Geostatistics for point processes

Predicting the intensity of partially observed point process data

Edith Gabriel[1,2] & Joël Chadœuf[2]

[1]Laboratory of Mathematics, Avignon University
[2]French National Institute for Agricultural Research

Stochastic Geometry and its Applications

Nantes, 7th April 2016

Motivations

Predicting the local intensity

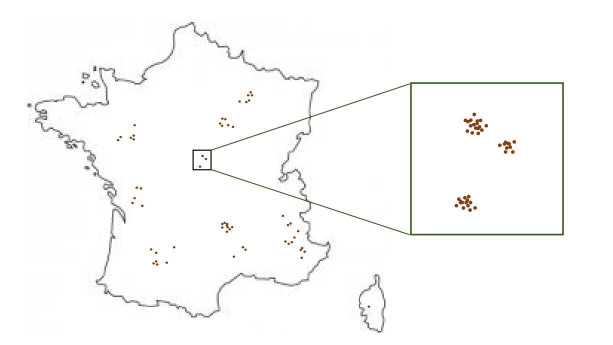
- Defining the predictor, similarly to a kriging interpolator
- Solving a Fredholm equation to find the weights
  - $\Rightarrow$ approximated solutions
- Illustrative results

Discussion

A well-known issue

The issue

How to extensively map the intensity of a point process in a large window when observation methods are available at a much smaller scale only?



Motivating examples

- Estimating spatial repartition of a bird species at a national scale from observations made in windows of few hectares.
- Detecting plant disease at the field scale from observations defined as spots of few square millimetres on leaves.
- Mapping the presence of plant species at the catchment scale when the observation scale is the square metre.

$\Rightarrow$ The intensity of the process must be predicted from data issued out of samples spread over the window of interest.

Context

Let Φ a point process assumed to be

- stationary and isotropic,

$$\lambda = \frac{\mathbb{E}\left[\Phi(S_{obs})\right]}{\nu(S_{obs})} \; ; \; g(r) = \frac{1}{2\pi r}\frac{\partial K^*(r)}{\partial r}$$

  with $K^*(r) = \frac{1}{\lambda}\mathbb{E}\left[\Phi(b(0,r)) - 1 | 0 \in \Phi\right]$.

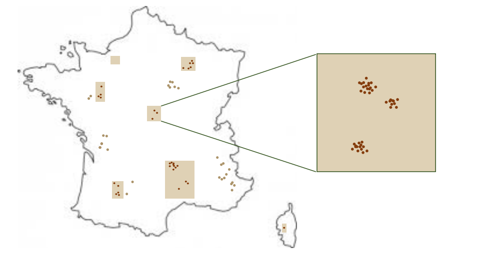
- observed in S_{obs},

- driven by a stationary random field, Z.

Our aim

Local intensity

We call *local* intensity of the point process Φ, its intensity given the random field, Z: $\lambda(x|Z)$.



Window of interest:

$$S = S_{obs} \cup S_{unobs} = (\cup \blacksquare) \cup (\cup \square)$$

$$\Phi = \{\circ, \bullet\}; \; \Phi_{S_{obs}} = \{\bullet\}$$

Our aim

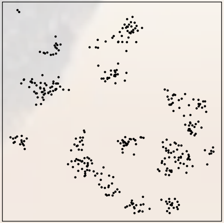
To predict the local intensity in an unobserved window S_{unobs}.

Example

Thomas process:

- κ: intensity of the Poisson process parents, Z,
- μ: mean number of offsprings per parent,
- σ: standard deviation of Gaussian displacement.

$\kappa = 15$, $\mu = 15$, $\sigma = 0.025$



This process is stationary with intensity $\lambda = \kappa\mu$.

The local intensity corresponds to the intensity of the inhomogeneous Poisson process of offsprings,
i.e. the intensity conditional to the parent process Z.

$\star$ More generally, we consider any process driven by a stationary random field $\star$

Existing solutions

- From the reconstruction of the process
  - Reconstruction method based on the 1^{st} and 2^{d}-order characteristics of Φ (see e.g. Tscheschel & Stoyan, 2006).
  - Get the intensity by kernel smoothing.

  A simulation-based method $\Rightarrow$ long computation times.

- Intensity driven by a stationary random field
  - Diggle *et al.* (2007, 2013): Bayesian framework
  - Monestiez *et al.* (2006, 2013): Close to classical geostatistics.

  Models constrained within the class of Cox processes.

Our alternative approach

We want to predict the local intensity $\lambda(x|Z)$

- outside the observation window,
- without precisely knowing the underlying point process
  $\Rightarrow$ we only consider the 1^{st} and 2^{d}-order characteristics,
- in a reasonable time.

We define an unbiased linear predictor

- which minimizes the error prediction variance (as in the geostatistical concept).
- whose weights depend on the structure of the point process.

Our predictor

Proposition

The predictor $\widehat{\lambda}(x_o|Z) = \sum_{x\in\Phi\cap S_{obs}} w(x)$ is the BLUP of $\lambda(x_o|Z)$.

The weights, $w(x)$, are solution of the Fredholm equation of the 2^d kind:

$$w(x)+\lambda\int_{S_{obs}} w(y)\left(g(x-y)-1\right)\,\mathrm{d}y-\frac{1}{\nu(S_{obs})}\left[1+\int_{S_{obs}^2} w(y)\left(g(x-y)-1\right)\,\mathrm{d}x\,\mathrm{d}y\right]$$
$$=\lambda\left(g(x_o-x)-1\right)-\frac{\lambda}{\nu(S_{obs})}\int_{S_{obs}}\left(g(x_o-x)-1\right)\,\mathrm{d}x$$

and satisfy $\int_{S_{obs}} w(x)\,\mathrm{d}x = 1$.

Elements of proof

Linearity:

- By definition, $\lambda(x_o|Z) = \lim_{\nu(B)\to 0} \dfrac{\mathbb{E}\left[\Phi(B \oplus x_o)|Z\right]}{\nu(B)}$.

- Furthermore, $\widehat{\mathbb{E}}\left[\Phi(B \oplus x_o)|Z\right] = \sum_{c_j \in \mathcal{G}(S_{obs})} \alpha(c_j; B, x_o)\Phi(B \oplus c_j)$ is the BLUP of $\Phi(B \oplus x_o)$[1], where $\mathcal{G}(S_{obs})$ is a grid superimposed on S_{obs}.

- Thus, we propose

$$\widehat{\lambda}(x_o|Z) = \lim_{\nu(B)\to 0} \sum_{c_j \in \mathcal{G}(S_{obs})} \frac{\alpha(c_j; B, x_o)}{\nu(B)} \Phi(B \oplus c_j) = \sum_{x \in \Phi \cap S_{obs}} w(x).$$

[1] Gabriel *et al.* (2016) Adapted kriging to predict the intensity of partially observed point process data.

Elements of proof

Unbiasedness:

$$\mathbb{E}\left[\widehat{\lambda}(x_o|Z) - \lambda(x_o|Z)\right] = 0$$

$$\iff \int_{S_{obs}} \lambda w(x)\,\mathrm{d}x - \mathbb{E}\left[\lim_{\nu(B)\to 0} \frac{\mathbb{E}\left[\Phi(B \oplus x_o)|Z\right]}{\nu(B)}\right] = 0$$

$$\iff \lambda\left(\int_{S_{obs}} w(x)\,\mathrm{d}x - 1\right) = 0$$

$$\iff \int_{S_{obs}} w(x)\,\mathrm{d}x = 1.$$

Elements of proof

Minimum error prediction variance:

For any Borel set B,

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

and for $B_o = B \oplus x_o$ with $x_o \notin S_{obs}$,

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

Then minimizing $\mathbb{V}\mathrm{ar}\left(\widehat{\lambda}(x_o|Z) - \lambda(x_o|Z)\right)$ is equivalent to minimize

$$\lambda \int_{S_{obs}} w^2(x)\,\mathrm{d}x + \lambda^2 \int_{S_{obs}\times S_{obs}} w(x)w(y)\,(g(x-y)-1)\,\mathrm{d}x\,\mathrm{d}y - 2\lambda^2 \int_{S_{obs}} w(x)\,(g(x_o-x)-1)\,\mathrm{d}x$$

Elements of proof

Using Lagrange multipliers under the constraint on the weights, we set

$$T(w(x)) = \lambda \int_{S_{obs}} w^2(x)\,\mathrm{d}x + \lambda^2 \int_{S_{obs}\times S_{obs}} w(x)w(y)\,(g(x-y)-1)\,\mathrm{d}x\,\mathrm{d}y$$
$$-2\lambda^2 \int_{S_{obs}} w(x)\,(g(x_o - x)-1)\,\mathrm{d}x + \mu\left(\int_{S_{obs}} w(x)\,\mathrm{d}x = 1\right)$$

Then, for $\alpha(x) = w(x) + \varepsilon(x)$,

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

Elements of proof

From variational calculation and the Riesz representation theorem,

$$\begin{aligned}
T(\alpha(x)) - T(w(x)) = 0 \quad &\Leftrightarrow \quad \int_{S_{obs}} \varepsilon(x) \left[w(x)x + \lambda \int_{S_{obs}} w(y)\,(g(x-y)-1)\,\mathrm{d}y \right. \\
&\qquad \left. - \lambda\,(g(x_o - x) - 1) + \frac{\mu}{2\lambda} \right] \mathrm{d}x = 0 \\
&\Leftrightarrow \quad w(x) + \lambda \int_{S_{obs}} w(y)\,(g(x-y)-1)\,\mathrm{d}y \\
&\qquad -\lambda\,(g(x_o - x) - 1) + \frac{\mu}{2\lambda} = 0
\end{aligned}$$

Thus,

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

from which we obtain μ and we can deduce the Fredholm equation

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

Solving the Fredholm equation

Any existing solution already considered in the literature can be used!

Our aim is to map the local intensity in a given window
$\Rightarrow$ access to fast solutions.

Several approximations can be used to solve the Fredholm equation.

The weights $w(x)$ can be defined as

- step functions $\rightsquigarrow$ direct solution,
- linear combination of known basis functions, e.g. splines
  $\rightsquigarrow$ continuous approximation.
- ...

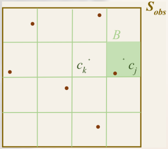
Here, we illustrate the ones with the less heavy calculations and implementation.

Step functions

Let consider the following partition of S_{obs}: $S_{obs} = \cup_{j=1}^{n} B_j$, with

B: elementary square centered at 0,
$B_j = B \oplus c_j$: elementary square centered at c_j,
$B_k \cap B_j = \emptyset$,

n: number of grid cell centers lying in S_{obs}.



For $w(x) = \sum_{j=1}^{n} w_j \frac{\mathbf{1}_{\{x \in B_j\}}}{\nu(B)}$, we get $\widehat{\lambda}(x_o|Z) = \sum_{j=1}^{n} w_j \frac{\Phi(B_j)}{\nu(B)}$,

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

- $C = \lambda \nu(B) \mathbb{I} + \lambda^2 \nu^2(B)(G - 1)$: covariance matrix with $G = \{g_{ij}\}_{i,j=1,\ldots,n}$, $g_{ij} = \frac{1}{\nu^2(B)} \int_{B \times B} g(c_i - c_j + u - v)\, du\, dv$, and $\mathbb{I}$ the $n \times n$-identity matrix.
- $C_o = \lambda \nu(B) \mathbb{I}_{x_o} + \lambda^2 \nu^2(B)(G_o - 1)$: covariance vector with $\mathbb{I}_{x_o}$ the n-vector with zero values and one term equals to one where $x_o = c_i$, and $G_o = \{g_{io}\}_{i=1,\ldots,n}$.

Step functions: variance of the predictor

We consider the Neuman series to invert the covariance matrix, $C = \lambda\nu(B)\mathbb{I} + \lambda^2\nu^2(B)(G-1)$, when $\lambda\nu(B) \to 0$:

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

where a generic element of the matrix J_λ is given by

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

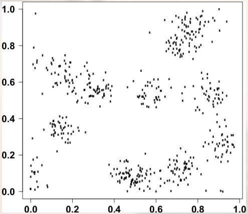
This leads to

$$\mathbb{V}\mathrm{ar}\left(\widehat{\lambda}(x_o|Z)\right) = \lambda^3\nu^2(B)(G_o-1)^T(G_o-1) + \lambda^4\nu^3(B)(G_o-1)^T J_\lambda(G_o-1)$$
$$+\frac{1-\left[\lambda\nu(B)\mathbf{1}^T(G_o-1) + \lambda^2\nu^2(B)\mathbf{1}^T J_\lambda(G_o-1)\right]^2}{\frac{\nu(S_{obs})}{\lambda} + \nu^2(B)\mathbf{1}^T J_\lambda \mathbf{1}}.$$

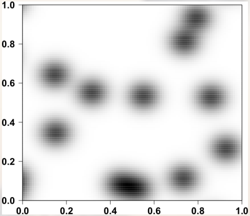
Step functions: illustrative results about prediction

Simulated Thomas process

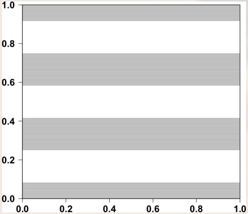
$\kappa = 10, \mu = 50, \sigma = 0.05$



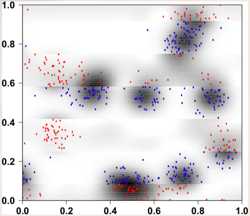
Theoretical local intensity



$$S = S_{obs} \cup S_{unobs} = (\cup\square) \cup (\cup\square)$$



Prediction within S_{unobs}



$\{\bullet\}: \Phi \cap S_{obs}$; $\{\bullet\}: \Phi \cap S_{unobs}$

Spline basis

Let consider that the weights of $\widehat{\lambda}(x_o|Z) = \sum_{x\in\Phi\cap S_{obs}} w(x)$ are defined as a degree d spline curve:

$$w(x) = \sum_{i=1}^{k} h_{i,d}(x),$$

where $h_{i,d}$ denotes the ith B-spline of order d.

A simplistic toy example in $\mathbb{R}$:

- $S_{obs} = [0, L) \subset [0, L'] = S$
- Linear spline defined from equally-spaced knots x_i:

$$\begin{aligned} w(x) &= \begin{cases} a_0 + b_0 x, & x \in \Delta_0 = [x_0, x_1) = [0, \frac{L}{k}), \\ a_1 + b_1 x, & x \in \Delta_1 = [x_1, x_2) = [\frac{L}{k}, \frac{2L}{k}), \\ \vdots & \\ a_{k-1} + b_{k-1} x, & x \in \Delta_{k-1} = [x_{k-1}, x_k) = [\frac{(k-1)L}{k}, L), \end{cases} \\ &= (a_i + b_i(x - x_i))\,\mathbf{1}_{\{x\in\Delta_i\}} \end{aligned}$$

Spline basis

From the continuity property and the constraint $\int_{S_{obs}} w(x)\,dx = 1$:

$$w(x) = \frac{1}{L} - \sum_{j=0}^{k-1} b_j P_j(x),$$

with $P_j(x) = \sum_{i=0}^{k-1} \left(\frac{1/2-k+j}{k^2} - \mathbf{1}_{\{j<i\}} - (x - \frac{iL}{k})\mathbf{1}\{i=j\} \right) \mathbf{1}_{\{x\in\Delta_i\}}$

The Fredholm equation becomes

$$\sum_{j=0}^{k-1} b_j \left[P_j(x) + \lambda \int_L P_j(y)(g(x-y)-1)\,dy - \frac{1}{L}\int_{L^2} P_j(y)(g(x-y)-1)\,dx\,dy \right]$$

$$= \frac{\lambda}{L}\int_L (g(x-y)-1)\,dy - \frac{1}{L^2}\int_{L^2}(g(x-y)-1)\,dx\,dy - \lambda(g(x_o-x)-1) + \frac{1}{L}\int_L (g(x_o-x)-1)\,dx$$

i.e. of the form $\sum_{j=0}^{k-1} b_j A_j(x) = Q(x)$,

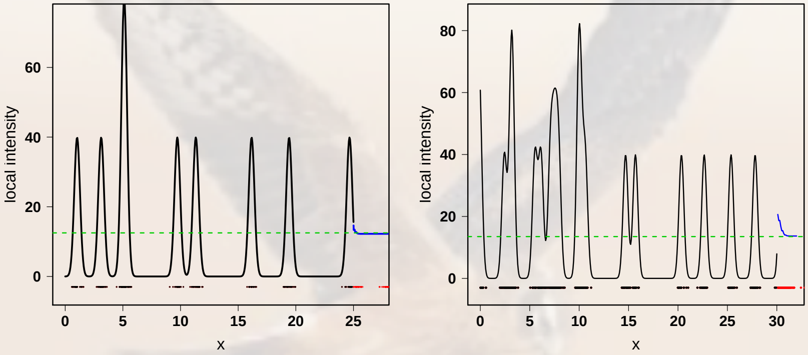
Then, $(b_0,\dots,b_{k-1}) = b$ is obtained from m control points and satisfy

$$b = (X^T X)^{-1} X^T Y,$$

with $X = (A_j(x_l))_{l=1,\dots,m}$ and $Y = (Q(x_l))_{l=1,\dots,m}$.

Spline basis: illustrative results

Thomas process in 1D ($\kappa = 0.5, \mu = 25, \sigma = 0.25$)



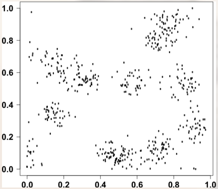
—— Theoretical local intensity on S_{obs} ; —— Predicted values ; —— Intensity of Φ

$\{\bullet\} = \Phi_{S_{obs}}$; $\{\bullet\} = \Phi_{S_{unobs}}$

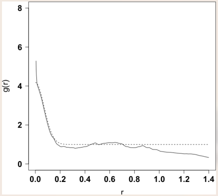
In practice: g must be estimated

Simulated Thomas process

$\kappa = 10, \mu = 50, \sigma = 0.05$

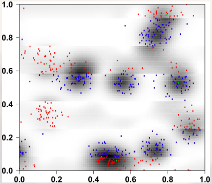


Estimated pcf



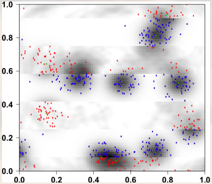
Prediction within S_{unobs}
(with the theoretical pcf)

$R^2 = 0.85$



(with the estimated pcf)

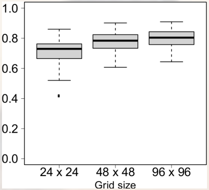
$R^2 = 0.8$



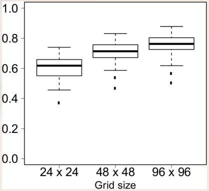
$\{\color{blue}\bullet\}$: $\Phi \cap S_{obs}$; $\{\color{red}\bullet\}$: $\Phi \cap S_{unobs}$

R^2 in linear regression
of predicted and theoretical values

(with the theoretical pcf)



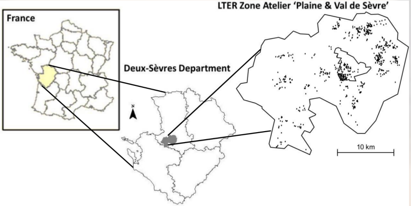
(with the estimated pcf)



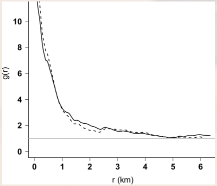
In practice

Application to Montagu's Harriers' nest locations

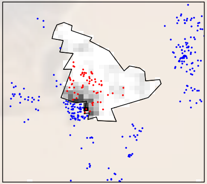
Data collection



Estimated pcf



Prediction



$\{\bullet\} = \Phi_{S_{obs}}$; $\{\bullet\} = \Phi_{S_{unobs}}$

Work in progress

- Take into account some covariates in the prediction.
- Get results with splines on the plane.
- Use finite elements method to solve the Fredholm equation.
- Determine the properties of the related predictor.
- Extend the approach to the spatio-temporal setting.

References

E. Bellier *et al.* (2013) Reducting the uncertainty of wildlife population abundance: model-based versus design-based estimates. *Environmetrics*, 24(7):476–488.

E. Gabriel *et al.* (2016) Adapted kriging to predict the intensity of partially observed point process data. *Spatial statistics*, in revision.

P. Diggle and P. Ribeiro (2007) *Model-based geostatistics*. Springer.

P. Diggle *et al.* (2013) Spatial and spatio-temporal log-gaussian cox processes: extending the geostatistical paradigm. Statistical Science, 28(4):542–563.

P. Monestiez *et al.* (2006) 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.

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