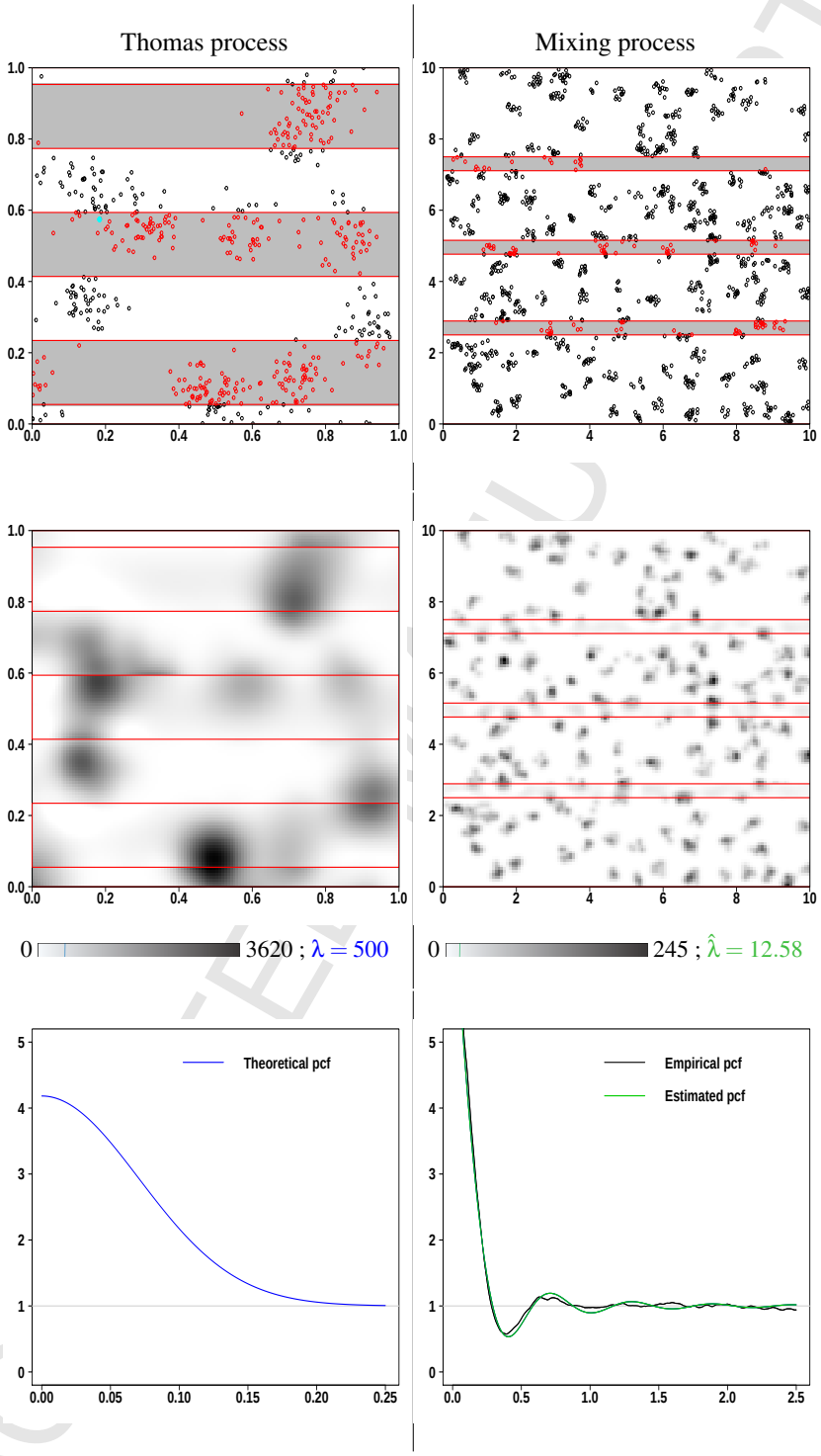


Figure 2: *Top* - Point process realization: the black dots correspond to $\Phi \cap W$, the red dots are removed for prediction. The prediction window is defined by the grey shade. *Middle* - prediction of the local intensity in W and kernel smoothing in the prediction window. *Bottom* - Pair correlation function.

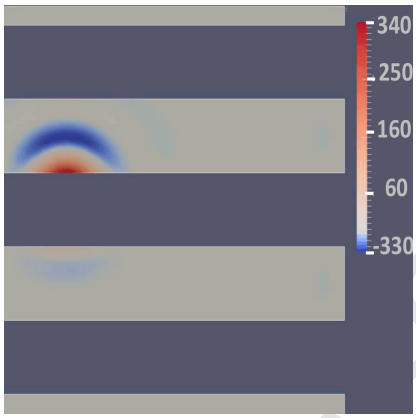


Figure 3: Weight function associated to $x_o = (0.18, 0.57)$

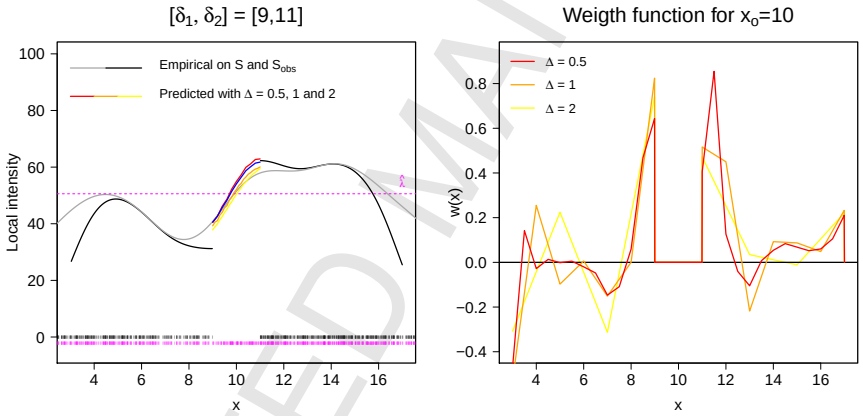


Figure 4: *Left* - Realization of the Thomas process in S (grey dots) and W (black dots) and prediction of the local intensity (red curve) when using different spacing distances in the spline approximation. *Right* - Weight function obtained for $x_o = 10$.

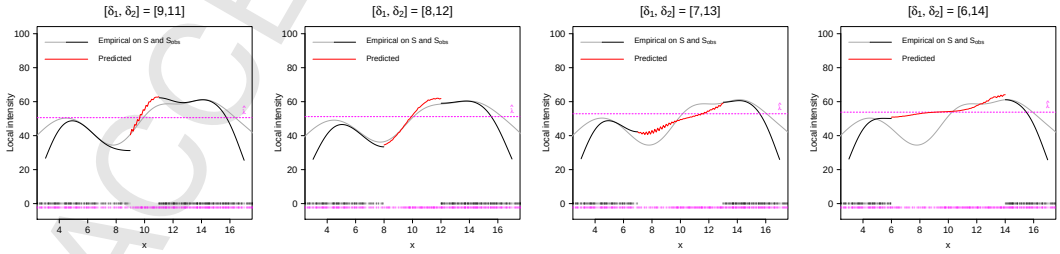


Figure 5: Prediction of the local intensity in increasing intervals $[\delta_1, \delta_2]$.

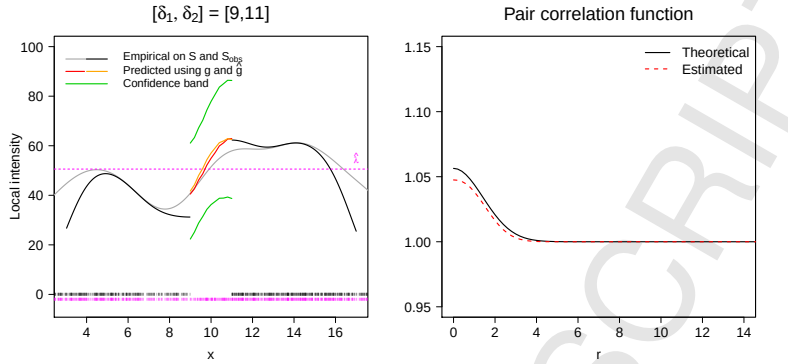


Figure 6: *Left* - Prediction of the local intensity and related confidence band. *Right* - Theoretical and estimated pair correlation function.

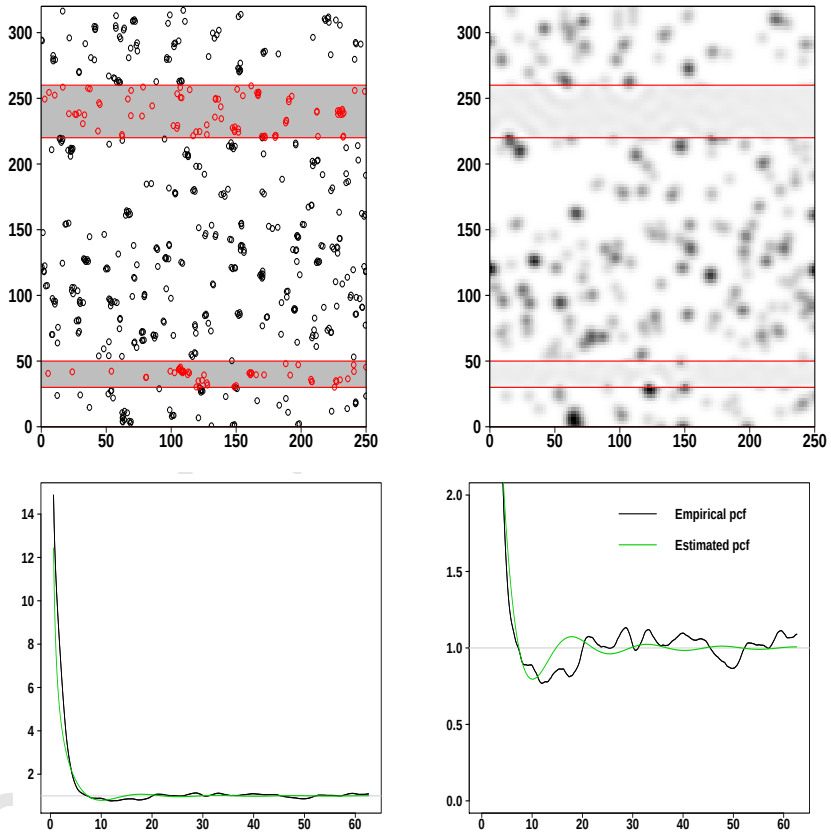


Figure 7: *Top left* - Point process realization: the black dots correspond to $\Phi \cap W$, the red dots are removed for prediction. The prediction window is defined by the grey shade. *Top right* - Prediction of the local intensity in W and kernel smoothing in the prediction window. *Bottom* - Empirical and estimated pair correlation function.

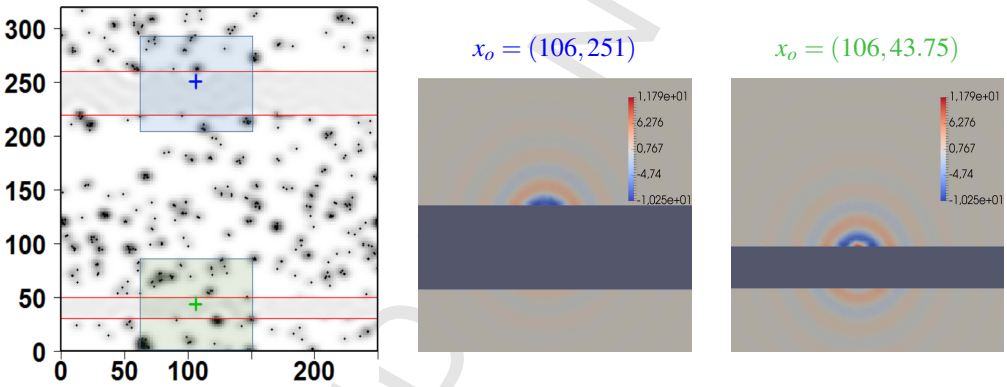


Figure 8: *Left* - Prediction of the local intensity in W and kernel smoothing in the prediction window. Blue and green squares delineate the zones where the weight function is plotted for $x_o =+$ (*Middle*) and $x_o =+$ (*Right*).