New Bridges between Mathematics and Data Science
Nov 8th - 11th 2021, Valladolid, Spain

Kernel methods to cope with the analysis of point processes on road networks

María Isabel Borrajo García

(joint work with C. Comas and J. Mateu)

modestya
Modelos de Optimización, Decisión, Estadística y Aplicaciones

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
DEPARTAMENTO DE ESTATÍSTICA, ANÁLISE MATEMÁTICA E OPTIMIZACIÓN

New Bridges between Mathematics and Data Science
Nov 8th - 11th 2021, Valladolid, Spain

Kernel methods to cope with the analysis of point processes on road networks using covariate information

María Isabel Borrajo García

(joint work with C. Comas and J. Mateu)

Modelos de Optimización, Decisión, Estadística y Aplicaciones

DEPARTAMENTO DE ESTATÍSTICA, ANÁLISE MATEMÁTICA E OPTIMIZACIÓN

New Bridges between Mathematics and Data Science
Nov 8th - 11th 2021, Valladolid, Spain

Kernel methods to cope with the analysis of point processes on road networks using covariate information

María Isabel Borrajo García

(joint work with C. Comas and J. Mateu)

modestya
Modelos de Optimización, Decisión, Estadística y Aplicaciones

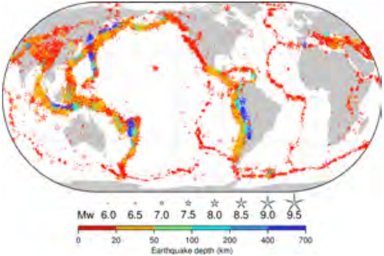
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
DEPARTAMENTO DE ESTATÍSTICA, ANÁLISE MATEMÁTICA E OPTIMIZACIÓN

What are spatial statistics?

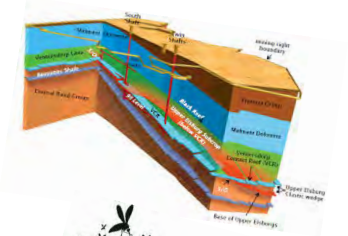
- Geostatistics
- Lattice/area data
- **Point processes**

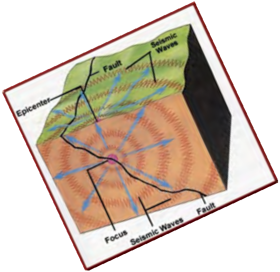
What are spatial statistics?

- Geostatistics
- Lattice/area data
- **Point processes**
  - Continuous space
  - Locations of objects (individuals) in space (typically 2D)
  - Examples: locations of trees in a forest, locations of earthquakes...



Where can we find point processes?





General introduction to point processes

- **Stochastic processes** governing the location of a finite number of events
- Number and location of events are random
- A **point pattern**, $X_1, \ldots, X_N$, is a realization of a point process on a bounded domain, $W \subset \mathbb{R}^2$

Basic elements

- Observation region: W
- Counting measure: $N : \mathcal{P}(W) \rightarrow \mathbb{Z}^{+}$
- Point pattern: $X_1, \ldots, X_N$ (events)
- Characteristic functions, such as first-order intensity: $\lambda(x)$

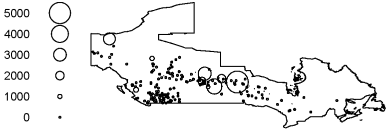
$$\lambda(x) = \lim_{|dx| \to 0} \frac{E[N(dx)]}{|dx|}, \quad x \in W$$

- It characterizes the point process
- Mean number of events per unit area
- Non-negative
- Closely related to the density function

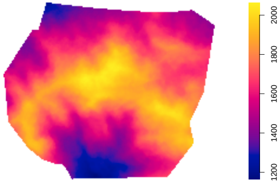
Framework with covariates

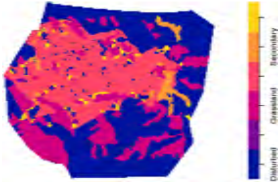
- $X \subset W \subset \mathbb{R}^2$ **inhomogeneous Poisson** point process
- $Z : W \subset \mathbb{R}^2 \to \mathbb{R}$ a spatial **continuous function** exactly known in every point of the region of interest W
- We assume $\lambda(u) = \rho(Z(u))$, $u \in W \subset \mathbb{R}^2$, with ρ an unknown real function
- Extra information: **covariates** (unmarked process)

Marks



Covariates





Framework with covariates

Advances in the euclidean space (mostly 2D)

- First-order kernel intensity estimator
- Definition of at least four ad-hoc designed bandwidth selection methods
- Good theoretical properties: consistency
- Significance covariate tests: parametric and non-parametric
- Two-sample tests

Borrajo, M.I, González-Manteiga, W. and Martínez-Miranda, M.D. (2020) Bootstrapping kernel intensity estimation for inhomogeneous point processes depending on spatial covariates. *Computational Statistics and Data Analysis*, 144

Borrajo, M.I, González-Manteiga, W. and Martínez-Miranda, M.D. (2020) Testing for significant differences between two spatial patterns using covariates. *Spatial Statistics*

Point processes on linear networks

Cáceres, L., Fernández, M. A., Gordaliza, A. and Molinero, A. (2021) etection of Geometric Risk Factors Affecting Head-On Collisions through Multiple Logistic Regression: Improving Two-Way Rural Road Design via 2+ 1 Road Adaptation. *International Journal of Environmental Research and Public Health*, 18(12)

- A **linear network** in the plane is defined as a union of line segments, $L = \cup_{i=1}^{D} l_i$
- A **line segment** l_i with endpoints u_i and v_i is $[u_i, v_i] = \{tu_1 + (1-t)v_i / 0 \leq t \leq 1\}$
- A **path** between any two points x, y along L
- The **shortest path distance** between any two points $x, y \in L$

Point processes on linear networks

BORRUSO, G. (2005). Network density estimation: analysis of point patterns over a network. In *International Conference on Computational Science and Its Applications* (pp. 126-132). Springer.

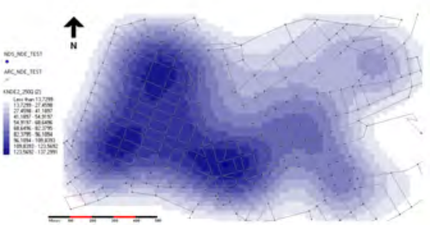
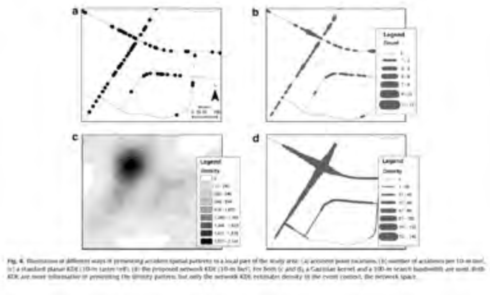


Fig. 2. Network Density Estimation computed over road network's nodes distribution in the Trieste (Italy) city centre. 250 m Bandwidth. (Density surface produced using CrimeStat 1.1 –

BORRUSO, G. (2005). Network density estimation: analysis of point patterns over a network. In *International Conference on Computational Science and Its Applications* (pp. 126-132). Springer.

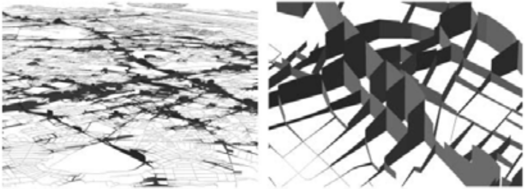
XIE, Z., AND YAN, J. (2008). Kernel density estimation of traffic accidents in a network space. *Computers, environment and urban systems*, 32(5), 396-406.



- Borruso, G. (2005). Network density estimation: analysis of point patterns over a network. In *International Conference on Computational Science and Its Applications* (pp. 126-132). Springer.

- Xie, Z., and Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. *Computers, environment and urban systems*, 32(5), 396-406.

- Sugihara, K., Satoh, T., and Okabe, A. (2010). Simple and unbiased kernel function for network analysis. In the *10th International Symposium on Communications and Information Technologies* (pp. 827-832). IEEE.

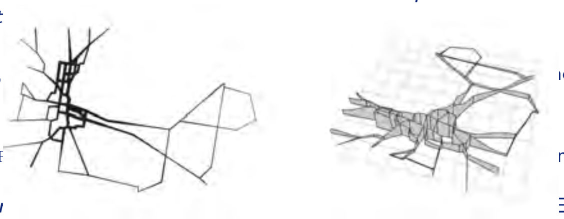


Point processes on linear networks

Borruso, G. (2005). Network density estimation: analysis of point patterns over a network. In *International Conference on Computational Science and Its Applicat*

Xie, Z., ccidents in a network 396-406.

Sugihar nbiased kernel function *on Commu* E.



McSwiggan, G., Baddeley, A., and Nair, G. (2017). Kernel density estimation on a linear network. *Scandinavian Journal of Statistics*, 44(2), 324-345.

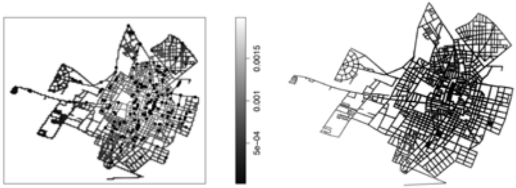
Point processes on linear networks

Borruso, G. (2005). Network density estimation: analysis of point patterns over a network. In *International Conference on Computational Science and Its Applications* (pp. 126-132). Springer.

Xie, Z., and Yan, J. (2008). Kernel density estimation of traffic accidents in a networ… …96-406.

Sugih… …iased kernel functic… …n *Comm…*

McSw… …nsity estima… …(2), 324-345.



(a) grayscale style (b) line-thickness style

Moradi, M. M., Rodríguez-Cortés, F. J., and Mateu, J. (2018). On kernel-based intensity estimation of spatial point patterns on linear networks. *Journal of Computational and Graphical Statistics*, 27(2), 302-311.

Kernel intensity estimator

$$\lambda(x) = \rho(Z(x))$$

Borrajo, M. I., Comas, C., Costafreda-Aumedes, S., and Mateu, J. (2021) Stochastic smoothing of point processes for wildlife-vehicle collisions on road networks. *Stochastic Environmental Research and Risk Assessment*

$$\lambda(x) = \rho(Z(x))$$

Theorem

Let X be a point process on a linear network $L \subset \mathbb{R}^2$ with intensity function of the form $\lambda(u) = \rho(Z(u))$ $u \in L$, where $Z : L \subset W \subset \mathbb{R}^2 \to \mathbb{R}$ is a differentiable function with non-zero gradient in every point of its domain.

Then, the transformed point process $Z(X)$ is a one-dimensional point process with intensity function $\rho g^\star$, where $g^\star = (G^\star)'$ and $G^\star(z) = \int_L 1_{\{Z(u) \leq z\}} d_1 u$ is the unnormalised version of the spatial cumulative distribution function of the covariate.

Borrajo, M. I., Comas, C., Costafreda-Aumedes, S., and Mateu, J. (2021) Stochastic smoothing of point processes for wildlife-vehicle collisions on road networks. *Stochastic Environmental Research and Risk Assessment*

$$\lambda(x) = \rho(Z(x))$$

Theorem

Let X be a point process on a linear network $L \subset \mathbb{R}^2$ with intensity function of the form $\lambda(u) = \rho(Z(u))$ $u \in L$, where $Z : L \subset W \subset \mathbb{R}^2 \to \mathbb{R}$ is a differentiable function with non-zero gradient in every point of its domain.

Then, the transformed point process $Z(X)$ is a one-dimensional point process with intensity function $\rho g^\star$, where $g^\star = (G^\star)'$ and $G^\star(z) = \int_L 1_{\{Z(u) \leq z\}} d_1 u$ is the unnormalised version of the spatial cumulative distribution function of the covariate.

Corollary

Furthermore, if the original point process is Poisson, this property is inherited and the transformed one is also Poisson.

Borrajo, M. I., Comas, C., Costafreda-Aumedes, S., and Mateu, J. (2021) Stochastic smoothing of point processes for wildlife-vehicle collisions on road networks. *Stochastic Environmental Research and Risk Assessment*

Kernel intensity estimator

$$\lambda(x) = \rho(Z(x))$$

BORRAJO, M.I, GONZÁLEZ-MANTEIGA, W. AND MARTÍNEZ-MIRANDA, M.D. (2020) Bootstrapping kernel intensity estimation for inhomogeneous point processes depending on spatial covariates. *Computational Statistics and Data Analysis*, 144

$$\lambda(x) = \rho(Z(x))$$

Idea: to build an **artificial density function** $f(\cdot) = \frac{\rho(\cdot)g^{\star}(\cdot)}{m}$

$$\hat{f}_h(z) = g^{\star}(z)\frac{1}{N}\sum_{i=1}^{N}\frac{1}{g^{\star}(Z_i)}K_h\left(z - Z_i\right)1_{\{N\neq 0\}}$$

Borrajo, M.I, González-Manteiga, W. and Martínez-Miranda, M.D. (2020) Bootstrapping kernel intensity estimation for inhomogeneous point processes depending on spatial covariates. *Computational Statistics and Data Analysis*, 144

Kernel intensity estimator

$$\lambda(x) = \rho(Z(x))$$

Idea: to build an **artificial density function** $f(\cdot) = \frac{\rho(\cdot)g^\star(\cdot)}{m}$

$$\hat{f}_h(z) = g^\star(z)\frac{1}{N}\sum_{i=1}^{N}\frac{1}{g^\star(Z_i)}K_h\left(z - Z_i\right)1_{\{N\neq 0\}}$$

$$\hat{\lambda}_h(u) = \sum_{i=1}^{N}\frac{1}{g^\star(Z_i)}K_h\left(Z(u) - Z_i\right)$$

BORRAJO, M.I, GONZÁLEZ-MANTEIGA, W. AND MARTÍNEZ-MIRANDA, M.D. (2020) Bootstrapping kernel intensity estimation for inhomogeneous point processes depending on spatial covariates. *Computational Statistics and Data Analysis*, 144

Bandwidth selectors

- Bias
- Variance
- MISE
- $AMISE = (1 - e^{-m})^2 \frac{h^4}{4} R\left(\frac{\rho'' g^\star}{m}\right) \mu_2^2(K) + \frac{A(m)}{h} R(K)$

BORRAJO, M. I., COMAS, C., COSTAFREDA-AUMEDES, S., AND MATEU, J. (2021) Stochastic smoothing of point processes for wildlife-vehicle collisions on road networks. *Stochastic Environmental Research and Risk Assessment*

Bandwidth selectors

- Bias
- Variance
- MISE
- $AMISE = (1 - e^{-m})^2 \frac{h^4}{4} R\left(\frac{\rho'' g^\star}{m}\right) \mu_2^2(K) + \frac{A(m)}{h} R(K)$

$$h_{AMISE} = \left(\frac{A(m)R(K)}{\mu_2^2(K)(1 - e^{-m})^2 R\left(\frac{\rho'' g^\star}{m}\right)} \right)^{1/5}$$

Borrajo, M. I., Comas, C., Costafreda-Aumedes, S., and Mateu, J. (2021) Stochastic smoothing of point processes for wildlife-vehicle collisions on road networks. *Stochastic Environmental Research and Risk Assessment*

Bandwidth selectors

- **Rule-of-thumb** (RT)

- **Bootstrap bandwidth** (Boot)

- **Non-model based approach** (NM)

Bandwidth selectors

- **Rule-of-thumb** (RT)
  - Based on Silverman's idea of assuming normality
  - Estimate the parameters
- **Bootstrap bandwidth** (Boot)
- **Non-model based approach** (NM)

Bandwidth selectors

- **Rule-of-thumb** (RT)
  - Based on Silverman's idea of assuming normality
  - Estimate the parameters

- **Bootstrap bandwidth** (Boot)
  - Define a smoothed bootstrap
  - Replicate the scenario in the bootstrap world
  - Compute AMISE and optimal bandwidth
  - Estimate unknown elements

- **Non-model based approach** (NM)

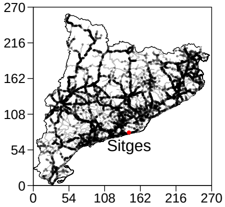
Bandwidth selectors

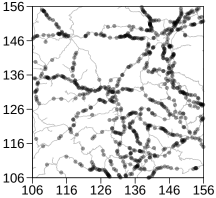
- **Rule-of-thumb** (RT)
  - Based on Silverman's idea of assuming normality
  - Estimate the parameters
- **Bootstrap bandwidth** (Boot)
  - Define a smoothed bootstrap
  - Replicate the scenario in the bootstrap world
  - Compute AMISE and optimal bandwidth
  - Estimate unknown elements
- **Non-model based approach** (NM)
  - $\int_L \frac{1}{\lambda(x)} \lambda(x) dx = |L|$
  - $\hat{h}_{\text{NM}} = \arg\min_{h>0} \left(T(h) - |L| \right)$
  - where $T(h) = \sum_{i=1}^{N} \hat{\rho}_h(Z_i)^{-1}$ inside L, and $|L|$ elsewhere

Cronie, O., and Van Lieshout, M. N. M. (2018). A non-model-based approach to bandwidth selection for kernel estimators of spatial intensity functions. Biometrika, 105(2), 455-462.

Simulations

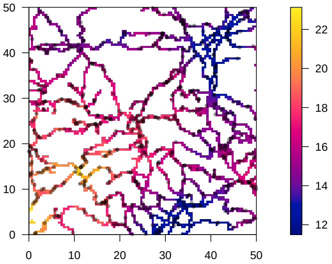
- Square area of 50 $\times$ 50 km^2 in the center of Catalonia (North-East of Spain)
- 11790 km of roads for three distinct categories





Simulations

- Inhomogeneous Poisson point process
- $\lambda(u) = \exp(\beta_0 + \beta_1 Z(u)),\ u \in L,\ \beta_0 = 3$ and $\beta_1 = 1$
- Covariate: Gaussian random field, with zero mean and exponential covariance structure $C(r) = \sigma^2 \exp(r/s)$ with parameters $\sigma = 0.316$ and $s = 150$
- Construct covariate on the network: average in circle of $r = 500$m



Simulations

	$\hat{h}_{MISE}$	$\hat{h}_{RT}$	$\hat{h}_{Boot}$	$\hat{h}_{NM}$
		$m = 100$		
$e1$	28.638	54.919	40.652	44.005
$e2$	18.173	29.520	26.557	23.617
$e3$	–	-0.605	-0.424	-0.504
		$m = 300$		
$e1$	15.938	25.865	18.758	24.979
$e2$	8.147	11.422	9.308	10.439
$e3$	–	-0.534	-0.318	-0.518
		$m = 700$		
$e1$	10.308	16.749	11.468	16.312
$e2$	4.860	6.651	5.221	6.461
$e3$	–	-0.485	-0.234	-0.475
		$m = 1000$		
$e1$	8.610	14.164	9.548	12.682
$e2$	3.812	4.628	3.881	4.298
$e3$	–	-0.466	-0.202	-0.416

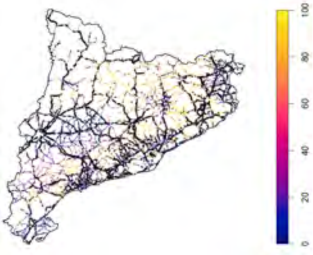
Simulations - without network structure

Simulations - without network structure

	$\hat{h}_{Silv}$	$\hat{h}_{RT}$	$\hat{h}_{Boot}$	$\hat{h}_{NM}$
		$m = 100$		
$e1$	200.776	174.060	155.336	337.542
$e2$	111.962	104.191	103.467	1045.431
		$m = 300$		
$e1$	121.865	109.318	98.966	523.572
$e2$	52.005	49.681	47.131	811.909
		$m = 700$		
$e1$	95.679	87.436	80.681	813.417
$e2$	32.014	30.309	29.239	708.880
		$m = 1000$		
$e1$	92.675	85.129	78.817	997.838
$e2$	30.605	29.694	28.339	721.105

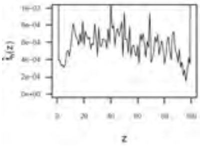
Case study

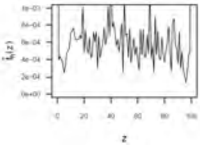
- Road network of Catalonia
- 11790 km of roads
- 6590 wildlife-vehicle collision points in 2010-2014
- Covariate: surface of forest

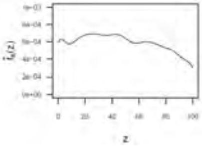


Case study

RT **Boot** **NM**







Ongoing/future work

- Pure internal covariates ($Z : W \subset \mathbb{R}^2 \to \mathbb{R}$)
- Increase covariate dimension
- Develop testing procedures
- Point processes on the sphere

Acknowledgements

Thanks!

Grant PID2020-116587GB-I00 funded by MCIN/AEI/10.13039/501100011033 and the European Union

MINISTERIO DE CIENCIA E INNOVACIÓN

AGENCIA ESTATAL DE INVESTIGACIÓN

New Bridges between Mathematics and Data Science
Nov 8th - 11th 2021, Valladolid, Spain

Kernel methods to cope with the analysis of point processes on road networks

María Isabel Borrajo García

(joint work with C. Comas and J. Mateu)

Modelos de Optimización, Decisión, Estadística y Aplicaciones

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
DEPARTAMENTO DE ESTATÍSTICA, ANÁLISE MATEMÁTICA E OPTIMIZACIÓN