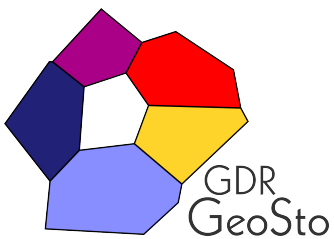
Regularization techniques for spatial point processes intensity estimation

Jean-François Coeurjolly
(joint work with A. Choiruddin, F. Letué)



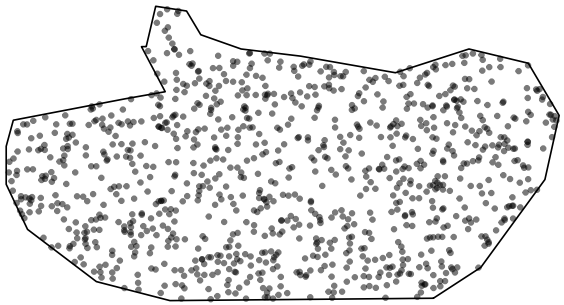
UQÀM | Département de mathématiques
FACULTÉ DES SCIENCES
Université du Québec à Montréal

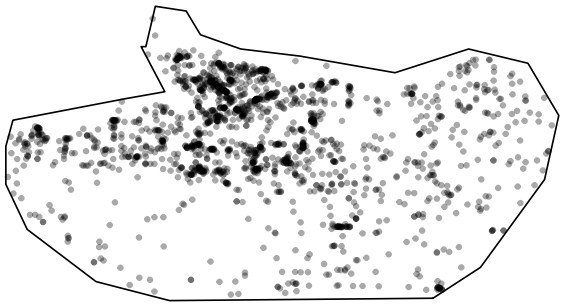
Point processes / Point patterns

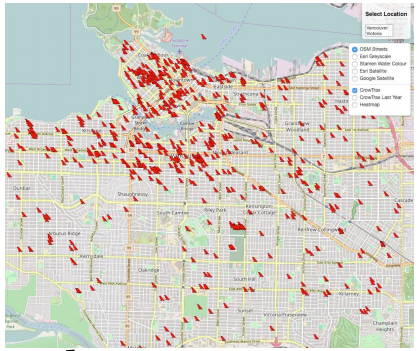
- Spatial point pattern : (locally finite random measure)

  $\mathbf{x} = \{x_1, \ldots, x_n\}$, $x_i \in W \subset \mathbb{R}^d$ (e.g. $d = 2, 3$), n = random.

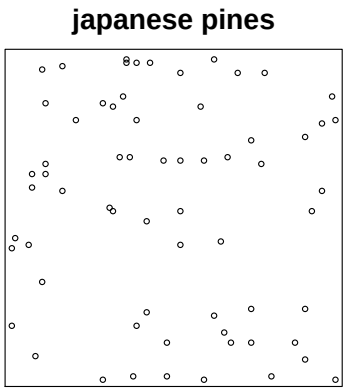
- Patterns [1] can be : homogeneous or inhomogeneous

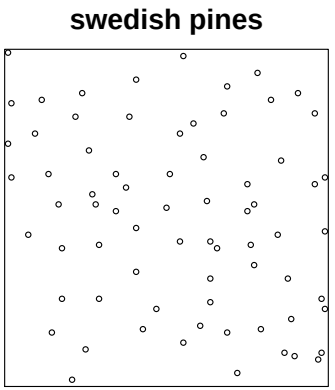


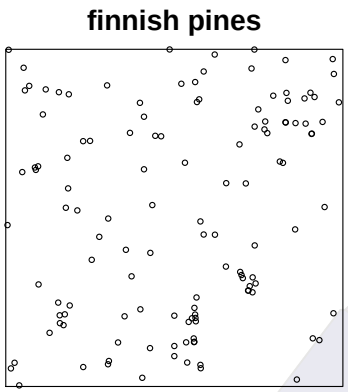




- . . .can exhibit independence, neg. or pos. dependence







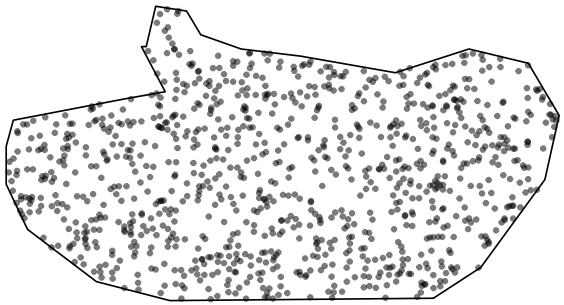
1. courtesy Jim O'Leary, crows attacks in Vancouver in 2017

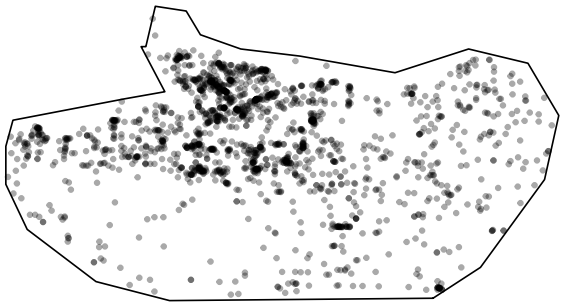
Point processes / Point patterns

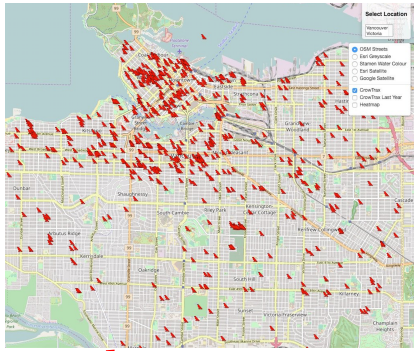
- Spatial point pattern : (locally finite random measure)

  $\mathbf{x} = \{x_1, \ldots, x_n\}$, $x_i \in W \subset \mathbb{R}^d$ (e.g. $d = 2, 3$), n = random.

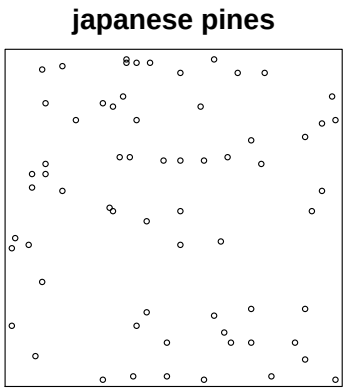
- Patterns [1] can be : homogeneous or inhomogeneous

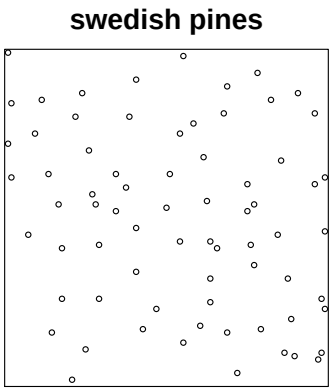


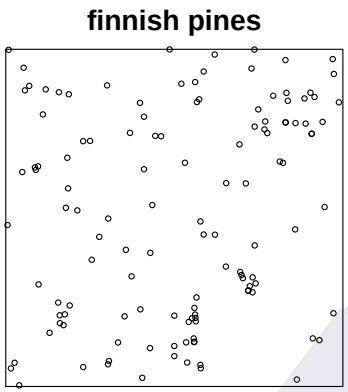




- . . .can exhibit independence, neg. or pos. dependence







1. courtesy Jim O'Leary, crows attacks in Vancouver in 2017

How to characterize or summarize a point process ?

- finite-dimensional distributions of counting variables ;
- void provabilities : $\mathrm{P}(N(B) = 0)$, for any compact set B ;
- via a density (with respect to a Poisson process) ;
- intensity functions : $\boxed{\rho}$, $\rho^{(2)}$, $\rho^{(3)}$,... .
- summary statistics : Ripley's K function, functions $F, G, J, L, \ldots$

Intensity function : $\rho : W \to \mathbb{R}^+$

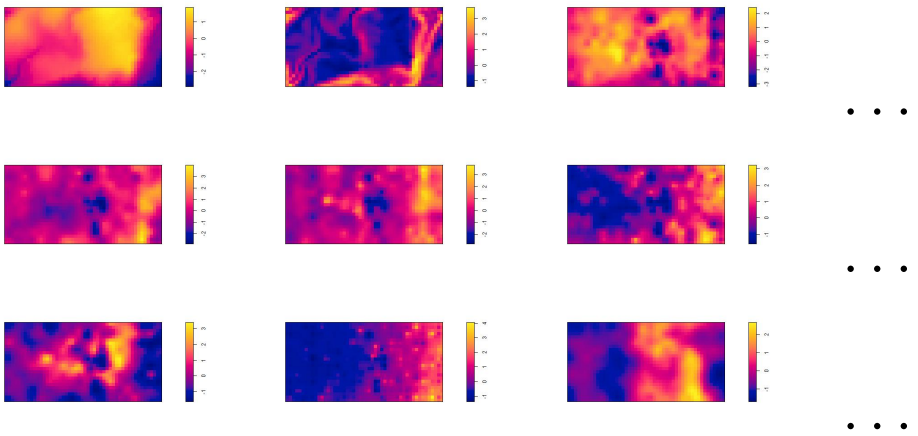
$$\rho(u) \approx \mathrm{P}\left(\mathbf{X} \text{ has a point in } B(u, \mathrm{d}u)\right)$$

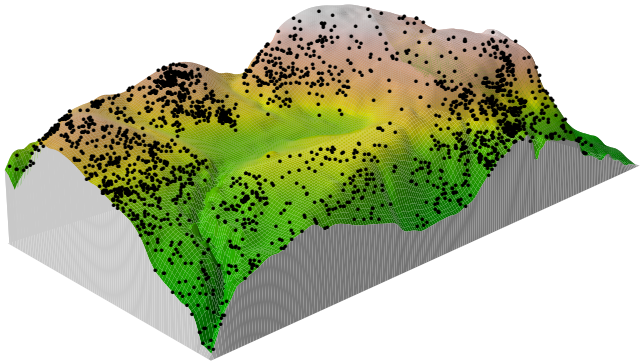
$$\mathrm{E} \sum_{u \in \mathbf{X}} h(u) = \int h(u)\rho(u)\mathrm{d}u, \qquad \text{[Campbell Theorem]}$$

Objective : parametric modeling ρ

Data : Barro Colorado Island (Hubell et al., 1999, 2005)

- $W = [0, 1000m] \times [0, 500m]$
- $>$ 300,000 locations of trees
- $\approx$ 300 species
- $\approx$ 100 spatial covariaties observed at fine scale (altitude, nature of soils,...)





Modeling ρ : for one species e.g.

$$\log \rho(u; \boldsymbol{\beta}) = \exp(\boldsymbol{\beta}^\top \mathbf{z}(u)),$$

$\boldsymbol{\beta} \in \mathbb{R}^p$, $\mathbf{z}(u) = (z_1(u), \ldots, z_p(u))^\top$

Problem : p large, covariates very correlated.

Estimation of $\boldsymbol{\beta}$ when p is moderate (1)

$$\hat{\boldsymbol{\beta}} = \operatorname{argmax}_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}), \quad \ell(\boldsymbol{\beta}) = \sum_{u \in \mathbf{X} \cap W} \underbrace{\boldsymbol{\beta}^\top \mathbf{z}(u)}_{\log \rho(u;\boldsymbol{\beta})} - \int_W \underbrace{\exp\left(\boldsymbol{\beta}^\top \mathbf{z}(u)\right)}_{\rho(u;\boldsymbol{\beta})} \mathrm{d}u$$

- ℓ = log-likelihood if $\mathbf{X}$ is a Poisson point process ;
- but $\ell^{(1)}(\boldsymbol{\beta})$ remains an *unbiased estimating equation* for general point processes. Indeed,

$$\ell^{(1)}(\boldsymbol{\beta}) = \sum_{u \in \mathbf{X}} \mathbf{z}(u) - \int_W \mathbf{z}(u) \rho(u;\boldsymbol{\beta}) \mathrm{d}u$$

  so $\mathrm{E}\ell^{(1)}(\boldsymbol{\beta}) = 0$ by Campbell Theorem (under the true model).

-

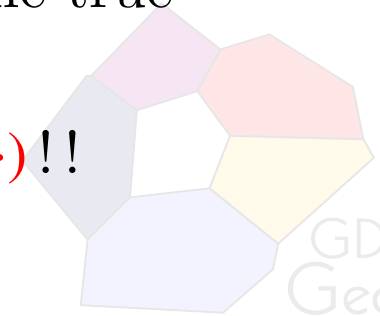
Estimation of $\boldsymbol{\beta}$ when p is moderate (1)

$$\hat{\boldsymbol{\beta}} = \operatorname{argmax}_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}), \quad \ell(\boldsymbol{\beta}) = \sum_{u \in \mathbf{X} \cap W} w(u) \underbrace{\boldsymbol{\beta}^\top \mathbf{z}(u)}_{\log \rho(u;\boldsymbol{\beta})} - \int_W w(u) \underbrace{\exp\left(\boldsymbol{\beta}^\top \mathbf{z}(u)\right)}_{\rho(u;\boldsymbol{\beta})} \mathrm{d}u$$

- ℓ = log-likelihood if $\mathbf{X}$ is a Poisson point process ;
- but $\ell^{(1)}(\boldsymbol{\beta})$ remains an *unbiased estimating equation* for general point processes. Indeed,

$$\ell^{(1)}(\boldsymbol{\beta}) = \sum_{u \in \mathbf{X}} w(u)\mathbf{z}(u) - \int_W w(u)\mathbf{z}(u)\rho(u;\boldsymbol{\beta})\mathrm{d}u$$

  so $\mathrm{E}\ell^{(1)}(\boldsymbol{\beta}) = 0$ by Campbell Theorem (under the true model).
- Nothing changes if we add a weight surface $w(\cdot)$! !



Estimation of $\boldsymbol{\beta}$ when p is moderate (2)

- $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}$ and is asympt. normal ($W = W_n \to \mathbb{R}^d$) for ... ;
- asympt. cov. matrix can be optimized in terms of $w(\cdot)$.
- Bermann-Turner approximation : discretize

$$\int_W w(u)\rho(u;\boldsymbol{\beta})\mathrm{d}u \approx \sum_{i=1}^{n+m} q_i\,\rho(u_i;\boldsymbol{\beta})$$

m : # dummy points ; n : # data points ; q_i quadrature weights. Then, with $y_i = q_i^{-1}\mathbf{1}(u_i \in \mathbf{X})$

$$\ell(\beta) \approx \sum_{i=1}^{n+m} q_i \Big\{ y_i \log \rho(u_i;\boldsymbol{\beta}) - \rho(u_i;\boldsymbol{\beta}) \Big\} \overset{\texttt{R}}{=} \underbrace{\texttt{glm(...,family=quasipoisson)}}_{\texttt{spatstat package}}$$

- BT approx. can be avoided (but $\mathrm{Var}(\hat{\boldsymbol{\beta}}') > \mathrm{Var}(\hat{\boldsymbol{\beta}})$)

$$\ell'(\boldsymbol{\beta}) = \sum_{u \in \mathbf{X} \cap W} \log \frac{\rho(u;\boldsymbol{\beta})}{\delta + \rho(u;\boldsymbol{\beta})} - \delta \int_W \log \frac{\rho(u;\boldsymbol{\beta}) + \delta}{\delta} \mathrm{d}u \approx \ldots \overset{R}{=} \texttt{glm(...,binomial)}$$

When p is large (even $p = p_n,\ W = W_n \to \mathbb{R}^d$)

- Penalization techniques : $\hat{\boldsymbol{\beta}} = \text{argmax}_{\boldsymbol{\beta}}\ Q(\boldsymbol{\beta})$ where

$$Q(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}) \ - \ |W_n| \sum_{j=1}^{p_n} \pi_{\lambda_{n,j}}(|\beta_j|)$$

  $\hookrightarrow$ $\lambda_{n,j} \geq 0$ are regularization parameters

  $\hookrightarrow$ $\pi_\lambda(\cdot)$: penalty function $\underbrace{\text{convex}}_{\|\cdot\|_1, \|\cdot\|_2, \ldots}$ or $\underbrace{\text{non-convex}}_{\text{SCAD, MC+}, \ldots}$

  $\hookrightarrow$ Example : adaptive Lasso $\pi_{\lambda_{n,j}} = \lambda_{n,j}|\beta_j|$.

When p is large (even $p = p_n,\ W = W_n \to \mathbb{R}^d$)

- Penalization techniques : $\hat{\boldsymbol{\beta}} = \text{argmax}_{\boldsymbol{\beta}}\ Q(\boldsymbol{\beta})$ where

$$Q(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}) \ - \ |W_n| \sum_{j=1}^{p_n} \pi_{\lambda_{n,j}}(|\beta_j|)$$

  $\hookrightarrow$ $\lambda_{n,j} \geq 0$ are regularization parameters

  $\hookrightarrow$ $\pi_\lambda(\cdot)$: penalty function $\underbrace{\text{convex}}_{\|\cdot\|_1, \|\cdot\|_2, \ldots}$ or $\underbrace{\text{non-convex}}_{\text{SCAD, MC+}, \ldots}$

  $\hookrightarrow$ Example : adaptive Lasso $\pi_{\lambda_{n,j}} = \lambda_{n,j}|\beta_j|$.

- Computational point of view : thanks to the BT approximation

$$\begin{aligned} \min\left(-Q(\boldsymbol{\beta})\right) &= \min\left(-\ell(\boldsymbol{\beta}) + \text{ penalty }\right) \\ &= \text{convex} \ + \ \text{convex/non-convex} \\ &\overset{\texttt{R}}{=} \texttt{spatstat + glmnet / ncvreg} \end{aligned}$$

What can we prove ? (well expected results !)

- $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, $\boldsymbol{\beta}_1 \in \mathbb{R}^s$, $\boldsymbol{\beta}_2 = 0 \in \mathbb{R}^{p_n - s}$.
- for the *adaptive lasso* : let

$$a_n = \max_{j=1,\dots,s} \lambda_{n,j}, \qquad b_n = \min_{j=s+1,\dots,p_n} \lambda_{n,j}.$$

Theorem ($|W_n| \to \infty$, $p_n \to \infty$) [Choiruddin, C., Letué'18]

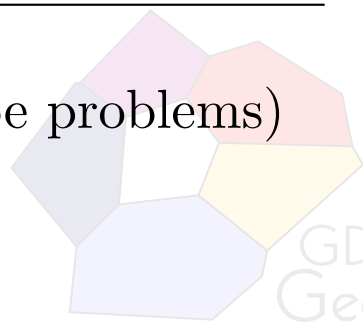
- Under some assumptions such that it works ...
- $p_n^3/|W_n| \to 0$, $a_n \sqrt{|W_n|} \to 0$, $b_n \sqrt{|W_n|/p_n^2} \to \infty$.

Then, as $n \to \infty$

$$\boxed{\mathrm{P}\left(\hat{\boldsymbol{\beta}}_2 = 0\right) \to 1 \quad \text{and} \quad \hat{\Sigma}_n^{-1/2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1) \xrightarrow{d} N(0, I_s)}$$

Technical difficulties : (compared to standard regr. type problems)

- asymptotic is different.
- most importantly : points are **dependent**.

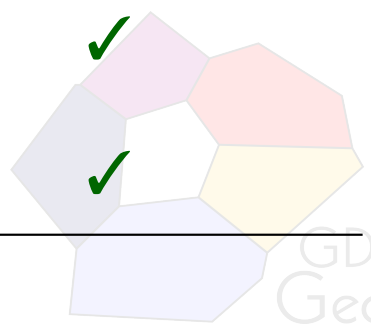


For other penalties

Possible ? $\Leftrightarrow$ $a_n \sqrt{|W_n|} \to 0$ and $b_n \sqrt{|W_n|/p_n^2} \to \infty$

Method	a_n	b_n	Possible ?
Ridge	$\lambda_n \max_{j=1,\dots s} \{\lvert\beta_{0j}\rvert\}$	0	✗
Lasso	λ_n	λ_n	✗
Enet	$\lambda_n \left[(1-\alpha) \max_{j=1,\dots s} \{\lvert\beta_{0j}\rvert\} + \alpha\right]$	$\lambda_n \alpha$	✗
ALasso	$\max_{j=1,\dots s} \{\lambda_{n,j}\}$	$\inf_{j=s+1,\dots p} \{\lambda_{n,j}\}$	✓
Aenet	$\max_{j=1,\dots s} \{\lambda_{n,j}((1-\alpha)\lvert\beta_{0j}\rvert + \alpha)\}$	$\alpha \inf_{j=s+1,\dots p} \{\lambda_{n,j}\}$	✓
SCAD	0^*	$\lambda_n{}^*$	✓
MC+	0^*	$\lambda_n - \frac{K}{\gamma\sqrt{\lvert D_n\rvert}}{}^*$	✓

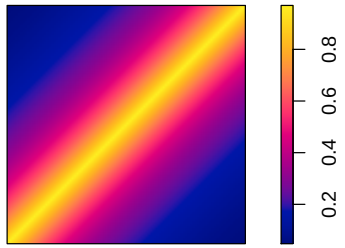
* if $\lambda_n \to 0$ and $\lambda_n \sqrt{|W_n|/p_n^2} \to \infty$.

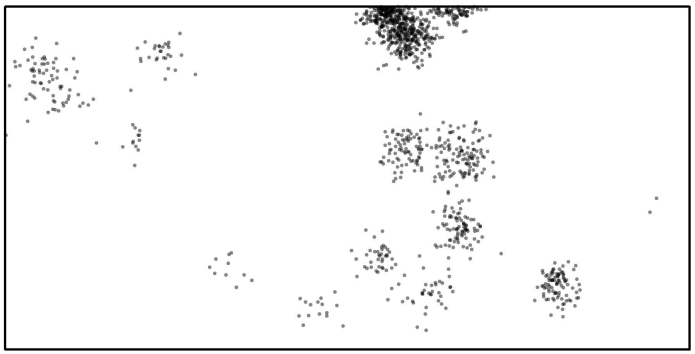


Short simulation

- $W = [0, 1000] \times [0, 500]$; $\mathbf{X}$: Thomas process (generates clustered patterns) ; $m = 5{,}000$ replications.
- $\lambda_{n,j} = \lambda / |\tilde{\beta}_j|$, $\tilde{\beta}$ is the non-regularized estimator (or ridge) ; λ is chosen using BIC-type criterion (CV is dangerous in this setting).

- $\mathbf{z}(u) = (\underbrace{z_1(u), z_2(u)}_{\text{true BCI cov.}}, \underbrace{z_3(u), \ldots, z_{100}(u)}_{\text{noisy correlated cov.}})^\top$
- the z_i's are then centered and reduced ; $\beta_1 = 2, \beta_2 = .75$.





1600 points in average

	FPR (%)	FNR (%)	RMSE
Lasso	97	23	0.73
A. Lasso	95	3	0.63
SCAD	96	7	0.65

Additional remarks and perspectives

- Applied to the `bei` dataset - selection of 8 (relevant) covariates (among 93) ; higher AUC.
- current work : compare AL with **Dantzig type selector**

$$\text{Minimize} \sum_j \lambda_{n,j}|\beta_j| \quad \text{subject to} \quad \|\ell^{(1)}(\boldsymbol{\beta})\| \leq \lambda_{n,j}$$

- 300 species of trees ? **high-dimensional multivariate LGCP** (much more challenging) : for $i = 1, \ldots, 300$

$$\Lambda_i(u) = \rho(u;\boldsymbol{\beta}_i) \exp Z_i(u) \quad Z_i(u) = \sum_{l=1}^{q} \alpha_{il} E_l(u) + U_i(u)$$

$\hookrightarrow$ E_l, U_i are indep. stationary Gaussian random fields ;
$\hookrightarrow$ typically approx. 10,000 interaction parameters

---

Choiruddin , **C.**, Letué. *Convex and non-convex regularization methods for spatial point processes intensity estimation*, EJS, 12(1) :1210–1255, 2018.