

Foundations of geospatial data: analysis, prediction and monitoring of space-time point pattern data

Jorge Mateu

Department of Mathematics
Universitat Jaume I, Castellon (Spain)
mateu@uji.es

CARTO bootcamp, Madrid, 06 September 2023

CONTENTS -Epidemiology

1. Introduction and motivation

2. A mechanistic spatio-temporal modeling framework

- 3. Spatio-temporal point process models based on neural kernels
- 4. Velocities for spatial growth models
- 5. Closing remarks

CONTENTS-Crime

INTRODUCTION

DETECTING FOCUSES AND GENERATORS OF CRIMINALITY

PROBABILISTIC MODELS FOR CRIMINAL PATTERNS

PREDICTION AND CLASSIFICATION USING ML TECHNIQUES

REDUCING DIMENSIONALITY: BARYCENTERS

STOCHASTIC INTEGRO-DIFFERENTIAL EQUATIONS

MODELING ORIGIN-DESTINATION POINT PATTERNS

1. Introduction and motivation to uncertainty

INTRODUCTION

COVID-19 has spread rapidly across the world in a short period of time following largely heterogeneous patterns

Mathematical and statistical modelling has played a central role in guiding interventions

Many studies follow the compartmental models in **epidemiology**, partitioning the population into subpopulations (compartments) of susceptible (S), exposed (E), infectious (I) and recovered (R), and fit several variations of the classical **deterministic SIR and SEIR** epidemiological models

Given the highly contagious nature, the spatial pattern of the spread of the disease changes rapidly over time. Thus, understanding the **spatio-temporal dynamics** of the spread is undoubtedly critical, and can result in informed and timely public health policies

INTRODUCTION

Considering **stochastic components** is essential to explain the complexity and heterogeneity of the spread over time and space

Literature has focussed mainly in **lattice data analysis**, i.e. aggregated incidence over small areas or regions to provide spatio-temporal models that try to explain and predict the **evolution of incidence and risk** in both space and time simultaneously

However, **point pattern analysis** for general infectious diseases is rather uncommon due to the lack of georeferenced data

We argue that with such type of analysis we can better dig into the evolution of the disease by understanding the disease spread pattern, performing accurate joint space-time prediction, and generate interpretable results

POINT PATTERN CASE

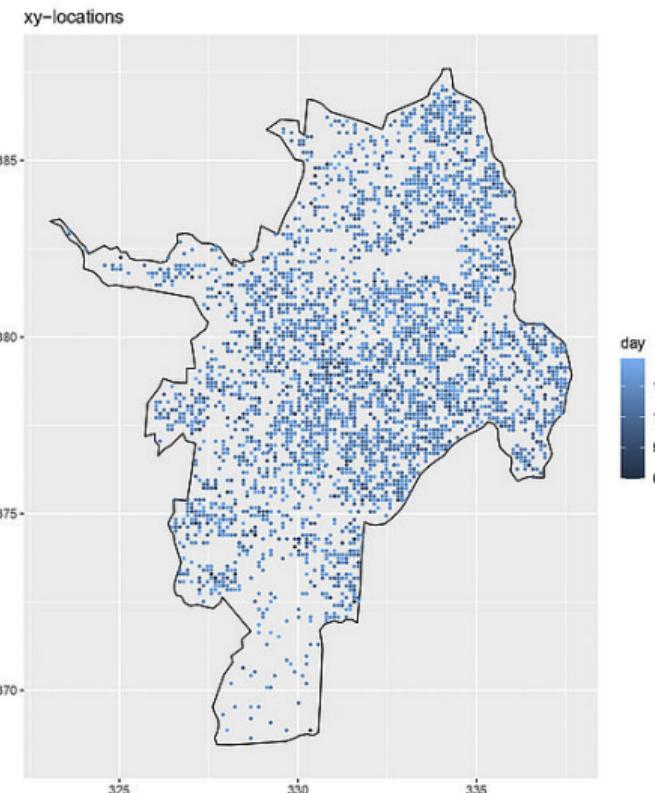


Figure: 38600 observed cases in Cali from March to September 2020

POINT PATTERN CASE

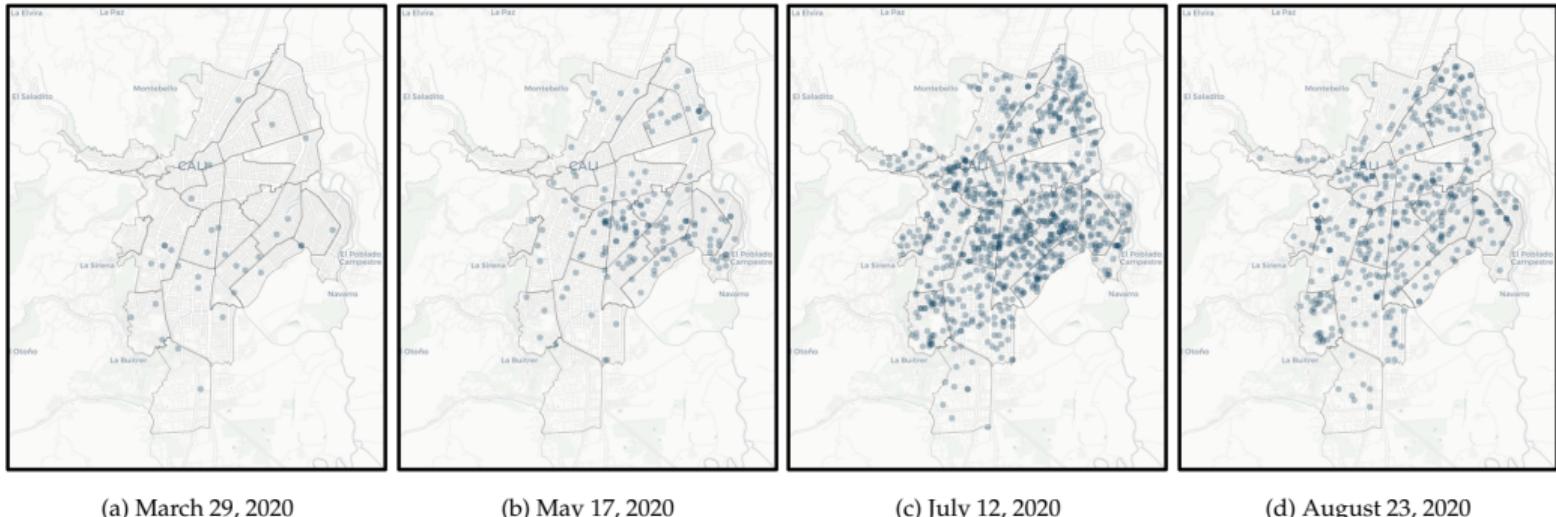


Figure: Snapshots of confirmed COVID-19 cases at four particular weeks. Each dot represents the location of a confirmed case. Note that darker dots indicate multiple dots being overlapped.

IN CONTRAST, THE LATTICE CASE

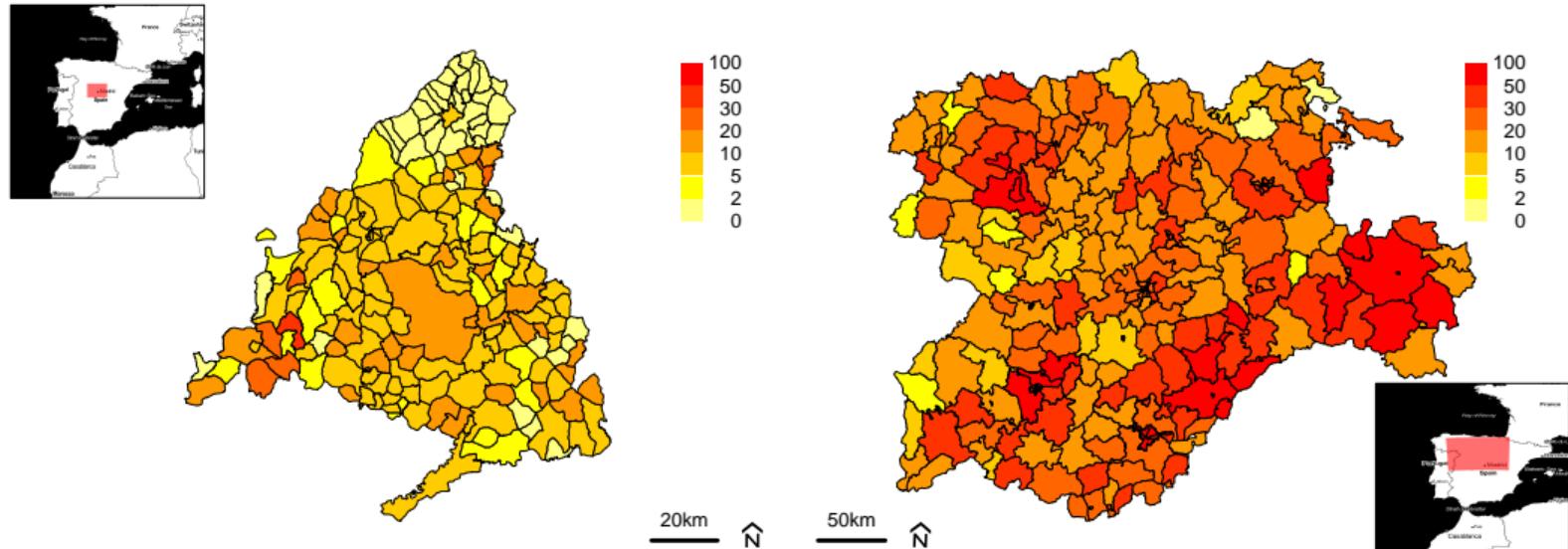


Figure: COVID-19 cases per thousand, up to May 31 2020 for two communities in Spain.

PRESENTATION BASED ON THESE PAPERS

- ▶ Briz, A., Iftimi, A., Mateu, J. and Romero, C. (2022). A mechanistic spatio-temporal modeling of COVID-19 data. Biometrical Journal. Accepted.
- ▶ Dong, Z., Zhu, S., Xie, Y., Mateu, J. and Rodriguez-Cortes, F. (2021). Non-stationary spatio-temporal point process modeling for high-resolution COVID-19 data. Under revision with Journal of the Royal Statistical Society, C.
- ▶ Park, J., Yi, S., Y., Chang, W. and Mateu, J. (2022). A spatio-temporal dirichlet process mixture model for coronavirus disease. Ongoing research.
- ▶ Platero, J., Mateu, J. and Gelfand, A. (2022). Velocity analysis for spatial growth curve models. Ongoing research.

2. A mechanistic spatio-temporal modeling framework

REGION SETTINGS

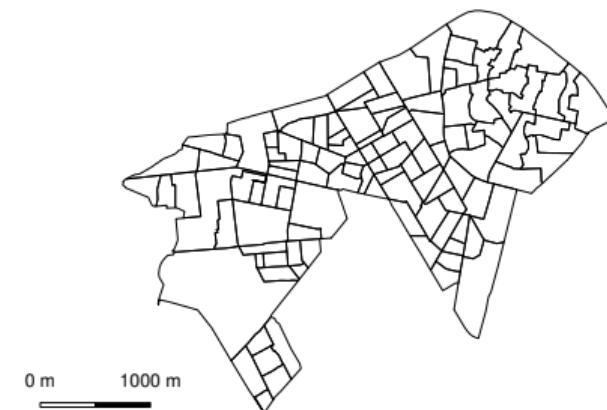
4 districts within the city of Valencia. Target population of about 350000. 6513 buildings and it is divided into 111 census tracts, spatial units of population size ranging from 500 to 2500 inhabitants

Building locations in the study area



(a)

Census tracts in the study area



(b)

Figure: Locations of the residential buildings within the study area (a), and subdivision of the study area into census tracts (b)

COVID-19 DATA

Home addresses of 6647 COVID-19 patients treated in the hospital from February 27, 2020, to January 31, 2021
Geocoded addresses (6647) and buildings (6513): buildings with multiple cases. These cases sharing the same location happen in different times

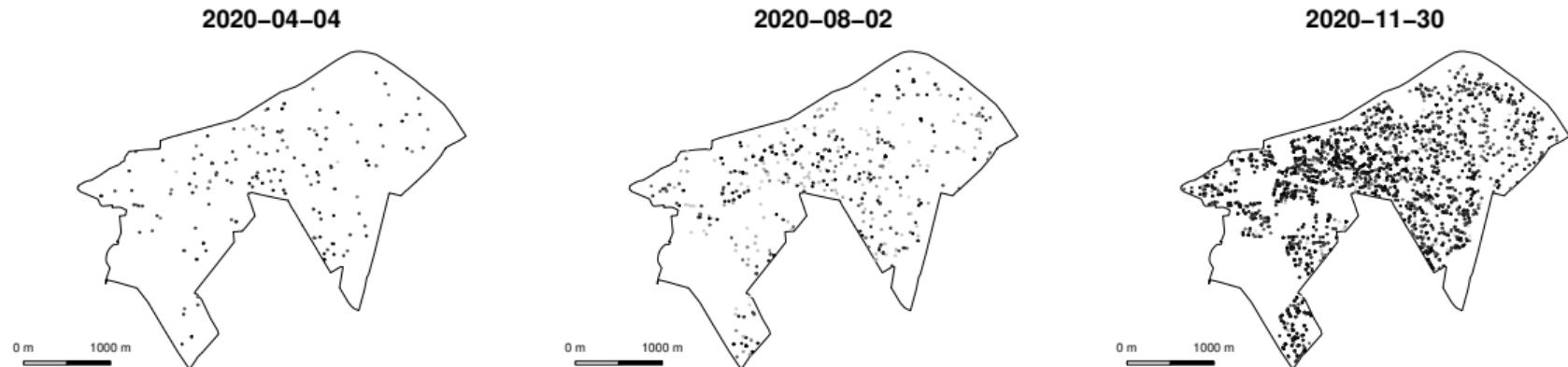


Figure: Evolution of the spatio-temporal point pattern of COVID-19 cases. The darker the color of the point, the more recent the detection of the case

COVID-19 DATA: MOBILITY DATA

The Spanish Statistical Office (INE) has provided estimates of the daily number of people that move from their home mobility area to a different mobility area from 10am to 4pm since the start of the COVID-19 pandemic

Mobility flows are only registered in the case that a person's cell phone is located in a mobility area other than the area considered to be the person's area of residence for more than two hours

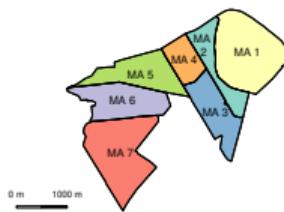
If a person stays for more than two hours in several mobility areas within a single day, only the destination where the person has stayed the longest is taken into account...so mobility flows estimated by INE represent an **underestimation of true mobility patterns**

We estimated the probability of movement from MA_i to MA_j during week w , denoted by $\pi_{MA_i \rightarrow MA_j}^w$, as

$$\pi_{MA_i \rightarrow MA_j}^w = \frac{\text{Mobility flow from } MA_i \text{ to } MA_j \text{ (in number of individuals) on week } w}{\text{Population size of } MA_i} \quad (1)$$

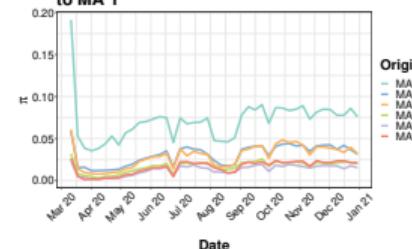
COVID-19 DATA: MOBILITY DATA

Mobility areas in the study area



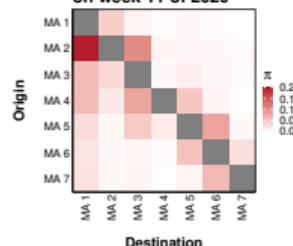
(a)

Evolution of the probability of movement to MA 1



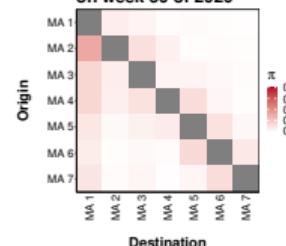
(b)

Matrix of movement probability on week 11 of 2020



(c)

Matrix of movement probability on week 50 of 2020



(d)

Figure: Subdivision of the study area in seven mobility areas (a), evolution of the (weekly) probability of movement to mobility area 1 (MA1) from each of the other six mobility areas (b), and matrices of movement probability for weeks 11 and 50 of 2020 (d)

COVID-19 DATA: SOCIO-DEMOGRAPHIC DATA

Proportion of the population aged 65 years and over (*oldpop*, in percentage), proportion of the population aged less than 18 years (*youngpop*, in percentage), average household income (*avginc*, in euros), and population density (*popdens*, in inhab/km²)



Figure: Spatial distribution of the variables *oldpop* (a), *youngpop* (b), *avginc* (c), and *popdens* (d) at the census tract level

MODELING FRAMEWORK

Spatio-temporal point pattern $\{(x_i, t_i)\}_{i=1}^n$, considered as a realization of a stochastic spatio-temporal point process, X , defined over the spatio-temporal window $W \times T$, $W \subset \mathbb{R}^2$ and $T \subset \mathbb{R}^+$. The point pattern can be ordered according to the temporal locations of the events ($t_1 < t_2 < \dots < t_n$)

The **first-order spatio-temporal intensity** represents the rate at which events occur in space and time

$$\lambda(x, t) = \lim_{|dx \times dt|} \frac{\mathbb{E}[N(dx \times dt)]}{dxdt},$$

where $\mathbb{E}[N(dx \times dt)]$ represents the expected number of events to be observed in the infinitesimal spatio-temporal region $dx \times dt$

The **conditional intensity** of a spatio-temporal point process is the expected rate that an event takes place around a spatio-temporal location (x, t) , conditionally on the history of the process up to time t , $\mathcal{H}_t = \{(x, s) : s \leq t\}$

Conditional intensity-based spatio-temporal models are called **mechanistic models**. Mechanistic models allow estimating the intensity in a realistic way, enabling to account explicitly for the spatio-temporal location of the most recent events.

In contrast, **empirical models** are considered less realistic because they do not necessarily admit a specific scientific interpretation, although they fit spatio-temporal data suitably in many real-life situations

MODELING FRAMEWORK: MECHANISTIC APPROACH

For $\{(x_i, t_i)\}_{i=1}^n$ ($n = 6647$), the spatio-temporal locations of COVID-19 cases, the following conditional intensity is considered (framework of Cox processes)

$$\lambda(x, t | \mathcal{H}_t) = \mu_0(t) \lambda_\beta(x) \rho(x, t) \quad (2)$$

Here, we have an overall temporal intensity ($\mu_0(t)$), an overall spatial intensity ($\lambda_\beta(x)$), and a residual spatio-temporal term which is the one that endows the model with a mechanistic nature ($\rho(x, t)$)

Rather than following a non-separable fashion, recalling arguments in Diggle et al (2013), we treat spatially averaged time-trends and temporally averaged spatial trends as first-order, non-stochastic effects, and any residual spatio-temporal structure as a second-order, stochastic effect

MODELING FRAMEWORK: TEMPORAL AND SPATIAL INTENSITIES

Overall temporal intensity: KDE for the times t_i at which cases were observed

$$\mu_0(t) = \frac{1}{n\sigma} \sum_{i=1}^n K\left(\frac{t - t_i}{\sigma}\right)$$

where $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$ is a Gaussian kernel function, and $\sigma > 0$ is the kernel's bandwidth

Overall spatial intensity: a parametric method at the census tract level. Case counts, denoted by $\mathbf{Y} = \{Y_i\}_{i=1}^{111}$, are described through the NB model

$$Y_i \sim \text{NB}(E_i r_i, \psi) \quad \log(r_i) = \beta_0 + \beta_1 \text{oldpop}_i + \beta_2 \text{youngpop}_i + \beta_3 \text{avginc}_i + \beta_4 \text{popdens}_i + u_i + v_i \quad (3)$$

E_i (offset) denotes the number of expected cases in census tract i , r_i is the relative risk at census tract i , and ψ is the parameter that allows for overdispersion

Terms u_i and v_i are the spatial random effects given by the BYM model, and the spatially-unstructured effect over the areas that follows a Gaussian distribution, $v_i \sim N(0, \sigma_v^2)$, where σ_v^2 is the variance of the effect

If $\lambda_\beta(x)$ is a parametric estimation of the spatial intensity, we compute $\hat{\lambda}_\beta(x) = \hat{Y}_{c(x)}/\text{area}_{c(x)}$, where $c(x)$ indicates the census tract where x is located. Rescale $\lambda_\beta(x)$ to integrate to 1

MODELING FRAMEWORK: STOCHASTIC TERM

$\rho(\mathbf{x}, t)$ attempts to capture the specific influence that a detected case has on new cases

$$\rho(\mathbf{x}, t) = \prod_{j: t_j \leq t} g(\mathbf{x}, \mathbf{x}_j), \quad (4)$$

i.e., each detected case by time t contributes to the value of $\rho(\mathbf{x}, t)$ according to some function $g(\mathbf{x}, \mathbf{x}_j)$

$g(\mathbf{x}, \mathbf{x}_j)$ is the transmission kernel (Diggle, 2006)

$$g(\mathbf{x}, \mathbf{x}_j) = f(||\mathbf{x} - \mathbf{x}_j||) = 1 + \theta e^{-||\mathbf{x} - \mathbf{x}_j||/\varphi} \quad (5)$$

The greater the distance, the lower the effect becomes. Parameter θ measures the strength of the effect of each case on the term $\rho(\mathbf{x}, t)$ ($1 + \theta$ is the upper bound of this effect), $\varphi > 0$ permits scaling the effect of the distance $||\mathbf{x} - \mathbf{x}_j||$

If $-1 < \theta < 0$ then there is some degree of inhibition in the transmission process, $\theta = 0$ indicates the absence of interaction, and $\theta > 0$ represents attraction

MODELING FRAMEWORK: STOCHASTIC TERM

We restrict the product in (4) to the most recent cases. If $N_c(t) = \{i \in \{1, \dots, n\} : t - c \leq t_i \leq t\}$ denotes the set of cases occurred in the last c days, we can write $\rho(\mathbf{x}, t) = \prod_{j \in N_c(t)} g(\mathbf{x}, \mathbf{x}_j)$

We modify the definition of $g(\mathbf{x}, \mathbf{x}_j)$ to account for the possibility that cases leave their corresponding mobility areas and might have an impact on people that live some distance apart

$$g(\mathbf{x}, \mathbf{x}_j) = \begin{cases} f(||\mathbf{x} - \mathbf{x}_j||) & \text{MA}(\mathbf{x}_j) = \text{MA}(\mathbf{x}) \\ (1 - \pi_{\mathbf{x}_j \rightarrow \mathbf{x}}^{w(t)})f(||\mathbf{x} - \mathbf{x}_j||) + \pi_{\mathbf{x}_j \rightarrow \mathbf{x}}^{w(t)}f(||\mathbf{x} - C_{\text{MA}(\mathbf{x})}||) & \text{MA}(\mathbf{x}_j) \neq \text{MA}(\mathbf{x}) \end{cases}$$

$\text{MA}(\mathbf{x})$ is the mobility area ($\text{MA } k, k = 1, \dots, 7$) where \mathbf{x} is located

$\pi_{\mathbf{x}_j \rightarrow \mathbf{x}}^{w(t)}$ is an estimate of the probability that case j visited the mobility area where \mathbf{x} is located during the week $w(t)$

$C_{\text{MA}(\mathbf{x})}$ denotes the centroid of the mobility area where \mathbf{x} belongs to

$g(\mathbf{x}, \mathbf{x}_j)$ term consists of a mixture between the contribution of each case to the intensity on \mathbf{x} by **physical proximity**, and its contribution as a consequence of **human mobility** between areas (imported case)

MODELING FRAMEWORK: LIKELIHOOD

Log-likelihood function of the conditional intensity as a function of the parameters θ and φ

$$L(\theta, \varphi) = \sum_{i=1}^n \log \lambda(\mathbf{x}_i, t_i | \mathcal{H}_{t_i}) - \int_T \int_W \lambda(\mathbf{x}, t | \mathcal{H}_t) d\mathbf{x} dt$$

Partial likelihood

$$L_p(\theta, \varphi) = \sum_{i=1}^n \log \frac{\lambda(\mathbf{x}_i, t_i | \mathcal{H}_{t_i})}{\int_W \lambda(\mathbf{x}, t_i | \mathcal{H}_{t_i}) d\mathbf{x}} \quad (6)$$

If the locations of the events are restricted to the locations of the buildings, the underlying process is said to be **spatially discrete** and the integral becomes a finite sum

$$L_p(\theta, \varphi) = \sum_{i=1}^n \log \frac{\lambda(\mathbf{x}_i, t_i | \mathcal{H}_{t_i})}{\sum_{j=1}^{6513} \lambda(\mathbf{x}_j, t_i | \mathcal{H}_{t_i})} \quad (7)$$

Differential evolution algorithm for global optimization to have an idea of the plausible values for the parameters θ and φ . Then, a Bayesian approach was followed to infer the posterior distribution of θ and φ

RESULTS: OVERALL TEMPORAL INTENSITY

Optimal temporal bandwidth parameter 14.72 days, yielding the estimate of $\mu_0(t)$:

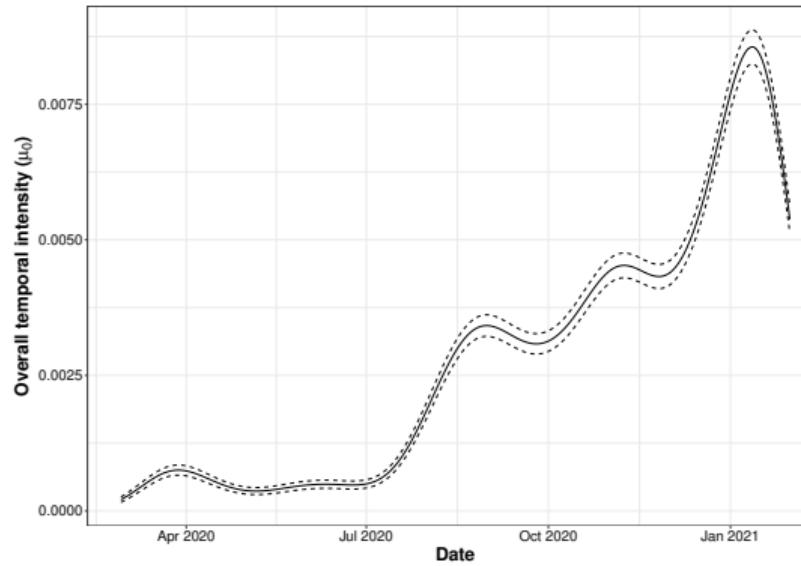


Figure: Non-parametric estimate of the overall temporal intensity of the process ($\mu_0(t)$ in (2)) over the study period, starting at the end of February 2020, and ending at the end of January 2021. The dashed lines represent the uncertainty around each estimate $\hat{\mu}_0(t)$, computed as $\hat{\mu}_0(t) \pm 2\sqrt{\mathbb{V}(\hat{\mu}_0(t))}$

RESULTS: OVERALL SPATIAL INTENSITY

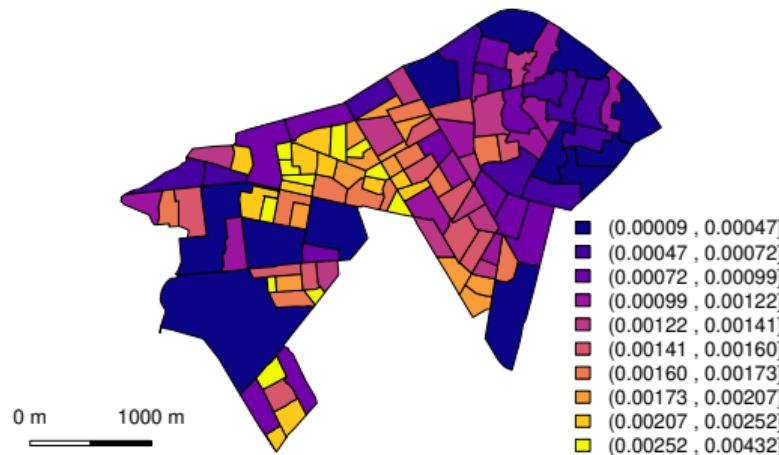
Overall spatial intensity (λ_β)

Figure: Parametric estimate of the overall spatial intensity of the process ($\lambda_\beta(x)$) across the study area, considering tract-level covariates and spatial random effects

RESULTS: SPATIAL RANDOM EFFECT ESTIMATES

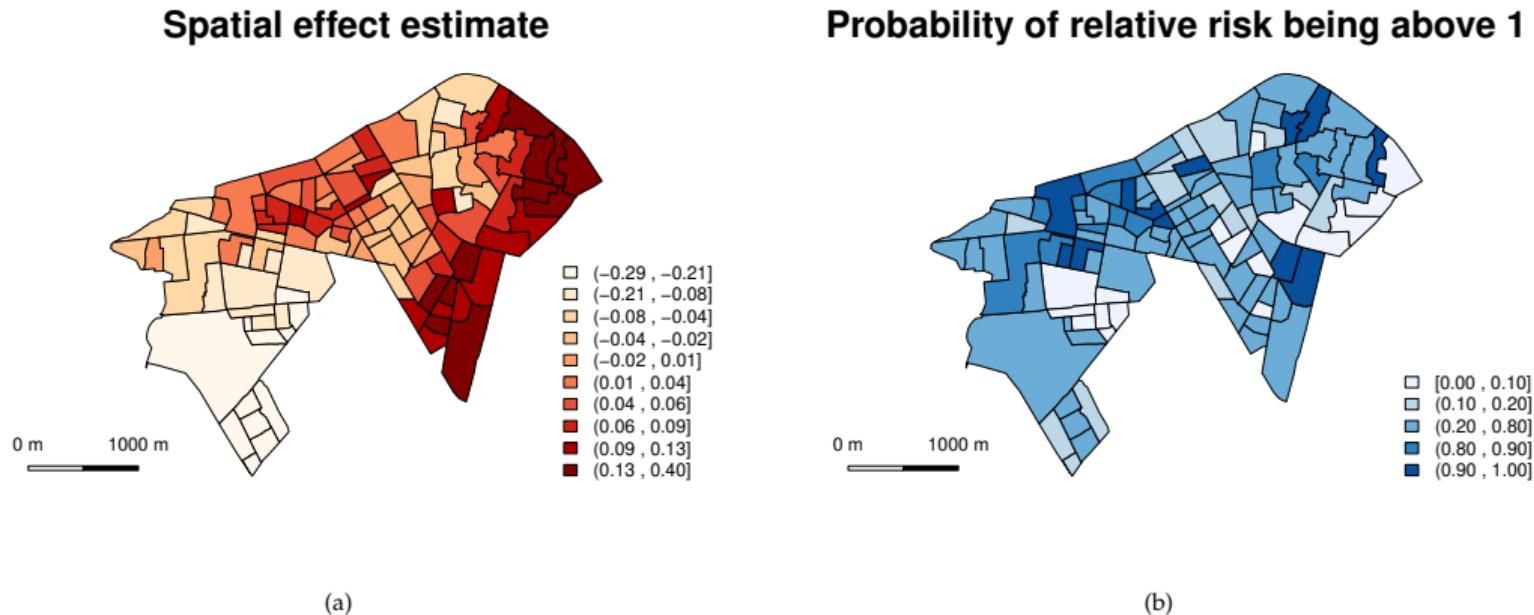


Figure: Spatial random effects estimates at the census tract level, which correspond to $u_i + v_i$ in (3), and posterior probabilities $P(r_i > 1|Y)$

RESULTS: SPATIO-TEMPORAL INTERACTION

c	Parameter	Estimate	Lo	Up	L_p
7	θ	0.47	0.45	0.48	-54571.87
	φ	2.04	0.10	7.03	
14	θ	0.26	0.25	0.26	-55375.03
	φ	1.90	0.10	5.33	
21	θ	0.20	0.19	0.20	-55874.56
	φ	1.40	0.10	4.01	

Table: Mean values of the posterior distributions of θ and φ , 0.025 (Lo) and 0.975 (Up) quantiles, and L_p values considering $c = 7, 14, 21$

The estimates of φ are small, which suggests that the spatial range of the influence of each case is quite short. Cases mostly increase COVID-19 risk of people that live in the same residential areas, or in close-by residential locations

RESULTS: SPATIO-TEMPORAL INTERACTION

Plot of the transmission kernel function for $\theta = 0.26$ and $\varphi = 1.90$ ($c = 14$). Most of the contribution to $\rho(x, t)$ occurs in the immediate vicinity of each case, with cases being in residential locations, with more than 10 meters away becoming almost irrelevant in this regard

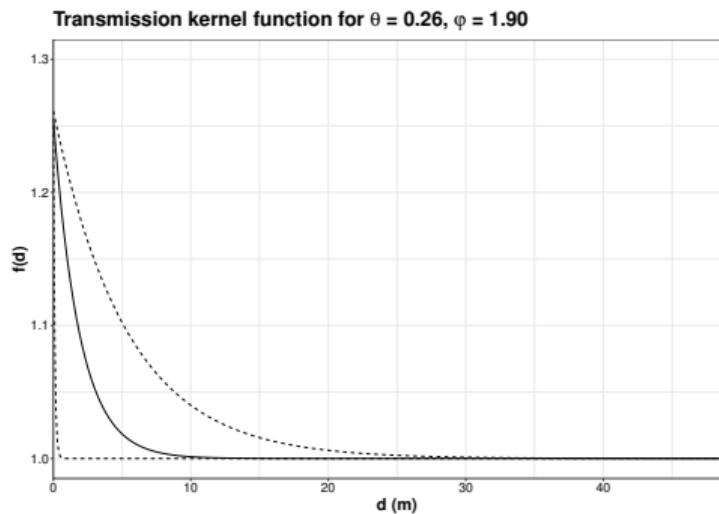


Figure: Transmission kernel function (solid line), defined as $f(d) = 1 + \theta e^{-d/\varphi}$, for $\theta = 0.26$ and $\varphi = 1.90$, which correspond to the mean value of the posterior distributions of both parameters considering $c = 14$. The dashed lines represent the transmission kernel functions that are obtained by using the 0.025 and 0.975 quantiles of the posterior distributions of θ and φ (Table 1)

3. Non-stationary spatio-temporal point process models based on neural kernels

CALI, COLOMBIA

Cali: Most populated city in southwest Colombia – 2.2 million inhabitants. 2nd largest in the country and the only major city with access to Pacific coast. Leading industrial and economic center in country's south

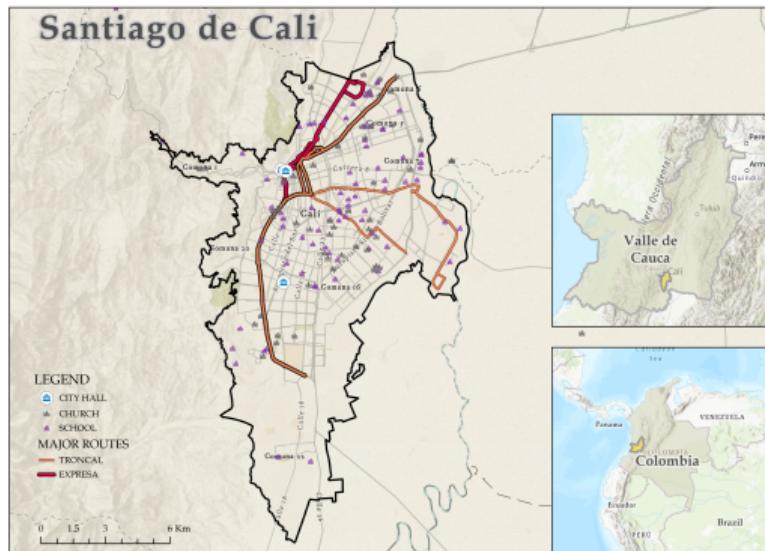
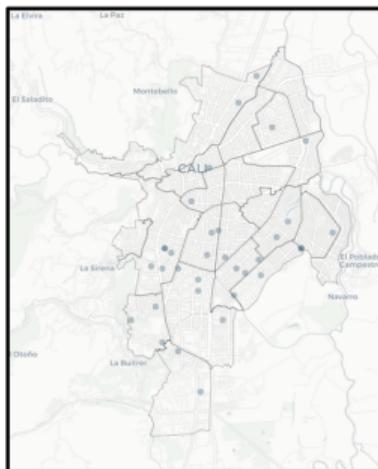


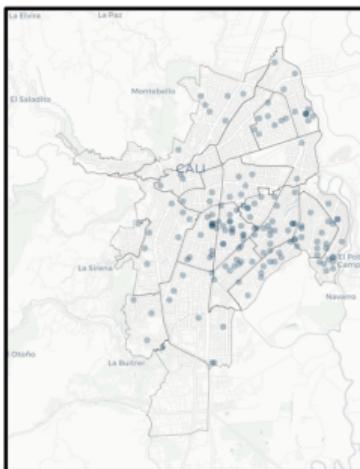
Figure: Geographical location of Cali

COVID CASES

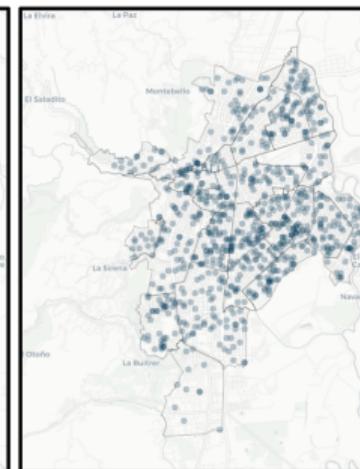
A unique high-resolution individual dataset: 38611 individual confirmed COVID-19 cases for 6 months (28 weeks), from March 15 to September 30, 2020. Each case: diagnosed date, geographic location (long/lat) of residence, with similar test rate in each comuna (22 comunas)



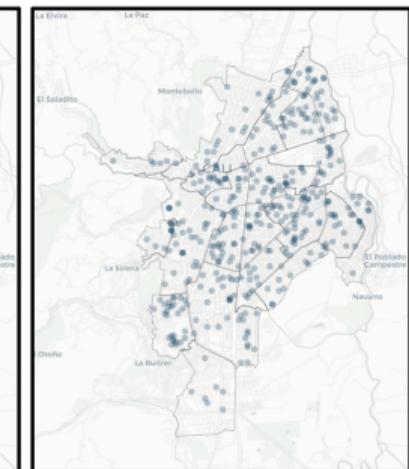
(a) March 29, 2020



(b) May 17, 2020

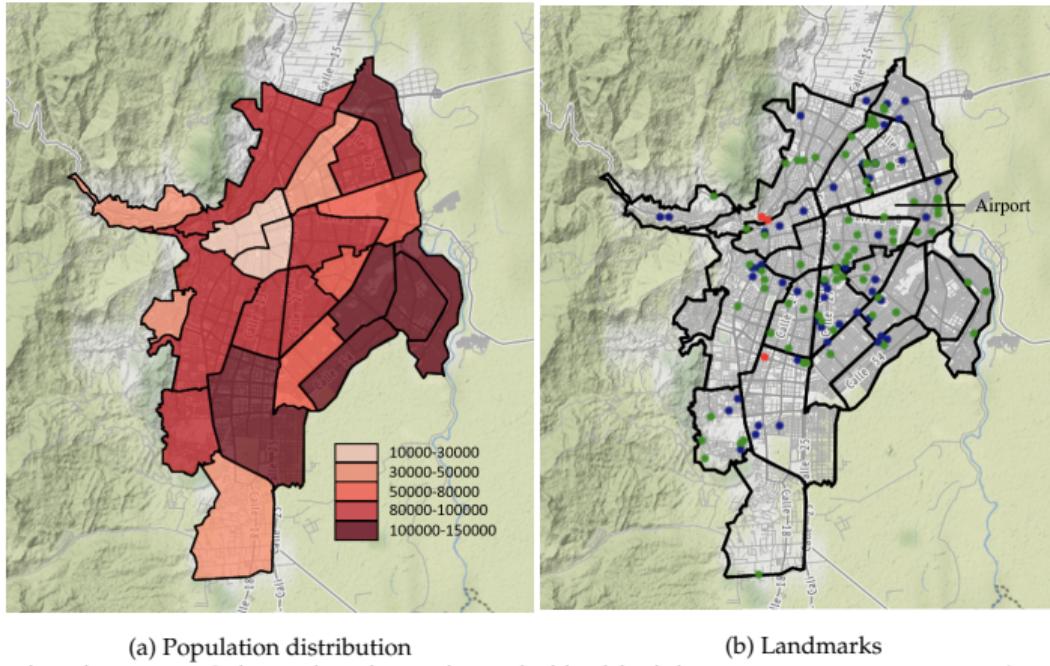


(c) July 12, 2020



(d) August 23, 2020

Figure: Snapshots of confirmed COVID-19 cases at four particular weeks



(a) Population distribution

(b) Landmarks

Figure: (a) Population distribution in Cali. Each polygon bounded by black lines represents a comuna (22). (b) Landmarks in Cali. Each dot represents the landmark's location, and its color indicates the type of the landmark, where the red dot is a town hall, the blue dot is a church, and the green dot is a school

SPATIO-TEMPORAL POINT PROCESSES

Spatio-temporal point processes (STPPs) over the observation space $\mathcal{X} = [0, T] \times \mathcal{S} \subseteq \mathbb{R}^+ \times \mathbb{R}^2$, where T is the time horizon and \mathcal{S} represents the space of geographic coordinate system

Each confirmed case is a *discrete event* defined by a data tuple $x := (t, s)$, where $t \in [0, T]$ is the time when the individual was diagnosed with COVID-19, and $s \in \mathcal{S}$ represents the location of residence of confirmed case

Let $\mathcal{H}_t := \{x_i = (t_i, s_i) | t_i < t\}$ denote the events' history before time t

Given the observed history \mathcal{H}_t , the probability structure of the point process is characterized by the **conditional intensity function** $\lambda(t, s)$, defined as $\lambda(t, s)dt \cdot |B(s, ds)| = \mathbb{E}[\mathbb{N}(t + dt, s + ds) | \mathcal{H}_t]$

HAWKES SPATIO-TEMPORAL POINT PROCESSES

Hawkes processes (a type of **self-exciting point process**) capture the triggering effects between events. The conditional intensity takes the form

$$\lambda(t, s) = \lambda_0 + \int_0^t \int_{\mathcal{S}} k(t, \tau, s, u) d\mathbb{N}(\tau, u), \quad (8)$$

where $\lambda_0 > 0$ denotes the background intensity, and $k(t, t', s, s')$ is a triggering kernel function that captures the influence of past events on the likelihood of event occurrence at the current time

Given the observed point pattern x , we can write the log-likelihood as

$$\ell(x) = \sum_{i=1}^{\mathbb{N}([0, T] \times \mathcal{S})} \log \lambda(t_i, s_i) - \int_0^T \int_{\mathcal{S}} \lambda(\tau, u) du d\tau, \quad (9)$$

where $\mathbb{N}([0, T] \times \mathcal{S})$ is the number of observed events

TRIGGERING KERNEL

Stationary. Epidemic Type Aftershock-Sequences (ETAS) model, that uses Gaussian diffusion kernel

$$k(t, t', s, s') = \frac{Ce^{-\beta(t-t')}}{2\pi\sqrt{|\Sigma|(t-t')}} \cdot \exp\left\{-\frac{(s-s' - \mu)^\top \Sigma^{-1}(s-s' - \mu)}{2(t-t')}\right\},$$

where $\Sigma \equiv \text{diag}(\sigma_x^2, \sigma_y^2)$ is a two-dimensional diagonal matrix representing the covariance of the spatial correlation, β is the decaying rate, μ is the mean shift, and C is a constant. **Diffusion kernel is stationary** and assumes the spatial correlation is isotropic - unable to capture complex spatial dependencies

Non-stationary. Triggering kernel that plays a vital role in modeling the heterogeneous spatial correlation across different regions. We adopt that the triggering effect of a past event is separable in space and time

$$k(t, t', s, s') = \nu(t, t') \cdot v(s, s'),$$

with **stationary temporal kernel**: $\nu(t, t') = Ce^{-\frac{1}{2\sigma_0^2}(t-t')^2}$, $t > t'$

NON-STATIONARY SPATIAL KERNEL

A non-stationary GP $Y(s) = \int k_s(u)W(du)$ with given correlogram, and W a white noise process

Higdon (1988) proposes a discrete approximation to a non-stationary GP $Y(s) = \sum_{i=1}^R k_s(u_i)x_i$ where x_i are iid zero-mean normal rv associated to each knot u_i

Our approach is inspired by Higdon (1998) in modeling a non-stationary kernel used for Gaussian processes. But, instead of using knot locations, we can **estimate the kernel function over the whole space represented using neural networks**

NON-STATIONARY SPATIAL KERNEL: CONSTRUCTION

Given two arbitrary locations $s, s' \in \mathcal{S}$, we define the spatial kernel $v(s, s')$ as an inner product between two feature mappings ϕ_s and $\phi_{s'}$, $v(s, s') = \langle \phi_s, \phi_{s'} \rangle$, $s, s' \in \mathcal{S}$ with $\langle f, g \rangle := \int_{\mathbb{R}^2} f(u)g(u)du$

Also, ϕ_s is a weighted sum of a set of R independent kernel-induced feature functions $\{\kappa_s^{(r)} := \kappa^{(r)}(s, \cdot)\}_{r=1}^R$,
 $\phi_s = \sum_{r=1}^R w_s^{(r)} \kappa_s^{(r)}$

$\kappa^{(r)} : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}_+$ is a general kernel and $w_s^{(r)}$ is the corresponding weight of that feature function at location s .
The location-dependent weight satisfies $\sum_{r=1}^R w_s^{(r)} = 1$ at any arbitrary location s

Hence the spatial kernel can be rewritten as

$$v(s, s') = \sum_{1 \leq r_1, r_2 \leq R} w_s^{(r_1)} w_{s'}^{(r_2)} \left\langle \kappa_s^{(r_1)}, \kappa_{s'}^{(r_2)} \right\rangle.$$

NON-STATIONARY SPATIAL KERNEL: CONSTRUCTION

Following Higdon et al (1988), we choose κ_s to be a Gaussian function centered at s with covariance matrix Σ_s . Thus

$$v(s, s') = \sum_{1 \leq r_1, r_2 \leq R} \frac{w_s^{(r_1)} w_{s'}^{(r_2)}}{2\pi |\Sigma_s^{(r_1)} + \Sigma_{s'}^{(r_2)}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (s - s')^\top (\Sigma_s^{(r_1)} + \Sigma_{s'}^{(r_2)})^{-1} (s - s') \right\}. \quad (10)$$

There exists a **one-to-one mapping between a bivariate normal distribution specified by Σ_s and its one standard deviation ellipse**

Specify the ellipse by a pair of focus points and the fixed area A . The focus points are denoted by $\psi_s = (\psi_x(s), \psi_y(s))$ and $-\psi_s = (-\psi_x(s), -\psi_y(s))$, where $\psi_s \in \Psi \subset \mathbb{R}^2$

Σ_s can be written as

$$\Sigma_s = \tau_z^2 \begin{pmatrix} Q + \frac{\|\psi_s\|^2}{2} \cos 2\alpha & \frac{\|\psi_s\|^2}{2} \sin 2\alpha \\ \frac{\|\psi_s\|^2}{2} \sin 2\alpha & Q - \frac{\|\psi_s\|^2}{2} \cos 2\alpha \end{pmatrix}$$

where $Q = \sqrt{4A^2 + \|\psi_s\|^4 \pi^2}/2\pi$, $\alpha = \tan^{-1}(\psi_y(s)/\psi_x(s))$, $\tau_z > 0$ is a scaling parameter that controls the overall level of the covariance.

CONSIDERING CITY LANDMARKS

Consider each landmark as a constant exogenous promotion to the virus spread at their locations. Introduce an additional term to the conditional intensity function $\lambda(t, s)$

$$\lambda(t, s) = \lambda_0 + \sum_{l=1}^L \gamma_l g(s|s_l, \Sigma_l) + \sum_{t' < t} k(t, t', s, s') . \quad (11)$$

The second term represents the exogenous promotion at location s , L denotes the number of landmarks, and γ_l indicates the significance of landmark l . The influence of landmark l located at s_l is modeled by a Gaussian function $g(s|s_l, \Sigma_l)$ centered at location $s_l \in \mathcal{S}$ with covariance $\Sigma_l := \sigma_l^2 \mathbf{I}$

RESULTS: COVID-19 DATA IN CALI

Investigate model's explanatory power by evaluating the in-sample performance and visualize the estimated kernel-induced feature functions and their corresponding spatial kernel

Study the exogenous effects of the city landmarks

Compare the out-of-sample predictive performance of the proposed method with four baseline approaches

$\{\text{MAEQ}_q^{in}, \text{MAEQ}_q^{out}\}$ denote to the lower q -quantile of the mean absolute error (MAE) for the in-sample and out-of-sample estimation, respectively.

RESULTS: IN-SAMPLE PERFORMANCE FOR ONE-WEEK AHEAD CASES

Table: Performance of in-sample estimation. The numbers in the brackets are one standard deviation

Models	Log-likelihood($\times 10^4$)	MAE $Q_{0.25}^{\text{in}}$	MAE $Q_{0.5}^{\text{in}}$	MAE $Q_{0.75}^{\text{in}}$
Random	/	5.000	11.000	18.000
SIR	/	1.862	3.759	7.391
AR(3)	/	1.307	2.880	6.496
ETAS	4.868 (0.0058)	1.486	4.737	14.895
NSSTPP–Exo ($R=1$)	8.671 (0.0772)	0.834	3.145	7.922
NSSTPP–Exo ($R=2$)	9.138 (0.0886)	0.806	2.728	7.119
NSSTPP–Exo ($R=3$)	9.190 (0.0906)	0.853	2.613	7.000
NSSTPP ($R=3$)	9.331 (0.0937)	0.797	2.620	6.757

RESULTS: IN-SAMPLE PERFORMANCE FOR ONE-WEEK AHEAD CASES

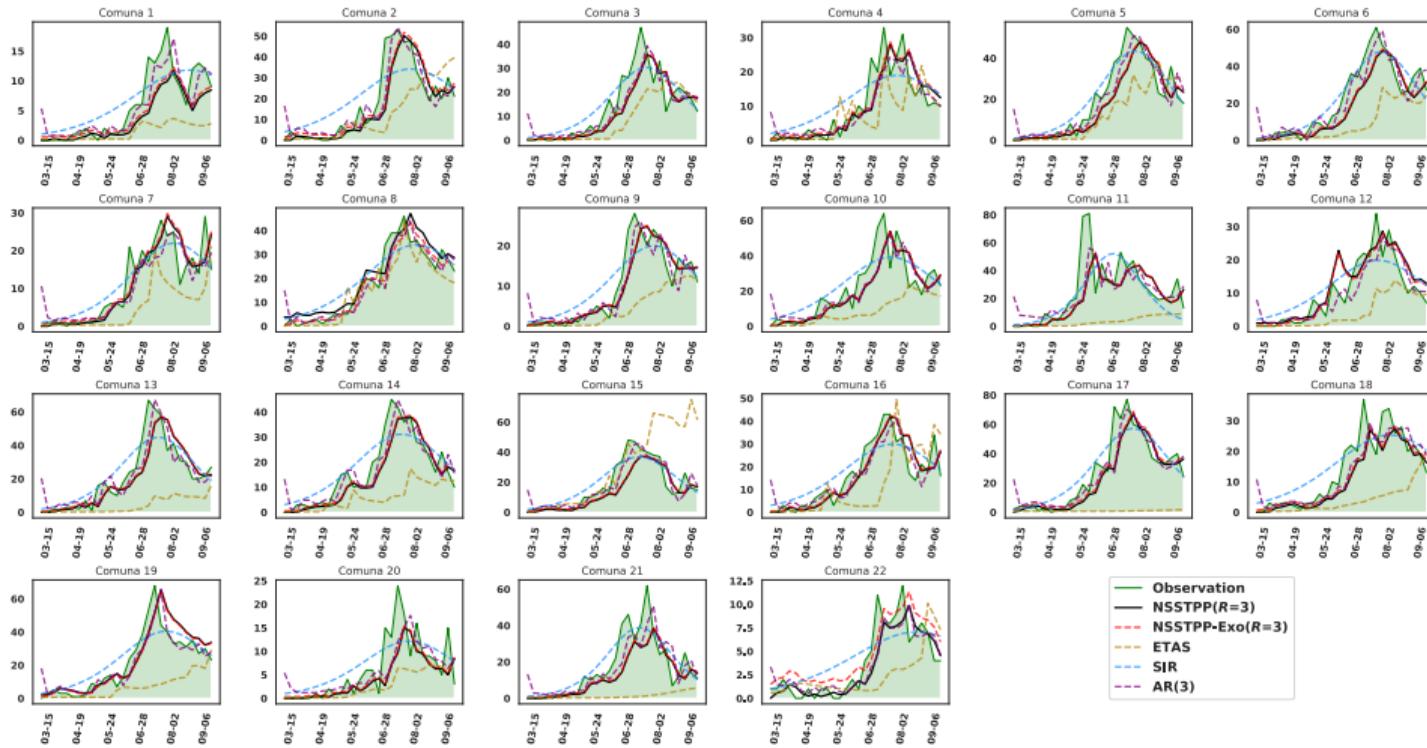


Figure: Comparison with baselines. Each line represents an in-sample estimation of one model. Green lines and shaded areas represent the ground truth. Black and red lines indicate NSSTPP and NSSTPP without exogenous promotion. Yellow, blue, and purple lines represent ETAS model, SIR model, and AR(3) model, respectively.

RESULTS: INTERPRETABLE SPATIAL KERNEL

Three learned spatial kernel-induced feature functions, which reveal the underlying spatio-temporal transmission dynamics. At any location s , $\kappa_s^{(r)}$ is a Gaussian kernel with a spatially varying covariance matrix represented by two focus points of its one standard deviation ellipse. The angle and length of each red line can be interpreted as the direction and strength of influence at the particular location

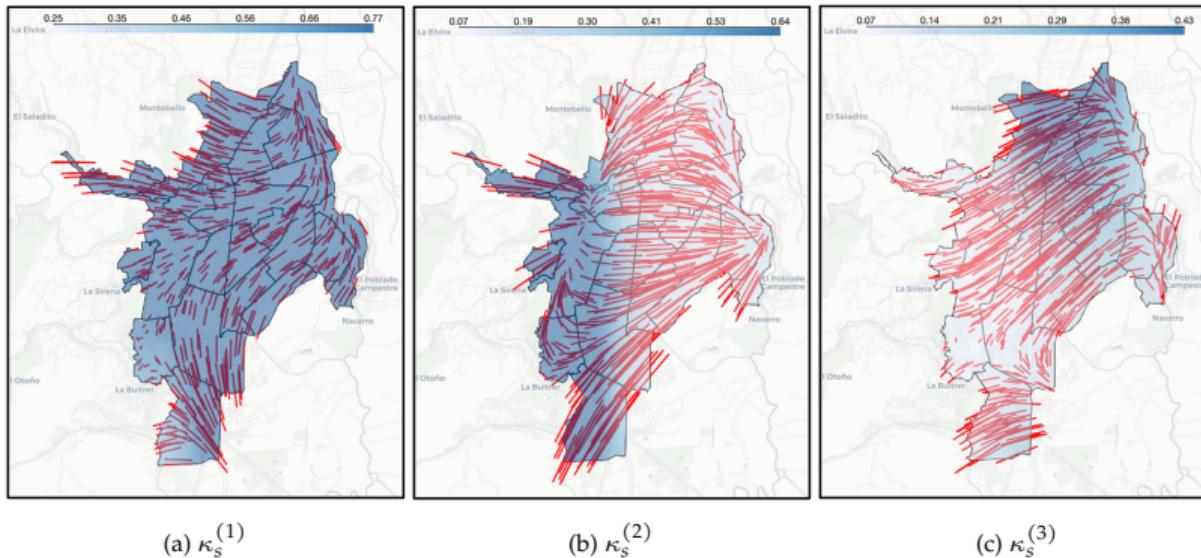
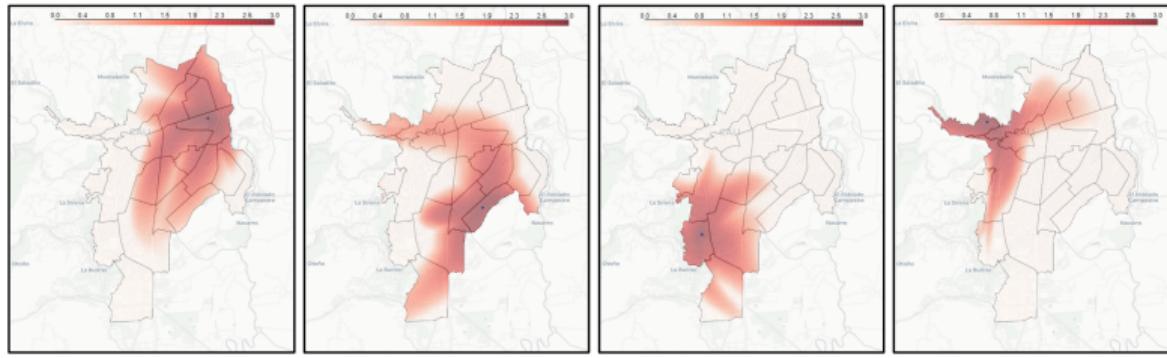


Figure: Learned kernel-induced feature functions. Red bars connect two focus points of location s . The blue shaded area shows the intensity of weight $w_s^{(r)}$ of each $\kappa_s^{(r)}$ over space. Darker colors mean larger weights

RESULTS: EVALUATION OF THE SPATIAL INFLUENCE



(a) Airport

(b) Comuna center 15

(c) Comuna center 18

(d) Comuna center 1

Figure: Evaluation of the spatial kernel $v(s, \cdot)$ with s fixed at four typical locations over space. These panels intuitively show the spatial influence of the regional hubs located in different parts of the city. The dots represent the fixed location. Color depth indicates the intensity of the kernel value. Darker color represents a higher kernel value.

RESULTS: EXOGENOUS EFFECTS OF LANDMARKS

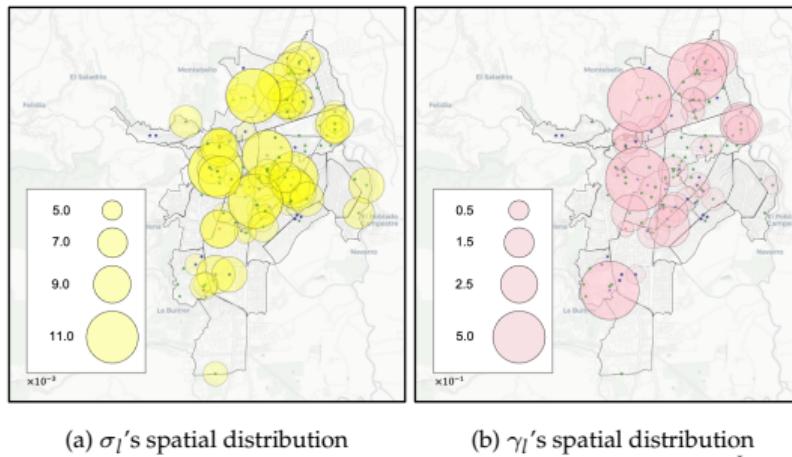


Figure: Estimated exogenous effects of landmarks in Cali. (a),(b) visualize the learned $\{\sigma_l\}_{l=1}^L$ and $\{\gamma_l\}_{l=1}^L$ on the map of Cali, respectively

RESULTS: PREDICTED CONDITIONAL INTENSITY FUNCTION, PREDICTIVE PERFORMANCE

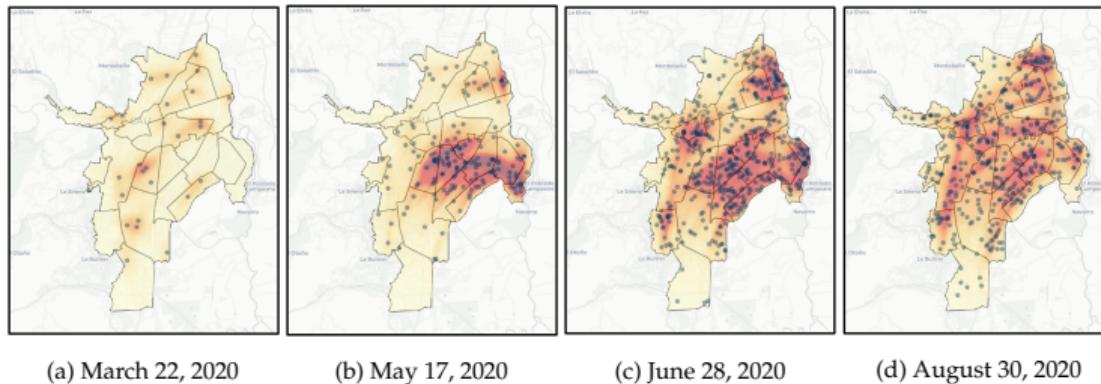


Figure: Predicted conditional intensity at four different weeks. The black dot represents an actual case. The color depth indicates the conditional intensity value. A darker color means a higher risk for citizens to be infected

4. Velocities for spatial growth models

OVERVIEW

Growth curve models have been widely studied in the literature. The intent for such models is to capture change in "size" per change in "time". Here, we view growth in a spatial context

For a point pattern setting, we have data in the form $\mathcal{Y} = \{(\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2), \dots, (\mathbf{s}_n, t_n)\}$, i.e., a point pattern of events at locations \mathbf{s}_i at time t_i , $i = 1, 2, \dots, n$. A space-time intensity driven by a growth curve model is used to explain the spatio-temporal point pattern. Examples include evolution of an epidemic or evolution of incidence of crimes

Idea is to bring the notion of velocity to spatial growth curve modeling. That is, all differentiable growth curve models have an implicit velocity, i.e., the rate of growth...with growth over space and time, we ask about the rate of growth in space relative to the change in time

Thus, assuming differentiability, there will be a temporal gradient at time t and a spatial gradient in any direction from \mathbf{s} . The ratio of the former to the latter will result in a distance/time, hence a velocity at location \mathbf{s} at time t in the given direction

OVERVIEW

The customary log-Gaussian Cox process for the space-time point pattern adds a Gaussian process to the log intensity of a nonhomogeneous Poisson process. For us, the space-time intensity $\lambda(\mathbf{s}, t)$ satisfies a SPDE over $D \times (0, T]$

$$d\lambda(\mathbf{s}, t) = r(\mathbf{s}, t)\lambda(\mathbf{s}, t)dt$$

, with $r(\mathbf{s}, t) = r(r(\mathbf{s}), t)$

This results in a $\lambda(\mathbf{s}, t)$ which is parametric in t and differentiable, with a function of a GP in \mathbf{s} for which we can obtain directional derivatives of process realizations

We have a PDE to which we add stochasticity. We need $\partial\lambda(\mathbf{s}, t)/\partial t$, and with $\mathbf{s} = (x, y)$, we also need $\partial\lambda(\mathbf{s}, t)/\partial x$ and $\partial\lambda(\mathbf{s}, t)/\partial y$

SOME GROWTH CURVE MODELS

Linear ODE: $dx(t) = (r_1x(t) + r_2)dt$. Here, r_1 provides the rate and r_2 allows for drift. **Solution:** $x(t) = e^{r_1 t}(x(0) + r_2/r_1) - r_2/r_1$. **Spatial analogue:** $d\lambda(\mathbf{s}, t) = (r(\mathbf{s})\lambda(\mathbf{s}, t) + \gamma)dt$

Logistic ODE: $dx(t) = rx(t)(1 - \frac{x(t)}{K})dt$, where K is a carrying capacity. **Solution:** $x(t) = \frac{x(0)Ke^{rt}}{K - x(0) + x(0)e^{rt}}$. **Spatial analogue:** $d\lambda(\mathbf{s}, t) = r(\mathbf{s})\lambda(\mathbf{s}, t)(1 - \frac{\lambda(\mathbf{s}, t)}{K})dt$

Weibull ODE: $dx(t) = (\frac{r}{t} - (r+1)t^r)x(t)dt$. This model extends the multiplicative model when $r(t) = r/t$. **Solution:** $x(t) = x(t_0)(\frac{t}{t_0})^r e^{-(t^{r+1} - t_0^{r+1})}$. **Spatial analogue:** $d\lambda(\mathbf{s}, t) = (\frac{r(\mathbf{s})}{t} - r(\mathbf{s})t^{r(\mathbf{s})})\lambda(\mathbf{s}, t)dt$

FORMALIZING VELOCITIES (I)

Develop velocities associated with a realization of a stochastic process, $\lambda(x, y, t)$ at $\mathbf{s} = (x, y)$ by assuming that $\lambda(x, y, t)$ is differentiable in each of its arguments

$\frac{\partial \lambda(\mathbf{s}, t)}{\partial x}$ is the instantaneous rate of change in intensity in direction x , and $\frac{\partial \lambda(\mathbf{s}, t)}{\partial t}$ is the instantaneous rate of change of intensity in time

The ratio $\frac{\frac{\partial \lambda(\mathbf{s}, t)}{\partial t}}{\frac{\partial \lambda(\mathbf{s}, t)}{\partial x}} \equiv \frac{\partial x}{\partial t}$ is the **velocity at location (\mathbf{s}, t) in direction x**

Let $\partial(\mathbf{s}, t)$ represent a product set neighborhood of (\mathbf{s}, t) , i.e., $\partial\mathbf{s} \times (t - \epsilon, t + \epsilon)$ with area $2\epsilon|\partial\mathbf{s}|$. For the expected number of points in $\partial(\mathbf{s}, t)$, written $N(\partial(\mathbf{s}, t))$, we have the approximation

$$P(N(\partial(\mathbf{s}, t)) = 1) \approx E(N(\partial(\mathbf{s}, t))) \approx \lambda(\mathbf{s}, t)2\epsilon|\partial\mathbf{s}|. \quad (13)$$

FORMALIZING VELOCITIES (II)

$$\frac{\frac{P(N(\partial(\mathbf{s}, t + \Delta t)) = 1) - P(N(\partial(\mathbf{s}, t)) = 1)}{\Delta t}}{\frac{P(N(\partial(\mathbf{s} + h\mathbf{u}), t) = 1) - P(N(\partial(\mathbf{s}, t)) = 1)}{h}} \quad (14)$$

where \mathbf{u} is a unit vector and h is a positive scalar

As $\Delta t \rightarrow 0$, the numerator is the instantaneous change in the probability of an event at (\mathbf{s}, t) in time t . As $h \rightarrow 0$, the denominator is the instantaneous change in the probability of an event at (\mathbf{s}, t) in direction \mathbf{u} . The ratio is the change in probability of an event per unit time over the change in probability of an event per unit distance in direction \mathbf{u} , i.e., a **velocity**

$$\frac{(\lambda(\mathbf{s}, t + \Delta t) - \lambda(\mathbf{s}, t))/\Delta t}{(\lambda(\mathbf{s} + h\mathbf{u}, t) - \lambda(\mathbf{s}, t))/h} \rightarrow \frac{\partial \lambda(\mathbf{s}, t)/\partial t}{D_{\mathbf{u}}\lambda(\mathbf{s}, t)} \quad (15)$$

where $D_{\mathbf{u}}\lambda(\mathbf{s}, t)$ is the directional derivative of the intensity at (\mathbf{s}, t) in direction \mathbf{u} . Then $D_{\mathbf{u}}\lambda(\mathbf{s}, t) = \mathbf{u}^T \nabla \lambda(\mathbf{s}, t)$ with $\nabla \lambda(\mathbf{s}, t) = (\partial \lambda(\mathbf{s}, t)/\partial x, \partial \lambda(\mathbf{s}, t)/\partial y)'$

$\frac{\partial \lambda(\mathbf{s}, t)/\partial t}{D_{\mathbf{u}}\lambda(\mathbf{s}, t)}$ is the **instantaneous relative change for an event in direction \mathbf{u} in units of distance per time**
 the direction of the **maximum change for an event over space** is $\mathbf{u}_{max} = \nabla \lambda(\mathbf{s}, t)/||\nabla \lambda(\mathbf{s}, t)||$, and the **magnitude** is $||\nabla \lambda(\mathbf{s}, t)||$

Claim: Direction of minimum velocity arises from the direction of maximum spatial gradient

MODEL SPECIFICATION, FITTING AND VELOCITIES (I)

Observed data $\mathcal{Y} = \{(\mathbf{s}_1, t_1), \dots, (\mathbf{s}_n, t_n)\}$, living in $D \times [0, T]$, and assume $d\lambda(\mathbf{s}, t) = g(r(\mathbf{s}), t)\lambda(\mathbf{s}, t)dt$, with $r(\mathbf{s})$ a spatial rate or growth process, and $\log r(\mathbf{s})$ a GP, and having a closed form solution

If Y_{D,t_0} denotes the initial point pattern with intensity $\lambda_0(\mathbf{s}) = \int_0^{t_0} \lambda(\mathbf{s}, \tau)d\tau$, we have $Y_{D,T} | \lambda(\mathbf{s}, t) \sim \text{Pois}(D \times T, \lambda(\mathbf{s}, t))$, $t \in (t_0, T]$ and $Y_{D,t_0} | \lambda_0 \sim \text{Pois}(D, \lambda_0)$

Assume the time interval begins at $t_0 > 0$, and we thus work in $[t_0, T = 1]$, with an stochastic intensity following a **Weibull model**

$$\lambda(\mathbf{s}, t) = \lambda_0(\mathbf{s}) \left(\frac{t}{t_0} \right)^{r(\mathbf{s})} e^{-\left(t^{r(\mathbf{s})+1} - t_0^{r(\mathbf{s})+1} \right)} \quad (16)$$

This $\lambda(\mathbf{s}, t)$ is the solution of the Weibull ODE given by $d\lambda(\mathbf{s}, t) = \left(\frac{r(\mathbf{s})}{t} - (r(\mathbf{s}) + 1)t^{r(\mathbf{s})} \right) \lambda(\mathbf{s}, t)dt$

Also, the spatial rate $r(\mathbf{s}) = \mu_r + \theta_r(\mathbf{s})$, $\theta_r(\mathbf{s}) \sim N(0, C_r)$ and $\lambda_0(\mathbf{s}) = \mu_\lambda + \theta_\lambda(\mathbf{s})$, $\theta_\lambda(\mathbf{s}) \sim N(0, C_\lambda)$

C_λ and C_r are **Matérn covariance functions** depending on parameters ($\nu = 3/2, \sigma, \phi$). $C_r(\mathbf{s}, \mathbf{s}') = \sigma_r^2(1 + \phi_r |\mathbf{s} - \mathbf{s}'|) e^{-\phi_r |\mathbf{s} - \mathbf{s}'|}$.

MODEL SPECIFICATION, FITTING AND VELOCITIES (II)

Inference will be performed through Markov chain Monte Carlo (MCMC) using slice sampling to approximate the marginal posterior distributions of each of the parameters of the model

Vague priors for the priori distributions of the parameters in the mean trend, $\pi(\mu) \stackrel{ind}{\sim} N(0, 10^8)$. Natural conjugate priors for the precision (inverse of variance) and the spatial range parameters, $\pi\left(\frac{1}{\sigma^2}\right), \pi(\phi) \stackrel{ind}{\sim} \text{Gamma}(2, 1)$

Slice sampling needs the full joint likelihood function

$$\mathcal{L}\left(\mu_r, \mu_\lambda, \frac{1}{\sigma_r^2}, \frac{1}{\sigma_\lambda^2}, \phi_r, \phi_\lambda | \mathcal{Y}\right) = \exp\left(-\int_{D \times T} \lambda(s, t) ds dt\right) \prod_{(s,t) \in \mathcal{Y}}^n \lambda(s, t)$$

PREDICTIVE PROCESS APPROXIMATION (I)

For velocities, we need the partial derivative with respect to time, and the gradient; in the Weibull case we have

$$\frac{\partial \lambda(\mathbf{s}, t)}{\partial t} = \left(\frac{r(\mathbf{s})}{t} - (r(\mathbf{s}) + 1)t^{r(\mathbf{s})} \right) \lambda(\mathbf{s}, t), \quad (17)$$

and

$$\nabla \lambda(\mathbf{s}, t) = \lambda(\mathbf{s}, t) \begin{pmatrix} \frac{\lambda_{0x}(\mathbf{s})}{\lambda_0(\mathbf{s})} + r_x(\mathbf{s}) \left(\ln \left(\frac{\frac{t_0^{r(\mathbf{s})+1}-1}{t_0^{r(\mathbf{s})+1}-1}}{\frac{t^{r(\mathbf{s})+1}-1}{t^{r(\mathbf{s})+1}-1}} \right) \right) \\ \frac{\lambda_{0y}(\mathbf{s})}{\lambda_0(\mathbf{s})} + r_y(\mathbf{s}) \left(\ln \left(\frac{\frac{t_0^{r(\mathbf{s})+1}-1}{t_0^{r(\mathbf{s})+1}-1}}{\frac{t^{r(\mathbf{s})+1}-1}{t^{r(\mathbf{s})+1}-1}} \right) \right) \end{pmatrix}, \quad (18)$$

BUT, we do not have analytical expressions for $\lambda_0(\mathbf{s})$ and $r(\mathbf{s})$, and resort these derivatives to using a predictive process approximation

PREDICTIVE PROCESS APPROXIMATION (II)

Recall

$$\begin{aligned} r(\mathbf{s}) &= \mu_r + \theta_r(\mathbf{s}), \theta_r(\mathbf{s}) \sim N(0, C_r), \\ \lambda_0(\mathbf{s}) &= \mu_\lambda + \theta_\lambda(\mathbf{s}), \theta_\lambda(\mathbf{s}) \sim N(0, C_\lambda), \end{aligned} \tag{19}$$

$$\text{with } \text{Cov}(\theta_k(x_i, y_i), \theta_k(x_j, y_j)) = \sigma_k^2(1 + \phi_k d_{ij}) \exp(-\phi_k d_{ij}), \quad k = r, \lambda.$$

Predictive process approximation for r . Let $\theta_r^* = [\theta_r(x_1^*, y_1^*), \dots, \theta_r(x_m^*, y_m^*)]$ be the spatial process evaluated in a lower dimension, where $(x_j^*, y_j^*), j = 1, \dots, m$ are the knot locations, and $m \ll n$

The predictive process is given by $\tilde{\theta}_r = E(\theta_r | \theta_r^*) = C_{\theta_r, \theta_r^*}^T \left(C_{\theta_r^*} \right)^{-1} \theta_r^*$

Using the predictive process, we can rewrite r and λ_0

$$\begin{aligned} r(\mathbf{s}) &= \mu_r + \tilde{\theta}_r(\mathbf{s}), \quad \tilde{\theta}_r(\mathbf{s}) \sim N(0, C_{\theta_r, \theta_r^*}^T \left(C_{\theta_r^*} \right)^{-1} C_{\theta_r, \theta_r^*}), \\ \lambda_0(\mathbf{s}) &= \mu_\lambda + \tilde{\theta}_\lambda(\mathbf{s}), \quad \tilde{\theta}_\lambda(\mathbf{s}) \sim N(0, C_{\theta_\lambda, \theta_\lambda^*}^T \left(C_{\theta_\lambda^*} \right)^{-1} C_{\theta_\lambda, \theta_\lambda^*}) \end{aligned}$$

CALI CASE

$15 \times 15 = 225$ grid of cells spanning the spatial region, and rescale the time variable to the interval $[0, 1]$. For the Weibull model we need an initial instant $t_0 > 0$, use $t_0 = \frac{1}{3}$, which is around the third week of April and represents a temporal instant before the large increase in the number of reported cases

Bayesian inference through MCMC by slice sampling with 50000 iterations, the first 25000 as a burn-in period and taking samples every 100 iterations to provide marginal posterior distributions of all parameters

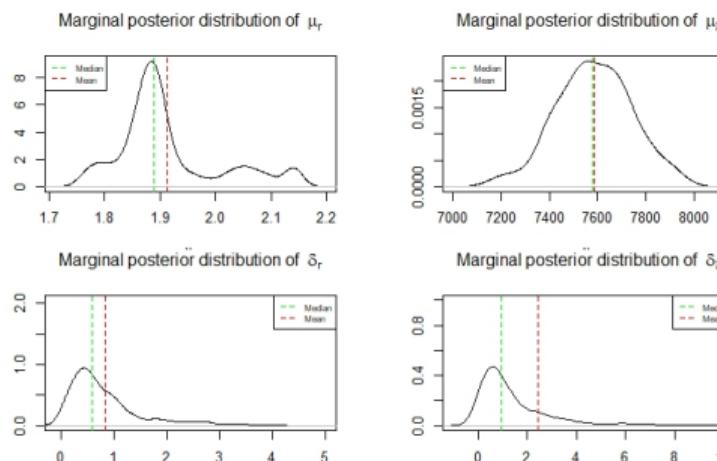


Figure: Marginal posterior distribution of the parameters

CALI CASE: ESTIMATED INTENSITIES

Estimated intensity $\lambda(\mathbf{s}, t)$ at four different time intervals corresponding to four selected days

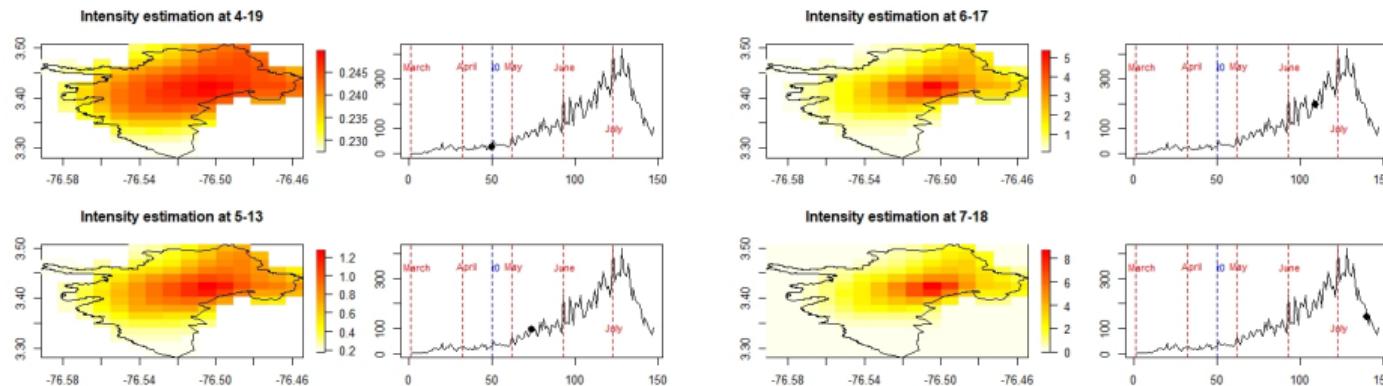


Figure: Intensity estimation at four time instants

CALI CASE: PARTIAL DERIVATIVES IN TIME

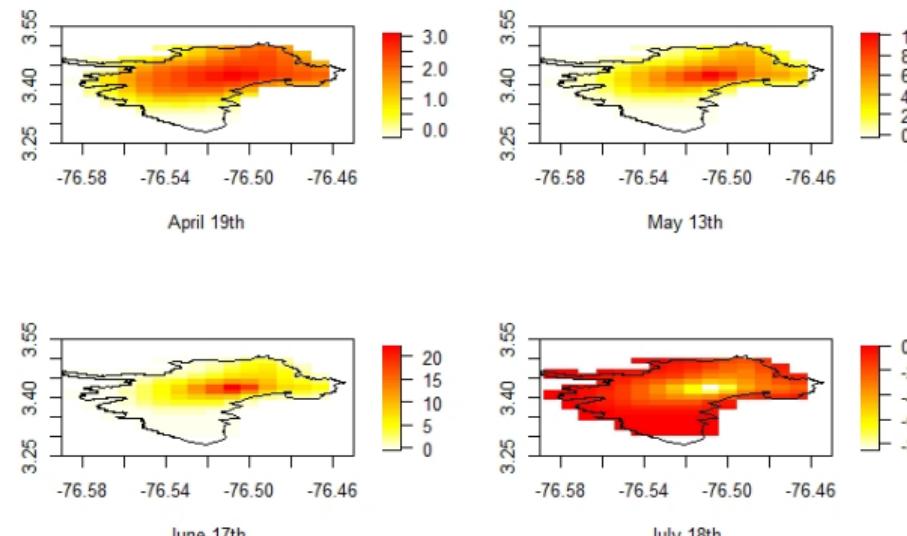


Figure: $\frac{\partial \lambda(s,t)}{\partial t}$ at the beginning of the four selected days

CALI CASE: PARTIAL DERIVATIVES IN X-DIRECTION

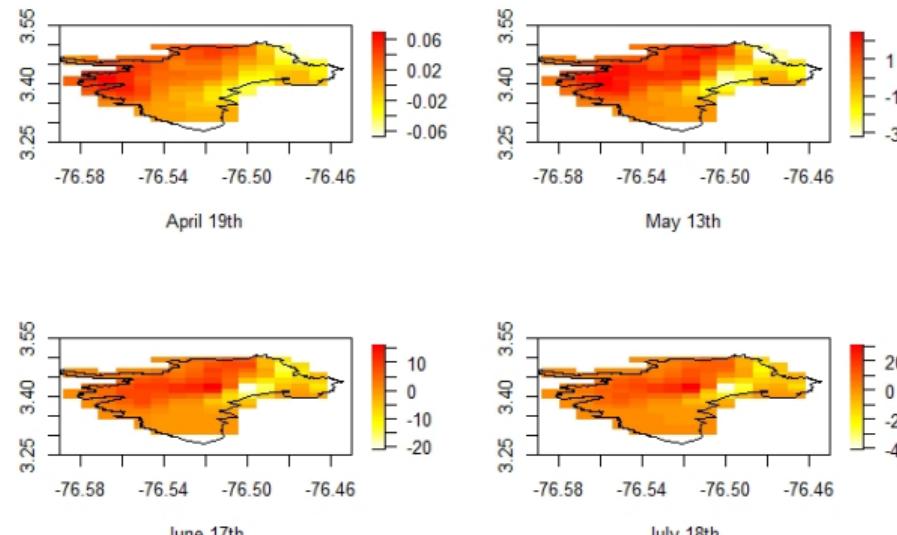


Figure: $\frac{\partial \lambda(s,t)}{\partial x}$ at the beginning of the four selected days

CALI CASE: : PARTIAL DERIVATIVES IN Y-DIRECTION

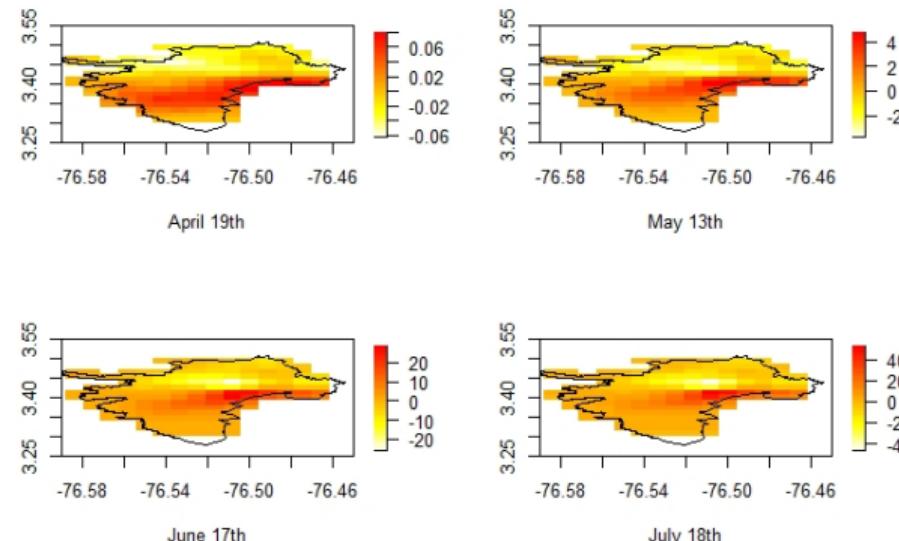


Figure: $\frac{\partial \lambda(s,t)}{\partial y}$ at the beginning of the four selected days

CALI CASE: VELOCITIES

Posterior mean of the minimum velocity surface at the four selected days. The minimum velocity and corresponding arrows show the slowest speed of change in the intensity and direction in which this speed is obtained

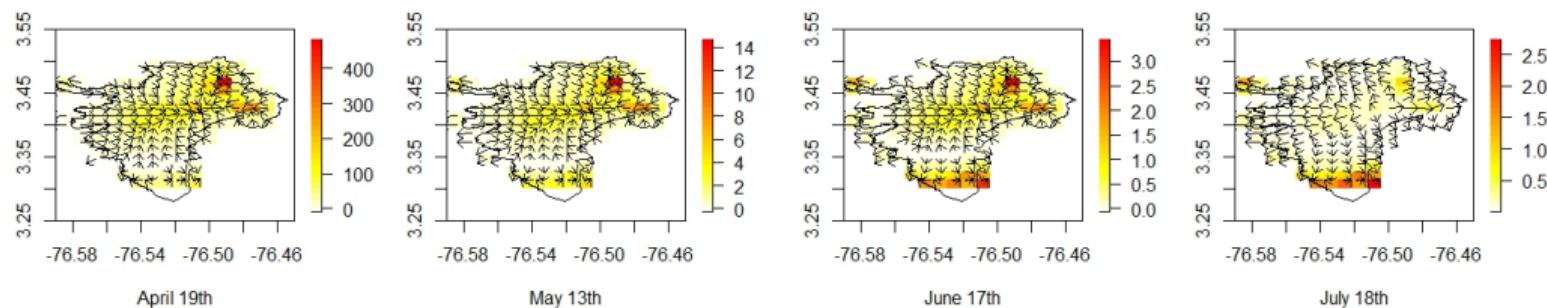


Figure: Posterior mean of minimum velocity surface at the four selected days

CALI CASE: DISTRIBUTION OF DIRECTIONS

Distribution function that indicates the direction in which we obtain the minimum velocity. Directions of minimum velocity point directly to the focus in each of the cells as well as the point estimate

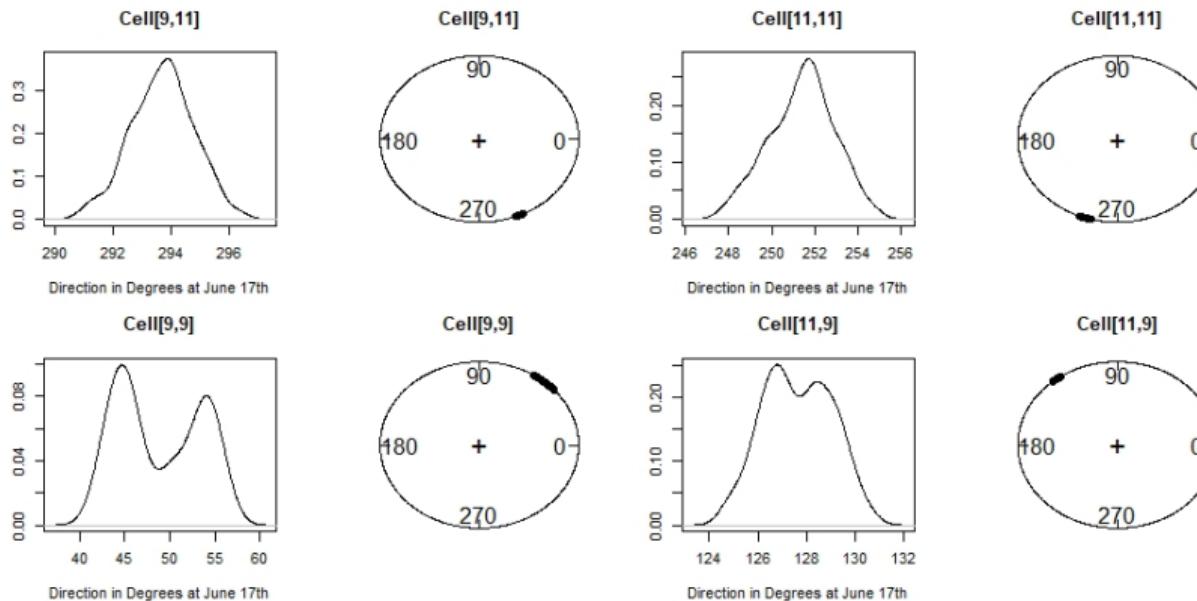


Figure: Distribution of minimum velocity directions in the four cells around the focus on June 17

CONTENTS-Crime

INTRODUCTION

DETECTING FOCUSES AND GENERATORS OF CRIMINALITY

PROBABILISTIC MODELS FOR CRIMINAL PATTERNS

PREDICTION AND CLASSIFICATION USING ML TECHNIQUES

REDUCING DIMENSIONALITY: BARYCENTERS

STOCHASTIC INTEGRO-DIFFERENTIAL EQUATIONS

MODELING ORIGIN-DESTINATION POINT PATTERNS

Introduction and Motivation

INTRODUCTION

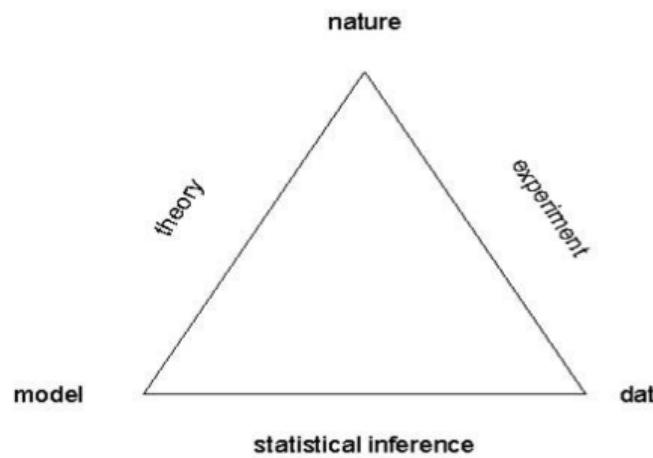
- ▶ A wealth of contributions and institutions in the context of spatial and spatio-temporal statistical analysis of crime data
- ▶ The **UCL Jill Dando Institute of Security and Crime Science** is the first Institute in the world devoted to Crime Science.
- ▶ Research is concentrated on new ways to cut crime and increase security, drawing upon UCL's vast experience in related disciplines, including architecture, economics, engineering, geography, medicine, psychology, **statistics** and town planning.
- ▶ **Smart, effective, and proactive policing** is clearly preferable to **simply reacting** to criminal acts.

INTRODUCTION

Data-Statistics-Police

INTRODUCTION

Analyse problems, not data



A statistical model is:

- a device to answer a question
- a bridge between scientific theory and empirical evidence
- a framework to enable principled inference in the presence of uncertainty

Detecting focuses and generators of criminality

- ▶ A multiplicative model including non-linear effects for modelling crime risk in relation to distances to multiple facilities
- ▶ The model is defined as follows

$$Y_i \sim \text{NB}(\mu_i, \psi)$$

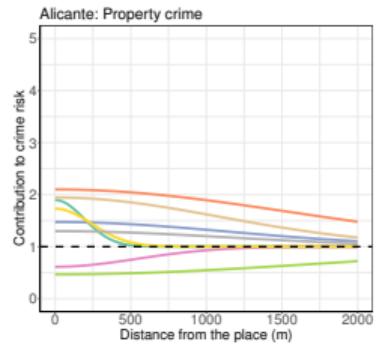
$$\log(\mu_i) = \log(E_i) + \sum_{k=1}^q \beta_k X_{ik} + \sum_{j=1}^P \log(f(d_{ij}; \alpha_j, \beta_j)) + u_i + \phi_i \quad (1)$$

- ▶ Y_i : number of crime events located in cell i , E_i : population living in cell i (exposure at cell i), X_{ik} : k th socio-demographic covariate on cell i , β_k : effect that this covariate produces on crime counts, d_{ij} : distance between the centroid of cell i and the place of type j , u_i : unstructured Gaussian random effect, ϕ_i : structured effect (with a conditional autoregressive structure) accounting for the spatial dependence between the cells forming the grid
- ▶

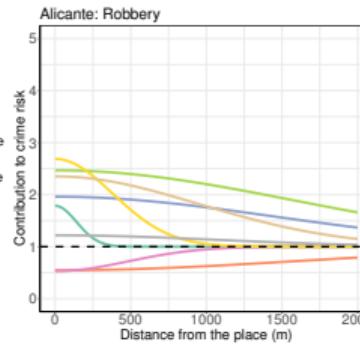
$$f(d_{ij}; \alpha_j, \beta_j) = 1 + \alpha_j \exp(-(d_{ij}/\beta_j)^2)$$

where α_j represents an increase ($\alpha_j > 1$) or decrease ($\alpha_j < 1$) in crime risk that is produced by places of type j . This effect dissipates according to parameter β_j .

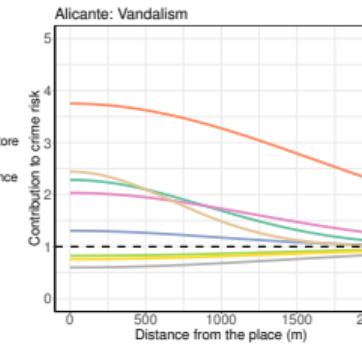
RISK FUNCTIONS



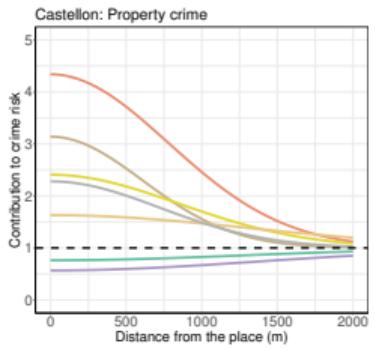
(a)



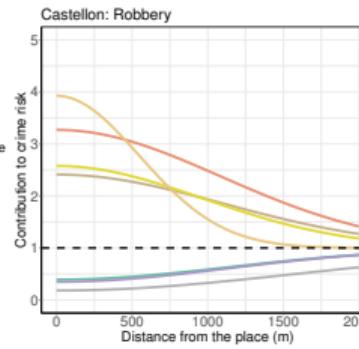
(b)



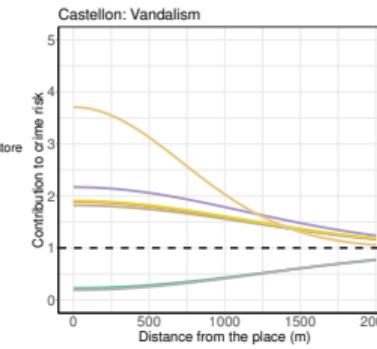
(c)



(d)



(e)



(f)

RISK FUNCTIONS (VALENCIA)

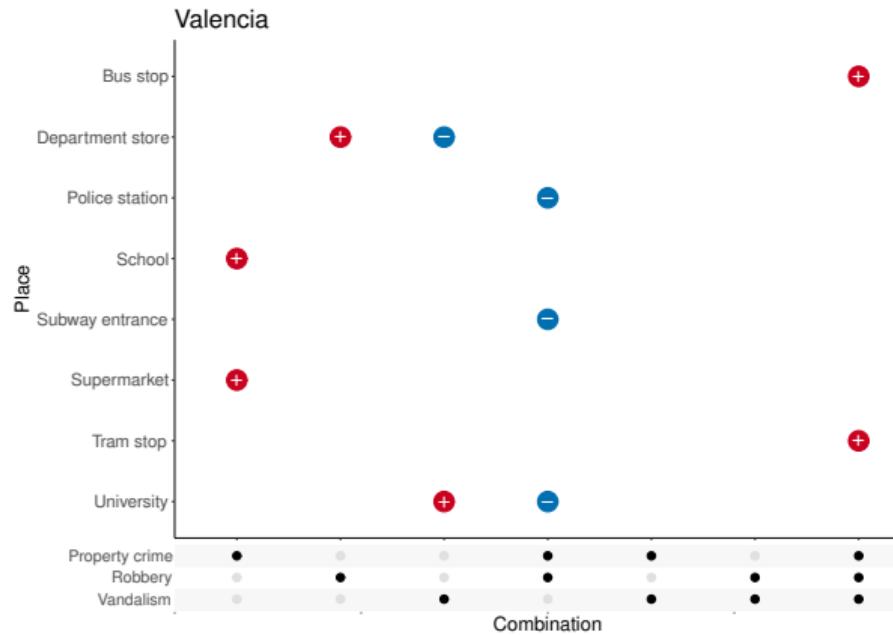


Figure: Results for Valencia in terms of the estimated α parameters

HIGH-RISK FUNCTIONS (VALENCIA)

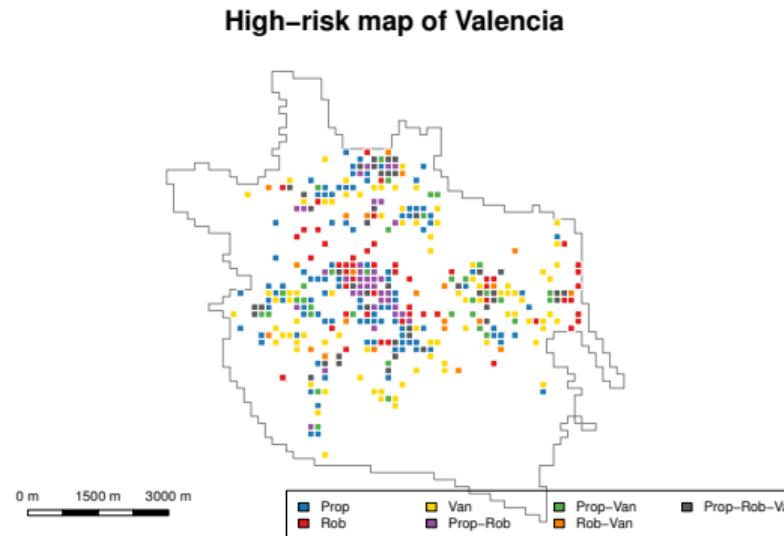


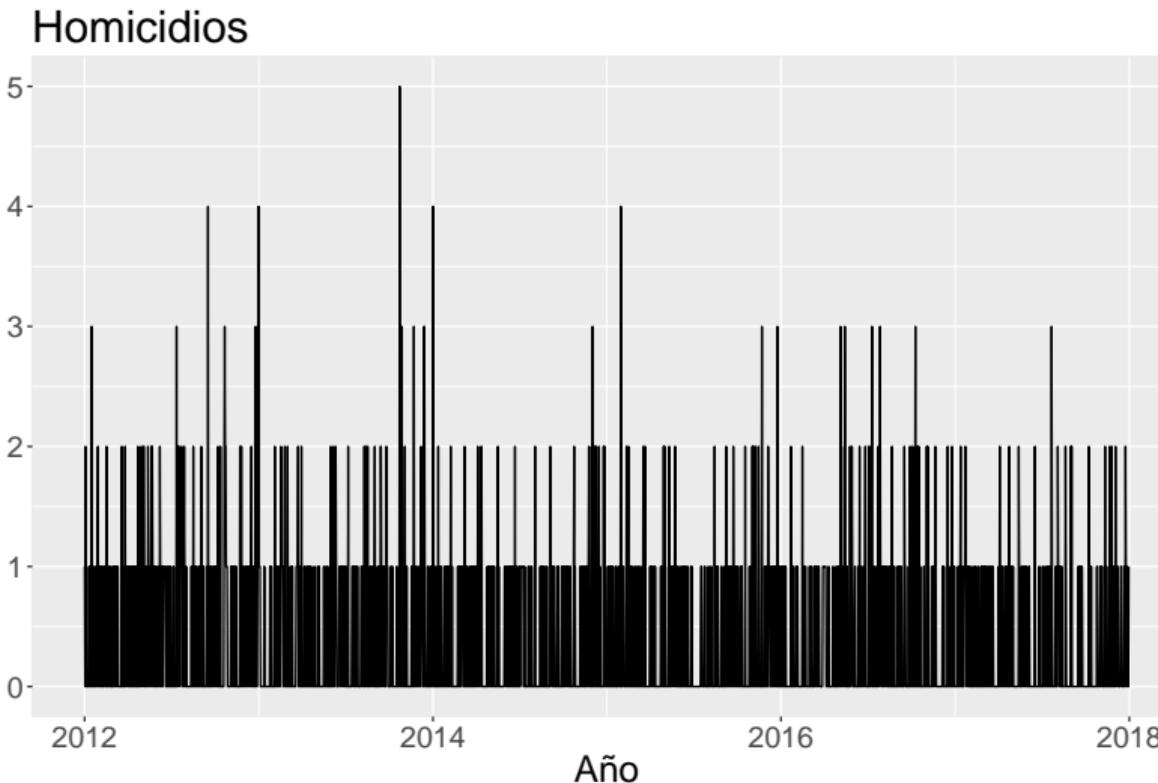
Figure: High-risk cells of Valencia in terms of property crime (Prop), robbery (Rob), and vandalism (Van). The colour of each cell indicates the crime or crimes that present a higher risk in the cell

Probabilistic models for criminal patterns

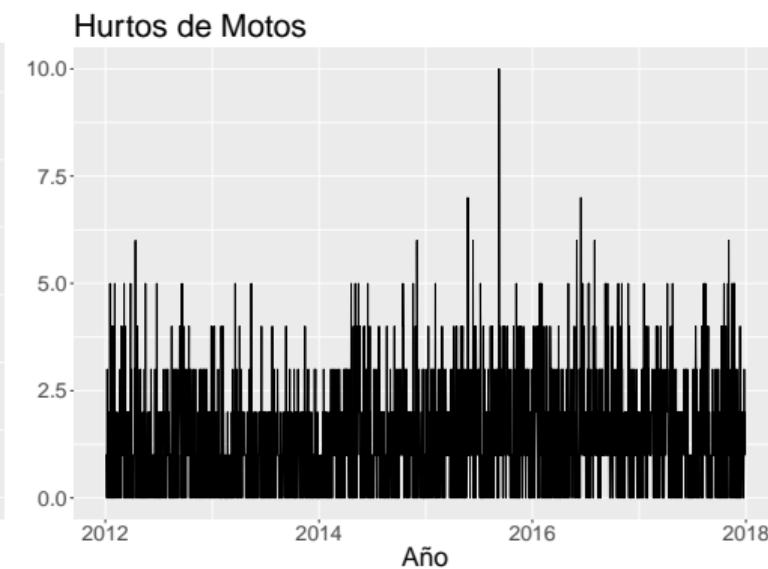
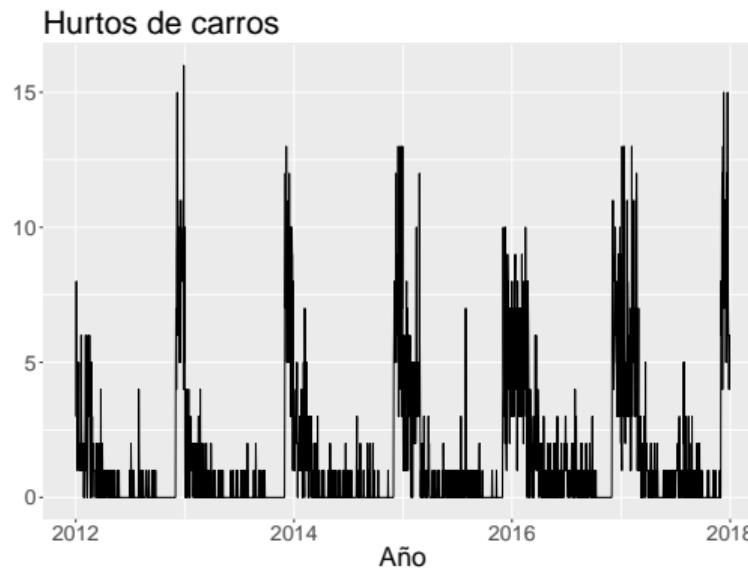
Crime data in Bogota (Colombia)

TIME SERIES FOR HOMICIDES

TIME SERIES OF DAILY CRIME COUNTS BY CATEGORY

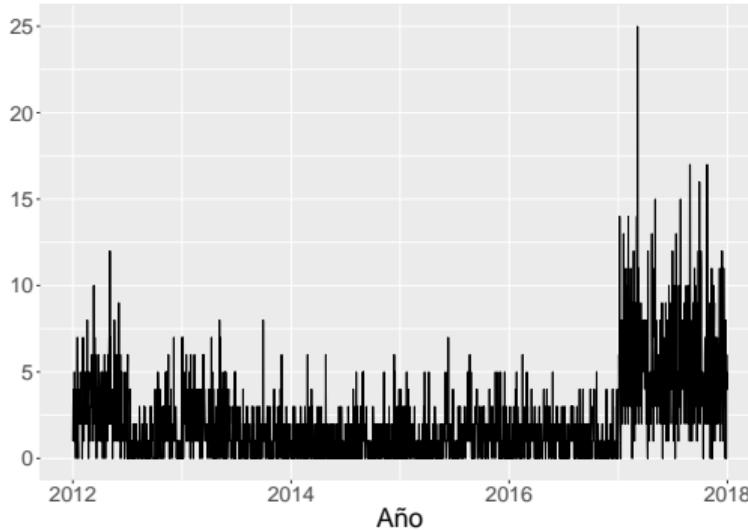


TIME SERIES FOR CAR AND MOTO ROBBERIES

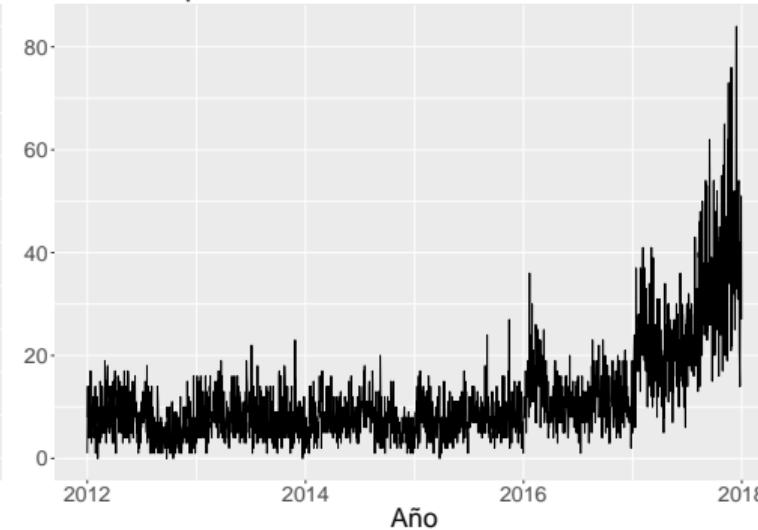


TIME SERIES FOR COMMERCIAL AND PERSONAL ROBBERIES

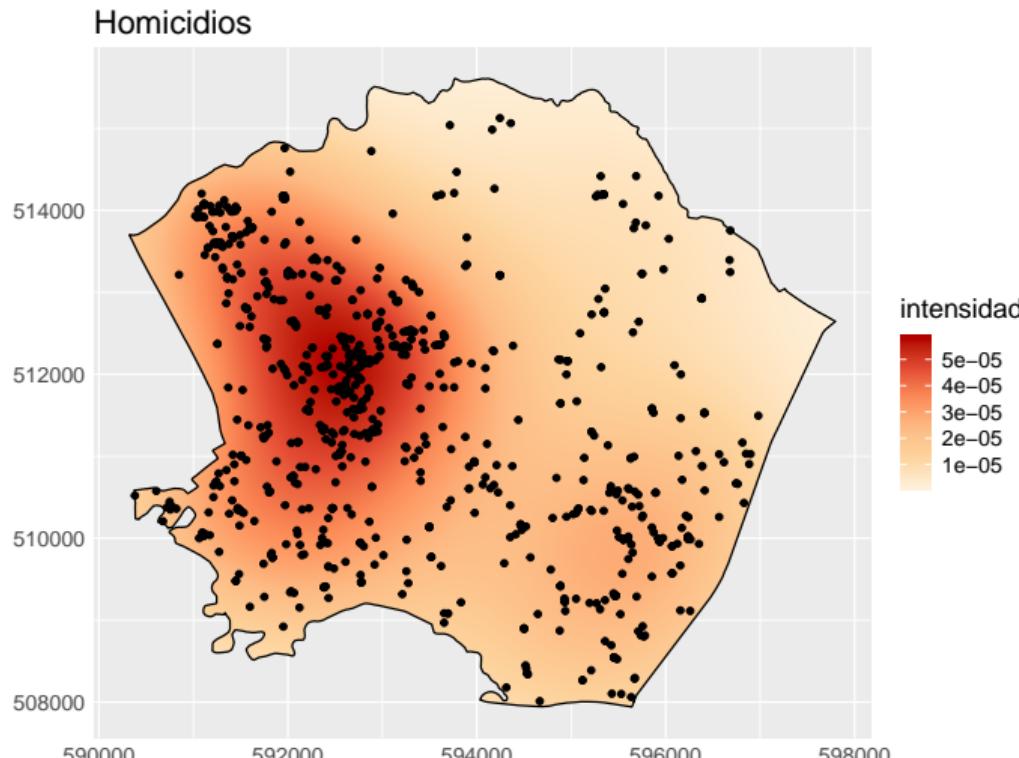
Hurtos a comercios



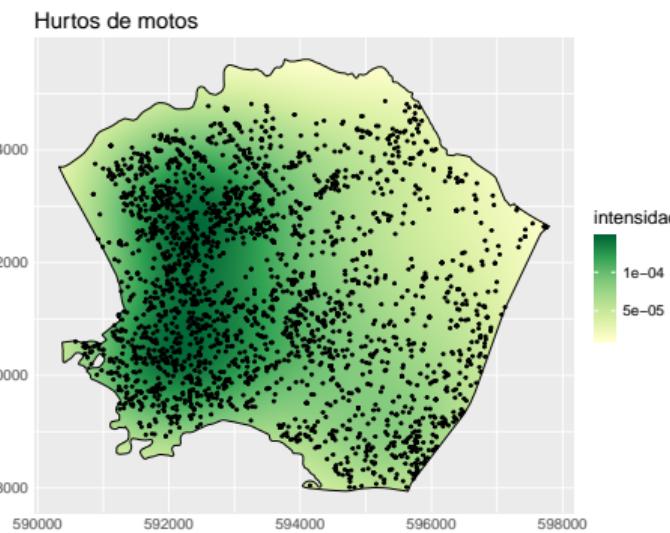
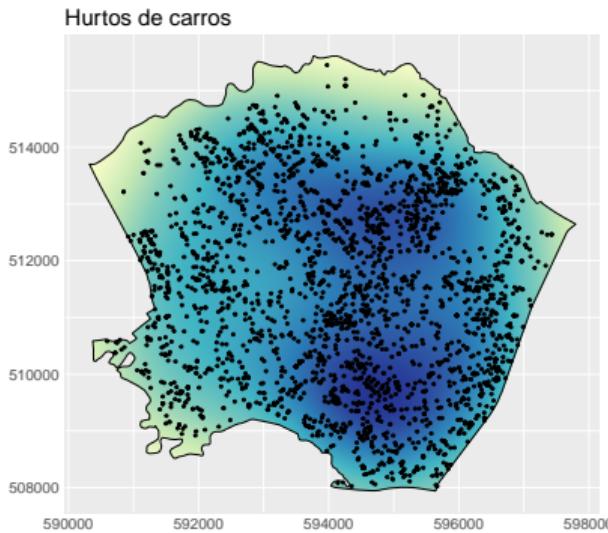
Hurtos a personas



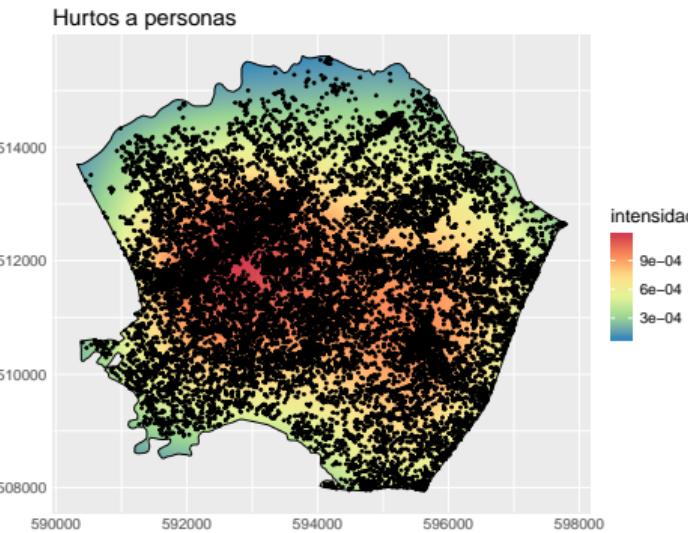
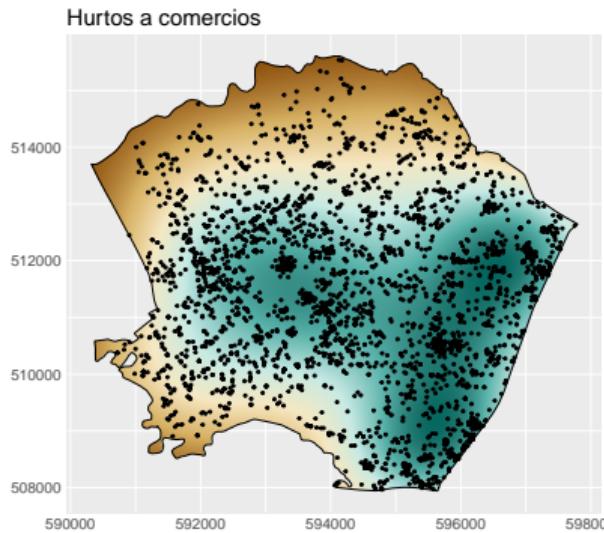
SPATIAL DENSITIES: HOMICIDES



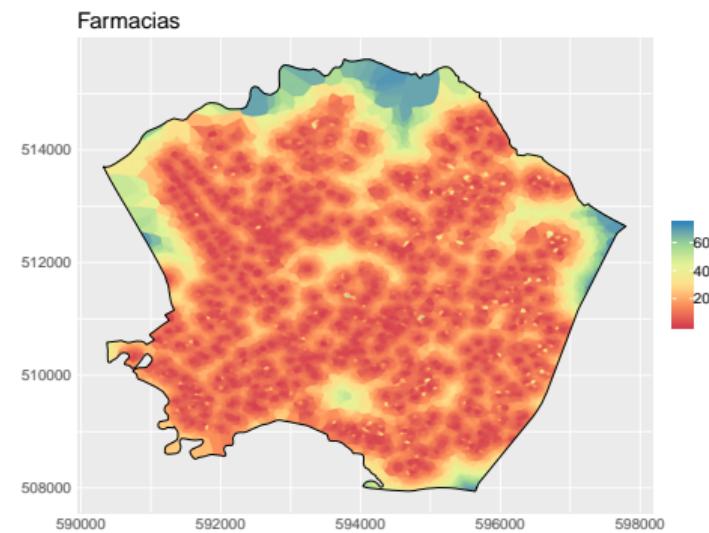
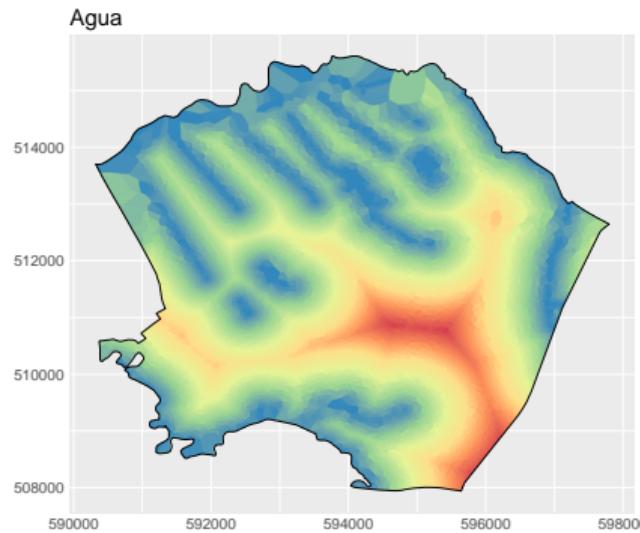
SPATIAL DENSITIES: CAR AND MOTO ROBBERIES



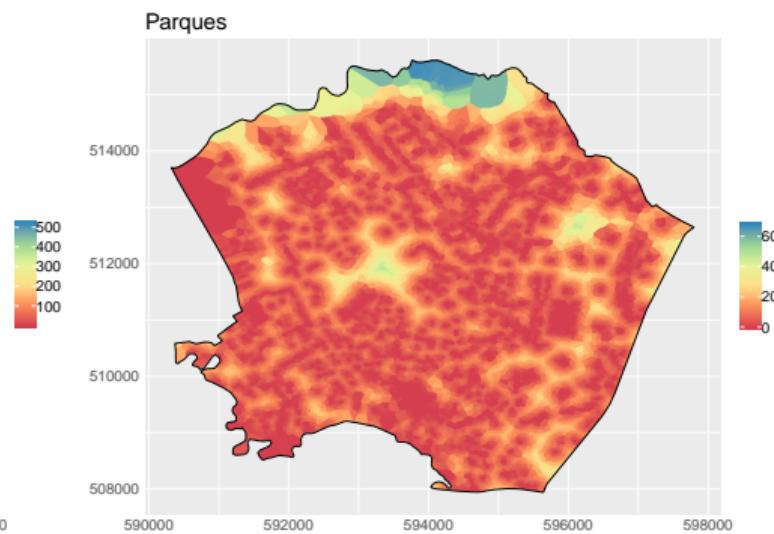
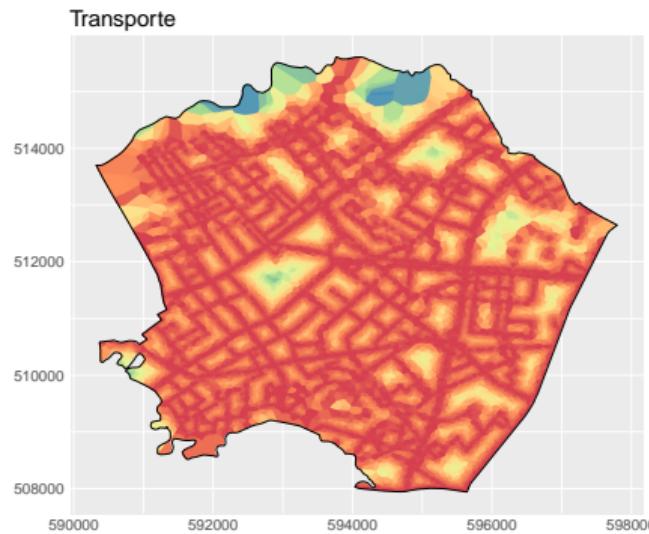
SPATIAL DENSITIES: COMMERCIAL AND PERSONAL ROBBERIES



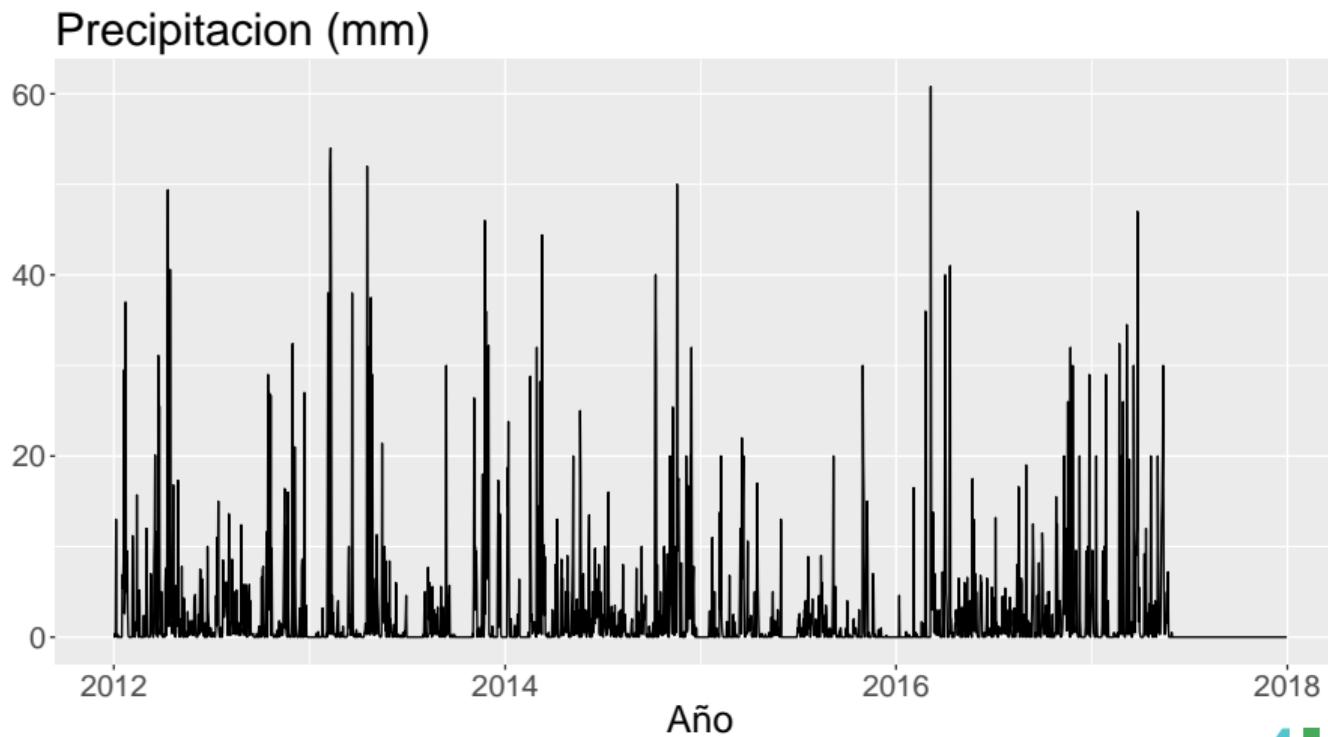
SPATIAL COVARIATES



SPATIAL COVARIATES



TEMPORAL COVARIATE



Spatio-temporal point process models

COX POINT PROCESSES AND LGCP

- ▶ **Cox process:** cases occur at spatio-temporal positions $(x, t) \in W \times T$ according to an inhomogeneous spatio-temporal Poisson process with a stochastic intensity $\Lambda(x, t)$.
- ▶ If $\Lambda(x, t) = \exp(S(x, t))$, the exponential of a Gaussian random field S , we have a **log-Gaussian Cox process**.
- ▶ Conditional on $R(x, t) = \exp(S(x, t))$, the number of events Y_{iT_j} within the spatio-temporal region $W_i \times T_j$ has a Poisson distribution

$$Y_{iT_j} | R(\cdot) \sim \text{Poisson}\left(\int_{T_j} \int_{A_i} \Lambda(x, t) dx dt \right)$$

A MODEL FOR THE SPATIO-TEMPORAL INTENSITY (I)

- ▶ The spatio-temporal intensity function can be given by the product between a deterministic spatio-temporal variation and a log-Gaussian stochastic process. For the deterministic part the spatial and temporal effects operate independently. The spatial intensity integrates to 1 over the study-region W .
- ▶

$$\begin{aligned}\Lambda(x, t) &= \mu(x, t)R(x, t) \\ &= \mu(x, t)\exp(S(x, t)) \\ &= \lambda(x)\mu(t)\exp(S(x, t))\end{aligned}$$

- ▶ The expectation of $\Lambda(x, t)$ is $\lambda(x)\mu(t)$. The function S is a second order stationary Gaussian process, continuous in both space and time, such that $\mathbb{E}(R(x, t)) = \mathbb{E}(\exp S(x, t)) = 1$, and $\mathbb{E}(S(x, t)) = -\frac{\sigma^2}{2}$, and $\text{Var}(S(x, t)) = \sigma^2$.

A MODEL FOR THE SPATIO-TEMPORAL INTENSITY (II)

- ▶ Conditional on $R(\mathbf{x}, t) = \exp(S(\mathbf{x}, t))$, the number of events within the spatio-temporal region $W_i \times T_j$ has a Poisson distribution

$$\begin{aligned} Y_{iT_j}|R(\cdot) &\sim \text{Poisson}\left(\int_{T_j} \int_{A_i} \Lambda(\mathbf{x}, t) dx dt\right) \\ &\sim \text{Poisson}\left(\int_{T_j} \int_{A_i} \mu(\mathbf{x}, t) R(\mathbf{x}, t) dx dt\right) \\ &\sim \text{Poisson}\left(\int_{T_j} \int_{A_i} \mu(t) \lambda(\mathbf{x}) R(\mathbf{x}, t) dx dt\right) \\ &\sim \text{Poisson}\left(\int_{T_j} \int_{A_i} \mu(t) \lambda_i R(\mathbf{x}, t) dx dt\right) \\ &\sim \text{Poisson}\left(\lambda_i \int_{T_j} \int_{A_i} \mu(t) R(\mathbf{x}, t) dx dt\right) \end{aligned}$$

FIRST MOMENT FOR THE UNCONDITIONAL COUNTS

► T continuous

$$\begin{aligned}
 \mathbb{E}(Y_{iT_j}) &= \mathbb{E}_R\{\mathbb{E}(Y_{iT_j} | R(\mathbf{x}, t))\} \\
 &= \mathbb{E}_R\left\{\lambda_i \int_{T_j} \int_{A_i} \mu(t) R(\mathbf{x}, t) d\mathbf{x} dt\right\} \\
 &= \lambda_i \int_{T_j} \int_{A_i} \mu(t) \mathbb{E}_R\{R(\mathbf{x}, t)\} d\mathbf{x} dt \\
 &= \lambda_i |A_i| \int_{T_j} \mu(t) dt
 \end{aligned}$$

► T discrete

$$\begin{aligned}
 \mathbb{E}(Y_{it_n}) &= \mu(t_n) \lambda_i \int_{A_i} \mathbb{E}_R\{R(\mathbf{x}, t_n)\} d\mathbf{x} \\
 &= \mu(t_n) \lambda_i \int_{A_i} 1 d\mathbf{x} \\
 &= \mu(t_n) \lambda_i |A_i| = \mu(t_n) p_i
 \end{aligned}$$

SECOND MOMENT FOR THE UNCONDITIONAL COUNTS

► T continuous

$$\begin{aligned}
 \text{Cov}(Y_{i,T_n}, Y_{j,T_{n-v}}) &= \\
 &= \text{Cov}_R(\mathbb{E}(Y_{i,T_n} | R), \mathbb{E}(Y_{j,T_{n-v}} | R)) + \mathbb{E}_R(\text{Cov}(Y_{iT_n}, Y_{j,T_{n-v}} | R)) \\
 &= \text{Cov}_R\left(\lambda_i \int_{T_n} \int_{A_i} \mu(t)R(x,t)dxdt, \lambda_j \int_{T_{n-v}} \int_{A_j} \mu(t)R(x,t)dxdt\right) \\
 &\quad + \mathbf{1}(v=0)\mathbf{1}(i=j)\lambda_i | A_i | \int_{T_j} \mu(t)dt \\
 &= \lambda_i \lambda_j \int_{T_n, A_i, T_{n-v}, A_j} \mu(h_1)\mu(h_2)(\exp(\gamma(\|l_1 - l_2\|, |h_1 - h_2|)) - 1) dl_1 dl_2 dh_1 dh_2 \\
 &\quad + \mathbf{1}(v=0)\mathbf{1}(i=j)\lambda_i | A_i | \int_{T_j} \mu(t)dt \\
 &= \lambda_i \lambda_j \left\{ \int_{T_n, A_i, T_{n-v}, A_j} \mu(h_1)\mu(h_2)\exp(\gamma(\|l_1 - l_2\|, |h_1 - h_2|)) dl_1 dl_2 dh_1 dh_2 \right. \\
 &\quad \left. - |A_i||A_j| \int_{T_n, T_{n-v}} \mu(h_1)\mu(h_2)dh_1 dh_2 \right\} + \mathbf{1}(v=0)\mathbf{1}(i=j)\lambda_i | A_i | \int_{T_j} \mu(t)dt
 \end{aligned}$$

MOMENTS FOR Y_{t_n} - THE TEMPORAL COMPONENT

- Considering discrete time, $Y_{t_n} = \sum_{i=1}^M Y_{it_n}$ denotes the number of the point in the whole region at the time t_n . Then

$$\begin{aligned} Y_{t_n} | R(\cdot) &\sim \text{Poisson}\left(\mu(t_n) \int_A \lambda(x) R(x, t_n) dx\right) \\ &\sim \text{Poisson}\left(\mu(t_n) \sum_{i=1}^M \lambda_i \int_{A_i} R(x, t_n) dx\right) \end{aligned}$$

- The covariance between number of events in two different time periods is

$$\begin{aligned} \text{Cov}(Y_{t_n}, Y_{t_{n-v}}) &= \text{Cov}\left(\sum_{i=1}^M Y_{it_n}, \sum_{j=1}^M Y_{jt_{n-v}}\right) = \sum_{i=1}^M \sum_{j=1}^M \text{Cov}(Y_{it_n}, Y_{jt_{n-v}}) \\ &= \mu(t_n)\mu(t_{n-v}) \left\{ \sum_{i=1}^M \sum_{j=1}^M \lambda_i \lambda_j \exp(\gamma(\|c_i - c_j\|, v)) | A_i || A_j | - 1 \right\} + \mathbf{1}(v=0)\mu(t_n) \end{aligned}$$

ON THE STOCHASTIC SPATIO-TEMPORAL PROCESS

- ▶ Recall that $\mathbb{E}(R(x, t)) = \mathbb{E}(\exp S(x, t)) = 1$, and $\mathbb{E}(S(x, t)) = \frac{-\sigma^2}{2}$, and $\text{Var}(S(x, t)) = \sigma^2$.
- ▶ The correlation function of S is $\rho(u, v)$ such that $\text{Cor}(S(x, t), S(x', t')) = \rho(\|x - x'\|, |t - t'|; \theta)$.
- ▶ Within this modeling framework, the surveillance system operates by calculating and reporting predictive probabilities of the form $P(R(x, t) > c)$ given data up to and including t , where c is a specified threshold value that, in the context of a specific application, might prompt corrective action.

ESTIMATION OF PARAMETERS OF $R(x, t)$

- ▶ Assume that $S(x, t)$ is an Ornstein-Uhlenbeck process, which constitutes a flexible class of continuous space-time Gaussian processes. Then

$$S(\cdot, t) \sim N[\xi(k)S(\cdot, t_2) + (1 - \xi(k))\mu, (1 - \xi(k)^2)\Sigma]$$

where $\xi(k) = \exp(-\theta k)$

- Forecast beyond the last observation at time T , forecast distribution of $S(\cdot, t_2 + k)$ given the observed data. Then we need to forecast the Poisson intensity

$$A\lambda(x)\mu(t_2+k)\exp(S(x,t_2+k))$$



$$\begin{aligned} E[S(\cdot, t_2 + k) | X_{t_1:t_2}] &= \xi(k)E[(\cdot, t_2) | X_{t_1:t_2}] + (1 - \xi(k))\mu \\ \text{Var}[S(\cdot, t_2 + k) | X_{t_1:t_2}] &= \xi(k)^2 \text{Var}[(\cdot, t_2) | X_{t_1:t_2}] + (1 - \xi(k)^2)\Sigma \end{aligned}$$

- Use MCMC+MALA algorithms

Statistical modelling results (Bogota)

TEMPORAL VARIATION $\mu(t)$

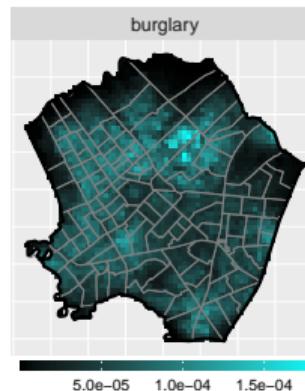
Semi-parametric

- ▶ $\log\{\mu(t)\} = \delta_{d(t)} + f(t)$
- ▶ $\log\{\mu(t)\} = \delta_{d(t)} + \alpha_1 \cos(\omega t) + \beta_1 \sin(\omega t) + \alpha_2 \cos(2\omega t) + \beta_2 \sin(2\omega t) + f(t)$

Parametric

- ▶ $\log\{\mu(t)\} = \delta_{d(t)} + \alpha_1 \cos(\omega t) + \beta_1 \sin(\omega t) + \alpha_2 \cos(2\omega t) + \beta_2 \sin(2\omega t) + \epsilon_1 t + \epsilon_2 t^2.$

SPATIAL INTENSITIES



TEMPORAL INTENSITIES

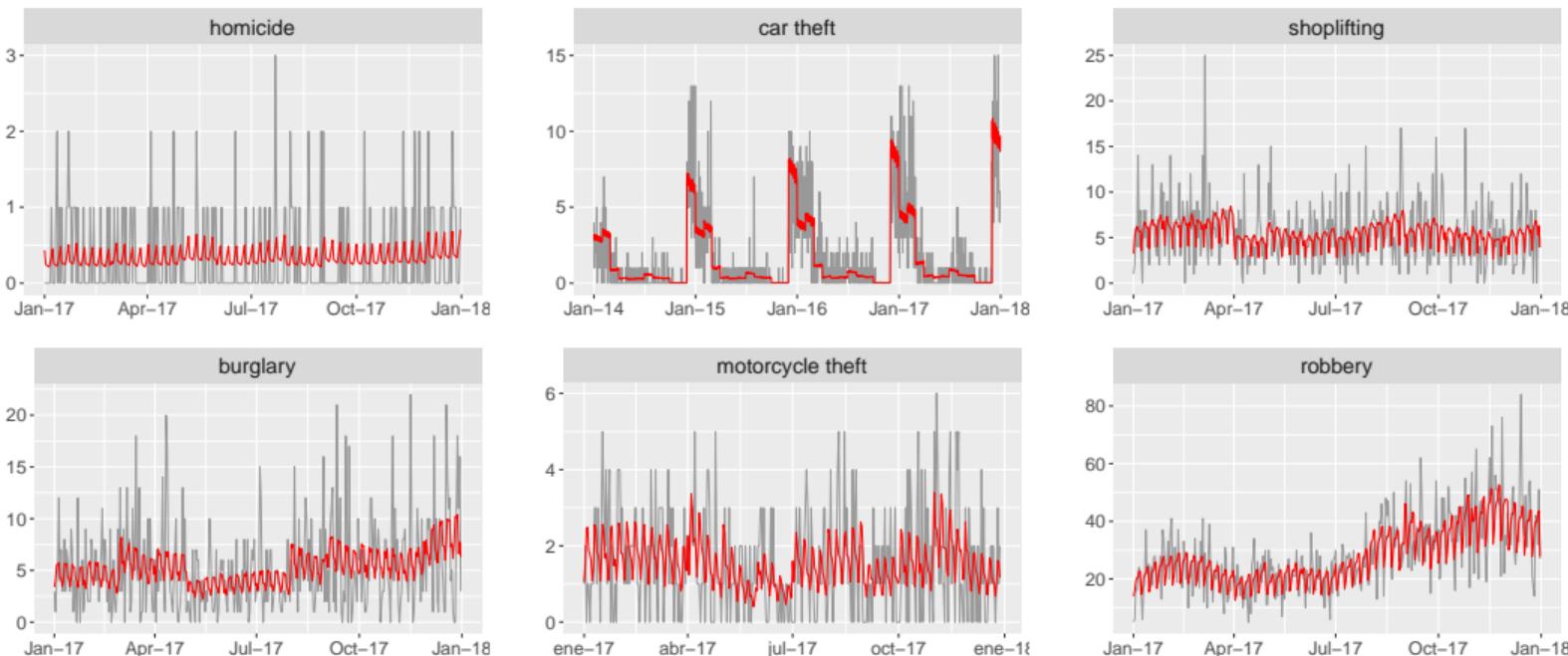
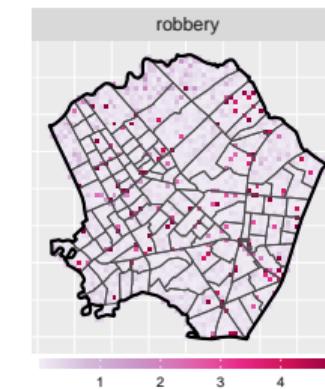
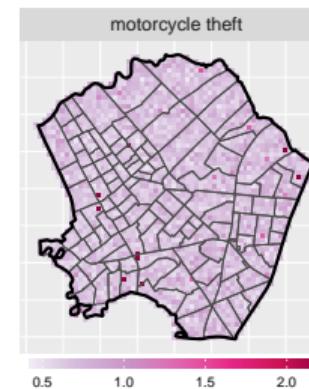
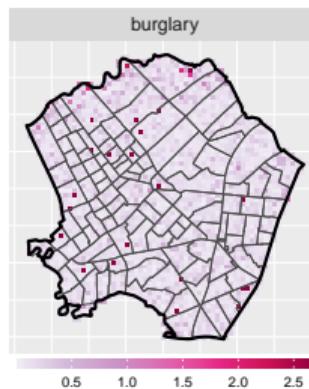
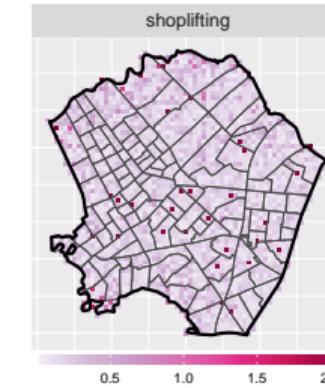
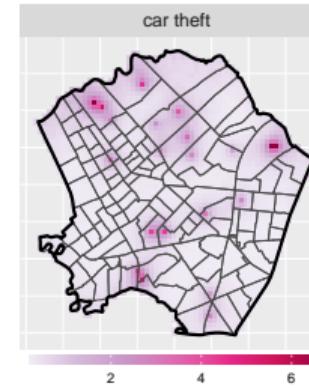
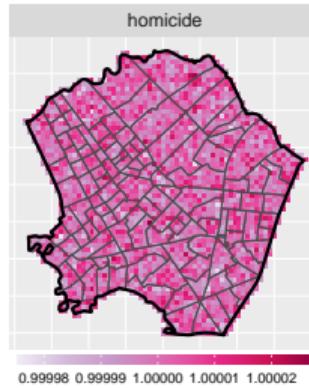
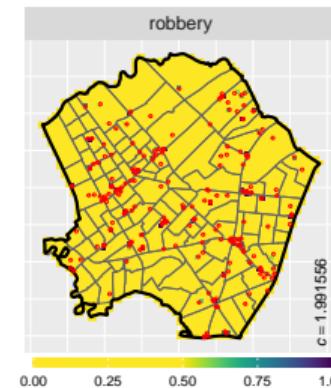
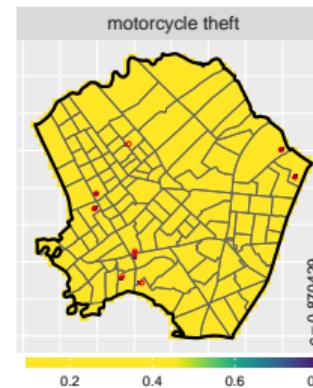
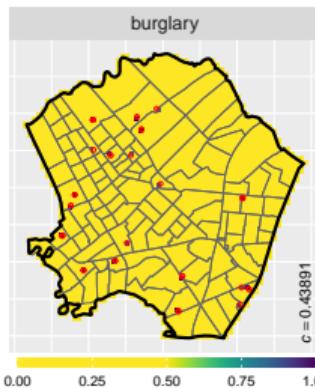
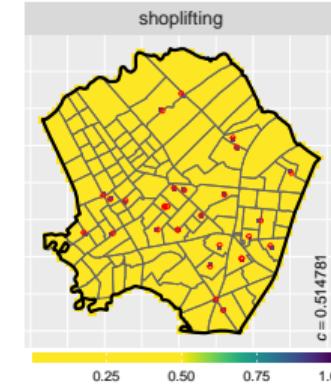
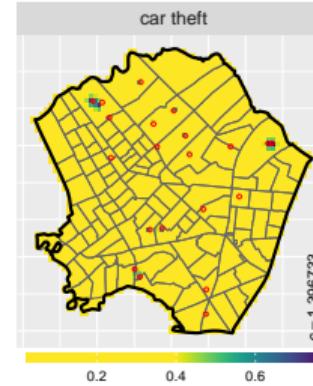
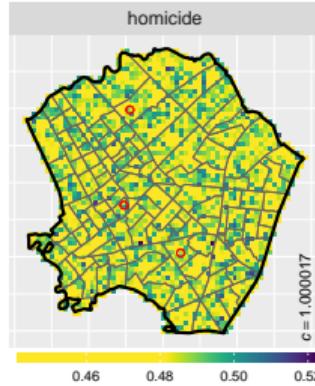


Figure: Observed daily crime temporal variation (grey lines) and the fitted variation from zero inflated Poisson regression models, or Poisson regression model in the case of robbery (red lines).

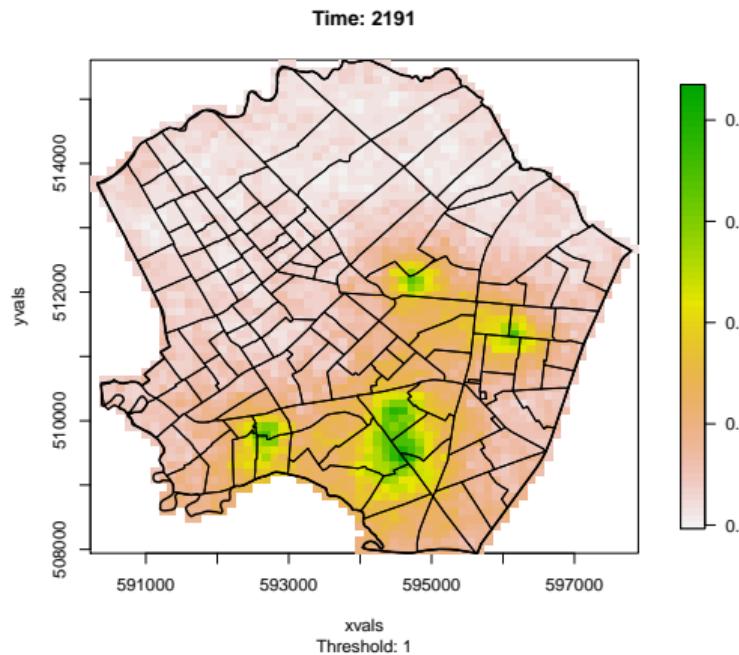
RISK MAPS



EXCEEDANCE PROBABILITIES

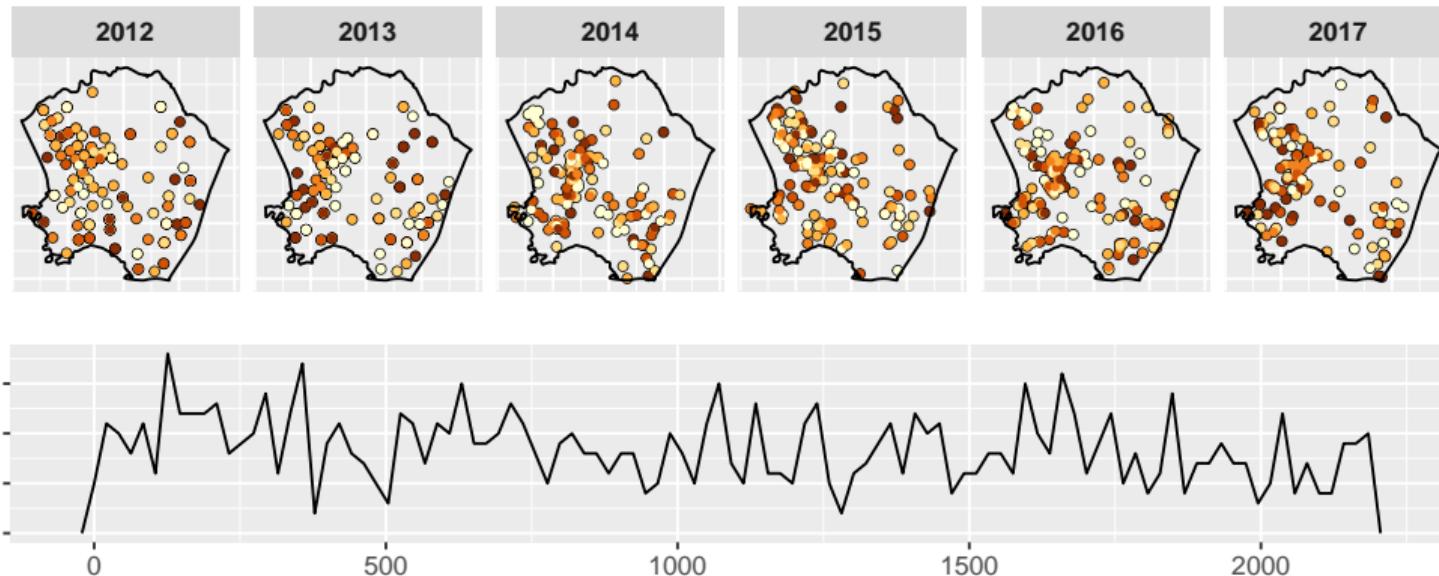


PREDICTION FOR CAR ROBBERIES

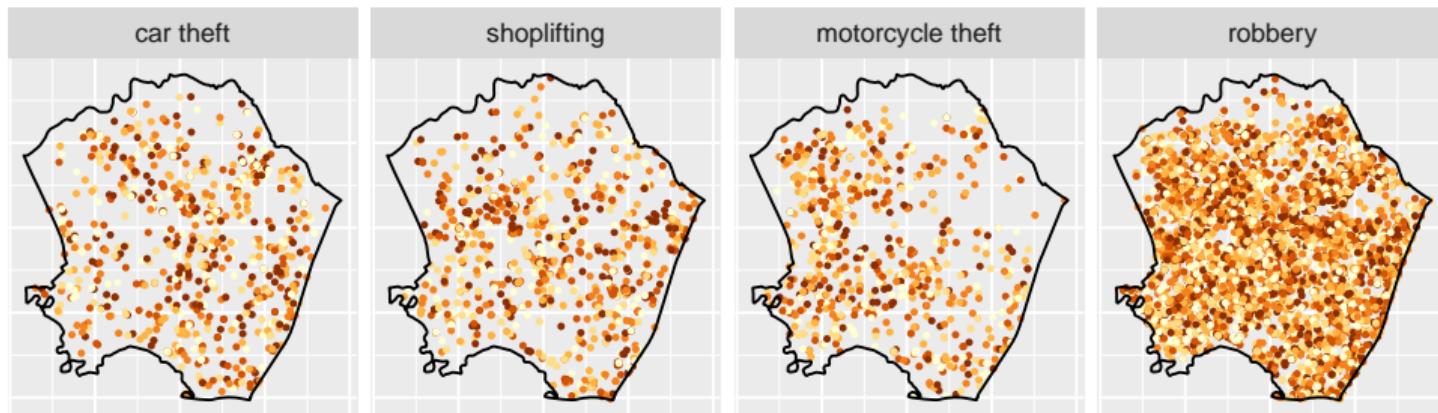


Prediction and classification using machine learning

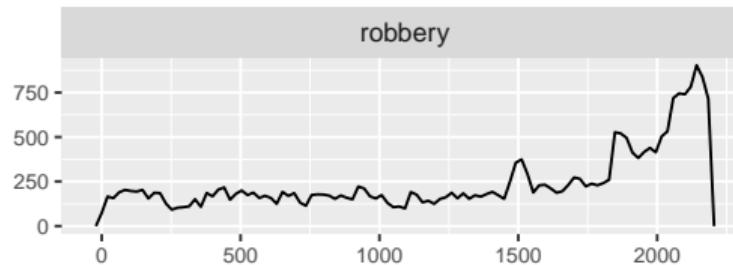
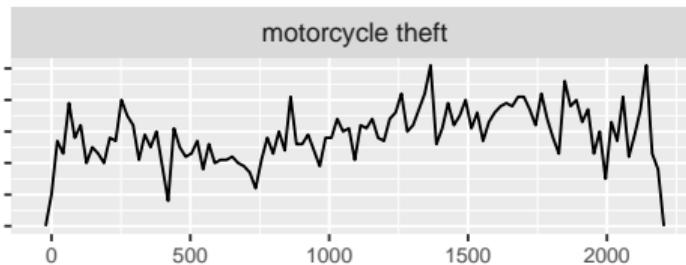
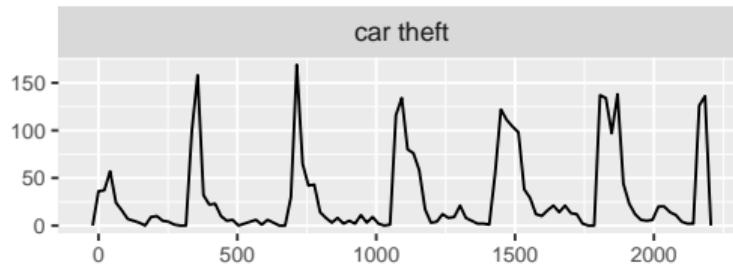
HOMICIDES



REST OF CRIMES



REST OF CRIMES



PAST-FUTURE: LEARNING PROCESS

- ▶ To link future events with historical data, we build several spatio-temporal covariates related to the point patterns
 - $\{d_1^1, d_2^1, d_3^1\}$, Minimum, median and accumulated distances from every homicide in a given temporal instant t to homicides from the previous temporal instant $t - 1$.
 - $\{d_1^2, d_2^2, d_3^2\}$, Minimum, median and accumulated distances from every homicide in a given temporal instant t to homicides from the temporal instant $t - 2$.
 - $\{d_4^1, d_5^1\}$, Average and accumulated distances from every homicide in a given temporal instant t to its neighbours in the temporal instant $t - 1$ in two neighbourhoods of radii 1.5 and 0.5km, respectively.
 - $\{d_4^2, d_5^2\}$, Average and accumulated distances from every homicide in a given temporal instant t to its neighbours in the temporal instant $t - 2$ in two neighbourhoods of radii 1.5 and 0.5km, respectively.
- ▶ Consider the point pattern given by the superposition of every other type of crimes, consider the statistics now based on distances related to crimes from the last two temporal instants, considering together $t - 1$ and $t - 2$:
 - $\{d_6, d_7\}$, Minimum and average distances from every homicide in a given temporal instant t to other previous types of crimes.
 - $\{d_8\}$, Accumulated distance from every homicide in a given time t to neighbours from other previous types of crimes in a neighbourhood of radius 1.0km.

POINT PROCESS METHODOLOGY (I)

- ▶ Objective: predict the intensity function of a point pattern based on observed past events and related covariates.
- ▶ Consider two spatio-temporal Poisson processes with intensities, for every time t , $\rho_t^0 \lambda_t^0(\mathbf{x})$ and $\rho_t^1 \lambda_t^1 \propto \lambda_t^0(\mathbf{x}) f(||\mathbf{x} - \mathbf{x}_{-t}||, ||\mathbf{x} - \mathbf{x}_0||)$, where \mathbf{x}_{-t} denotes the observed point process up to time t (representing the moving sources), and \mathbf{x}_0 is pure spatial (standing for the fixed sources).
- ▶ Considering that n_t^0 and n_t^1 are the numbers of events of each type of points that occurred at time t , an event that has happened at location \mathbf{x} has a probability of being from the case process

$$p_t(\mathbf{x}) = \frac{n_t^1 f(||\mathbf{x} - \mathbf{x}_{-t}||, ||\mathbf{x} - \mathbf{x}_0||)}{n_t^1 f(||\mathbf{x} - \mathbf{x}_{-t}||, ||\mathbf{x} - \mathbf{x}_0||) + n_t^0}. \quad (2)$$

POINT PROCESS METHODOLOGY (II)

- ▶ Under these assumptions, the case and control point processes comprise a dataset of binary variables.
- ▶ This problem can be addressed as a classification problem, and **machine learning algorithms (Random Forest)** are flexible and computationally feasible methods to estimate $p_t(\mathbf{x})$ and, consequently, one can estimate $f(||\mathbf{x} - \mathbf{x}_{-t}||, ||\mathbf{x} - \mathbf{x}_0||)$ by applying the reverse transformation (3)

$$f(||\mathbf{x} - \mathbf{x}_{-t}||, ||\mathbf{x} - \mathbf{x}_0||) = \frac{n_t^0 p_t(\mathbf{x})}{n_t^1 - n_t^1 p_t(\mathbf{x})} \quad (3)$$

- ▶ Consider the intensity of the control process constant over the study area ($\lambda_t^0(\mathbf{x}) = \lambda_0$);

RESULTS (I)

Predicted intensity of the last two temporal instants. The point patterns are also included to provide a visual assessment of the quality of predictions

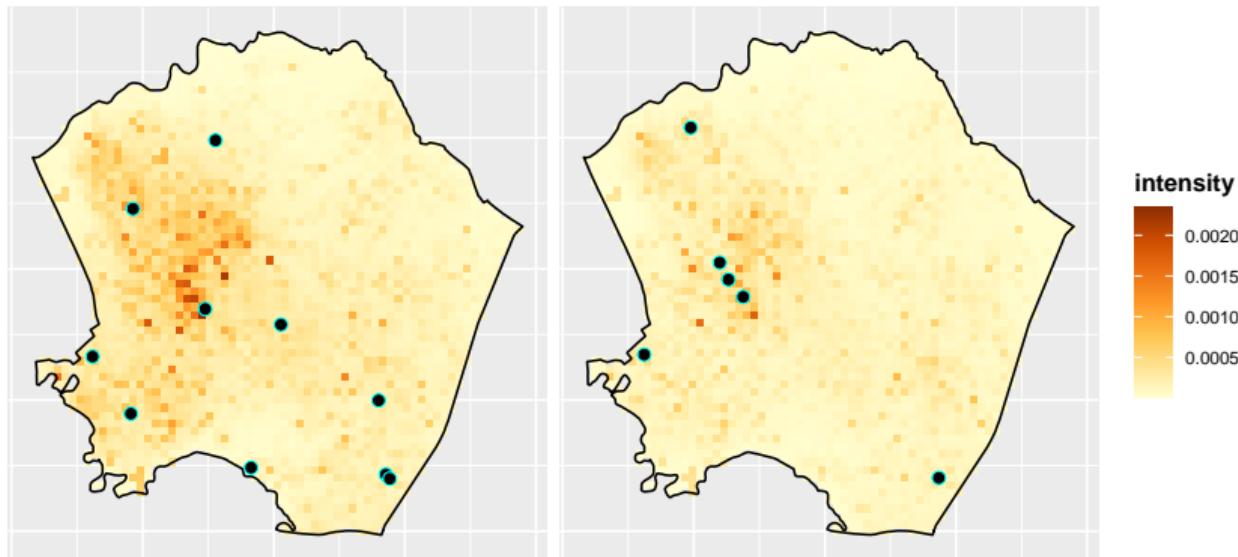


Figure: Random Forest estimates of the intensity of homicides in Kennedy for two periods of 21 days. Black highlighted points represent the locations of the current homicides in these periods.

RESULTS (II)

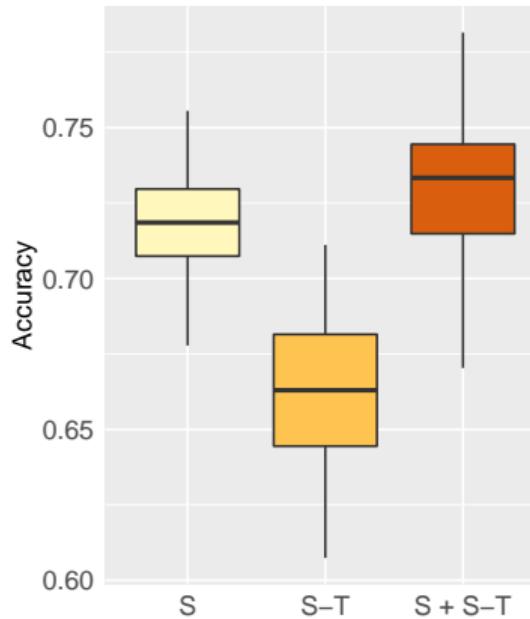


Figure: Boxplots of the accuracy for RF according to spatial (S), spatio-temporal (S-T) or the combination of spatial and spatio-temporal covariates (S+ S-T).

The results suggest the pure spatial covariates are the components that contribute most to improve the accuracy. However, the inclusion of the spatio-temporal covariates slightly increases the performance of the method.

Reducing dimensionality: Barycenters

TT METRIC AND BARYCENTERS

- ▶ Consider the set $\mathfrak{N}_{\text{fin}}$ of finite counting measures on some metric space (\mathcal{X}, d) . Equip $\mathfrak{N}_{\text{fin}}$ with a metric τ that reflects the concept of distance between point patterns in an appropriate problem-related way
- ▶ In the metric space $(\mathfrak{N}_{\text{fin}}, \tau)$ we can furthermore define a Fréchet mean of order $q \geq 1$; that is, for data $\xi_1, \dots, \xi_k \in \mathfrak{N}_{\text{fin}}$ any $\zeta \in \mathfrak{N}_{\text{fin}}$ minimizing

$$\sum_{j=1}^k \tau(\xi_j, \zeta)^q. \quad (4)$$

- ▶ Such a q -th order mean may serve as a “typical” element of $\mathfrak{N}_{\text{fin}}$ to represent the data, and gives rise to more complex statistical analyses, such as Fréchet regression
- ▶ Two metrics on the space of point patterns that have been widely used are the **spike time metric**, (Victor and Purpura, (1997) for one dimension, and Diez et al. (2012) for higher dimension), and the **optimal subpattern assignment (OSPA) metric** (see Schuhmacher and Xia, 2008).

TT METRIC AND BARYCENTERS

- ▶ Both the *transport–transform (TT) metric* and its *normalized version (RTT)* are based on matching the points between two point patterns on \mathcal{X} optimally in terms of some power p of d and penalizing points that cannot be reasonably matched. We may interpret these metrics as unbalanced p -th order Wasserstein metrics
- ▶ Denote by $\mathfrak{N}_{\text{fin}}$ the space of finite point patterns (counting measures) on a complete separable metric space (\mathcal{X}, d) , equipped with the usual σ -algebra \mathcal{N}_{fin} generated by the point count maps $\Psi_A : \mathfrak{N}_{\text{fin}} \rightarrow \mathbb{R}$, $\xi \mapsto \xi(A)$ for $A \subset \mathcal{X}$ Borel measurable
- ▶ Let $C > 0$ and $p \geq 1$ be two parameters, referred to as *penalty* and *order*, respectively.
 - For $\xi = \sum_{i=1}^m \delta_{x_i}$, $\eta = \sum_{j=1}^n \delta_{y_j} \in \mathfrak{N}_{\text{fin}}$ define the *transport-transform (TT) metric* by

$$\tau(\xi, \eta) = \tau_{C,p}(\xi, \eta) = \left(\min_{(i_1, \dots, i_l; j_1, \dots, j_l) \in S(m,n)} \left((m+n-2l)C^p + \sum_{r=1}^l d(x_{i_r}, y_{j_r})^p \right) \right)^{1/p}, \quad (5)$$

where the minimum is taken over equal numbers of pairwise different indices of $[m]$ and $[n]$, respectively, i.e.

$$\begin{aligned} S(m, n) = & \{(i_1, \dots, i_l; j_1, \dots, j_l) ; l \in \{0, 1, \dots, \min\{m, n\}\}, \\ & i_1, \dots, i_l \in [m] \text{ pairwise different}, j_1, \dots, j_l \in [n] \text{ pairwise different}\}. \end{aligned}$$

- (b) For $\xi, \eta \in \mathfrak{N}_{\text{fin}}$ define the *relative transport-transform (RTT) metric* by

$$\bar{\tau}(\xi, \eta) = \bar{\tau}_{C,p}(\xi, \eta) = \frac{1}{\max\{|\xi|, |\eta|\}^{1/p}} \tau_{C,p}(\xi, \eta).$$

TT METRIC AND BARYCENTERS

- ▶ For data on quite general metric spaces, **barycenters** can formalize the idea of a center element representing the data.
- ▶ In the case of $\mathfrak{N}_{\text{fin}}$ we are thus looking for a center point pattern that gives a good first order representation of a set of data point patterns ξ_1, \dots, ξ_k . More formally we may define a barycenter as the (weighted) q -th order Fréchet mean with respect to τ
- ▶ For $k \in \mathbb{N}$ let $\xi_1, \dots, \xi_k \in \mathfrak{N}_{\text{fin}}$ be data point patterns and $\lambda_1, \dots, \lambda_k > 0$ with $\sum_{j=1}^k \lambda_j = 1$ be weights. Let furthermore $q \geq 1$. Then we call any

$$\zeta_* \in \arg \min_{\zeta \in \mathfrak{N}_{\text{fin}}} \sum_{j=1}^k \lambda_j \tau(\xi_j, \zeta)^q \quad (7)$$

a (*weighted*) *barycenter of order q*. If no weights are specified we tacitly assume that $\lambda_j = 1/k$ for $1 \leq j \leq k$, leading to an “unweighted” barycenter

TT METRIC AND BARYCENTERS

The reduction to the underlying location problem allows us to treat the case of point patterns on a network equipped with the **shortest-path metric** and $p = 1$. We give an example for crime data in Valencia, Spain.

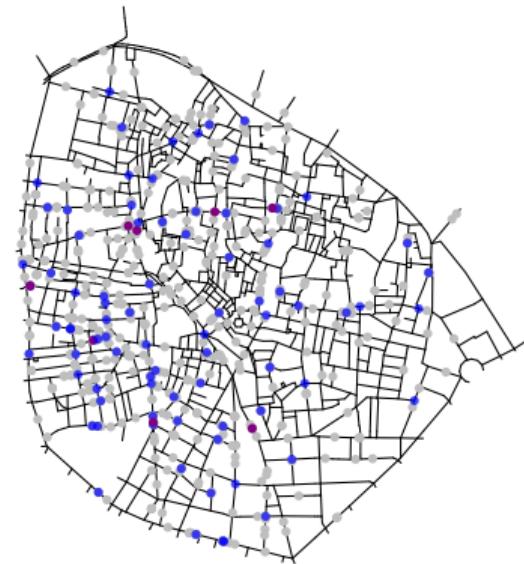


Figure: An example of a barycenter on a street network. Shown are 8 patterns of assault crimes during the summer months of 2010–2017 in the old town of Valencia (all in grey for better overall visibility). The resulting barycenter with respect to shortest-path distance along the streets is given in blue, with multipoints in purple ($p = q = 1$).

STREET THEFTS IN BOGOTA

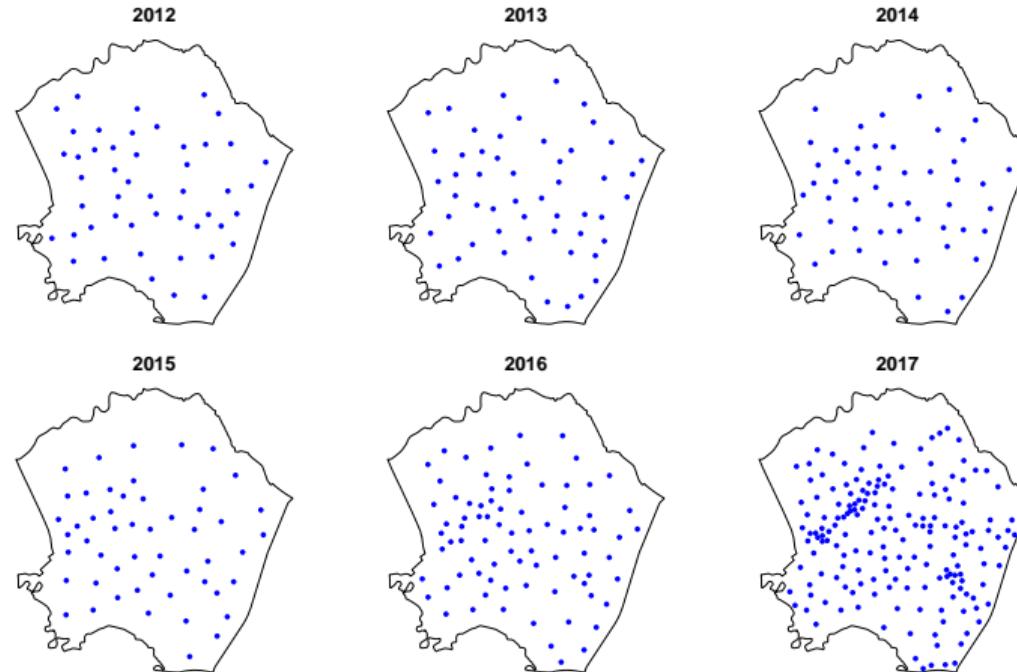


Figure: Barycenters of weekly street thefts in the localidad of Kennedy in Bogotá. The cardinalities are 48, 53, 52, 52, 80 and 175, respectively.

ASSAULT CASES IN VALENCIA

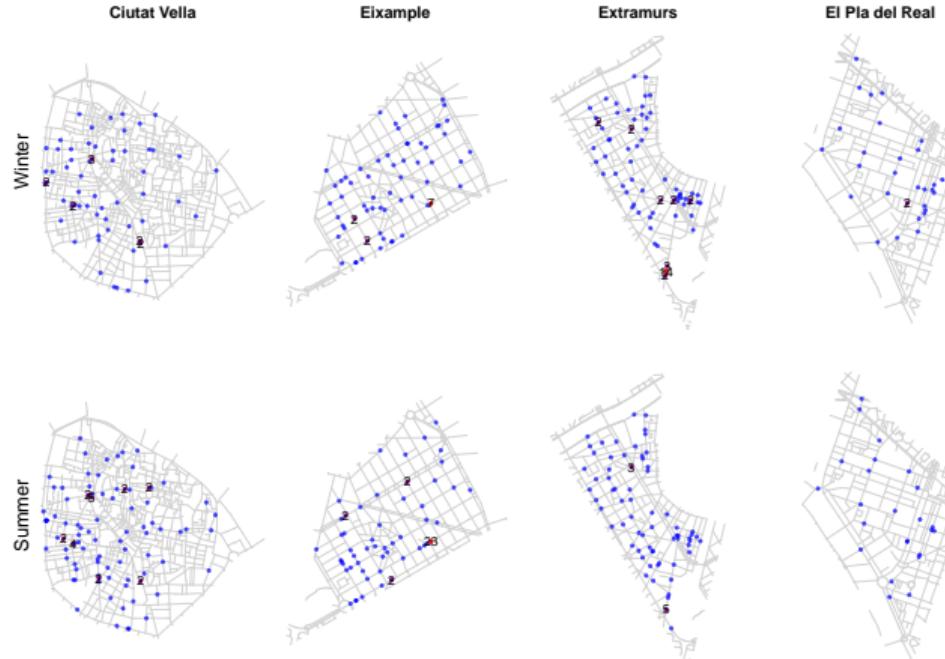


Figure: Barycenters of cases of assault for different districts of Valencia in winter and in summer. The numbers indicate multiplicities if there are several points at a single location. The cardinalities of the barycenters are 68, 69, 88, 30 (winter) and 103, 79, 74, 24 (summer).

Stochastic integro-differential equations and point processes

MOTIVATING EXAMPLE

- ▶ Geo-coded locations and times of the calls to the emergency number (112) related to individual crimes from 2010 to 2020 in the city of Valencia, Spain
- ▶ During these 10 years (2010-2019), 83379 calls were received. Among all these calls, 51533 (assault), 23282 (robbery), 388 (woman alert), 8176 (other causes)
- ▶ Spatial division of the city of Valencia in 81 neighborhoods or districts
- ▶ Covariate information: shortest distances to different points of interest such as police stations, bars, restaurants, pubs, coffee shops, supermarkets, taxi stops, industries, nightclubs, and ATMs

MOTIVATING EXAMPLE

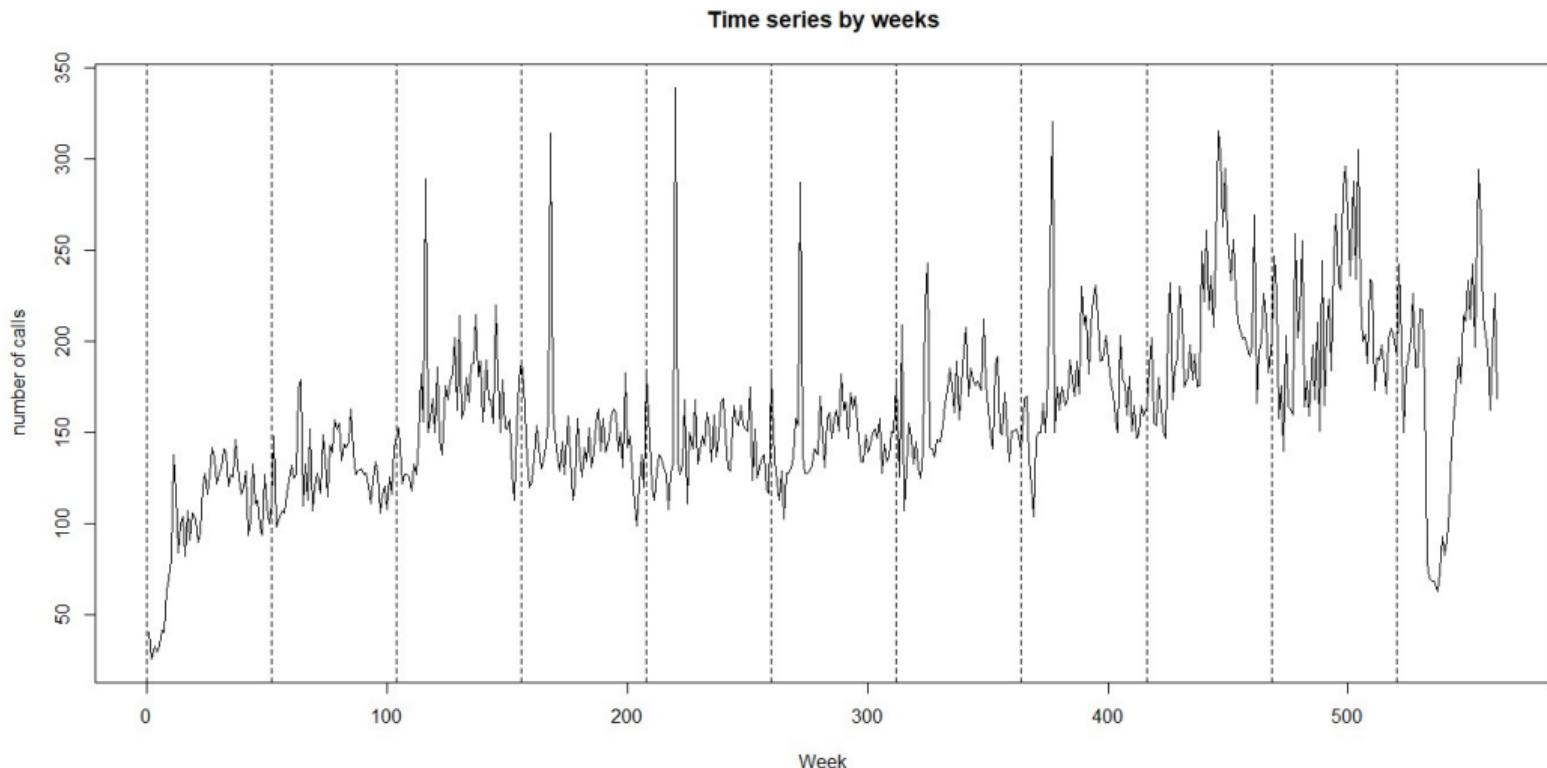


Figure: Temporal series by weeks

MOTIVATING EXAMPLE

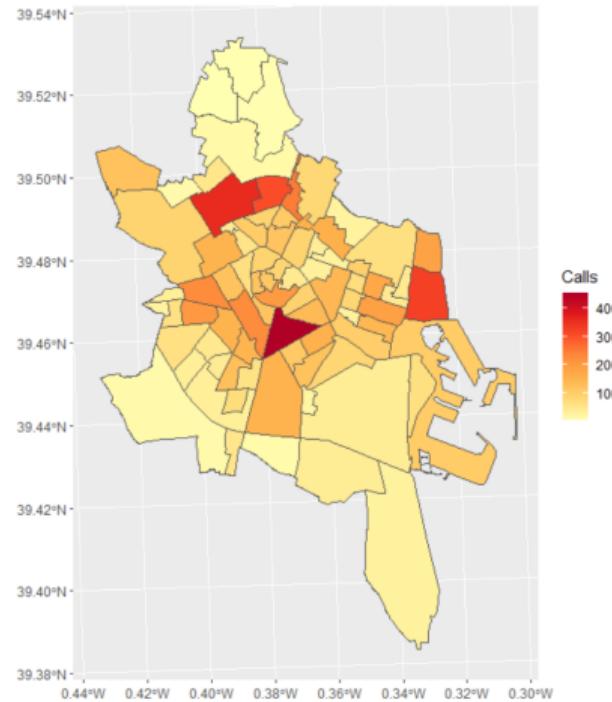


Figure: Total number of calls per district

MOTIVATING EXAMPLE

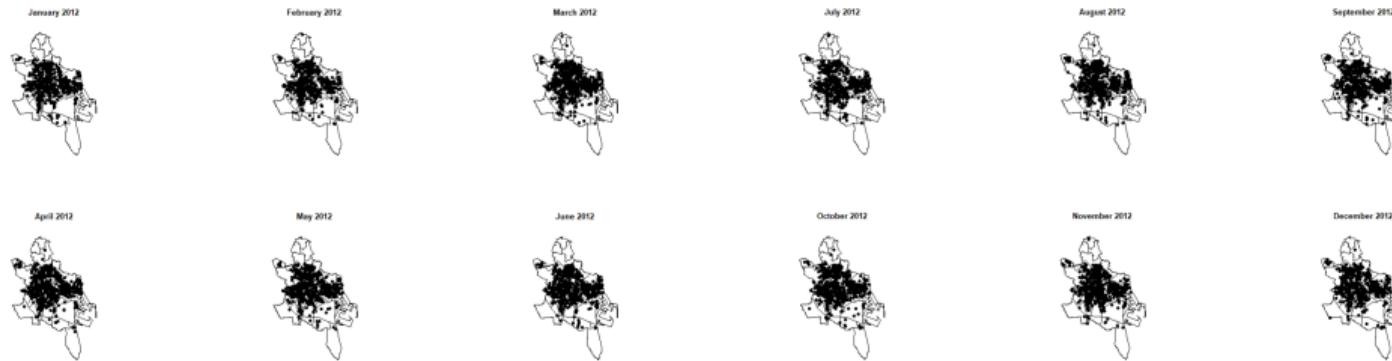


Figure: Spatial patterns per months in 2012

MOTIVATING EXAMPLE

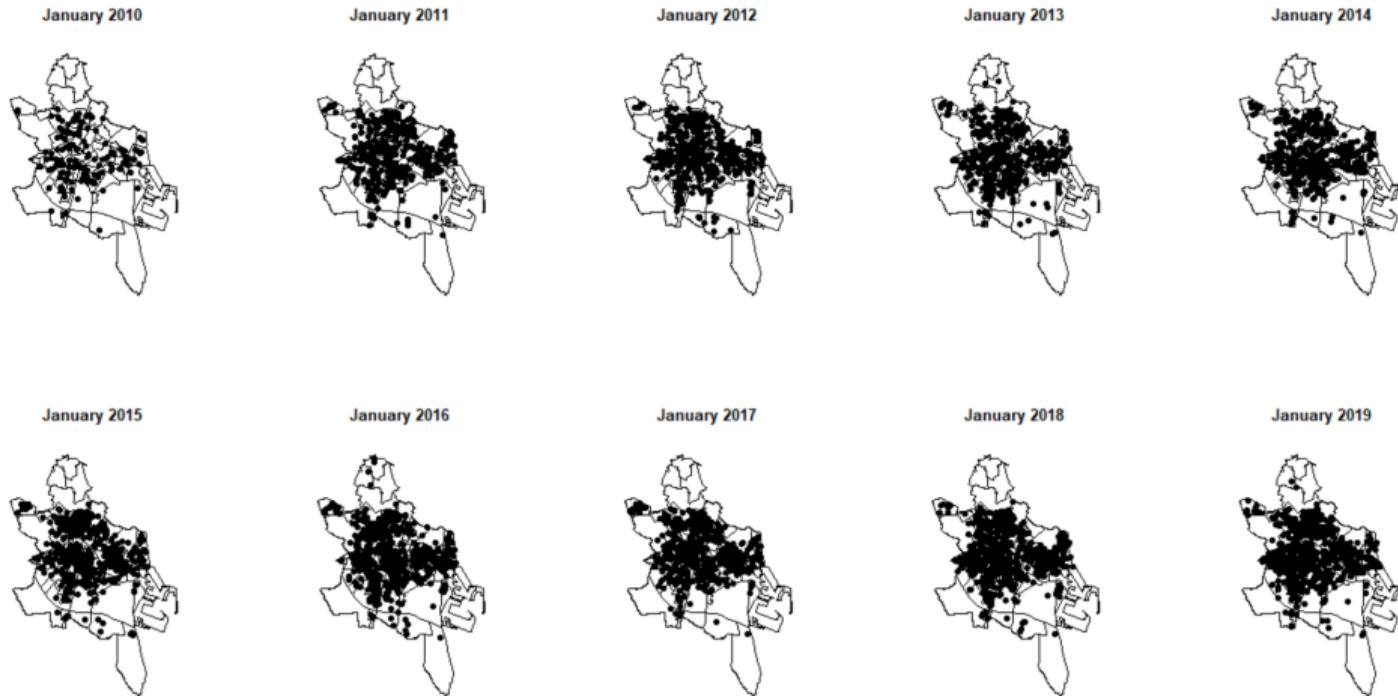


Figure: Spatial patterns in January over the 10 years

METHODOLOGY

- ▶ Consider a discrete-time series of continuous-space LGCPs
- ▶ Let $k \in \mathcal{K}, \mathcal{K} = \{1, \dots, K\}$ denote a discrete-time index set
- ▶ $z_k(\mathbf{s}) \sim \text{GP}(\mu_k(\mathbf{s}), \sigma_k^2 \Psi_k(\mathbf{s}, \mathbf{r}))$, a set of temporally correlated spatial Gaussian Processes (GPs), each with mean $\mu_k(\mathbf{s})$ and covariance function $\sigma_k^2 \Psi_k(\mathbf{s}, \mathbf{r})$
- ▶ For each k , the point process intensity function is defined as $\lambda_k(\mathbf{s}) = \exp(z_k(\mathbf{s}))$
- ▶ Alternatively, if $\mathbf{d}(\mathbf{s})$ be a vector of spatially referenced covariates and \mathbf{b}^T the corresponding regression coefficients, $\lambda_k(\mathbf{s}) = \exp(\mathbf{b}^T \mathbf{d}(\mathbf{s}) + z_k(\mathbf{s}))$

METHODOLOGY

- ▶ The temporal dynamics of the intensity functions through $z_k(\mathbf{s})$ can be defined under the stochastic integro-difference equation (SIDE) framework
- ▶ SIDE is a flexible modelling tool that represents temporal dynamic effects such as diffusion and dispersal
- ▶ SIDE associates the spatio-temporal dependent variable $z_k(\mathbf{s})$ to $z_{k+1}(\mathbf{s})$ through the following integral equation

$$z_{k+1}(\mathbf{s}) = \int_D k_I(\mathbf{s}, \mathbf{r}) f_1(z_k(\mathbf{r})) d\mathbf{r} + e_k(\mathbf{s}), \quad (8)$$

- ▶ $k_I(\mathbf{s}, \mathbf{r})$ is the mixing kernel in the integral
- ▶ $e_k(\mathbf{s}) \sim \text{GP}(\mu_Q(\mathbf{s}), k_Q(\mathbf{s}, \mathbf{r}))$ is an added disturbance, modeled as a Gaussian field with mean $\mu_Q(\mathbf{s})$ and covariance function $k_Q(\mathbf{s}, \mathbf{r})$
- ▶ The nonlinear mapping $f_1(\cdot)$ distorts the field; we use the identity case

METHODOLOGY

- ▶ The pair auto-correlation function (PACF) $g_{k,k}(\mathbf{s}, \mathbf{r})$, and the pair cross-correlation function (PCCF) $g_{k,k+1}(\mathbf{s}, \mathbf{r})$ quantify the correlation between the probability of an event at location \mathbf{r} given that an event has occurred at \mathbf{s} within the same time frame k or at previous time frame $k - 1$
- ▶ The fundamental lemma of LGCPs states that the log PACF is proportional to the field auto-correlation function,

$$g_{k,k}(\mathbf{s}, \mathbf{r}) = \exp(\sigma_k^2 \Psi_k(\mathbf{s}, \mathbf{r})),$$

- ▶ the Fourier transform of the auto-correlation function is the spectrum of the signal. Use this to select a set of representative basis functions
- ▶ Compact Gaussian Radial Basis functions (CGRBF)

- The likelihood function is used for inference

$$p(y_k | \lambda_k(\mathbf{s})) = \prod_{\mathbf{s}_j \in y_k} \lambda_k(\mathbf{s}_j) \exp \left(- \int_D \lambda_k(\mathbf{s}) d\mathbf{s} \right),$$

- and each $\lambda_k(\mathbf{s})$ is approximated using the same basis representation

$$\lambda_k(\mathbf{s}) = \exp(\mathbf{b}^T \mathbf{d}(\mathbf{s}) + z_k(\mathbf{s})) \approx \exp(\mathbf{b}^T \mathbf{d}(\mathbf{s}) + \phi(\mathbf{s})^T \mathbf{x}_k).$$

- Use variational Bayes inference for parameter estimation

RESULTS

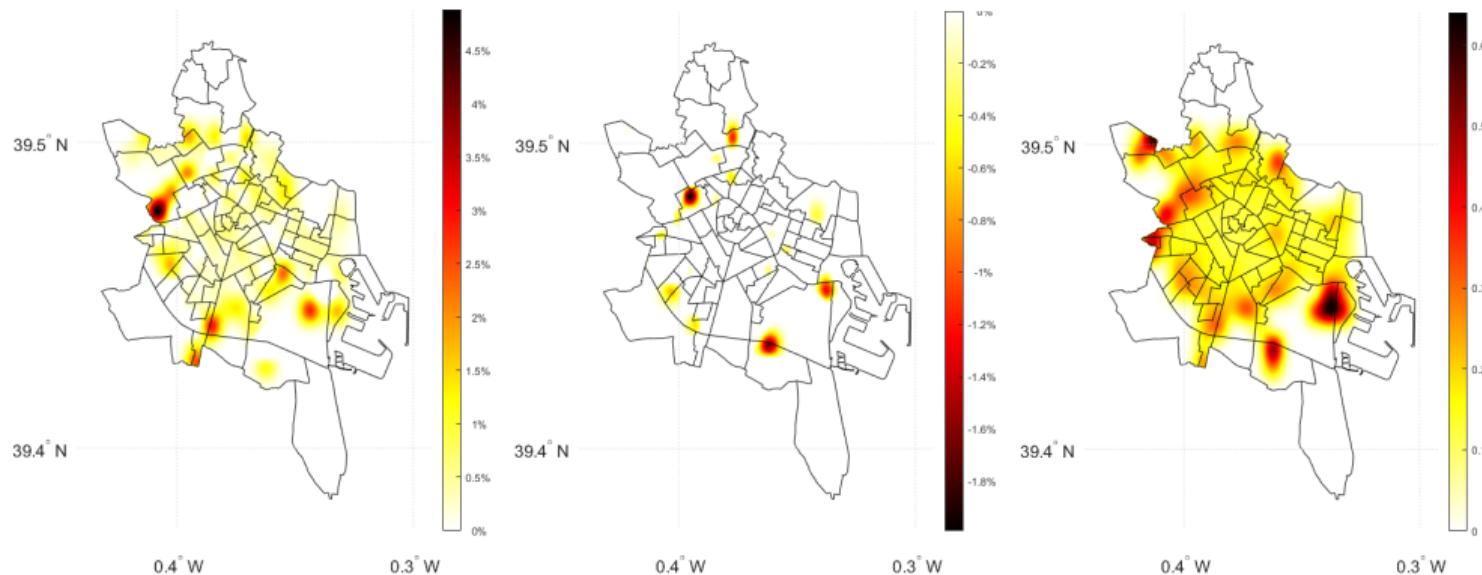


Figure: Average growth, average decrease and volatility of the predictions for 2010-2019

RESULTS

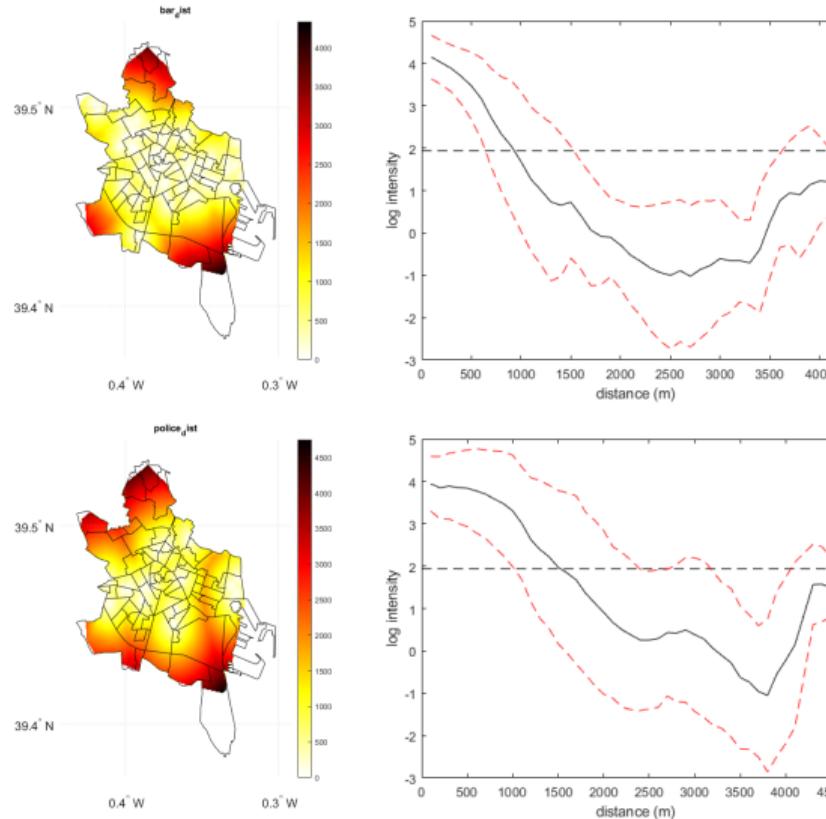


Figure: Empirical relationship between independent variable and log spatial intensity

RESULTS

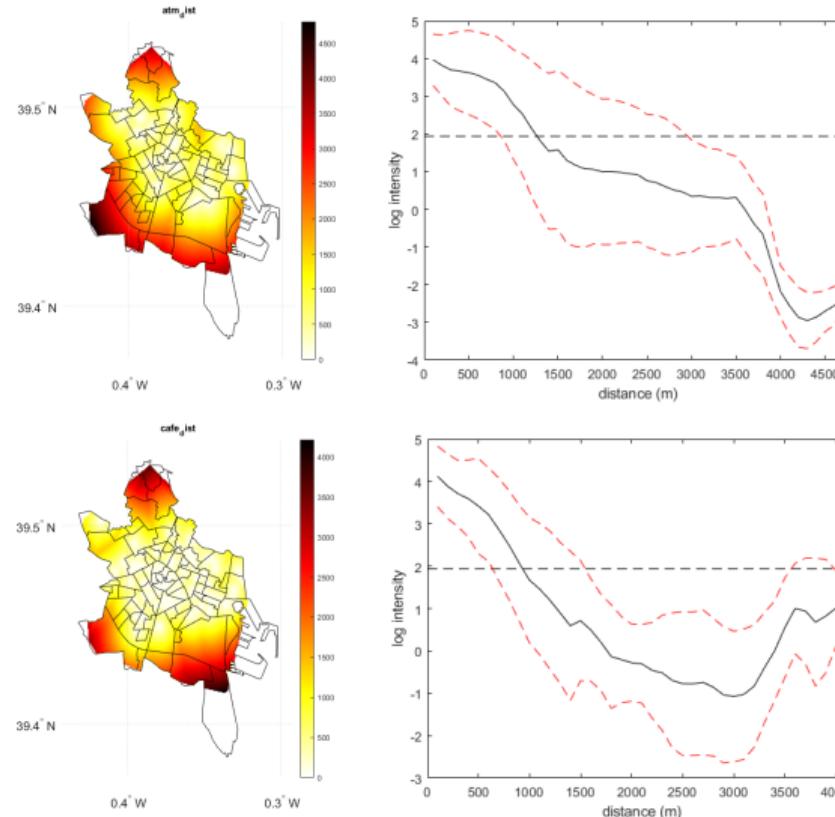


Figure: Empirical relationship between independent variable and log spatial intensity

RESULTS

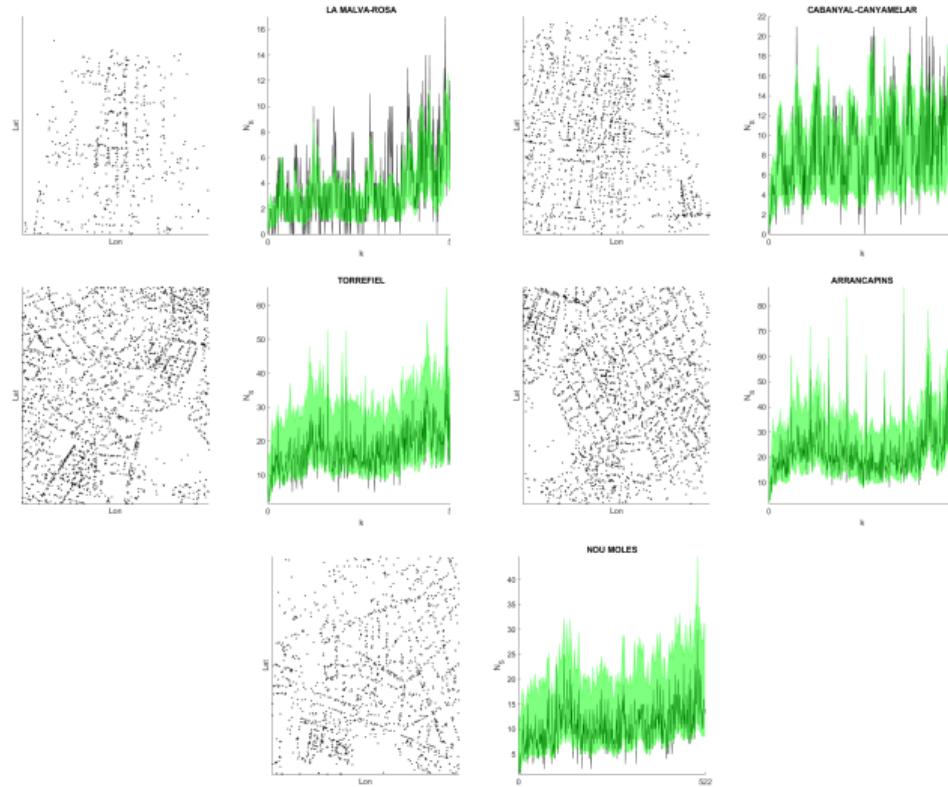


Figure: Model fit for five different neighborhoods in Valencia. The green bandwidth shows the variance of the prediction

RESULTS

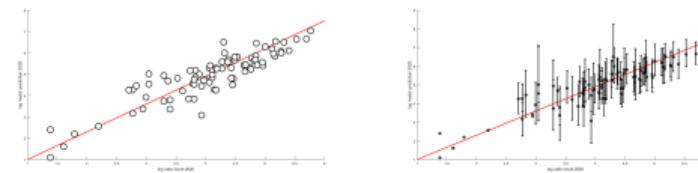
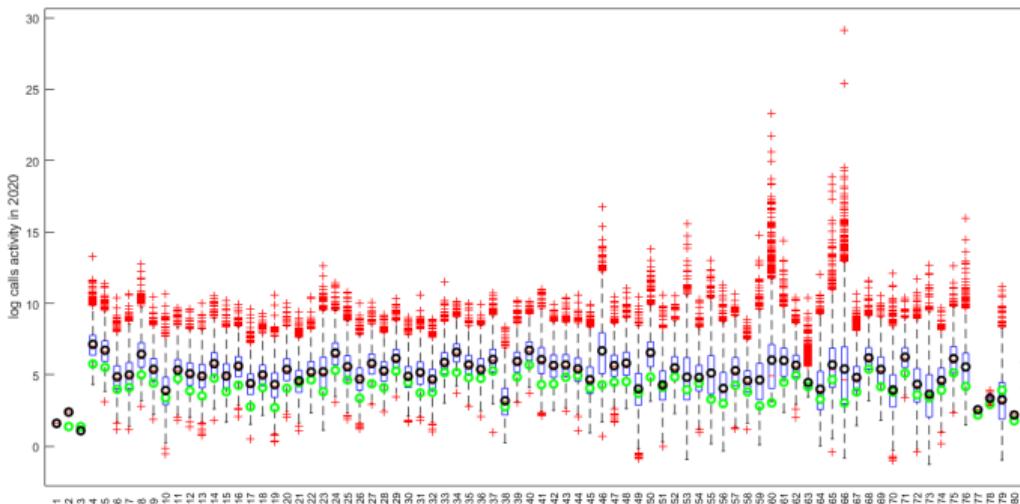


Figure: Log predictions for 2020

Modeling Origin-Destination Point Patterns

- ▶ Dataset composed of car theft locations along with a linked dataset of recovery locations which, due to partial recovery, is a relatively small subset of the set of theft locations
- ▶ Understand the behavior of car thefts and recoveries in the region
- ▶ Viewing the set of theft locations as a point pattern, what types of predictive models can be built to learn about recovery location given theft location? Can the dependence between the point pattern of theft locations and the point pattern of recovery locations be formalized?
- ▶ Origin–destination modeling offers a natural framework for such problems.

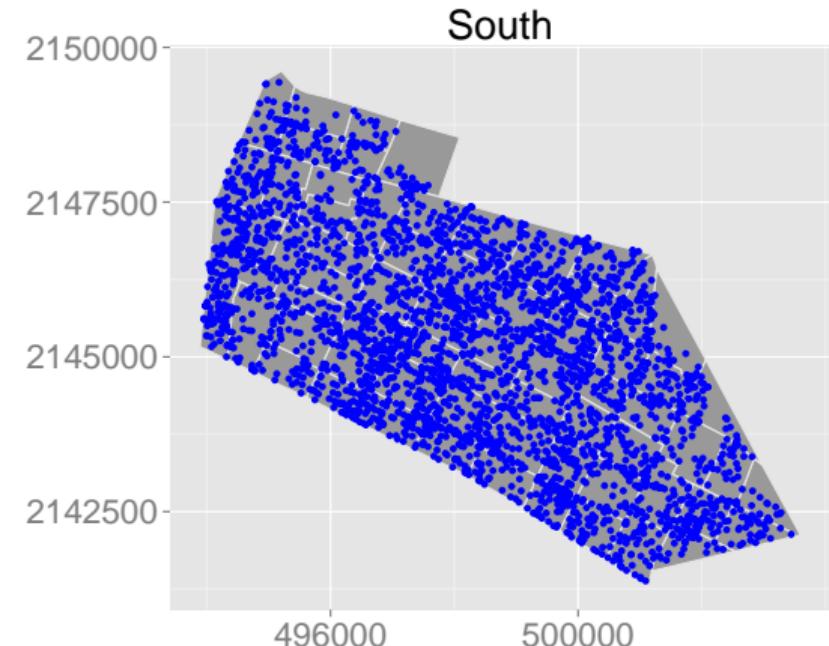
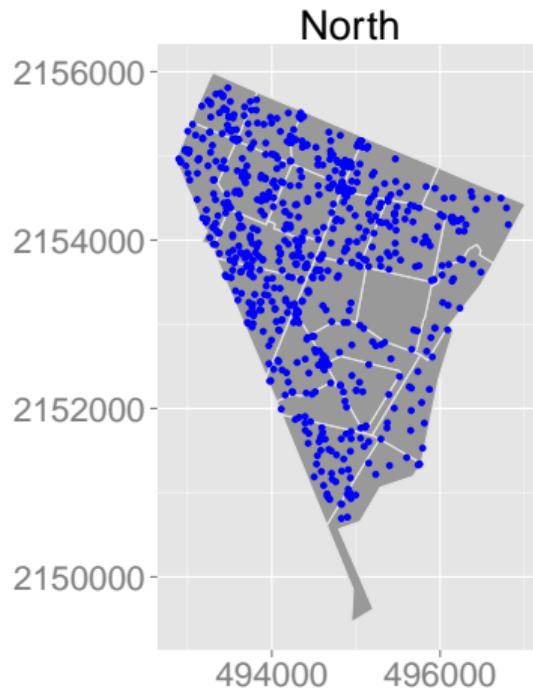


Figure: Car theft locations in the North region (left) and the South region (right) in Neza (x -axis (easting) and y -axis (northing) are at km scale).

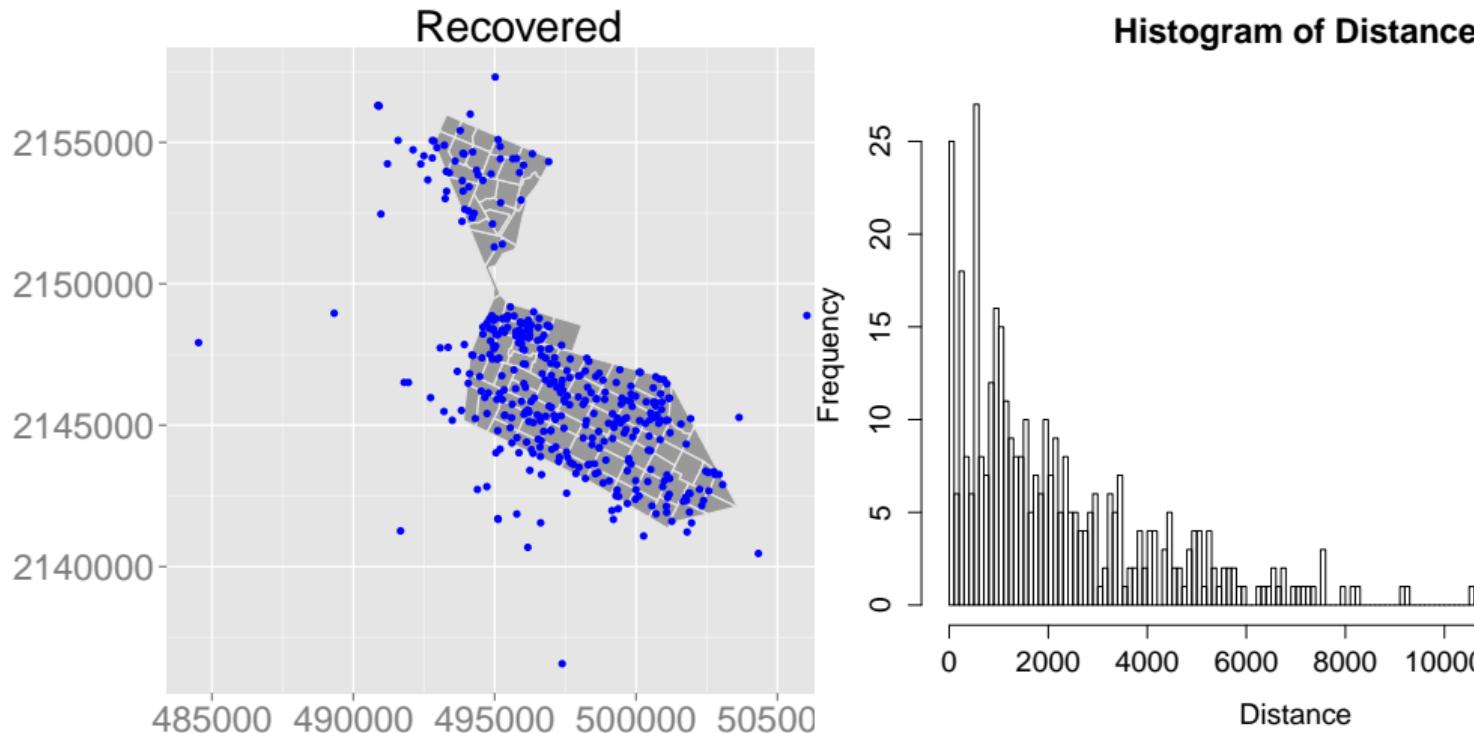


Figure: Recovery locations (left, x-axis (easting) and y-axis (northing) are at km scale) and histogram of the distance between theft and recovery locations (right, x-axis (distance) also at km scale).

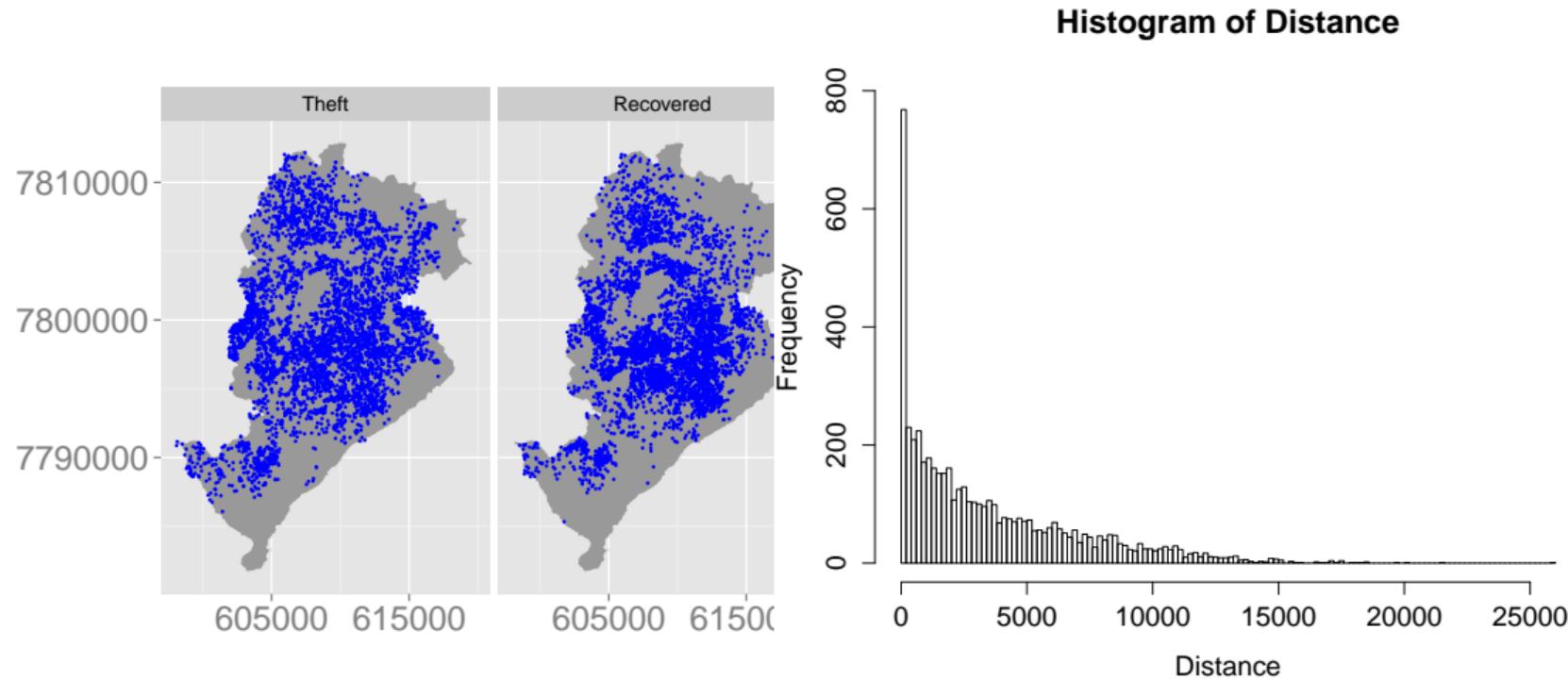


Figure: Car theft and recovery locations (left, x -axis (easting) and y -axis (northing) are at km scale) and histogram of the distance between theft and recovery locations (right, x -axis (distance) also at km scale) in Belo Horizonte

- ▶ First, the set of theft locations is modeled by using both a nonhomogeneous Poisson process as well as a log-Gaussian Cox process
- ▶ Second, a conditional regression specification is proposed to provide the distribution of recovery location given theft location
- ▶ Third, we investigate the dependence between the theft location point pattern and the recovery location point pattern. We consider a joint model, viewing the data as an origin–destination pair of points, and treating the point pattern data as consisting of random pairs of locations

LGCP AND NHPP MODELS FOR VEHICLE THEFT

- We seek to provide an investigator with understanding of the nature of the intensity surface that is driving the point pattern of thefts. This surface can be viewed as a risk surface for theft, enabling clarification of where risk is high, where it is low
- Let $\mathcal{S} = \{s_1, \dots, s_n\}$ denote the observed point pattern over the study region $D \subset \mathbb{R}^2$. We view the theft events as conditionally independent given the intensity and therefore consider a non-homogeneous Poisson process (NHPP) and a log-Gaussian Cox processes (LGCP) for modeling theft events
- The LGCP is defined so that the log of the intensity is a Gaussian process (GP), i.e.,

$$\log \lambda(s) = X(s)\beta + z(s), \quad z(\mathcal{S}) \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_z), \quad s \in D. \quad (9)$$

where $X(s)$ is a covariate vector at s and $z(s)$ is a Gaussian process.

- The component spatial random effects for the intensity surface provide local pushing up and pulling down the surface, as appropriate. We assume an exponential covariance function, i.e.,
$$C(u, u') = \sigma^2 \exp(-\phi \|u - u'\|)$$

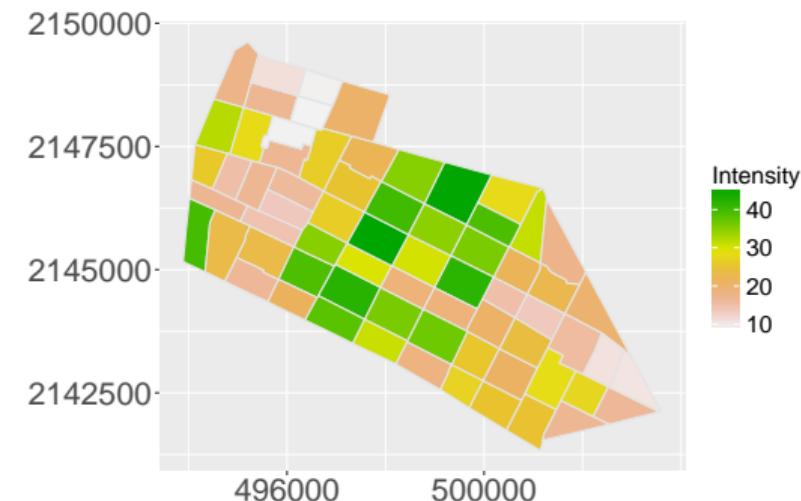
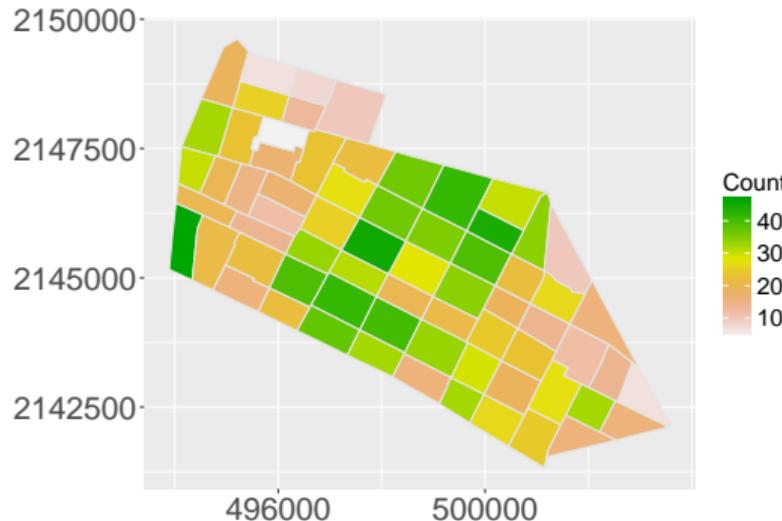


Figure: Testing data counts (left) and posterior predictive intensity surface, LGCP (right) in the South region.

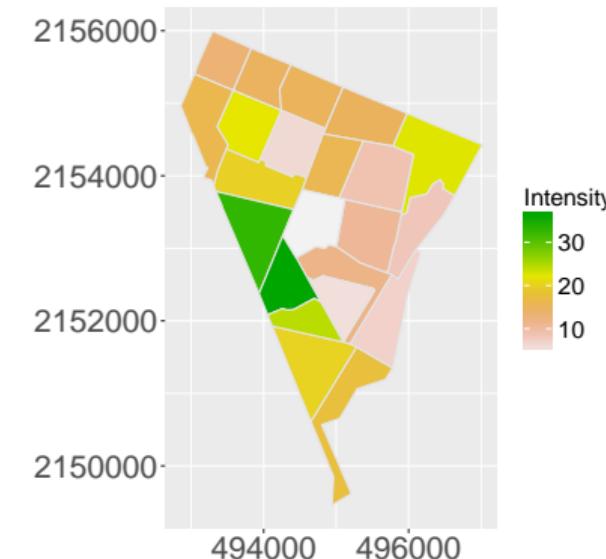
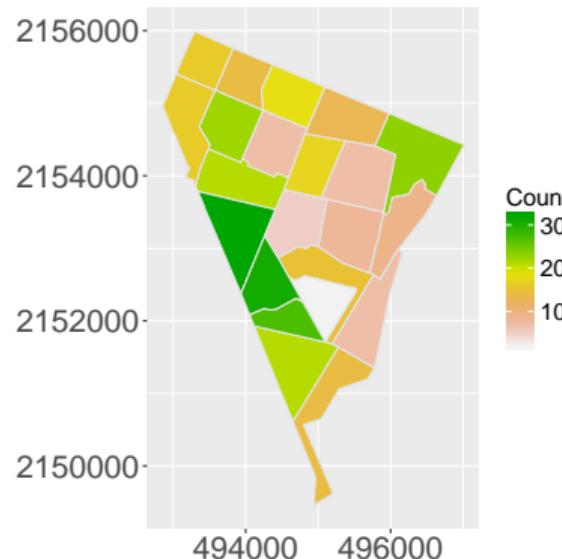


Figure: Testing data counts (left) and posterior predictive intensity surface, LGCP (right) in the North region.

CONDITIONING RECOVERY LOCATION ON THEFT LOCATION

- ▶ Denote by s_R a recovery location and by s_T a theft location with the set of theft locations denoted by $\mathcal{S}_T = \{s_{T,1}, \dots, s_{T,n}\}$ and the set of recovery locations denoted by $\mathcal{S}_R = \{s_{R,1}, \dots, s_{R,m}\}$ where $m < n$
- ▶ The object of interest is the conditional density specification for recovery location s_R given a theft location s_T , denoted as $f_R(s_R|s_T)$
- ▶ Let $\mathcal{S}_T^* = \{s_{T,1}^*, \dots, s_{T,m}^*\}$ be the set of theft locations corresponding to recovery points, i.e., $s_{T,j}^*$ is the corresponding theft location for the recovery point $s_{R,j}$ for $j = 1, \dots, m$. For $j = 1, \dots, m$,

$$f_R(s_{R,j}|s_{T,j}^*) \propto |\Sigma(s_{T,j}^*)|^{-1/2} \exp\left(-(s_{R,j} - s_{T,j}^*)' \Sigma(s_{T,j}^*)^{-1} (s_{R,j} - s_{T,j}^*)\right), \quad (10)$$

$\Sigma(s_{T,j}^*)$ is 2×2 covariance kernel dependent on theft location $s_{T,j}^*$

- ▶ A benchmark specification would assume a constant covariance kernel across theft locations, but we enrich this specification using a locally adaptive covariance kernel, employing the spatially varying covariance kernel

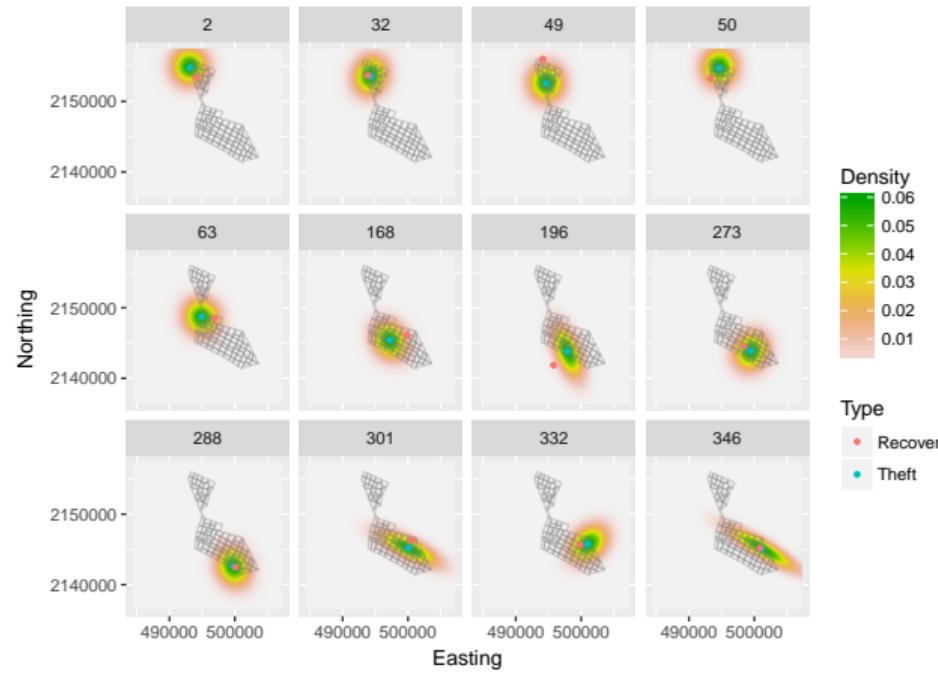


Figure: Predictive conditional density $f_R(\cdot|\cdot)$ for selected pairs in Neza with $\phi^* = 30$.

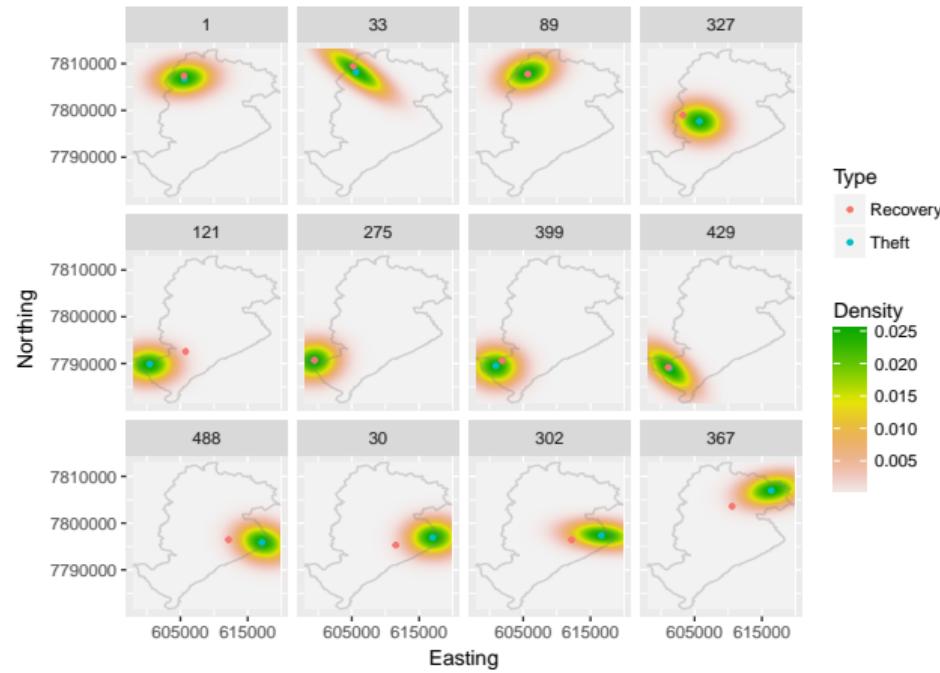


Figure: Predictive conditional density $f_R(\cdot | \cdot)$ for selected pairs in Belo Horizonte with $\phi^* = 10$.

JOINT POINT PATTERN MODELING

- We build a *joint* intensity of the form $\lambda(s_o, s_d)$ over pairs of locations $(s_o, s_d) \in D_o \times D_d$ where s_o is a theft location and s_d is a recovery location
- LGCP is introduced for *pairs* of locations as a joint point process model over $D_o \times D_d \subset \mathbb{R}^2 \times \mathbb{R}^2$. We denote observed pairs as $\mathcal{S}_P = \{s_{P,1}, \dots, s_{P,m}\} = \{(s_{R,1}, s_{T,1}^*), \dots, (s_{R,m}, s_{T,m}^*)\}$; R denotes recovery, T denotes theft. The intensity function for observed pairs is specified as

$$\log \lambda(s_R, s_T^*) = X_R(s_R)\beta_R + X_T(s_T^*)\beta_T + \eta(s_R - s_T^*)'\Sigma(s_T^*)^{-1}(s_R - s_T^*) + z_R(s_R) + z_T(s_T^*), \quad (11)$$

$$z_R \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{z_R}), \quad z_T \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{z_T}). \quad (12)$$

- $z_R(s)$ and $z_T(s)$ are mean 0 GP's with covariance functions C_R and C_T , respectively. Exponential covariance functions are assumed for C_R and C_T , i.e., $C_R(u, u') = \sigma_R^2 \exp(-\phi_R \|u - u'\|)$ and $C_T(u, u') = \sigma_T^2 \exp(-\phi_T \|u - u'\|)$

CRIME PAPERS

- ▶ FORERO, A.M., BOHORQUEZ, M., RENTERIA, R.R. and MATEU, J. (2022). Identification of patterns for space-time event networks. **Applied Network Science**. doi: 10.1007/s41109-021-00442-y.
- ▶ BRIZ, A., MATEU, J. and MONTES, F. (2022). Modeling the influence of places on crime risk through non-linear effects: a comparison with risk terrain modeling. **Applied Spatial Analysis and Policy**. doi: 10.1007/s12061-021-09410-6.
- ▶ BRIZ, A., MATEU, J. and MONTES, F. (2022). Modeling accident risk at the road level through zero-inflated negative binomial models: a case study of multiple road networks. **Spatial Statistics**. doi: 10.1016/j.spasta.2021.100503.
- ▶ SHIROTA, S., GELFAND, A.E. and MATEU, J. (2022). Analyzing car thefts and recoveries with connections to modeling origin-destination point patterns. **Spatial Statistics**. doi: 10.1016/j.spasta.2020.100440.
- ▶ BRIZ, A., MATEU, J. and MONTES, F. (2022). Identifying crime generators and spatially overlapping high-risk areas through a non-linear model: a comparison between three cities of the Valencian region (Spain). **Statistica Neerlandica**, 76 (1), 97-120.
- ▶ RODRIGUES, A., GONZALEZ, J.A. and MATEU, J. (2021). A conditional machine learning classification approach for spatio-temporal prediction of crime data. **Submitted**.
- ▶ GONZALEZ, J.A., MATEU, J., CESPEDES, N., CAMACHO, E.A. and CERVANTES, L.C. (2021). A double stochastic point process approach for spatio-temporal dynamics and prediction of crime data. **Submitted**.
- ▶ MULLER, R., SCHUHMACHER, D. and MATEU, J. (2020). Metrics and barycenters for point pattern data. **Statistics and Computing**, 30 (4), 953-972.
- ▶ ZHUANG, J. and MATEU, J. (2019). A semi-parametric spatiotemporal Hawkes-type point process model with periodic background for crime data. **Journal of the Royal Statistical Society A**, 182 (3), 919-942.

Final comments

- ▶ Real-time surveillance in space-time
- ▶ Network large data: new challenges
- ▶ Computing challenge
- ▶ Solving real problems for society
- ▶ Thanks for your attention!!