Inspire Create Transform

UNIVERSIDAD EAFIT®

PH.D IN MATHEMATICAL ENGINEERING

DOCTORAL SEMINAR IV

May 17, 2019

CRIME PREDICTION USING MAHALANOBIS DISTANCE APPLIED IN VILLAVICENCIO (META)

Andrés Pérez-Coronado

Thesis Advisor:
Henry Laniado and Gustavo Canavire

EAFIT University
School of Sciences - Department of Mathematics Sciences
Ph.D in Mathematical Engineering
May 17, 2019

Outline

Research introduction

Data source
- The Metropolitan Police of Villavicencio
- Data collection

Mahalanobis distance
- Formulation
- Partial results

Outlook
- Model adjustment
- Next objectives

Research introduction

Criminal approach

- **Crime:** an action or omission which constitutes an offence and it is punishable by law.
- **Criminal:** an individual who has committed a crime.
- **Organized crime:** a structured network (criminals) whose primary objective is to obtain money through illegal activities (crimes).

Research question

Can a statistical model to predict criminal events, and the disruption of the criminal networks?

Research objectives

Main objective:

To design a non-parametric statistical model, for the space-time prediction of criminal events and the disruption of the criminal networks.

First specific objective:

To propose a non-parametric space-time model to predict the occurrence of criminal events, based on police data as calls to the emergency telephone number and documented crimes.

Data source

Data source: The Metropolitan Police of Villavicencio

Characteristics

- Gathers the municipalities of Villavicencio, Restrepo, Acacias and Cumaral.
- 452.472 population estimated.
- The main crime categories are thief, personal injuries, domestic violence, threats, burglary, illegal constraint, and shoplifting.

Data source: Data collection

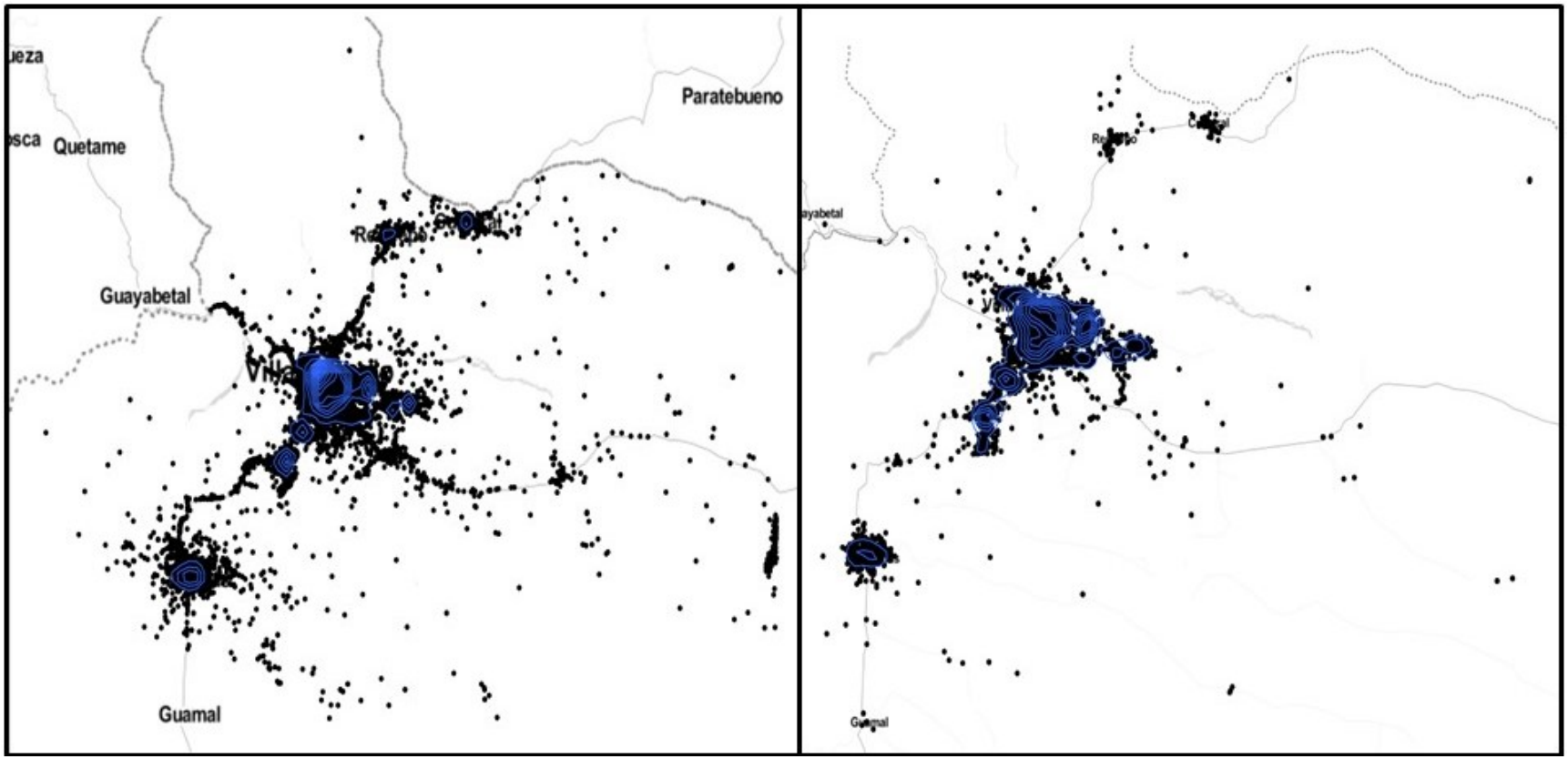
Information system of Statistics, Crime, Contravention and Operation (SIEDCO)

- SIEDCO is the biggest data source of reported crimes in Colombia.
- The data was collected information on all crimes documented between January 1, 2018 and February 3, 2019 ($n = 39.951$).
- Each crime record in our subset contained a time-stamp of occurrence, latitude/longitude coordinates of the crime at the city block level, and one of 32 types.

Information System for Case Tracking and Control (SECAD)

- It has the calls that report possible crime events or related issues about public safety.
- It is managed by Command Center of Citizen Control of the same city.
- The data was collected information on all crimes documented between October 1, 2018 and February 3, 2019 ($n = 9.985$).
- Each call records in our subset contained a time-stamp of occurrence, latitude/longitude coordinates of the crime at the city block level.

Crimes (SIEDCO) vs Calls (SECAD)



Model formulation

$$Pr(Label_p = T | f_1(p), ..., f_n(p)) = \frac{1}{1 + e^{-(\beta_0 + \prod_{i=1}^{n} \beta_i f_i(p))}} \qquad (1)$$

- $T =$ type of crime.
- $f_1(p) =$ density function.
- $f_2(p), \ldots, f_n(p) =$ topic modeling.
- $i = 1$, f_i equals the KDE.
- $i > 1$, f_i equals $Pr(i - 1|r)$.
- $r =$ is the unique topic neighborhood that spatially contains p.
- $\beta_i =$ coefficients.

Mahalanobis distance

Mahalanobis distance: Formulation

Definition

$$D_{Mh} = \sqrt{\left(x_i - x_j\right)^T \sum^{-1} \left(x_i - x_j\right)} \tag{2}$$

- x_i Known point
- x_j Unknown point
- $\sum^{-1}$ Covariance matrix

Distance-weighted spatial interpolation (IDW)

$$Pr_1(Label_p = T, W) =$$

$$\frac{\sum_{i=1}^{|N(p,W)|}(W - D_{Mh}) * Pr(Label_{n_i} = T)}{\sum_{j=1}^{|N(p,W)|}(W - D_{Mh})} \quad (3)$$

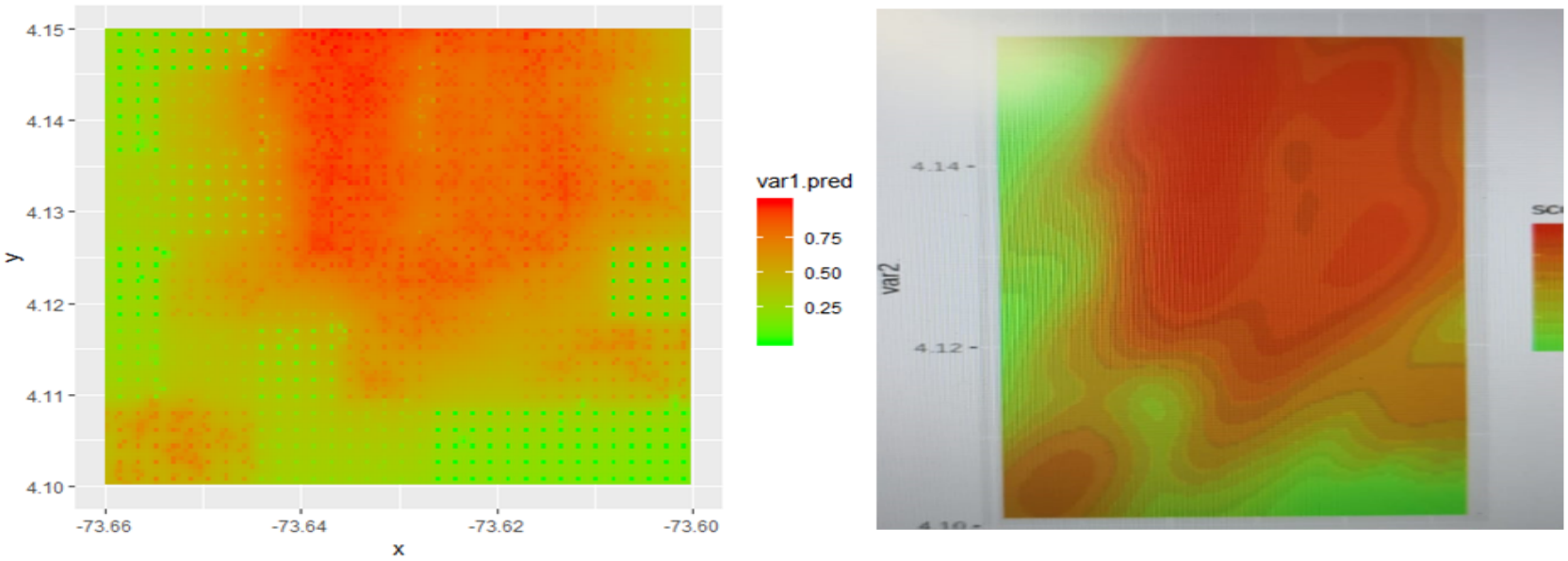
- Pr_1 = probability interpolation function.
- W = is a windowing parameter.
- $N(p, W)$ = is the set of p's neighbors within a distance of W (this set includes p itself).
- $D(p, n_i)$ = is the straight-line distance between p and one of its neighbors n_i.
- $(Label_{ni}) = T$ is the non-interpolated probability.

Advantages of the measure

- To reduce the outliers impact on the estimation.
- To include the dependence structure between variables longitude and latitude.
- This measure is invariant at scale.
- It is a statistical distance.

Mahalanobis distance: Partial results

Euclidean vs Mahalanobis



Outlook

Outlook: Model adjustment

Calibration of the model

- To try different alternatives for estimating the covariance matrix.
- To make a spatial cluster of the crimes by concentration zones.
- Reduce computation times and memory use.
- To explore the use of the GPU for calculations.

Outlook: Next objectives

Research objectives

Second specific objective:

To develop a predictive framework based on social network analysis to make disruption in criminal networks.

Third specific objective:

To propose a making decision process based on artificial intelligence for policing.

Thanks