



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY



MÉTODOS PARA LA PREDICCIÓN DE ROBOS VIOLENTOS

Ejercicio comparado para Montevideo, Uruguay

Federico Alejandro Veneri Guarch

Programa de Posgrado en Ingeniería Matemática
Facultad de Ingeniería
Universidad de la República

Montevideo – Uruguay
Julio de 2019



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY



MÉTODOS PARA LA PREDICCIÓN DE ROBOS VIOLENTOS

Ejercicio comparado para Montevideo, Uruguay

Federico Alejandro Veneri Guarch

Tesis de Maestría presentada al Programa de Posgrado en Ingeniería Matemática , Facultad de Ingeniería de la Universidad de la República, como parte de los requisitos necesarios para la obtención del título de Magister en Ingeniería Matemática .

Director de tesis:

D.Sc. Prof. Ricardo Fraiman

Director académico:

Mathias Bourel D.Sc.

Montevideo – Uruguay

Julio de 2019

Veneri Guarch, Federico Alejandro

MÉTODOS PARA LA PREDICCIÓN DE ROBOS
VIOLENTOS / Federico Alejandro Veneri Guarch. -
Montevideo: Universidad de la República, Facultad de
Ingeniería, 2019.

XIV, 87 p. 29, 7cm.

Director de tesis:

Ricardo Fraiman

Director académico:

Bourel D.Sc.

Tesis de Maestría – Universidad de la República,
Programa de Ingeniería Matemática , 2019.

Referencias bibliográficas: p. 53 – 59.

1. Policiamiento predictivo, 2. Procesos puntuales auto-
excitados, 3. Concentración del delito, 4. Policiamiento
en puntos calientes. I. Fraiman, Ricardo. II. Universidad
de la República, Programa de Posgrado en Ingeniería
Matemática . III. Título.

INTEGRANTES DEL TRIBUNAL DE DEFENSA DE TESIS

D.Sc. Prof. Alejandro Cholaquidis

D.Sc. Prof. José León

D.Sc. Prof. Ernesto Mordecki

Montevideo – Uruguay
Julio de 2019

A mi viejo, quien me contagió el
interés por las políticas públicas.

Agradecimientos

Quisiera agradecer a las autoridades del Ministerio del Interior y de la Agencia Nacional de Investigación e Innovación quienes apoyaron este proyecto ¹. En particular a Charles Carrera, Fabricio Fagundez, Ricardo Fraiman, Federico Laca y a todo el equipo de la Unidad de Análisis Criminal quienes pacientemente respondieron innumerables dudas.

Al Dr. Ricardo Fraiman, por aceptar embarcarse en este largo viaje como director de tesis y al Dr. Mathias Bourel, quien me guió como director académico.

Al Dr. Rodrigo Serrano-Berthet, quien me abrió las puertas para comenzar a trabajar en el sector de seguridad ciudadana e incentivó a continuar trabajando en esta línea.

A mi familia y amigos, a quienes aburrí durante casi dos años compartiendo los pequeños avances de este proyecto en cada reunión.

Especialmente a Cecilia y a Tau, mis coautores, quienes me acompañaron durante largas noches de programación, pruebas y redacción.

Este proyecto no podría haberse llevado adelante sin su apoyo. Infinitas gracias.

¹Este proyecto fue apoyado mediante la beca de Posgrado Nacional en su edición 2016 (POS NAC 2016 1 130833)

While the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician

Arthur Conan Doyle

*It makes about as much sense to
have police patrol routinely in
cars to fight crime as it does to
have firemen patrol routinely in
fire trucks to fight fire.*

Carl Klockars, 1983

RESUMEN

Los delitos son eventos que no se distribuyen de manera homogénea en las ciudades. Estos tienden a concentrarse en algunas unidades geográficas denominados puntos calientes. Por este motivo, las agencias policiales han adoptado estrategias focalizadas de patrullaje intentando priorizar ciertas zonas. La piedra angular de este tipo de estrategia es la correcta identificación de estas zonas.

Concentrándose en los robos violentos (rapiñas), este trabajo tiene como objetivo contribuir en dos aspectos. En primer lugar, caracterizar la evolución del delito y su comportamiento espacio temporal, y en segundo lugar comparar distintos métodos para la predicción y selección de zonas a patrullar.

Los resultados muestran que los delitos de rapiña presentan un alto nivel de concentración y un comportamiento de aglomeración espacio temporal. Una vez que un delito de rapiña es cometido, es probable que se observe otro a una distancia pequeña y en un breve periodo de tiempo. Esto puede responder a que algunas zonas son más atractivas para los ofensores (mecanismo endémico) y a la existencia de un patrón de contagio a zonas cercanas (mecanismo epidémico).

Basado en estos dos mecanismos, se presentan y compararan cinco métodos para la selección de zonas a patrullar: conteo, estimación de densidad, un método prospectivo y dos modelos endémico epidémico, uno que utiliza la presencia de factores de riesgo (eg: bancos, cajeros, bares) que vuelvan más atractivas ciertas zonas.

La comparación de los métodos en un día fijo y sobre una muestra aleatorias de días permitió identificar sus fortalezas y desventajas.

El método basado en conteo presentó mejores resultados en cuanto a su tasa de éxito, sin embargo, la dispersión de celdas seleccionadas para patrullar puede dificultar una estrategia de patrullaje necesitando una mayor cantidad de personal.

El método basado en la estimación de densidad seleccionó áreas más compactas, pero estas tienden a repetirse en el tiempo y presentó un menor nivel de tasa de éxito.

Los métodos prospectivos presentaron los mejores niveles de cobertura en

cuanto a celdas patrulladas, aunque también presentaron bajos niveles de tasa de éxito.

Los modelos endémico epidémicos aplicados en este trabajo se encuentran en una situación intermedia. Proporcionaron áreas de patrullaje compactas y celdas alejadas, con un nivel de tasa de éxito intermedio, por lo que pueden considerarse como una alternativa de compromiso entre eficacia predictiva y eficiencia de patrullaje. Es importante destacar que la introducción de covariables no represento una mejoría respecto a un modelo simple de contagio.

Este trabajo es el primero en realizar una comparación de distintos métodos para seleccionar zonas de patrullaje en Uruguay, abriendo una línea de investigación a futuro para desarrollar otros métodos que puedan contribuir a mejorar la seguridad ciudadana uruguaya.

Palabras claves:

Policiamiento predictivo, Procesos puntuales auto-excitados, Concentración del delito, Policiamiento en puntos calientes.

ABSTRACT

Crimes are not distributed homogeneously in cities. These events tend to concentrate on some geographic units called hot spots. For this reason, law enforcement agencies have adopted focalized patrol strategies trying to prioritize these areas. The cornerstone of this type of strategy is the correct identification of these areas.

Focusing on violent robberies (rapiñas), this thesis aims to contribute in two aspects. First, describing the evolution rapiñas and its spatiotemporal behavior, and secondly, compare different methods for predicting and selecting areas to patrol.

The results show that violent robberies have a high level of concentration and tends to cluster in space and time. Once a crime is committed, it is likely that another is committed at a small distance and in a short period. This may be because some areas are more attractive to the offenders (endemic mechanism) and there is a pattern of contagion to nearby areas (epidemic mechanism).

Based on these two mechanisms, five methods for selecting patrol zones are presented and compared: counting, density estimation, a prospective method and two endemic-epidemic models, one of which uses the presence of risk factors (eg: banks, ATMs, bars) that could make certain areas more attractive.

The comparison of these methods, first for a fixed day and then using a random sample of days, allowed the identification of their strengths and disadvantages.

The method based on counting presented better results in terms of its success rate; however, the dispersion of selected cells to patrol can hinder a patrol strategy because it may require a larger number of officers. The method based on density estimation selected compact areas, which is beneficial for a patrolling strategy, but these zones tend to repeat over time and presented lower success rate.

The prospective methods presented the best levels of coverage in terms of patrolled cells, although they also presented a low success rate.

The endemic epidemic models presented in this thesis provided compact

patrol areas and remote cells, with an intermediate level of success rate, so they could be consider a compromise between predictive effectiveness and patrol efficiency. It is important to note that the introduction of covariates did not imply a substantial improvement over a simple contagion model.

This document is the first to compare prediction methods for selecting patrol areas to prevent violent robberies in Uruguay, opening a line of research in the future to develop other methods that can contribute to improving Uruguayan public safety.

Keywords:

Predictive policing, Self-exciting point processes, Crime concentration, Hot spot policing.

Tabla de contenidos

Lista de símbolos	XIII
Lista de siglas	XIII
1 Introducción	1
2 Antecedentes	3
2.1 Concentración del delito	3
2.2 El policiamiento en puntos calientes	5
2.3 Policiamiento predictivo y su aplicación en Uruguay	6
2.4 Procesos Puntuales aplicados al delito	8
3 Metodología	11
3.1 Naturaleza de los datos y el problema de predicción	11
3.2 Métodos basados en conteo y estimación de densidad	12
3.2.1 Conteo por cuadrante (Conteo)	13
3.2.2 Estimación de densidad (KDE)	13
3.2.3 Métodos prospectivo (PHM)	14
3.3 Proceso Puntuales	15
3.3.1 Procesos puntuales espaciales	15
3.3.2 Proceso puntuales espacio temporales	19
3.3.3 Proceso puntuales espacio temporal auto-excitados	22
3.3.4 Estimación modelo	24
3.4 Evaluación de las predicciones	26
4 Aplicación para Montevideo	30
4.1 Fuentes de Información y variable de interés	30
4.2 Análisis preliminar de los datos	32
4.3 Estrategia empírica y estimación	37

4.4	Resultados	39
4.4.1	Predicción para un día pre fijado	39
4.4.2	Predicción sobre una muestra aleatoria de días	45
4.5	Discusión	46
5	Consideraciones finales	50
	Referencias bibliográficas	53
	Glosario	59
	Apéndices	60
Apéndice 1	Librerías y software utilizado.	61
Apéndice 2	Curvas de concentración del delito	63
Apéndice 3	Huellas de patrullaje	65
Apéndice 4	Código en R.	68
4.1	Preparación de base de datos	68
4.2	Análisis descriptivo	72
4.2.1	Funciones para descriptivos y utilitarias	72
4.2.2	Análisis descriptivos	72
4.3	Comparación de métodos.	75
4.3.1	Funciones para estimación y utilitarias	75
4.3.2	Funciones para estimación y utilitarias	79
4.3.3	Predicciones sobre días aleatorios	81
4.3.4	Código para ejemplos	86

Capítulo 1

Introducción

Los delitos son eventos que no se distribuyen de manera homogénea en las ciudades. Estos tienden a concentrarse en algunas unidades geográficas denominados puntos calientes, los cuales tienden a perdurar en el tiempo. Uruguay ha sido históricamente considerado un país seguro en el contexto regional. Sin embargo, la seguridad pública ha ganado terreno en la agenda como una de las principales preocupaciones de los ciudadanos. De acuerdo a la primera Encuesta Nacional de Victimización (ENV), en 2017, el 21.7% declaraba a la inseguridad como su principal preocupación. Aproximadamente 1 de cada 4 uruguayos declaró haber sido víctima de un delito durante 2016 y un 52% considera que el país es inseguro.

Las últimas comparaciones internacionales mostraron que en 2015, Uruguay presentaba una tasa de homicidio (8.42 cada 100 mil habitantes) por debajo de la media regional (16.87 cada 100 mil habitantes) aunque presenta valores elevados en cuanto a delitos contra la propiedad¹. Con 3043.61 hurtos y 566.01 robos cada 100 mil habitantes, Uruguay se situaba primero y quinto en ambos delitos².

El robo con violencia o rapiña³ ha sido el delito que ha presentado mayor dinamismo durante la última década y su reducción es uno de los principales objetivos planteados por las autoridades del Ministerio del Interior (MI) y la

¹Las comparaciones internacionales se dificultan ya que los países toman distintas definiciones para los delitos contra la propiedad. En términos generales, es aceptado que el hurto refiere a sustraer ilegalmente un bien sin el consentimiento del propietario y sin incurrir en violencia o amenazas. Por el contrario, el robo refiere al uso de violencia o amenaza para obtener el bien.

²Elaboración propia en base a UNDOC, datos 2015

³El código penal Uruguayo en su artículo 344 establece que la rapiña es un delito contra la propiedad caracterizado por el uso de la violencia o amenazas para apoderarse del bien

Policía Nacional Uruguaya (PNU) .

Fruto de ese esfuerzo, se creó el Programa de Alta Dedicación Operativa (PADO), que comienza a funcionar en abril de 2016 en Montevideo, Canelones y San José. Este programa adopta la modalidad de policiamiento en puntos calientes, una estrategia policial que consiste en patrullar zonas de alta concentración delictivas y ha demostrado ser exitosa (Chainey et al.,2018)

Para seleccionar las zonas a patrullar, los analistas criminales utilizan diversas herramientas para identificar zonas de alta concentración como conteo por unidades geográficas o estimaciones de densidad.

Este trabajo tiene como objetivo la comparación y evaluación de distintas técnicas de predicción usadas habitualmente en la literatura, poniendo especial énfasis en la utilización de procesos puntuales. Mediante una aplicación en Montevideo, la capital de Uruguay, se espera aportar evidencia sobre cuáles son las mejores herramientas para predecir el delito y guiar una estrategia de despliegue en territorio, colaborando en el refinamiento del trabajo de selección de zonas a patrullar.

El documento se organiza en 5 capítulos. El capítulo 2 presenta un resumen de los antecedentes en cuanto a evolución de delitos, estrategias policiales y policiamiento predictivo. El capítulo 3 reseña las herramientas y métodos utilizados en este trabajo, mientras que el capítulo 4 presenta su comparación para el año 2018 en Montevideo. Finalmente, el capítulo 5, presenta una discusión sobre los resultados encontrados.

Capítulo 2

Antecedentes

En este capítulo se detallan algunos antecedentes y hallazgos empíricos sobre la distribución espacio temporal de los delitos (sección 2.1) así como su influencia en las estrategias policiales (sección 2.2). En la sección 2.3 se introduce el concepto de policiamiento predictivo y su estado en Uruguay. Si bien existe un amplia literatura en cuanto a la aplicación de herramientas de predicción del delito, la sección 2.4 presenta los principales antecedentes sobre la utilización de procesos puntuales para la predicción del delito y algunos referentes a las técnicas con las cuales serán comparadas.

2.1. Concentración del delito

Los delitos son eventos que no se distribuyen de manera homogénea en las ciudades. Estos tienden a concentrarse en algunas unidades geográficas denominados puntos calientes, los cuales tienden a perdurar en el tiempo. Este hecho empírico llevó a [Weisburd \(2015\)](#) a enunciar lo que denominada ‘ley de concentración del delito’, un pequeño porcentaje de territorios concentra una gran proporción de los delitos. En su revisión, [Lee et al. \(2017\)](#) analizan diversos estudios realizados durante 1970 y 2015. A pesar que los estudios difieren en la unidad geográfica seleccionada (e.g.: segmento de calles, manzanas censales, grilla), en su forma de medir el crimen (e.g.: llamados de emergencia o reportes policiales) y el tipo de crimen estudiado, los autores sostienen que no hay dudas que los delitos se concentran en una pequeña proporción de lugares. Sin embargo, en su meta análisis encuentran que estos valores tienden fluctuar según tipo de crimen, área geográfica y forma de medición del delito por lo que

debe interpretarse como una descripción de la distribución de la concentración más que una ley fuerte (i.e.: un porcentaje x de calles concentran el y por ciento de los delitos). Estudios regionales, han mostrado que esta ley también es válida para América Latina. [Ajzenman and Jaitman \(2016\)](#) estudian la concentración en diferentes ciudades (Belo Horizonte, Bogota, Montevideo, Sucre, Zapopan) estableciendo que aproximadamente el 50 % de los delitos se encuentra concentrado entre el 3 % y el 7.5 % de los segmentos de calle.

Esta concentración elevada toma la forma de puntos calientes crónicos y tiene una explicación en la teoría criminológica ([Gorr et al., 2017](#)). El enfoque de la elección racional ([Clarke and Cornish, 1985](#)) plantea que detrás de la actividad criminal existe una evaluación racional entre costos y beneficios. Mientras que la teoría de la actividad rutinaria desarrollada por los criminólogos [Cohen and Felson \(1979\)](#) establece que para que un delito ocurra, deben estar presentes 3 elementos: (1) delincuentes motivados, (2) objetivos adecuados (3) ausencia de personas que puedan intervenir (e.g.: policías, vecinos). La teoría de patrones de crimen presentada por [Brantingham and Brantingham \(1984\)](#), se sitúa en la intersección entre las dos teorías anteriores ([Chainey and Ratcliffe, 2013](#)) y plantea que en la búsqueda de oportunidades, los posibles ofensores tenderán a mantenerse en su zona confort. Bajo estos mecanismos, es esperable observar que algunas zonas puedan presentar mayores oportunidades y pueden convertirse en un área de confort para posibles ofensores.

Bajo un enfoque de prevención del delito, los puntos calientes son entonces un síntoma de otros factores los cuales deben ser el foco del análisis ([Kennedy et al., 2012](#)). La modelización del riesgo del terreno (Risk Terrain Modelling) trae estos factores al primer plano, buscando identificar cuáles son las características geográficas que contribuyen a un mayor riesgo de ocurrencia de un delito ([Kennedy and Caplan, 2010](#)). A modo de ejemplo, una concentración de agencias financieras y calles poco iluminadas puede contribuir a que esa zona presente una mayor cantidad de delitos.

La revisión de la literatura presentada por [Caplan and Kennedy \(2011\)](#) para robos en la vía pública establece algunos factores como agravantes del riesgo de ocurrencia del delito. La proximidad a zona de venta de droga y áreas de prostitución, incide positivamente por dos vías. La presencia de ofensores potenciales en búsqueda de dinero fácil y posibles víctimas que llevarán dinero u objetos valiosos. Por otro lado, de acuerdo a los autores, los ofensores perciben que las víctimas son menos propensas a reportar estos robos ya que

se involucraron con actividades ilegales.

Adicionalmente, la proximidad a centros de estudio, lugares de recreación y reunión presentan mayor movimiento de personas lo que puede atraer a ofensores. En particular lugares donde se sirva alcohol como pubs y bares atraen mayor cantidad de delito ya que los ofensores perciben menor riesgo al abordar una potencial víctima.

La proximidad a lugares de transporte público como terminales de ómnibus contribuye al incremento de delitos debido al mayor flujo de personas. Finalmente, la presencia de bancos, ATM o casas de cambios vuelven a las posibles víctimas más atractivas ya que los agresores pueden potencialmente obtener mayor rédito de un solo robo.

Un segundo hallazgo empírico de la literatura es lo que algunos autores definen como re victimización cercana en el tiempo (near-repeat victimization): el hecho de que un delito ocurra puede desencadenar un evento en el futuro. [Mohler et al. \(2011\)](#) realiza una analogía entre el crimen y los terremotos ya que un primer movimiento sísmico puede desencadenar réplicas cercanas en espacio y tiempo. En el mismo trabajo cita diferentes estudios criminológicos sobre hurtos en domicilios que plantean cómo el crimen se distribuye siguiendo un proceso similar al contagio de una enfermedad. Al registrarse un hurto en un domicilio eleva la probabilidad de que se cometa otro hurto en domicilios cercanos ([Johnson et al., 2007](#)), posiblemente porque los ofensores tienden a elegir las zonas que ya visitaron ([Bernasco et al., 2015](#)).

En síntesis y haciendo un paralelismo con la epidemiología, el fenómeno de no homogeneidad en el espacio y tiempo se debe a que existen factores asociados al entorno que constituyen el componente endémico del lugar y un segundo componente epidémico asociado a un proceso de contagio espacio temporal.

2.2. El policiamiento en puntos calientes

El enfoque de la economía del crimen plantea que los ofensores, desde una toma racional de decisiones, incorporan la información del costo asociado al cometer un delito, dado por la sanción que se les impondrá y la probabilidad de ser capturado ([Becker, 1968](#)). Desde este enfoque, el incremento de la probabilidad de ser capturados, ya sea por el incremento de la presencia policial o el aumento de su productividad, incrementará la percepción de riesgo observado por los posibles ofensores y contribuirá a la reducción del delito por medio de

la disuasión ([Chalfin and McCrary, 2017](#)).

La evidencia sobre la concentración del delito ha traído cambios en las estrategias policiales, llevando a enfocar recursos donde se concentra el delito, aumentando el riesgo percibido por los potenciales ofensores y protegiendo a los ciudadanos. Esta nueva estrategia toma el nombre de policiamiento en puntos calientes ([Weisburd and Telep, 2014](#))

[Braga et al. \(2014\)](#) realizan un meta análisis sobre 10 experimentos aleatorios de patrullaje en puntos calientes encontrando efectos positivos en la reducción del delito. En dicho trabajo se enumeran diversos estudios basados en metodologías menos robustas que también han identificado efectos positivos.

Este tipo de estrategias han sido incorporadas en Uruguay a través del Programa de Alta Dedicación Operativa (PADO), que comenzó en abril de 2016 en Montevideo y zona metropolitana. [Chainey et al. \(2018\)](#) establecen que durante el periodo de abril-diciembre, cuando se encontraba activo el programa se redujeron las rapiñas en Montevideo en un 7,2 % siendo las unidades tratadas las que contribuyeron en mayor medida al descenso del delito. Mediante una estrategia de diferencias en diferencias establecen el impacto del programa en una reducción en la tasa a la que ocurren las rapiñas del entorno del 22-23 % en los tratados.

En este contexto, donde las estrategias focalizadas contribuyen a la reducción del delito, surge la discusión de cuáles son las mejores herramientas para guiar la selección de las zonas a patrullar.

2.3. Policiamiento predictivo y su aplicación en Uruguay

La efectividad de los programas de puntos calientes está asociada a la capacidad de contar con métodos que puedan identificar los puntos de alta concentración del delito y sus modificaciones.

[Perry \(2013\)](#) define el policiamiento predictivo como la aplicación de técnicas cuantitativas para identificar posibles puntos de intervención policial, prevención de un nuevo delito o resolución de delitos que sucedieron en el pasado. El foco de este trabajo es la identificación de zonas para la intervención policial. Dado que los patrones de delito evolucionan con el tiempo, adoptar un enfoque de policiamiento predictivo basado en datos permite a las agencias

de policía contar con métodos adicionales para identificar patrones y cambios sutiles rápidamente (Fitzpatrick et al., 2019).

Existe una amplia gama de métodos que pueden clasificarse dentro de la taxonomía propuesta por Perry (2013): mapas de puntos calientes (hotspots), métodos de regresión, minería de datos (clusters, clasificación, reducción de dimensiones), repetición cercana de eventos (near-repeat), análisis espacio temporal y modelos de riesgo espacial. Si bien las estrategias son diversas, el objetivo buscado es el mismo, encontrar zonas donde la probabilidad de un evento futuro sea más alta.

Desde la Policía Nacional, en los últimos años se ha avanzado sustancialmente en el análisis criminal buscando focalizar recursos en el territorio de manera eficiente (Del Castillo, 2018). Si bien en el pasado los analistas utilizaban grandes unidades como departamentos o seccionales policiales, actualmente han incorporado análisis a nivel de mapa de grilla o agrupaciones más pequeñas como segmentos de calles, lo que permite obtener una mejor noción de la concentración del delito. Adicionalmente, se utilizan estimaciones de densidad basadas en kernels o mapas de calor, lo que permite subsanar los problemas de definir arbitrariamente una unidad geográfica. Es importante destacar que la policía nacional también ha incursionado en la utilización de herramientas más sofisticadas como el Predpol, un software basado en el desarrollo de Mohler et al. (2011) aunque ha dejado de utilizarlo dado que no encontró que este fuera más eficiente que el trabajo de sus analistas¹.

Desde fuera de la policía existen antecedentes nacionales sobre el estudio de la distribución espacial del delito. Veneri (2016) realiza una primera aplicación de modelos de riesgo espacial (RTM) para identificar factores que inciden en la concentración de rapiñas. Jacottet (2018) utiliza técnicas de reducción de dimensiones encontrando dos factores subyacentes a la distribución del delito, el primero asociado a los delitos que vulneran la integridad física y psicológica de las personas y el segundo asociados a los delitos contra la propiedad. El primer factor presenta valores más elevados en las zonas periféricas de la ciudad, donde se observa la presencia de hogares más vulnerables, menores resultados educativos en su población y menor inserción laboral, mayor proporción de adultos jóvenes, mayores niveles de fecundidad, mayor presencia de niños y adolescentes. Por otro lado, la prevalencia del segundo factor se encuentra asociado a los segmentos del sur de Montevideo donde se observa mayores niveles

¹https://www.180.com.uy/articulo/68790_la-sociedad-esta-conforme-con-el-pado

de bienestar. [Veneri and Regueira \(2018\)](#) estudian la intensidad de los delitos contra la propiedad a nivel de los segmentos censales diferenciando según su modalidad, rapiña o hurto, utilizando técnicas de regresión que incorporan la dimensión espacial. La densidad poblacional y una mayor proporción de jóvenes (14 y 29 años) están asociados a la intensidad de ambos delitos. Los mayores niveles educativos, que pueden estar funcionando como un proxy de ingresos, están asociados a una mayor intensidad de hurtos mientras que la desigualdad relativa está asociada a una mayor intensidad del delito de rapiña.

Si bien no se pretende hacer una revisión de todos los trabajos cuantitativos sobre crimen en Uruguay, es posible mencionar otros trabajos desde la óptica de la reincidencia criminal o modificaciones sobre las penas ([Bukstein and Montossi, 2009](#) , [Munyo and Rossi, 2015](#), [Gandelman and Munyo, 2015](#)), factores sociodemográficos que inciden en la evolución del delito ([Aboal et al., 2007](#)), el impacto de programas de transferencia condicionada sobre el delito ([Alonso, 2018](#)) y el efecto de intervenciones policiales ([Munyo et al. 2016](#), [Cid, 2018](#), [Chainey et al., 2018](#))

A pesar de no ser una revisión exhaustiva, cabe destacar que la evidencia sobre la efectividad de los métodos para predecir el delito en Uruguay es virtualmente inexistente por lo cual este trabajo significa un aporte a la literatura en Uruguay.

2.4. Procesos Puntuales aplicados al delito

La literatura sobre la aplicación de métodos para la identificación de puntos calientes es extensa y su revisión no es el objetivo de esta sección, sólo concentrándose en los procesos puntuales².

En cuanto a la utilización de procesos puntuales, el trabajo seminal de [Mohler et al. \(2011\)](#) estudia los hurtos a viviendas en el Valle de San Fernando (Los Ángeles), confirmando que en las viviendas cercanas a las unidades que sufrieron un delito se incrementa el riesgo de sufrir un delito futuro en un periodo corto de tiempo. Por este motivo proponen adaptar un modelo ETAS (Epidemic Type Aftershock-Sequences) para modelar el efecto de re-victimización cercana. Este tipo de modelos han sido utilizado en sismología para modelar el

²Para el lector interesado se sugiere [Eck et al. \(2005\)](#), [Chainey and Ratcliffe \(2013\)](#) [Perry \(2013\)](#) o la revisión reciente realizada por [Fitzpatrick et al. \(2019\)](#) sobre policiaamiento predictivo

comportamiento de terremotos, dividiendo los eventos entre eventos de fondo y otro tipo de eventos o replicas que son generados como consecuencia de los eventos iniciales.

Buscando evaluar la capacidad predictiva de su algoritmo, [Mohler et al. \(2015\)](#) presenta los resultados de dos experimentos aleatorios realizado en tres divisiones de la Policía de Los Ángeles (Estados Unidos) y dos divisiones de la policía de Kent (Reino Unido). El experimento comparó, bajo condiciones controladas, la capacidad de seleccionar zonas de patrullaje de los analistas y de los algoritmos, así como su capacidad para prevenir el delito.

En estas jurisdicciones, utilizaron su algoritmo para identificar veinte zonas a patrullar por turno y solicitaron a los analistas criminales que hicieran lo mismo. En el Reino Unido y una de las jurisdicciones de Estados Unidos, estas predicciones no fueron utilizadas para patrullar, evitando el sesgo asociado a la introducción del patrullaje, un efecto de delito prevenido.

En las restantes jurisdicciones, los oficiales recibieron aleatoriamente zonas identificadas por los analistas o por el algoritmo. La aleatorización permitió comparar el efecto sobre la reducción del delito asociadas cada herramienta. Considerando los delitos que sucedieron dentro de la zona seleccionada como predicho en la primera prueba, los resultados muestran que el modelo ETAS logro predecir entre 9.8 % (6.8 % analistas criminales) para la jurisdicción donde mejor se desempeñó y 4.7 % (2.1 %) para el peor caso. Esto indica que los algoritmos tienen una mejora de predicción³ de entre 1.4 y 2.2 veces que los analistas. En las jurisdicciones donde se implementó el patrullaje guiado por el algoritmo, se observó una reducción del 7.4 % en los delitos, mientras que el patrullaje guiado por los analistas no presentó reducciones significativas.

Es importante resaltar que este experimento fue el primero de su clase ya que no utilizaron información del pasado, lo que los autores denominan como una situación “estéril”, sino que compitieron en tiempo real con analistas criminales. En su revisión [Fitzpatrick et al. \(2019\)](#) indican que actualmente están realizando experimentos para la ciudad de Pittsburgh (Pennsylvania, Estados Unidos)

Además de estos dos trabajos, existen otros antecedentes relevantes de los mismos autores como la implementación de este tipo de modelos para predecir muerte de civiles en Irak ([Lewis et al., 2012](#)) o una ampliación que permite introducir como indicadores adelantado los delitos relacionados con el uso de

³El autor denomina mejora o Boost, al ratio entre ambas tasas de éxito

armas para predecir casos de homicidios ([Mohler, 2014](#)).

Otros autores también han implementado variantes para identificar zonas de mayor riesgo de hurtos a viviendas en Chicago y Nueva York ([Kajita and Kajita, 2017](#)) encontrando que estos modelos predicen mejor el delito. En la región [Dulce et al. \(2018\)](#), comparan la efectividad de estos modelos en la ciudad de Bogota (Colombia) y propone su implementación para seleccionar áreas de patrullaje. Finalmente, [Reinhart and Greenhouse \(2018\)](#) proponen utilizar estos modelos incorporando covariables socio demográficas. Un resumen de otras aplicaciones en otras áreas puede encontrarse en [Reinhart \(2018\)](#)

Capítulo 3

Metodología

En este capítulo se detalla la metodología utilizada en este trabajo. La sección 3.1 discute la naturaleza de los datos y el problema de predicción, la sección 3.2 presenta los métodos clásicos basadas en conteo y estimación de densidad, así como la propuesta de un método prospectivo. Finalmente, la sección 3.3 presenta una introducción a los procesos puntuales y los procesos puntuales auto excitados, herramientas descriptivas y su uso para la predicción del delito.

3.1. Naturaleza de los datos y el problema de predicción

Como se mencionó previamente, el policiamiento predictivo tiene como uno de sus objetivos la selección de zonas de intervención policial. Partiendo de un área de interés $S \subseteq R^2$ (latitud y longitud) y una ventana temporal $[0, T] \subseteq R^+$, el problema es seleccionar entre zonas $\{s_1, s_2, \dots, s_n\}$ de S ($\cup s_i = S$) un subconjunto para patrullar en algunos momentos de tiempo $\{t_1, t_2, \dots, t_n\}$ con $\cup t_j = T$ (Wang and Brown, 2012), siendo estos momentos una discretización del tiempo continuo en intervalos (e.g.: días, semanas, turnos)

La selección de estas zonas puede basarse en alguna medida de intensidad del delito esperado para ese momento de tiempo $\lambda(s_i, t_j)$, buscando cubrir las zonas que presenten mayor riesgo de ocurrencia de un delito.

Para la definición de las zonas, es posible utilizar como zonas agregaciones geográficas barrios, segmentos censales o manzanas. Sin embargo, la utilización de límites administrativos puede ofuscar la extracción de patrones crimi-

nales alterando los resultados obtenidos. El problema de la definición de áreas geográficas (modifiable areal unit problem o MAUP) es un problema de larga data en literatura de estadística espacial (Openshaw, 1983, Fotheringham and Wong, 1991) y ha sido reconocido como un desafío para el análisis de los patrones de delito (Chainey and Ratcliffe, 2013). Una posibilidad es introducir sistemas de grilla uniformes sobre el área de estudio para mitigar el problema. Un tamaño de grilla reducido puede favorecer la extracción de patrones, aunque trae aparejado el incremento del costo computacional (Eck et al., 2005, Chainey and Ratcliffe, 2013).

En cuanto a la definición de los intervalos de tiempo y momentos en los cuales se debe tomar la decisión de que unidades seleccionar, esto debe estar asociado a la necesidad de las agencias de policía pudiendo considerar intervalos mensuales, semanales, diarios o turnos policiales.

Para este trabajo se optó por un esquema de grillas uniforme (200 metros) y decisiones diarias de patrullaje. La sección 4.1 presenta detalles adicionales. A continuación se describen los métodos para establecer una medida de riesgo de ocurrencia de un delito $\lambda(s_i, t_j)$, distinguiendo los métodos tradicionales y los basados en procesos puntuales.

3.2. Métodos basados en conteo y estimación de densidad

Las policías han utilizado distintos métodos para identificar zonas de mayor intensidad criminal, entre los más populares se encuentran los mapas temáticos, los métodos de conteo por cuadrantes (Sección 3.2.1) o estimaciones de densidad (Sección 3.2.2). Estos dos métodos incorporan la dimensión temporal implícitamente mediante la elección del analista de una ventana temporal o pasado para evaluar la intensidad. Esta sección presenta una breve introducción a estos métodos y una versión reciente denominada mapeo prospectivo que incorpora pesos temporales, asignado mayor peso a los eventos recientes.

3.2.1. Conteo por cuadrante (Conteo)

El método más sencillo para establecer la intensidad criminal consiste en dividir la región de interés en una grilla homogénea o cuadrantes y contabilizar el número de eventos comprendidos en cada zona (Eck et al.,2005).

Baddeley et al. (2015) definen el método de cuadrantes para el estudio de la intensidad. Partiendo de una región de observación espacial S , el método consiste en dividir S en m sub regiones o cuadrantes idénticos (s_1, s_2, \dots, s_m) , la intensidad estará dada por:

$$\lambda(s_i) = n_j,$$

donde $n_j = N(s_j)$ indica el número de eventos que cayeron dentro del cuadrante.

La selección de zonas a patrullar estará dada por el ordenamiento de las zonas de acuerdo a la intensidad estimada. Corresponde notar que este método requiere una decisión del analista al determinar los eventos de que periodo desea considerar para la estimación.

3.2.2. Estimación de densidad (KDE)

La estimación de densidad o mapa de calor, es un método de interpolación que genera mapas suavizados que representan la distribución de densidad del delito en el área de estudio (Chainey et al., 2008).

Sea u un punto cualquiera en la zona de estudio, el método consiste en utilizar un radio de búsqueda para encontrar eventos cercanos y otorgarles un peso dado por una función. La intensidad estará dada entonces por:

$$\lambda(u) = \sum_{i=1}^n \kappa_h(u - x_i),$$

donde x_i son el vector de coordenadas geográficas de los eventos observados. La función $\kappa_h(\cdot)$ es el kernel que pondera los eventos, usualmente se utiliza el kernel gaussiano (Baddeley et al.,2015).

Este método se convirtió en una herramienta muy popular para las agencias de policía, demostrando su utilidad para identificar puntos calientes de concentración del delito (Chainey et al., 2008)

Para este trabajo se optó por evaluar la función $\lambda(\cdot)$ en los centroides de

cada grilla para asignar un valor a cada s_i y permitir una comparación entre los métodos.

La función tiene dos parámetros a definir, explícitamente plantea la utilización de un parámetro h que consiste en la ventana espacial o radio de búsqueda, así como la decisión de que periodo considerar para estimar la intensidad.

3.2.3. Métodos prospectivo (PHM)

Lo métodos anteriores presentan el inconveniente de no incorporar la dimensión temporal en su análisis, eventos que sucedieron el día o la semana anterior tienen el mismo peso si estos se encuentran a la misma distancia del punto donde se está evaluando la función $\kappa_h(\cdot)$. [Bowers et al. \(2004\)](#) proponen incorporar el tiempo desde que ocurrió el evento como un factor adicional al peso.

Partiendo de una grilla de puntos equiespaciados u , para cada punto se establece un radio de búsqueda. Los eventos que se encuentren a una distancia mayor de r' o hayan ocurrido a una distancia temporal mayor de t' se les otorga peso cero. Los puntos intermedios son pesados de acuerdo a una función $g(\cdot, \cdot)$, estando la intensidad dada por:

$$\lambda(u) = \sum_{i=1}^n g(t_i, r_i),$$

donde r_i es la distancia del evento al centroide de la grilla seleccionada y t_i es la distancia temporal del evento a la fecha donde se está estimando la función. Para este trabajo se utilizó la versión descrita en [Kajita and Kajita \(2017\)](#) para la función $g(t_i, r_i)$ para datos diarios:

$$g(t_i, r_i) = \frac{1}{(1 + t_i/(7\Delta_t))(1 + 2r_i/\Delta_x)}.$$

Siendo, Δ_x la distancia entre centroides o el ancho de la grilla y Δ_t los intervalos de tiempo utilizados. Adicionalmente se agrega la restricción que si $t_i > t'$ o $r_i > r'$ $g(t_i, r_i) = 0$ de manera de otorgar peso cero a los eventos muy alejados en el tiempo y espacio.

[Bowers et al. \(2004\)](#) proponen utilizar $r' = 400$ metros y $t' = 60$ días para los hurtos a vivienda. Para este trabajo exploramos puntos de corte a partir del estudio descriptivo de los datos.

3.3. Proceso Puntuales

Los crímenes pueden considerarse como realizaciones de un proceso puntual en el espacio. En lo que sigue se planteara la notación para procesos puntuales espaciales, así como su versión espacio temporal y sus propiedades. Finalmente se plantea el marco para modelar un proceso espacio temporal auto excitado y la introducción de covariables.

3.3.1. Procesos puntuales espaciales

Un proceso puntual espacial es un mecanismo aleatorio cuyo resultado es un patrón de puntos en el espacio, en particular un proceso puntual finito es un mecanismo cuyo resultado es un patrón de puntos con un número finito de elementos y para cualquier región B , el número de puntos que caen en esa región es una variable aleatoria $N(B)$ (Baddeley et al.,2015).

Estos procesos pueden caracterizarse a partir de sus propiedades de primer y segundo orden (Diggle, 2013).

Las propiedades de primer orden refieren a la intensidad del proceso, sea s un punto en la región S^1 la intensidad estará dada por

$$\lambda(s) = \lim_{\Delta_s \rightarrow 0} \frac{E(N(B(s, \Delta_s)))}{|B(s, \Delta_s)|}$$

Donde Δ_s es el tamaño de la región infinitesimal $B(s, \Delta_s)$ y $|B(s, \Delta_s)|$ el volumen de la región considerada. La intensidad puede interpretarse como la probabilidad que ocurra un evento en ese volumen infinitesimal.

Si un proceso fuera homogéneo, se esperaría que esta intensidad fuera constante para todos los s_i . Los métodos descritos en la sección anterior, en particular conteo por cuadrantes y estimaciones de densidad, son algunas de las herramientas que permiten identificar esta propiedad de un proceso (Baddeley et al., 2015).

La propiedad de segundo orden o correlación permite caracterizar la relación entre las intensidades de dos puntos s_i, s_j y se define como:

$$\lambda_2(s_i, s_j) = \lim_{\Delta_{s_i} \Delta_{s_j} \rightarrow 0} \frac{E(N(B(s_i, \Delta_{s_i})), N(B(s_j, \Delta_{s_j})))}{|B(s_i, \Delta_{s_i})||B(s_j, \Delta_{s_j})|}$$

¹En esta sección s refiere a un punto con coordenadas, no a una subregión de S

El estudio de estas propiedades permite caracterizar el proceso, indicando si los eventos tienen a agruparse, se repelen o son independientes.

Para su estudio es común utilizar la función de distribución radial (Pair correlation function) dada por:

$$g(s_i, s_j) = \frac{\lambda_2(s_i, s_j)}{\lambda(s_i)\lambda(s_j)}.$$

Un estimador de la función está dado por:

$$\hat{g}(r) = \frac{|S|}{2\pi r n(n-1)} \sum_{i=1}^n \sum_{j=1, i \neq j}^n \kappa_h(r - d_{i,j}) e_{ij}(r),$$

donde $d_{i,j} = |x_i - x_j|$ indica la distancia entre los eventos i y j , $\kappa_h(\cdot)$ es un estimador basado en kernel y $e_{ij}(\cdot)$ es un factor opcional de corrección por borde. Es posible utilizar como kernel la distribución normal multivariada aunque en la práctica es más habitual utilizar el kernel de Epanechnikov ([Baddeley et al., 2015](#)).

De la misma manera que la propiedad de primer orden puede interpretarse intuitivamente como la tasa de ocurrencia de un evento, la propiedad de segundo orden, indica la probabilidad de encontrar un evento a una distancia exacta de r , dividida por la probabilidad de un proceso homogéneo. Si el proceso fuera homogéneo, es decir equiespaciado, se esperaría que esta función fuera constante.

Alternativamente algunos autores sugieren utilizar la función $K(\cdot)$ que, en vez de la probabilidad a la distancia exacta ponderada, indica el número promedio de eventos en una bola de radio r . La función empírica está definida como:

$$\hat{K}(r) = \frac{|S|}{n(n-1)} \sum_{i=1}^n \sum_{j=1, i \neq j}^n \mathbb{1}\{d_{ij} \leq r\} e_{ij}(r).$$

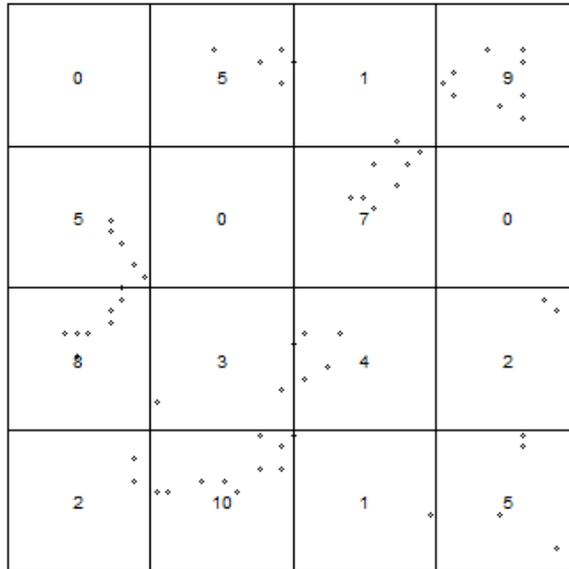
Para un proceso homogéneo la función teórica está definida por $K(r) = \pi r^2$, por lo cual es posible comparar la función teórica con la estimada para distinguir cómo evoluciona el proceso, siendo tres situaciones de particular interés. Si la empírica evoluciona a la par de la teórica, indica un proceso independiente. Si evoluciona por encima, indica un comportamiento de aglomeración, mientras que si evoluciona por debajo indica un patrón de regularidad o repulsión.

Otros autores prefieren utilizar una transformación de la K , $L(r) =$

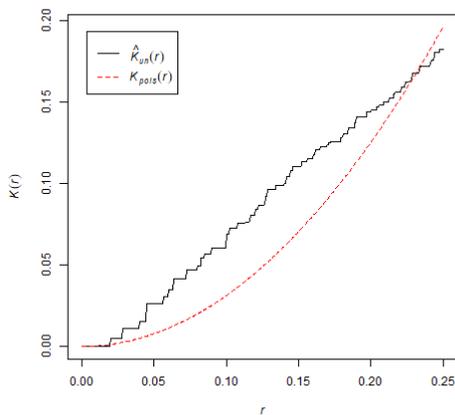
$\sqrt{K(r)/\pi}$, que para un proceso homogéneo de Poisson es una recta $L(r) = r$

Para mostrar el uso de estas herramientas, la figura 3.1 presenta un ejemplo sobre una base de datos clásica generada en Ripley (1977) en base al relevamiento de la ubicación de secuoyas disponible en Baddeley et al. (2004). Dado que estos árboles se reproducen mediante la dispersión de semillas en el viento, es esperable que se encuentren árboles cercanos en el espacio. El estudio de la intensidad revela que existen cuadrantes donde la presencia de árboles es mayor, lo que sugiere que el proceso puede no ser homogéneo. Las funciones \hat{g} y \hat{K} indican la probabilidad de encontrar otro árbol a una distancia exacta de r y el número esperado en una bola de radio r respectivamente. Se observa que ambas se distancian de las referencias teóricas homogéneas, y dado que \hat{K} presenta valores superiores a la teórica, presenta indicios de que los árboles tienden a agruparse en el espacio.

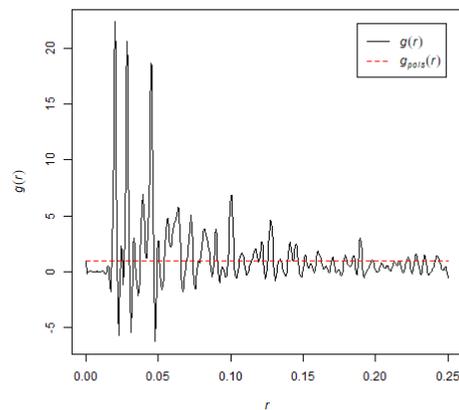
Figura 3.1: Ejemplo distribución espacial de secuoyas



(a) Distribución de Secuoyas y conteo por cuadrante



(b) Función \hat{K}



(c) Función \hat{g}

El conteo de cuadrantes permite estudiar la intensidad del proceso mientras que las funciones \hat{g} y \hat{K} indica la probabilidad de encontrar otro árbol a una distancia exacta de r y el número esperado una bola de radio r comparado a un proceso homogéneo respectivamente

3.3.2. Proceso puntuales espacio temporales

Los procesos espaciales antes descritos pueden expandirse para considerar la dimensión temporal y predecir la intensidad en un punto $s \in S \subseteq \mathbb{R}^2$ en el momento $t \in [0, T] \subseteq \mathbb{R}^+$. De la misma manera, pueden caracterizarse según sus propiedades de primer y segundo orden (Reinhart, 2018, Diggle, 2013).

La función de intensidad para un proceso espacio temporal toma la forma:

$$\lambda(s, t | H_t) = \lim_{(\Delta_s, \Delta_t) \rightarrow 0} \frac{E(N(B(s, \Delta_s) \times [t, t + \Delta_t]) | H_t)}{|B(s, \Delta_s)| |\Delta_t|}.$$

Donde $N(A)$ es el conteo de eventos del subconjunto $A \subseteq S \times [0, T]$ y $|B(s, \Delta_s)|$ es la medida de Lebesgue de la bola $B(s, \Delta_s)$ de centro s y radio Δ_s , H_t indica condicionalidad a la historia del proceso hasta el momento t . Intuitivamente la función de intensidad condicional indica el promedio de eventos por unidad de volumen en (s, t) .

La propiedad de segundo orden estará dada por:

$$\lambda_2((s_i, t_i), (s_j, t_j)) = \lim_{(D_s, D_j) \rightarrow 0} \frac{E(N(D_s)N(D_j))}{|D_i||D_j|}$$

Donde $N(D_i)$ el número de eventos en $B(s_i, \Delta_{s_i}) \times [t_i, t_i + \Delta_t]$ y $|D_i|$ su volumen. Análogamente se definen $N(D_j)$ y $|D_j|$. Esta función establece la relación entre pares de subconjuntos dentro de $S \times T$.

Al igual que en la versión espacial se puede utilizar la función de distribución radial y la función K (Gabriel et al., 2013).

La función de distribución radial está dada por:

$$g((s_i, t_i), (s_j, t_j)) = \frac{\lambda_2((s_i, t_i), (s_j, t_j))}{\lambda(s_i, t_i)\lambda(s_j, t_j)},$$

y un estimador:

$$\hat{g}(u, v) = \frac{1}{|S \times T|} \sum_{i=1}^n \sum_{i \neq j} \frac{1}{w_{ij} v_{ij}} \frac{k_s(u - ||s_i - s_j||) k_t(v - |t_i - t_j|)}{\lambda(s_i, t_i) \lambda(s_j, t_j)},$$

donde u y v indican la distancia espacial y temporal, con $k_u(\cdot)$, $k_t(\cdot)$ funciones kernel de espacio y tiempo respectivamente y w_{ij} v_{ij} son correcciones de borde. Al igual que su versión espacial puede interpretarse como la probabilidad de encontrar un evento a las distancias de u metros y v días.

[Gabriel and Diggle \(2009\)](#) extienden la función K a un proceso espacio temporal como:

$$K_{ST}(u, v) = \int_0^v \int_0^u g(u', v') u' du' dv'$$

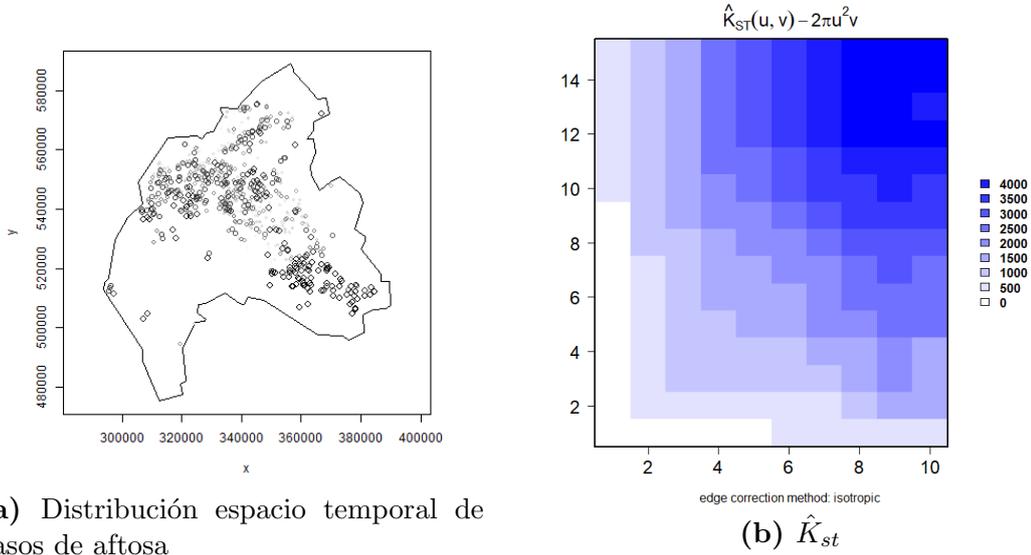
Una estimación no paramétrica está dada por:

$$\hat{K}_{ST}(u, v) = \frac{1}{|S \times T|} \frac{n}{n_v} \sum_{i=1}^{n_v} \sum_{j=1, j>i}^{n_v} \frac{1}{w_{ij}} \frac{1}{\lambda(s_i, t_i) \lambda(s_j, t_j)} \mathbb{1}\{|s_i - s_j| \leq u, |t_j - t_i| \leq v\},$$

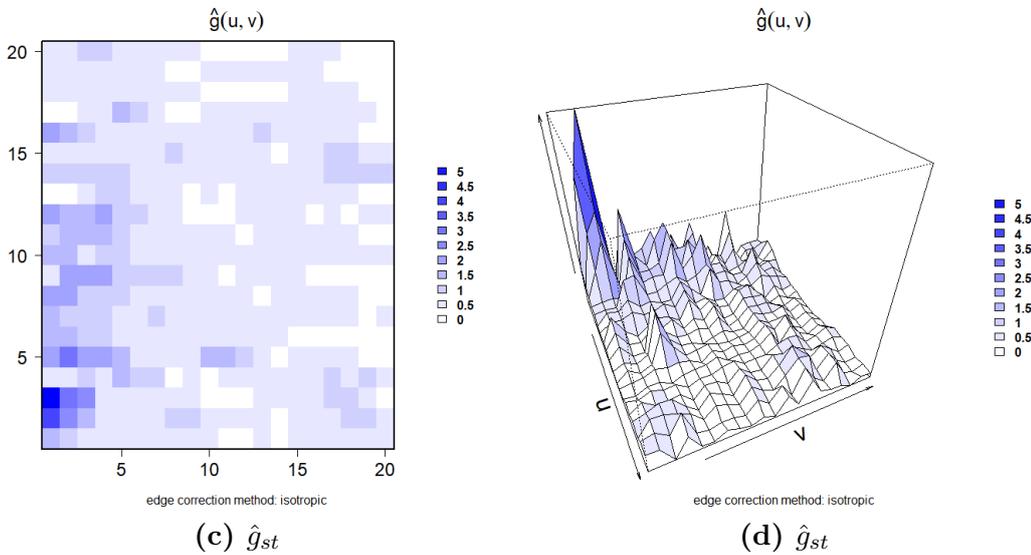
donde n_v indica la cantidad de eventos que cumplen $t_i \leq T_1 - v$, con $T = [T_0, T_1]$, y w_{ij} representa una corrección por borde.

Para un proceso homogéneo, $K_{ST}(u, v) = \pi u^2 v$. Es posible entonces estimar $\hat{K}_{ST}(u, v)$ y compararlo con el valor teórico al igual que en su versión espacial, si es mayor, esto indica que el proceso tiende a aglomerarse en el espacio y tiempo. La figura [3.2](#) reproduce el análisis presentado por [Gabriel et al. \(2013\)](#) donde los autores utilizan los reportes de fiebre aftosa registrados durante la epidemia de 2001 en el Reino Unido para ilustrar el uso de estas herramientas. Dado que se trata de una enfermedad altamente contagiosa es esperable observar concentración espacio temporal.

Figura 3.2: Ejemplo de distribución espacio temporal de casos de aftosa



(a) Distribución espacio temporal de casos de aftosa



(a) Indica la aparición de casos de aftosa, puntos claros indican que se trata de eventos más alejados en el tiempo. (b) Presenta la versión centrada de \hat{K}_{st} , su evolución por encima de la referencia indica un proceso de aglomeración. (c) Presenta la estimación de la función \hat{g}_{st} , se observa que la probabilidad de registrar otro evento de aftosa es más elevada para una distancia temporal corta siendo particularmente elevada para una distancia temporal de 2-3 días. (d) Presenta una variante del gráfico anterior.

3.3.3. Proceso puntuales espacio temporal auto-excitados

En 1971, Hawkes define un proceso temporal auto excitado a partir de su intensidad condicional.

$$\lambda(t|H_t) = v(t) + \sum_{i:t_i < t} g(t - t_i).$$

La aparición de eventos está asociada a una tasa $v(t)$ que puede evolucionar con el tiempo o ser constante y a una función $g(\cdot)$ que introduce la auto excitación del modelo. La aparición de un evento aumenta la intensidad del proceso y su contribución está dada por $g(\cdot)$, que en general se asume decreciente por lo que a medida que pasa el tiempo la intensidad disminuye. La figura 3.3 ilustra este mecanismos.

Figura 3.3: Ilustración de un proceso de Hawkes con caída exponencial

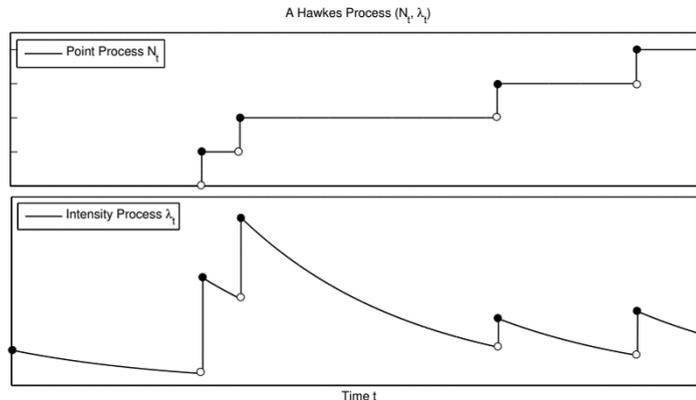


Figura extraída de Dassios et al., 2013. La figura superior presenta mediante puntos negros la llegada de un nuevo evento. La figura inferior presenta la evolución de la intensidad, la llegada de un evento tiene asociado el incremento temporal de la intensidad que luego decae exponencialmente.

La separación en intensidad de base y contribución mediante la función $g(\cdot)$ puede interpretarse desde la óptica de la sismología como un evento de base (terremoto) y sus réplicas. La figura 3.4 ilustra esta estructura de ramificación. En el caso de los delitos de rapiña, pueden separarse crímenes originales y reiteraciones asociados al mecanismo de re victimización cercana en el tiempo. Si bien no es posible identificar el origen de cada evento, es posible inferirlo mediante una modelización adecuada (Reinhart, 2018).

Figura 3.4: Ilustración de un proceso con eventos de base y replicas

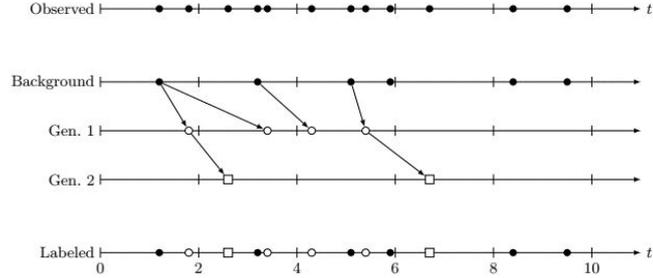


Figura extraída de [Reinhart, 2018](#). La línea superior presenta los eventos observado en un intervalo temporal $[0, 10]$. Estos eventos pueden separarse en eventos de base y sus réplicas, las flechas indican esta relación de contagio. La última línea presenta la estructura inferida.

En su revisión, [Reinhart \(2018\)](#) plantea distintos usos que se le ha dado a este modelo para eventos que tienden a ocurrir en rachas generando agrupaciones temporales como la llegada de transacciones bancarias, actividad neuronal o ataques terroristas.

La intensidad condicional de su versión espacio temporal auto excitado ([Mohler et al.,2011](#)) puede escribirse como:

$$\lambda(s, t|H_t) = \mu(s, t) + \sum_{i:t_i < t} g(s - s_i, t - t_i),$$

donde s_i indica la localización de los eventos y t_i el tiempo donde ocurrieron estos eventos. $\mu(s, t)$ indica entonces la intensidad de base del proceso en el punto s al momento t y $g()$ es la función de disparo o contagio que introduce la naturaleza auto excitada del proceso en el espacio y tiempo, es decir cómo los eventos inciden en futuros eventos. La función $g(s - s_i, t - t_i)$ puede considerarse separable en espacio y tiempo, es decir $g(s - s_i, t - t_i) = f(s - s_i)\nu(t - t_i)$ y en general se suponen que decaen con el tiempo ([Reinhart, 2018](#)).

[Meyer et al. \(2012\)](#), plantean un marco general para expresar la función intensidad condicional espacio temporal como dos componentes, uno endémico que opera como la intensidad de base del proceso y un segundo componente epidémico que representa el contagio y le da la naturaleza auto excitada al proceso.

$$\lambda(s, t|H_t) = h(t, s) + e(t, s).$$

El componente endémico $h(t, s)$ puede escribirse multiplicativamente, considerando un offset $\rho(t, s)$ que controle la intensidad espacio temporal. En el caso epidemiológico puede ser por ejemplo la población potencial en la región s al momento t y un predictor lineal $\beta'z(t, s)$, por ejemplo una tendencia o efectos estacionales, así como otras covariables asociadas al lugar s en el momento t .

$$h(t, s) = \rho(t, s) \exp(\beta'z(t, s)).$$

En el caso de la predicción de delitos, puede ser relevante plantear $\rho(t, s)$ como un efecto específico a la región y otro temporal por ejemplo día de la semana. Así mismo, es posible introducir la presencia de atractores del delito como presencia de bancos, bares u otros descritos en la sección 2.1. Para este trabajo se optó por la segunda versión como se describe en la sección 4.1.

El efecto epidémico $e(t, s)$ opera como el efecto contagio cuando se trate de modelar un proceso epidemiológico o la re-victimización cercana cuando se trata del modelado del delito. La función que dispara la aparición de un nuevo evento puede separarse en un efecto tiempo y otro espacio, quedando descripta para un evento j como:

$$e_j(t, s) = g(t - t_j)f(s - s_j) \text{ con } t > t_j.$$

Ambos se construyen de manera que el efecto contagio decaiga a medida que nos alejamos del evento original. El efecto temporal puede modelarse como un decaimiento exponencial $g(t - t_j) = e^{-\alpha t}$ o una función escalón, mientras que el efecto espacial puede establecerse como un kernel simétrico, por ejemplo una densidad normal bivariada u otra función similar. El efecto epidémico total está dado por la suma de las contribuciones de cada j evento.

3.3.4. Estimación modelo

El modelo antes descrito fue implementado por Meyer et al. (2017), definiendo *twistim* como un esquema para modelar la intensidad espacio temporal en base a dos componentes (endémico y epidémico).

La presencia de factores espaciales (i.e. presencia de bares, ATM, etc.) que inciden en la intensidad de base un efecto de contagio espacio-temporal, puede integrarse en la función de intensidad condicional como:

$$\lambda(t, s) = \exp(\beta^t z(s)) + \sum_{j \in I(s,t)} f(\|s - s_j\|)g(t - t_j).$$

El componente endémico es estimado por un predictor lineal $\exp(\beta^t z(s))$ que incorpora las características del lugar s , en el caso de la predicción del delito, la presencia de factores de riesgo como variables dicotómicas. El efecto epidémico o de contagio, está limitado por:

$$I(s, t) = \{j : t_j < t \wedge t - t_j < \tau \wedge \|s - s_j\| \leq \delta\}.$$

Es decir, un evento tiene un periodo de contagio potencial limitado a τ periodos desde su aparición y a una distancia máxima de δ unidades espaciales. Estos valores pueden establecerse de acuerdo a información previa de otros estudios o análisis descriptivos. Las funciones $f(\cdot)$ y $g(\cdot)$ regulan la intensidad del contagio y la contribución a la intensidad. Para este trabajo, se optó por un kernel gaussiano para el contagio espacial ($f(x) = e^{-\frac{x^2}{2\sigma^2}}$) y un decrecimiento exponencial para el contagio temporal ($g(t) = e^{-\alpha t}$).

La intensidad descrita anteriormente requiere la estimación de un vector de parámetros $\theta = \{\beta, \sigma, \alpha\}$ donde β es un vector de largo $p + 1$ asociado a los p factores de riesgo más una constante, σ y α los parámetros que regulan el decrecimiento del contagio. Partiendo del proceso puntual observado $x = \{(t_i, s_i) : i = 1, \dots, n\}$, siendo n la cantidad de eventos registrado, t_i el periodo cuando ocurrió el evento y s_i la posición geográfica, es posible estimar θ mediante máxima verosimilitud.

La log verosimilitud del proceso en $[0, T] \times S$ puede escribirse como²:

$$\sum_{i=1}^n \log \lambda_{\theta}(t_i, s_i) - \int_0^T \int_S \lambda_{\theta}(t, s) dt ds$$

La maximización se realiza mediante métodos de optimización numérica, en particular los autores implementaron la optimización de Broyden- Fletcher - Goldfarb - Shanno (BFGS).

Para este trabajo, se estimaron dos versiones de este modelo. Una primera versión homogénea en intensidad espacial con efectos de contagio y una segunda que incorporó la presencia de covariables que afectan la intensidad de base.

²El capítulo 7 de Vere-Jones (2003) una discusión sobre intensidad condicionales y verosimilitudes.

Luego de la estimación, la intensidad es evaluada en el centroide de cada celda.

3.4. Evaluación de las predicciones

Para evaluar la capacidad predictiva, se siguió una estrategia habitual en la literatura que busca comparar el porcentaje de delitos cubiertos en las zonas seleccionadas por los métodos. Formalmente para cada $t + 1$ se establecen como celdas a patrullar aquellas con mayor intensidad estimada y se calcula la tasa de éxito como:

$$TE_{t+1} = \frac{\text{Crímenes comprendidos en las zonas seleccionadas en } t + 1}{\text{Crímenes totales en } t + 1}$$

La métrica anterior, si bien es útil puede no ser suficiente ya que los métodos pueden requerir distinta cantidad de celdas para alcanzar la misma tasa de éxito. Una medida adicional debe considerar la precisión con lo que logran alcanzar ese indicador. Por este motivo, [Chainey et al. \(2008\)](#) propone utilizar el PAI (prediction accuracy index) que relaciona el número de eventos capturados por los puntos calientes identificados por el método y la precisión con la que se identificaron.

En primer lugar, plantea calcular el porcentaje de área cubierta (PA) para evidenciar la cantidad de patrullaje requerido por el método. En este trabajo, dado que se utilizó una discretización en grillas de Montevideo, es posible usar una simplificación asociado al porcentaje de celdas seleccionadas.

$$\begin{aligned} PA_{t+1} &= \frac{\text{Área de los puntos calientes identificados en } t + 1}{\text{Área total}} = \\ &= \frac{\text{número de celdas seleccionadas en } t + 1}{\text{Total de celdas}} \end{aligned}$$

Al combinar la tasa de éxito con el porcentaje de cobertura, deriva como indicador el PAI.

$$PAI_{t+1} = \frac{\text{Tasa de éxito } (TE_{t+1})}{\text{Porcentaje del área } (PA_{t+1})}$$

Este indicador tomara valores elevados en la medida que se logren identifi-

car con precisión los puntos calientes y que estos capturen un número elevado de eventos.

A modo de ejemplo, supongamos un área de estudio de 6x6 celdas y dos métodos A y B que seleccionaron zonas a patrullar marcadas en gris, en algunas de las cuales ocurrirá un delito denotado por X (Figura 3.5). El método A solamente logro capturar 2 de 4 delitos (TE=25 %), cubriendo 4 celdas (PA \approx 11 %). El método B, si bien logro una mayor tasa de éxito (TE=75 %) requirió también una zona mayor (PA \approx 17 %)). Utilizando el PAI, el método B presenta un PAI mayor (4.5) al del método A (2.25) por lo cual el método B sería seleccionado.

Figura 3.5: Ejemplo PAI

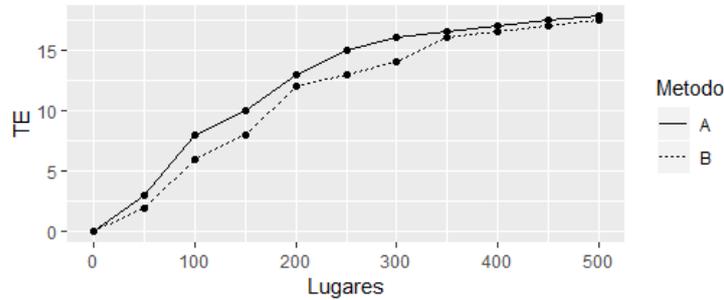
Método A						Método B							
	A	B	C	D	E	F		A	B	C	D	E	F
1	X						1	X					
2							2						
3			X	X	X		3			X	X	X	
4							4						
5							5						
6							6						

Ejemplo de aplicación de PAI sobre una grilla de 6x6. X indica que un delito fue cometido en la celda. El color gris indica que la celda fue seleccionada por el método

Otros trabajos han optado por herramientas de comparación que permiten incorporar otras dimensiones al análisis. La curva de retornos permite incorporar las restricciones de recursos de las agencias de policía relacionando la tasa de éxito y el porcentaje de cobertura o el número de lugares a patrullar. Brinda a los hacedores de política una representación gráfica de la relación costo beneficio de patrullar una celda adicional (Fitzpatrick et al. (2019), Gorr et al. (2017)). Es esperable que las curvas presenten rendimientos marginales decrecientes, es decir, cubrir un lugar adicional presente menores incrementos en la tasa de éxito. Dado un nivel fijo de lugares a patrullar, esta herramienta permite comparar distintos métodos. La figura 3.6 presenta un ejemplo para dos métodos A y B, en el caso de estar restringidos a 100 lugares, el método A se presenta como la mejor opción.

Otro factor importante en el diseño de una estrategia de patrullaje es la posibilidad de agrupar las celdas seleccionadas en zonas de patrullaje, de manera de reducir la cantidad de personal necesario, y las distancias entre ellas

Figura 3.6: Ejemplo de curva de retorno



La curva de retorno presenta en el eje de las ordenadas la tasa de éxito (TE) y en las abscisas el número de celdas seleccionadas. Permite comparar la tasa de éxito entre distintos métodos para un mismo número de celdas patrulladas.

tal que los oficiales puedan desplazarse rápidamente entre ellas. En el ejemplo, el método A, constituye una única zona de patrullaje compuesta de 4 celdas, mientras que el método B requiere de dos zonas. [Bowers et al. \(2004\)](#) proponen analizar la relación área/perímetro de estas agrupaciones, las zonas que presenten mayores ratios de este indicador pueden considerarse como más eficientes en términos policiales ya que permiten un desplazamiento más rápido a cualquier punto. Volviendo al ejemplo anterior, el método B tiene dos zonas de patrullaje, la primera con un área de 2 y un perímetro de 6 mientras que la segunda tiene un área de 4 y un perímetro de 8. Siguiendo esta lógica, la segunda zona es más eficiente ya que presenta un ratio de $1/2$ comparado al $2/6$ de la primera.

Para este trabajo proponemos utilizar el índice de contorno (Contour Index o CI) $CI = P/\sqrt{A}$ siendo P el perímetro y A el área. Esta métrica se ha sido utilizado en para el estudio de imágenes bi dimensionales en medicina ([Cuevas et al.,2007](#)). El valor mínimo se alcanza en el círculo y es $2\sqrt{\pi}$. En nuestro caso es deseable observar valores menores de este indicador lo que indicaría zonas de patrullaje más compactas. En el ejemplo anterior, la primera zona elegida por el método B tendría un CI de 4.24 mientras que la segunda tendría un CI de 4, indicando que es más compacta.

Los criterios anteriores pretenden evaluar los algoritmos de predicción desde el punto de vista de la eficacia y la eficiencia, es decir, que tan buenos son logrando sus objetivos y con qué recursos lo logran. Algunos autores plantean la necesidad de incorporar también la dimensión de la equidad. Por ejemplo [Gorr et al. \(2017\)](#) plantean que el patrullaje es un servicio que debe brindarse

a toda la comunidad. Para estudiar la distribución del patrullaje proponen utilizar la huella de patrullaje. Esta técnica consiste en mapear todas las zonas que han sido seleccionadas por el método durante el período de estudio.

Trabajos más recientes han sugerido la idea de utilizar un índice de entropía para puntos calientes como una medida de dispersión de las áreas seleccionadas Fitzppatrik (sin publicar). Para un método cualquiera, sea n_i el número de veces que una celda i es seleccionada a lo largo del periodo de estudio, se define p_i como la proporción que ocupa la celda i dentro del esfuerzo de patrullaje.

$$p_i = \frac{n_i}{\sum_j^N n_j}$$

Donde N es el número total de celdas. A partir de esta proporción el indicador de entropía se calcula como:

$$H = - \sum_{i=1}^N p_i \log_2 p_i$$

Este estadístico toma valores menores para métodos que seleccionen las mismas unidades en todos los periodos y valores más elevados cuanto menos tiendan a repetirse las zonas seleccionadas. Adicionalmente al componente de equidad, el hecho de seleccionar las mismas unidades de patrullaje puede tener aparejado otros aspectos negativos. Los ofensores podrían adaptarse al patrullaje policial rápidamente, reduciendo su poder disuasorio.

Capítulo 4

Aplicación para Montevideo

En este capítulo se detallan las fuentes de información utilizadas (sección 4.1) así como un análisis preliminar de los datos (sección 4.2). La sección 4.3 detalla la estrategia empírica seguida, mientras que la sección 4.4 presenta los resultados y la sección 4.5 las conclusiones .

4.1. Fuentes de Información y variable de interés

Para este trabajo se utilizó como fuente principal de información el Sistema de Gestión de Seguridad Pública (SGSP) del Ministerio del Interior. El SGSP conecta todas las agencias relacionadas a la seguridad pública en Uruguay, registrando todos los delitos denunciados en el territorio nacional.

Como aproximación al delito de robo violento, se consideró las denuncias de rapiña, es decir la sustracción de un bien mediante violencia o amenaza ¹. Los reportes del [Observatorio Nacional Sobre Violencia y Criminalidad \(2019\)](#) indican que “Se clasifican como Rapiñas todos los incidentes en que se sustrajo o intentó sustraer, por medio de la fuerza o amenaza de uso de la fuerza, cualquier objeto o propiedad, al cuidado o bajo la custodia de otra o varias personas”.

Corresponde notar que esta definición es más amplia que la presentada en otros estudios. [Caplan and Kennedy \(2011\)](#) plantea que deben considerarse únicamente los delitos cometidos en espacios públicos o cuasi públicos, excluyendo por ejemplo los robos al interior de vehículos. Adicionalmente plantea

¹La definición legal puede encontrarse en el artículo 344 del Código Penal Uruguayo

que los bienes robados deben pertenecer a personas y no a empresas u otro tipo de agencias.

La definición legal uruguaya no establece distinción respecto a la víctima o el lugar del hecho. Si bien cada registro cuenta con información complementaria y texto asociado al parte policial que podría ser explotada, no existe una sub categorización adecuada actualmente.

Para cada evento, el registro permite conocer el lugar aproximado del hecho, así como la fecha y hora, de acuerdo a la información proporcionada por el denunciante o policía interviniente. Esta información puede considerarse como una aproximación para construir coordenadas geográficas y una marca temporal asociada al evento.

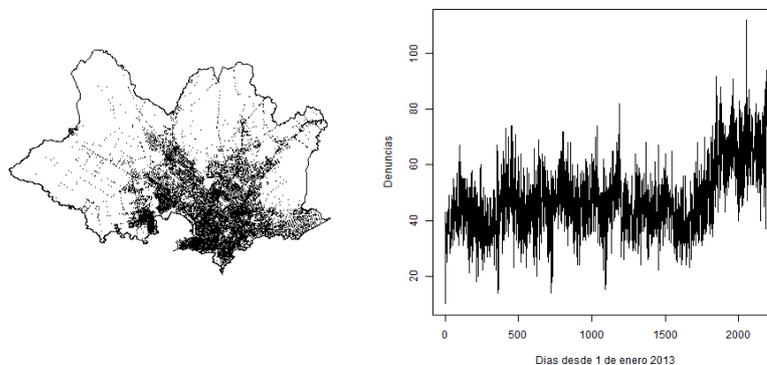
Partiendo de todas las denuncias asociadas a delitos de rapiña cometidos entre 2013 y 2018 en Uruguay, solo fueron considerados los eventos que contaban con coordenadas geográficas válidas pertenecientes al departamento de Montevideo. Dado que la hora reportada es una aproximación, para este trabajo se limita la marca temporal al día reportado por la víctima.

Este procedimiento generó una base de 103463 eventos para los cuales se cuenta con coordenadas geográficas y fecha del evento lo que permite una representación en el tiempo (Figura 4.1b) y en el espacio (Figura 4.1a). Se observa que los delitos se encuentran concentrados en el sur del departamento de Montevideo y presentan un incremento sobre el final del periodo de estudio. Un análisis descriptivo se presenta en la sección 4.2

Para la delimitación geográfica del estudio se utilizó información cartográfica generada por el Instituto Nacional de Estadística (INE) creada en el marco del censo 2011. A partir de esta información, se generó una grilla de 13740 celdas de 200 metros de ancho las cuales serán usadas en las siguientes secciones y se calculó el centroide de cada una de ellas.

Para la construcción de factores que pueden estar asociados a una mayor presencia de delitos, se utilizó la información relevada por la Fundación Open StreetMap ([OpenStreetMap Foundation contributors, 2019](#)). En particular se extrajo información sobre la presencia de actividad comercial (ATM, bancos, casas de cambios y estaciones de servicio) entretenimiento (bares, pubs y restaurantes), centros de salud (hospitales y clínicas), así como centros educativos distinguiendo su modalidad (Universitarios, terciarios y escuela/liceo). Adicionalmente se incorpora la presencia de industrias en zonas designadas como industriales.

Figura 4.1: Representación espacio y tiempo



(a) Representación Espacial (b) Representación temporal

(a) Cada punto presenta un evento de rapiña y su ubicación geográfica entre 2013 y 2018. (b) Presenta la evolución del número de rapiñas diarias denunciadas en Montevideo.

A partir de la información presentada en el Sistema de Información Geográfico de Montevideo², se incorporó la información asociada a la presencia de asentamientos.

Corresponde notar que lo anterior no pretende ser una lista taxativa de factores que pueden incidir en la mayor presencia de actividades criminales, simplemente se trata de un análisis exploratorio.

Para construir zonas influencia espacial de estos factores, se optó por un enfoque similar al utilizado por los modelos de riesgo espacial (RTM). Se consideró una zona de influencia de 300 metros a partir del centroide de cada celda, si alguno de los factores descritos anteriormente se encuentran en esta zona, se registró la presencia del factor como una variable dicotómica.

4.2. Análisis preliminar de los datos

Para una primera aproximación se consideró como unidad de análisis las celdas en las que se dividió Montevideo, calculándose algunos estadísticos de interés (Tabla 4.1). Entre 2013 y 2018 se observa un incremento del total de rapiñas denunciadas pasando de 14561 eventos registrados a 23490, se destaca la caída intermedia registrada en los años 2016 y 2017, así como el importante

²<http://sig.montevideo.gub.uy/>

aumento de 2018. A la par del aumento del delito, se incrementó el número de celdas que fueron afectadas por al menos un delito en 4 puntos porcentuales entre 2013 y 2018. En 2013, mientras por año se observaron 1.06 rapiñas por celda, en 2018 esto se incrementó a 1.71 así como se incrementó la dispersión medido por el desvío estándar.

La última columna presenta el índice de Gini considerando como ‘ingreso’ el número de rapiñas por celda. En caso de una perfecta igualdad en la distribución, este estadístico tomaría el valor 0, mientras que valdrá 1 si una celda concentra todas las rapiñas. Los valores de .87 en 2013 indican que el delito se encuentra más cercano al escenario de concentración, aunque en años recientes se ha reducido la concentración del delito.

Para verificar la disminución de la concentración se calcularon curvas de concentración del delito las cuales se presentan en el apéndice 2 y los resultados principales en la tabla 4.2. En 2013 el 25 % de los delitos reportados estaban concentrados en aproximadamente el 1.48 % de las celdas, mientras que el 50 % lo hacían en 4.45 %. Para 2018 estos valores se situaron en 1.73 % y 5.28 %. A pesar de una disminución de la concentración del delito de rapiña, se observa que se trata de un crimen altamente concentrado.

Tabla 4.1: Estadísticos descriptivos a nivel de celdas

	TR	PC (%)	P	DS	G
2013	14561	24.44	1.06	2.99	0.87
2014	16443	25.30	1.20	3.25	0.87
2015	17141	25.49	1.25	3.53	0.87
2016	16166	25.53	1.18	3.13	0.86
2017	15662	25.47	1.14	3.08	0.86
2018	23490	28.43	1.71	4.41	0.85

TR = Total de rapiñas registradas en el año. PC=Porcentaje de celdas afectadas por al menos un delito. P= Promedio de rapiñas por celda. DS= Desvío estándar de rapiñas por celda. G= Coeficiente del índice de Ginni utilizando las celdas como unidades y el número de rapiñas como “ingreso”.

A la par del aumento del delito, se observó un incremento de las celdas afectadas y una reducción de la concentración del delito.

Una primera aproximación al efecto de factores que pueden contribuir a la mayor concentración del delito se presenta en la tabla 4.3. Se destacan que el 30.72 % de los delitos de rapiñas denunciados en 2018 se cometieron en las zonas de influencia de una escuela o liceo, un 30.72 % en las inmediaciones de

Tabla 4.2: Concentración del delito a nivel de celdas.

	Rapiñas: 25 %	Rapiñas: 50 %
2013	1.48	4.45
2014	1.48	4.64
2015	1.40	4.53
2016	1.62	4.91
2017	1.55	4.80
2018	1.73	5.28

Rapiñas 25 %: indica el porcentaje de celdas que acumulaban el 25 % de las rapiñas. Rapiñas 50 %: indica el porcentaje de celdas que acumulaban el 50 % de las rapiñas. Se observa una reducción en la concentración del delito de rapiña en 2018 comparado a años anteriores.

un restaurante y un 25.42 % en las inmediaciones de asentamientos.

Buscando caracterizar el proceso se calcularon estadísticos que permiten aproximar sus propiedades como se describieron en la sección 3.3. Para esto, se utilizaron los delitos del año 2018.

Si las rapiñas respondieran a un proceso homogéneo es esperable que la intensidad fuera constante en el espacio, es decir, no existirían áreas de mayor intensidad. La figura 4.2a presenta la estimación de intensidad para el periodo mediante kernels, dando indicios de que no es constante en el espacio. Las restantes figuras (4.2b,4.2c,4.2d) presentan estimaciones de intensidad para tres momentos de 2018, lo que permite observar que estos puntos de mayor concentración varían con el tiempo.

Buscando describir las propiedades de segundo orden del proceso, se estimaron las funciones de correlación espacial K y L (ver sección 3.3.1). La función de correlación 4.3c indica que hay mayor probabilidad de encontrar un evento a una distancia muy pequeña, 50 y 100 metros comparado a un proceso homogéneo. Las figuras 4.3a y 4.3b presentan las estimaciones de la función K y L respectivamente. Se observa que en ambos casos la función estimada evoluciona por encima de la función teórica lo que indica un proceso de agrupamiento en el espacio.

Para validar la existencia de este comportamiento de agrupamiento tanto en espacio como en tiempo se estimaron las funciones K_{st} y g_{st} (ver sección 3.3.2). Se presenta la estimación para los primeros 180 días de 2018³.

³Por costos computacionales se optó por utilizar esta ventana temporal

Tabla 4.3: Delitos de rapiña en inmediaciones de factores de riesgo.

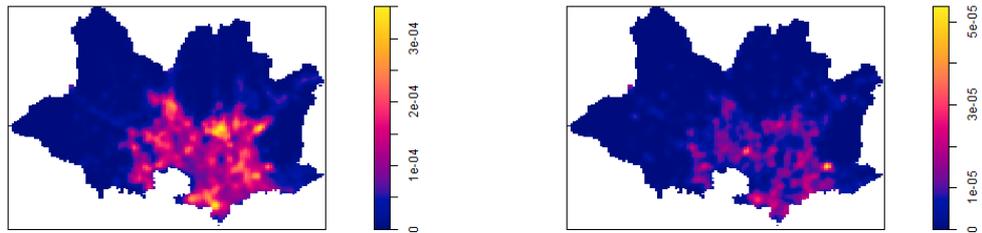
Agrupación	Factor	Porcentaje de los delitos (%)
Actividad comercial	ATM	15.38
	Bancos	10.72
	Estaciones de servicio	20.87
	Cambio	11.28
Centros de salud	Hospitales	5.63
	Clinicas	9.65
Entrenimiento	Bar	10.95
	Pub	3.89
	Restaurantes	29.76
Centros educativos	Escuela/liceo	30.72
	Terciario	2.38
	Universitario	7.19
Uso de tierra	Industrial	1.28
	Asentamientos	25.42

La columna porcentaje de delitos indica el porcentaje de rapiñas que se encontraban en la zona de influencia de estos factores. Algunos delitos pueden encontrarse en la zona de influencia de dos factores por lo que la columna no suma 100

La evolución de la función K_{st} por encima de la versión homogénea de un proceso espacio temporal indica que las rapiñas presentan un comportamiento de aglomeración (Figura 4.4) tanto en el espacio como en el tiempo. Por otra parte, la función de distribución g_{st} (Figura 4.5) permite caracterizar esta aglomeración. Se observa una correlación temporal elevada que perdura en el tiempo para una distancia muy corta, lo cual puede estar asociado a que la misma zona sufre un delito de rapiñas reiteradas veces, lo que concuerda con la elevada concentración observada previamente. El efecto de zonas cercanas también presenta una correlación positiva, siendo esta mayor en el entorno de los 300 metros, aunque parece disminuir con el pasaje del tiempo siendo más fuerte dentro de los 10 días del evento.

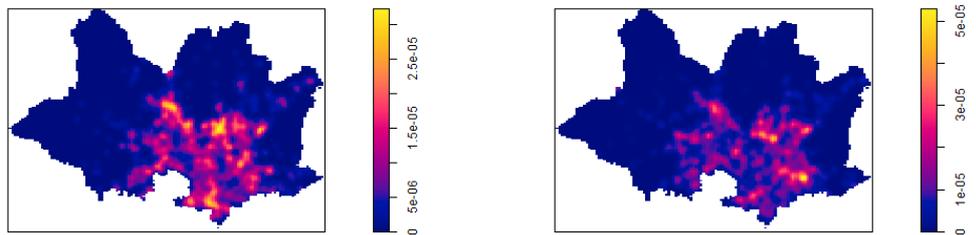
El análisis preliminar de los datos permitió verificar algunas reglas empíricas establecidas por la literatura internacional. Los delitos de rapiña tienden a estar concentrados en pocas zonas de Montevideo, las cuales pueden definirse como puntos calientes. Estas zonas parecen evolucionar con el tiempo lo que indicaría que estos puntos calientes son dinámicos. Los delitos no son un proceso homogéneo en el espacio tiempo, tienden a presentar aglomeración lo que podría estar asociado a un proceso de contagio. En particular, parece existir

Figura 4.2: Estimación de intensidad de rapiña



(a) 2018

(b) Enero 2018



(c) Junio 2018

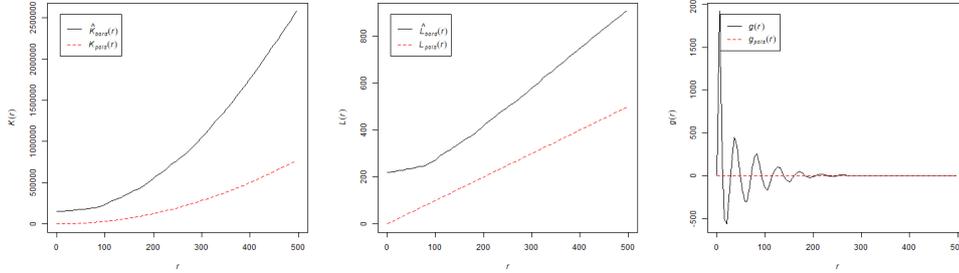
(d) Diciembre 2018

Las estimaciones de intensidad mediante kernels permiten establecer puntos de mayor intensidad del delito. Colores más oscuros indican una menor intensidad del delito mientras que colores más claros indican mayor intensidad.

una correlación elevada con el propio pasado de la zona y sus vecinos cercanos, aunque esta última decae rápidamente con el tiempo.

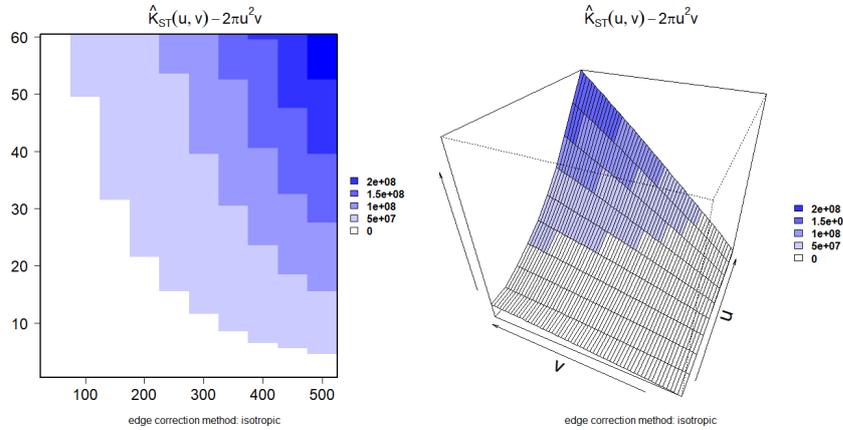
Figura 4.3: Estimaciones K, L y g espaciales

(a) Función K estimada (b) Función L estimada (c) Función g estimada



Las funciones K, L permiten establecer las propiedades de segundo orden del proceso, el eje de las ordenadas presenta el valor esperado de eventos adicionales en una bola de radio r partiendo de un evento cualquiera. La evolución por encima de la referencia homogénea indica que el proceso presenta agrupamiento espacial. La función $g(r)$ indica la probabilidad relativa a un proceso homogéneo en una distancia exacta de r .

Figura 4.4: Función K espacio temporal

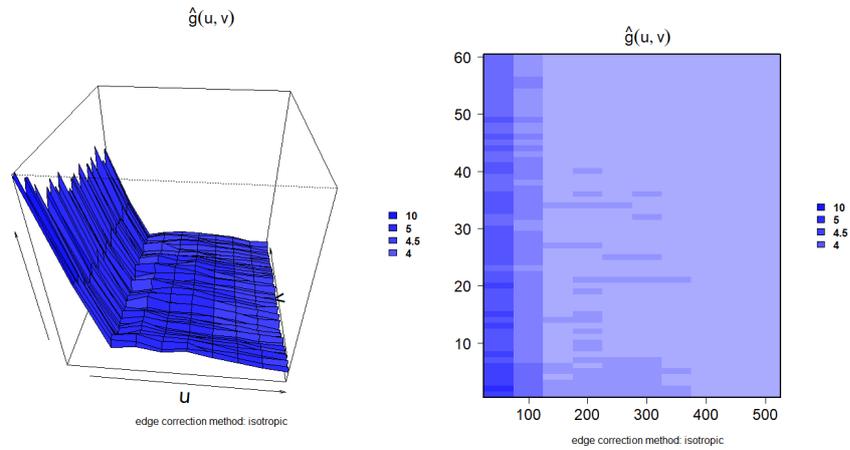


La función $\hat{K}_{st}(\cdot)$ centrada permite establecer la propiedad de segundo orden de un proceso puntual espacio temporal. Su evolución por encima del valor de referencia homogéneo indica la presencia de aglomeración espacio temporal.

4.3. Estrategia empírica y estimación

Para verificar la capacidad predictiva de las técnicas descritas en el capítulo 3 se utilizó una estrategia de validación sobre una muestra de testeo no conocida por los modelos. Utilizando una ventana móvil hasta el momento t ,

Figura 4.5: Función de correlación espacio temporal



La función $\hat{g}(\cdot)$ permite caracterizar la aglomeración. Se observa una correlación elevada que perdura en el tiempo para una distancia corta. Las zonas cercanas (300mts) prestan un valor positivo, aunque su efecto disminuye con el pasaje del tiempo más rápidamente.

se intentó realizar predicciones para el momento $t + 1$ sobre una grilla de s_j celdas. La predicción consiste en seleccionar n celdas a patrullar, siendo estas las que presentan mayor riesgo de acuerdo a cada modelo.

Si bien no existe una única definición del tamaño de la ventana temporal, tamaño de las regiones y cuantos días del pasado deben usarse para entrenar los distintos métodos, es posible basarse en otros trabajos o en estudios descriptivos. A modo de ejemplo en los trabajos de procesos puntuales, [Mohler et al. \(2011\)](#) utiliza datos diarios creando una grilla espacial de largo 200mts, [Kajita and Kajita \(2017\)](#) utiliza datos diarios utilizando una grilla 250mts. Otros autores, por ejemplo [Dulce et al. \(2018\)](#) han utilizado 10 semanas de delitos para predecir las siguientes cuatro semanas. [Reinhart and Greenhouse \(2018\)](#), utilizan en lugar de una estructura de grillas bloques censales de diferente tamaño. En cuanto a los trabajos de PHM, [Bowers et al. \(2004\)](#) sugiere utilizar hasta 60 días y un radio de 400 metros como referencia para los delitos de hurto en viviendas.

Buscando hacer las distintas metodologías comparables, se partió de una grilla de 200 metros sobre la cual se intentó identificar cuáles serían las más propensas a sufrir un delito al día siguiente. Siguiendo el análisis descriptivo, se consideró una distancia de 300 metros para los modelos de kernel, prospectivo

y endémico-epidémico. Para estos últimos dos se consideró como parámetro de contagio temporal un máximo de 30 días. En el caso de los endémico-epidémico se utilizó para la función de contagio espacial un kernel gaussiano bivariado y una función exponencial para el efecto temporal.

Para verificar que la capacidad predictiva, en primera instancia se seleccionó un día fijo (8 de junio de 2018) para ilustrar los resultados del modelo y se seleccionaron 20 días al azar de 2018 para comprobar que la capacidad predictiva no estuviera asociada a un día específico. La tabla 4.4 presenta la estrategia seguida para cada día.

Tabla 4.4: Estrategia de estimación

-
- Parámetros de entrada: fecha para la cual se desea la predicción ($t + 1$).
 1. Se peticionan los datos en un conjunto de entrenamiento definiendo 365 días anteriores como datos para estimar los métodos.
 2. Se retiene el día $t + 1$ para validar.
 3. Se aplican los distintos métodos.
 - I Conteo por cuadrante (Conteo)
 - II Densidad de kernel (KDE)
 - III Modelo prospectivo (PHM)
 - IV Modelo tipo endémico-epidémico (EP)
 - V Modelo tipo endémico-epidémico con covariables (EP cov)
 4. Salida: Base de datos donde a cada celda se le otorga un puntaje de riesgo de acuerdo al método utilizado.
-

Para comparar las técnicas utilizadas se propone utilizar las métricas descritas en la sección 3.4.

4.4. Resultados

La sección de resultados se estructura en dos sub secciones. En la primera parte se muestra la capacidad predictiva del modelo para un día fijo específico, mientras que la segunda discute los resultados al haber seleccionado aleatoriamente 20 días para evaluar el desempeño de los distintos métodos.

4.4.1. Predicción para un día pre fijado

Se definió como una primera fecha para probar los modelos el 8 de junio de 2018. Esta fecha es considerada como el $t + 1$, es decir el momento para el

cual se quiere predecir. Los distintos métodos utilizaron los 365 días previos al 8 de junio, estimando en cada caso una medida de intensidad asociada a cada unidad de la grilla.

En una primera instancia, para seleccionar celdas a patrullar se utilizaron los percentiles de las intensidades estimadas. Para cada método, se ordenan las celdas en función de su intensidad y se calcula el percentil p , es decir el valor para el cual el $p\%$ de las celdas presentan intensidades menores. Al seleccionar las celdas con valores mayores se concentra el esfuerzo en las zonas donde se observa mayor intensidad para cada método. La tabla 4.5 presenta la tasa de éxito, porcentaje del área seleccionada, PAI y el número de celdas seleccionadas para cada punto de corte y método. Utilizando el percentil 90, en todos los casos se cubre más de 1000 celdas. El método de conteo alcanza el mayor PAI (6.91) debido a que seleccionando únicamente 8.37% de las celdas, se logra cubrir el 57.8% de los delitos. Los métodos endémico epidémicos lo siguen en cuanto a PAI. Al utilizar el percentil 99, es decir, seleccionando menos unidades para patrullar, se destacan el método de conteo y el método endémico epidémico con covariables sobre los restantes métodos, alcanzando un PAI de 16.24 y 12.5 respectivamente. La figura 4.6 presenta la curva de retorno asociada a agregar una celda adicional a patrullar siguiendo el criterio de cada método. La columna 3 de la tabla 4.7 presenta la evaluación de la curva en 3 puntos de interés, 100 celdas (0.73% de las celdas de Montevideo), 500 celdas (3.64%) y 1000 celdas (7.28%).

Se observa que para el método de conteo la curva evoluciona por encima de los restantes, seguido de los modelos endémico epidémico con covariables hasta los 250 lugares. Luego este último tiene un comportamiento similar a su versión sin covariables y al método prospectivo. Se destaca que el método basado en estimación de densidad presenta rendimientos mucho menores.

Desde el punto de vista de la tasa de éxito, independientemente del número de unidades seleccionadas, sería recomendable utilizar un modelo de conteo seguido del endémico epidémico con covariables. Sin embargo, para establecer rutas de patrullaje es necesario considerar las necesidades de personal.

Para incorporar esta dimensión, se estudió que tan alejada se encuentran las celdas seleccionadas. Se consideró que, si dos celdas se encontraban cercanas, estas podrían agruparse generando así unidades de patrullaje (UP), reduciendo la cantidad de personal necesario.

La tabla 4.7 presenta algunas estadísticas sobre estas agrupaciones, mien-

Tabla 4.5: Estadísticos sobre percentiles de la intensidad estimada para el 8 de Junio según método

Percentil	Modelo	TE	PA	PAI	Celdas
percentil 90	Conteo	57.8	8.37	6.91	1150
	Kernel	42.2	10.00	4.22	1374
	PHM	50.0	10.00	5.00	1374
	EP Cov	54.7	10.00	5.47	1374
	EP	54.7	10.00	5.47	1374
percentil 95	Conteo	37.5	4.45	8.42	612
	Kernel	18.8	5.00	3.75	687
	PHM	31.3	5.00	6.25	687
	EP Cov	28.1	5.00	5.63	687
	EP	26.6	5.00	5.31	687
percentil 99	Conteo	14.1	0.87	16.24	119
	Kernel	3.1	1.00	3.11	138
	PHM	4.7	1.00	4.67	138
	EP Cov	12.5	1.00	12.45	138
	EP	6.3	1.00	6.22	138

Tabla 4.6: TE =Tasa de éxito, PA =Porcentaje de área cubierta. $PAI=TE/PA$. Celdas= Número de celdas seleccionadas por el método.

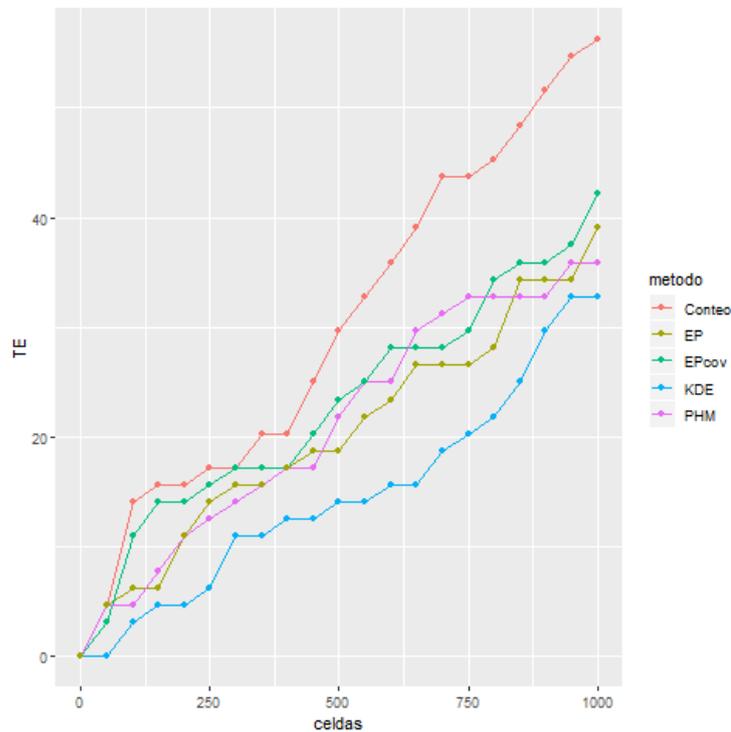
El método de conteo presenta el mejor desempeño seguido del endémico epidémico con covariables.

tras que las figuras 4.7 a 4.11 presenta el perímetro de las unidades de patrullaje y los delitos cometidos el 8 de junio.

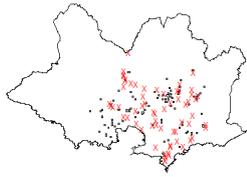
Los resultados indican que los métodos basados en estimación de densidad, aunque presentaron los peores resultados en cuanto a tasa de éxito, seleccionaron celdas conexas, lo que facilitaría una estrategia de patrullaje. El método de conteo, por el contrario, selecciona celdas más dispersas, generando zonas más pequeñas y en promedio más alejadas entre sí, por lo que cubrir estas unidades podría significar un desafío adicional para la policía.

El método prospectivo y los endémico epidémico parecen situarse en una situación intermedia, seleccionando algunas zonas cercanas permiten generar zonas de patrullaje y capturan celdas más alejadas que efectivamente presentaron delitos.

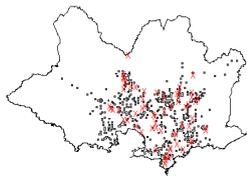
Figura 4.6: Curva de retorno: 8 de junio 2018



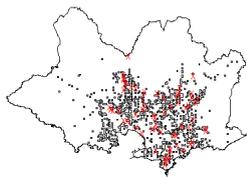
La curva de retorno presenta una comparación de los distintos métodos la misma cantidad de celdas seleccionadas y el retorno marginal de agregar una unidad adicional al patrullaje. El método de conteo presenta el mejor desempeño seguido del endémico epidémico con covariables.



(a)

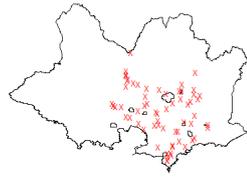


(b)

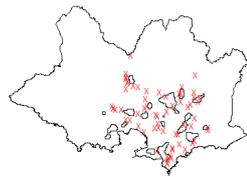


(c)

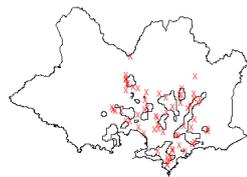
Figura 4.7: Conteo



(a)

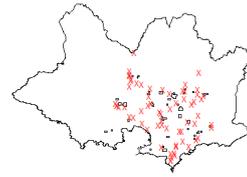


(b)

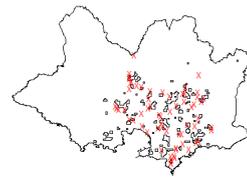


(c)

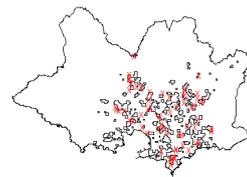
Figura 4.8: KDE



(a)



(b)

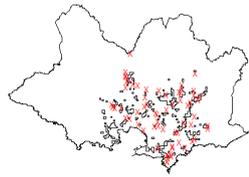


(c)

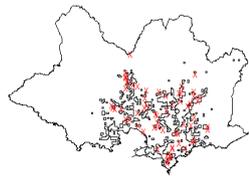
Figura 4.9: PHM



(a)

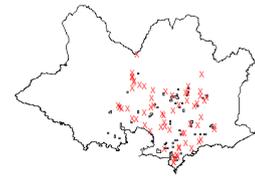


(b)

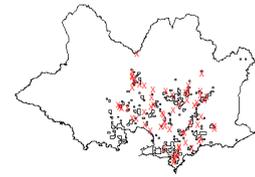


(c)

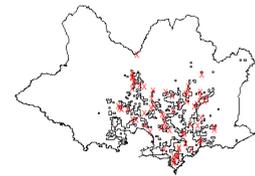
Figura 4.10: EP



(a)



(b)



(c)

Figura 4.11: EP cov

Perímetro de Unidades de Patrullaje en negro y delitos de rapiña para el día 8 de junio de 2018 marcadas como X. (a) Indica que se seleccionaron 100 celdas, (b) 500 celdas, (c) 1000 celdas.

El método KDE presenta unidades compactas y cercanas en el espacio mientras que el método de conteo selecciono celdas dispersas. El método prospectivo y los endémicos epidémicos seleccionaron celdas adyacentes que posibilitan construir unidades de patrullaje y algunas más alejadas que efectivamente presentaron delitos.

4.4.2. Predicción sobre una muestra aleatoria de días

Para verificar que la capacidad predictiva no se encuentra asociada a un día específico se seleccionaron al azar 20 días de 2018.

La figura 4.13 presenta el promedio de la tasa de éxito para cada número de celdas seleccionadas. Mediante barras se presentan el desvío estándar. La primera columna de la tabla 4.8 presenta estos promedios en algunos puntos de interés.

Los resultados muestran que el método de conteo continúa alcanzando la mayor tasa de éxito. Para pocos lugares seleccionados los endémicos epidémicos tienen un comportamiento similar al del método prospectivo, aunque este presenta retornos menores a mayor cantidad de lugares seleccionados.

La selección de días de manera aleatoria permite estudiar otro aspecto del comportamiento de los métodos. Idealmente, las zonas patrulladas deben abarcar la mayor cantidad de celdas brindando el servicio de patrullaje a la mayor cantidad de población. En segundo lugar, deben ser flexibles y adaptarse a cambios en el delito.

Para estudiar este aspecto se analizó la huella de patrullaje (Apéndice 3). Las columnas 2 y 3 de la tabla 4.8 presentan el número de celdas patrulladas en esos 20 días y un índice de entropía. Los valores más bajos de este estadístico indican que el método tiende a seleccionar las mismas unidades en todos los periodos.

El método prospectivo presenta valores más elevados de entropía en todos los escenarios asociado a que se eligen distintas celdas de patrullaje en cada periodo, ampliando el listado de zonas que se patrullan en esos 20 días. Sin embargo, este método presenta junto al kernel, en promedio, menores niveles de tasa de éxito en especial para valores elevados de celdas seleccionadas.

Si bien el método basado en conteo y el endémico epidémico presentan valores similares en cuanto a lugares patrullados y entropía. Estos resultados deben leerse a la luz de la sección anterior, las celdas seleccionadas por conteo tienden a encontrarse más dispersos en Montevideo lo que puede dificultar el patrullaje.

Figura 4.12: Curva de retorno promedio sobre muestra aleatoria de días

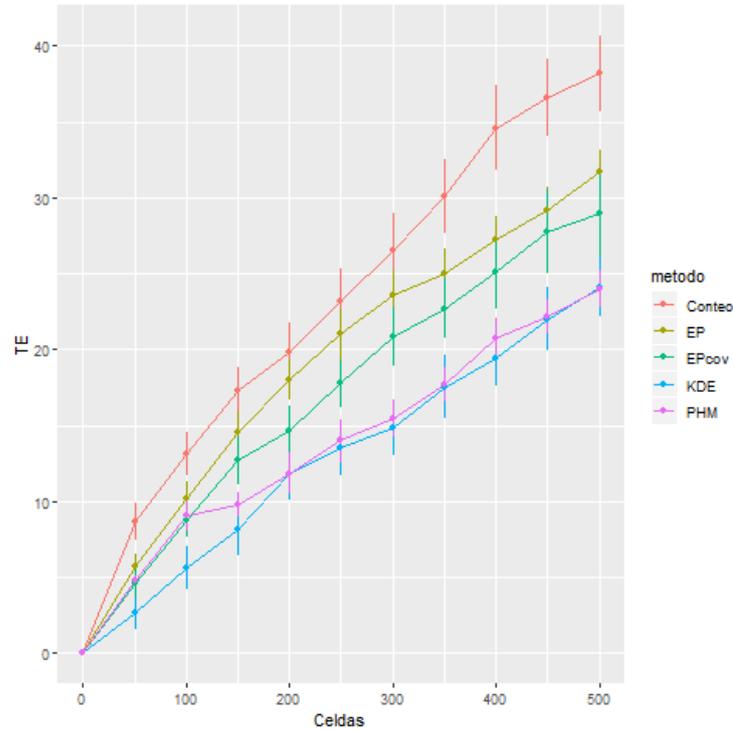


Figura 4.13: Promedio de tasa de éxito y desvió estándar (barras). El método de conteo presenta el mejor desempeño seguido del método endémico epidémico.

4.5. Discusión

Los delitos de rapiña en Montevideo presentan un alto grado concentración espacial. En 2018, el 25 % de los delitos se encontraba concentrado en el 1.73 % de las celdas, mientras que el 50 % lo hacía en el 5.28 %. Se destaca que pueden existir factores que contribuyan a la concentración del delito como la presencia de asentamientos, escuelas o liceos, restaurantes y estaciones de servicios. A pesar de esto, la estimación de densidad indica que la intensidad del delito evoluciona con el tiempo. La comparación de las funciones $\hat{K}(\cdot)$ y $\hat{K}_{st}(\cdot)$ con un proceso homogéneo muestra que las rapiñas tienden a aglomerarse en el espacio y tiempo. En particular, la función de distribución radial $\hat{g}_{st}(\cdot)$, presenta valores elevados para una distancia corta y tiende a perdurar en el tiempo, posiblemente porque las mismas unidades tienden a sufrir reiteradas veces rapiñas. El efecto de contagio espacial parece existir para zonas aledañas, pero disminuye luego de los 300 metros.

La comparación de los métodos evidencia que cada uno trae asociado ventajas y desventajas a considerar.

El método de conteo presenta mejores resultados en cuanto a la tasa de éxito, sin embargo, la dispersión de las celdas seleccionadas puede dificultar una estrategia de patrullaje.

La estimación de densidad, si bien presenta zonas de patrullaje más compactas, su tasa de éxito es significativamente menor a los restantes métodos.

El método prospectivo demostró cubrir una mayor área de Montevideo durante los 20 días seleccionados al azar, sin embargo, presentó menor retornos al agregar una celda adicional.

Los modelos endémico epidémicos se sitúan en una situación intermedia. El rendimiento de agregar unidades de patrullaje adicional los sitúa segundos en cuanto a tasa de éxito, siendo estos retornos bastante superiores a los métodos prospectivo y estimación de densidad en particular para valores elevados de celdas seleccionadas. Comparado con el modelo de conteo, presentan menores tasa de éxito y niveles levemente menores de entropía ya que tienden a seleccionar las mismas zonas, sin embargo, la selección de zonas compactas y algunas celdas alejadas de mayor intensidad delictiva puede facilitar a una estrategia de patrullaje. Corresponde notar que la incorporación de covariables no resultó en una mejora de la capacidad predictiva.

Tabla 4.7: Estadísticas para el 8 de junio 2018

Método	Celdas	TE	UP	Promedio:			Distancias entre UP
				Perímetro UP	Área UP	CI UP	
Conteo	100	14.06	79	940	51,079	4.2	6929
	500	29.69	243	1,357	82,886	4.6	7812
	1000	56.25	265	2,039	151,859	5.0	8646
KDE	100	3.125	9	2,758	445,823	4.5	6332
	500	14.06	17	5,510	1,176,332	5.3	5732
	1,000	32.81	20	7,345	2,001,055	5.6	5635
PHM	100	4.6875	33	1,445	121,934	4.3	6602
	500	21.88	76	2,403	264,364	4.9	6841
	1,000	35.94	99	3,074	404,825	5.0	7174
EP Cov	100	10.9375	51	1,180	78,239	4.3	6526
	500	23.44	103	2,029	194,228	4.8	7276
	1,000	42.10	108	3,126	370,964	5.3	7529
EP	100	6.25	41	1,331	97,253	4.4	6534
	500	18.75	86	2,323	232,538	5.0	6714
	1,000	39.06	101	3,150	396,487	5.1	7542

TE= Tasa de éxito. UP=Unidades de patrullaje, definidas como la unión de celdas adyacentes.

El método basado en KDE un número menor de unidades de patrullaje, cercanas en el espacio, aunque capturó menos delito que los otros métodos. El método de conteo si bien presenta la mayor tasa de éxito, seleccionó unidades dispersas y alejadas lo que dificulta las tareas de patrullaje. El método prospectivo y los endémico epidémico se sitúan en una situación intermedia.

Tabla 4.8: Estadísticos sobre muestra aleatoria de días

	Celdas	Promedio TE	Celdas patrulladas	Entropia
Conteo	100	12.35	180	7.23
	500	36.19	835	9.45
	1000	54.46	1585	10.40
Kernel	100	4.60	185	7.39
	500	22.10	748	9.41
	1000	39.94	1431	10.35
PHM	100	8.14	369	9.24
	500	23.20	1505	10.85
	1000	38.36	2546	11.41
EP Cov	100	8.19	174	7.40
	500	28.61	745	9.49
	1000	45.94	1433	10.45
EP	100	9.11	170	7.27
	500	29.98	760	9.44
	1000	47.83	1443	10.36

TE=Tasa de éxito. Celdas patrulladas refiere al número de celdas visitadas por patrullaje sin repetir

Capítulo 5

Consideraciones finales

Este trabajo tuvo como principal objetivo la comparación y evaluación de distintas técnicas de predicción de robos violentos o rapiñas, poniendo especial énfasis en la utilización de procesos puntuales. El delito ha tomado un rol protagónico en la agenda, convirtiéndose una de las mayores preocupaciones de los ciudadanos y autoridades uruguayas. En particular el delito de rapiña, debido al grado de violencia involucrada y su dinámico crecimiento en los últimos años.

Para este trabajo se consideraron las rapiñas registradas en Montevideo entre 2013 y 2018, comprobando algunas reglas empíricas establecidas en la literatura. Los delitos de rapiñas presentaron un alto grado de concentración espacial. En 2018, el 25 % de los delitos se encontraba concentrado en el 1.73 % de las celdas, mientras que el 50 % lo hacía en el 5.28 %. Las herramientas de análisis exploratorio para procesos puntuales mostraron que, si bien existen puntos de mayor concentración delictiva, no son estáticas en el tiempo. Se destaca que las rapiñas presentan un comportamiento de aglomeración espacio temporal, posiblemente porque algunas zonas son más atractivas para los ofensores (factor endémico) y existe un patrón de contagio a zonas cercanas (factor epidémico), una vez que se comete una rapiña es probable que se observe otra a una distancia pequeña y en un breve periodo de tiempo. Estos mecanismos tienen un respaldo en la literatura criminológica como por ejemplo la teoría de elección racional y actividad rutinaria.

El patrullaje en puntos calientes es una de las estrategias que se ha mostrado efectiva para reducir el delito. Uruguay ha avanzado en este sentido al incorporar el Programa de Alta Dedicación Operativa, un programa focalizado

en prevención de rapiñas. La piedra angular de este tipo de estrategia es la correcta identificación de zonas a patrullar. Este trabajo pretendió colaborar en este sentido mediante la comparación y evaluación de distintos métodos.

La comparación de los métodos evidencia que cada uno presenta ventajas y desventajas a considerar. El método basado en conteo presentó mejores resultados en cuanto a su tasa de éxito, sin embargo, la dispersión de celdas seleccionadas para patrullar puede dificultar una estrategia de patrullaje necesitando una mayor cantidad de personal.

El método basado en la estimación de densidad seleccionó áreas más compactas, pero estas tienden a repetirse en el tiempo y presentó un menor nivel de tasa de éxito.

Los métodos prospectivos presentaron los mejores niveles de cobertura en cuanto a celdas patrulladas, aunque también presentaron bajos niveles de tasa de éxito.

Los modelos endémico epidémicos aplicados en este trabajo se encuentran en una situación intermedia, proporcionando áreas de patrullaje compactas y celdas alejadas, con un nivel de tasa de éxito intermedio, por lo que pueden considerarse como una alternativa de compromiso entre eficacia predictiva y eficiencia de patrullaje. Es importante destacar que la introducción de covariables no represento una mejoría respecto a un modelo simple de contagio.

Corresponde notar algunas limitaciones de este trabajo y posibles ampliaciones.

En primer lugar, la información sobre el momento y lugar del hecho es extraída a partir de los reportes de denuncia. Puede existir un nivel de error que no es capturado por los métodos, además de un nivel de delito no denunciado.

En segundo lugar, se estableció una estrategia de pruebas alineada con la literatura internacional pero que puede ser una simplificación de la realidad. Para cada día seleccionado, se realizaron predicciones utilizando la información hasta el día anterior. Esto supone que la policía podrá contar con esta información al momento de realizar la predicción. En la realidad, las víctimas pueden optar por denunciar el delito en los días subsiguiente al hecho lo que limitaría la información disponible.

Finalmente, debe considerarse las limitaciones de la incorporación de covariables y efecto temporal. El listado de factores considerado es una primera aproximación y podría expandirse en dos aspectos fundamentales. En primer lugar, se consideró únicamente la información disponible en fuentes públicas

no incorporando información que puede manejar la policía. En particular, parece relevante incorporar el efecto de presencia policial pasada, dado que es esperable que al encontrarse un policía en la zona se reduzca la intensidad del delito. En segundo lugar, no se consideró la interacción de las covariables con efectos temporales. Es esperable que los atractores del delito evolucionen con el tiempo, por ejemplo, que algunos servicios como bares y pubs puedan atraer más crimen los fines de semana.

A pesar de las limitaciones, este trabajo es el primero en realizar una comparación de distintos métodos para seleccionar zonas de patrullaje en Uruguay, abriendo una línea de investigación a futuro para desarrollar otros métodos que puedan ser de utilidad para la mejora de la seguridad ciudadana.

Referencias bibliográficas

- Aboal, D., Lorenzo, F., and Perera, M. (2007). Crimen y violencia en uruguay. *Informe de investigación del proyecto de investigación financiado por la CSIC de la Universidad de la República*.
- Ajzenman, N. and Jaitman, L. (2016). Crime concentration and hot spot dynamics in latin america. Technical report, Inter-American Development Bank.
- Alonso, C. (2018). Transferencias monetarias y crimen. evidencia para la última década en montevideo.
- Appelhans, T., Detsch, F., Reudenbach, C., and Woellauer, S. (2018). *mapview: Interactive Viewing of Spatial Data in R*. R package version 2.3.0.
- Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial point patterns: methodology and applications with R*. Chapman and Hall/CRC.
- Baddeley, A. and Turner, R. (2005). spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42.
- Baddeley, A. J., Turner, R., et al. (2004). Spatstat: An r package for analyzing spatial point pattens.
- Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime*, pages 13–68. Springer.
- Bernasco, W., Johnson, S. D., and Ruiters, S. (2015). Learning where to offend: Effects of past on future burglary locations. *Applied Geography*, 60:120–129.
- Bivand, R. and Lewin-Koh, N. (2017). *maptools: Tools for Reading and Handling Spatial Objects*. R package version 0.9-2.

- Bivand, R. and Rundel, C. (2018). *rgeos: Interface to Geometry Engine - Open Source ('GEOS')*. R package version 0.3-28.
- Bowers, K. J., Johnson, S. D., and Pease, K. (2004). Prospective hot-spotting: The future of crime mapping? *British Journal of Criminology*, 44(5):641–658.
- Braga, A. A., Papachristos, A. V., and Hureau, D. M. (2014). The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice quarterly*, 31(4):633–663.
- Brantingham, P. J. and Brantingham, P. L. (1984). *Patterns in crime*. Macmillan New York.
- Brunsdon, C. and Chen, H. (2014). *GISTools: Some further GIS capabilities for R*. R package version 0.7-4.
- Bukstein, D. and Montossi, F. (2009). Un estudio sobre la criminalidad en montevideo.
- Caplan, J. M. and Kennedy, L. W. (2011). Risk terrain modeling compendium. *Rutgers Center on Public Security, Newark*.
- Chainey, S. and Ratcliffe, J. (2013). *GIS and crime mapping*. John Wiley & Sons.
- Chainey, S., Serrano, R., and Veneri, F. (2018). Evaluando el impacto del programa de alta dedicación operativa (pado) en la reducción de robos violentos en montevideo, uruguay. *Inter-American Development Bank and Uruguayan Interior Ministry, eds., ¿Cómo evitar el delito urbano*.
- Chainey, S., Tompson, L., and Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security journal*, 21(1-2):4–28.
- Chalfin, A. and McCrary, J. (2017). Criminal deterrence: A review of the literature. *Journal of Economic Literature*, 55(1):5–48.
- Cid, A. (2018). Evaluación de medidas aplicadas en la policía de montevideo durante el período 2013-2015. *Inter-American Development Bank and Uruguayan Interior Ministry, eds., ¿Cómo evitar el delito urbano*.

- Clarke, R. V. and Cornish, D. B. (1985). Modeling offenders' decisions: A framework for research and policy. *Crime and justice*, 6:147–185.
- Cohen, L. E. and Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American sociological review*, pages 588–608.
- Cuevas, A., Fraiman, R., Rodríguez-Casal, A., et al. (2007). A nonparametric approach to the estimation of lengths and surface areas. *The Annals of Statistics*, 35(3):1031–1051.
- Dassios, A., Zhao, H., et al. (2013). Exact simulation of hawkes process with exponentially decaying intensity. *Electronic Communications in Probability*, 18.
- Del Castillo, F. (2018). Génesis, desarrollo y profesionalización del análisis criminal en la policía nacional. *Inter-American Development Bank and Uruguayan Interior Ministry, eds., ¿Cómo evitar el delito urbano*.
- Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. Chapman and Hall/CRC.
- Dulce, M., Ramírez-Amaya, S., and Riascos, Á. (2018). Efficient allocation of law enforcement sources using predictive police patrolling. *arXiv preprint arXiv:1811.12880*.
- Eck, J., Chainey, S., Cameron, J., and Wilson, R. (2005). Mapping crime: Understanding hotspots.
- Evans, J. S. (2017). *spatialEco*. R package version 0.0.1-7.
- Fitzpatrick, D. J., Gorr, W. L., and Neill, D. B. (2019). Keeping score: Predictive analytics in policing. *Annual Review of Criminology*, 2:473–491.
- Fotheringham, A. S. and Wong, D. W. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and planning A*, 23(7):1025–1044.
- Gabriel, E. and Diggle, P. J. (2009). Second-order analysis of inhomogeneous spatio-temporal point process data. *Statistica Neerlandica*, 63(1):43–51.

- Gabriel, E., Diggle, P. J., Rowlingson, B., and Rodriguez-Cortes, F. J. (2018). *stpp: Space-Time Point Pattern Simulation, Visualisation and Analysis*. R package version 2.0-3.
- Gabriel, E., Rowlingson, B., Diggle, P., et al. (2013). *stpp: an r package for plotting, simulating and analyzing spatio-temporal point patterns*. *Journal of Statistical Software*, 53(2):1–29.
- Gandelman, N. and Munyo, I. (2015). Juvenile incarceration and crime after release: Evidence from a harsher law.
- Gorr, W., Lee, Y., Weisburd, D., and Eck, J. (2017). Chronic and temporary crime hot spots. In *Unraveling the Crime-Place Connection: New Directions in Theory and Policy-Advances in Criminological Theory*. Transaction Publishers.
- Grolemund, G. and Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3):1–25.
- Gómez-Rubio, V., Zheng, P., Diggle, P., Sterratt, D. C., Peng, R. D., Murdoch, D., and Rowlingson, B. (2017). *spatialkernel: Non-Parametric Estimation of Spatial Segregation in a Multivariate Point Process*. R package version 0.4-23.
- Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90.
- Hijmans, R. J. (2017). *raster: Geographic Data Analysis and Modeling*. R package version 2.6-7.
- Hope, R. M. (2013). *Rmisc: Rmisc: Ryan Miscellaneous*. R package version 1.5.
- Jacottet, J. (2018). Configuraciones territoriales de las violencias y el delito en la ciudad de montevideo. *Inter-American Development Bank and Uruguayan Interior Ministry, eds., ¿Cómo evitar el delito urbano*.
- Johnson, S. D., Bernasco, W., Bowers, K. J., Elffers, H., Ratcliffe, J., Rengert, G., and Townsley, M. (2007). Space–time patterns of risk: A cross national assessment of residential burglary victimization. *Journal of Quantitative Criminology*, 23(3):201–219.

- Kajita, M. and Kajita, S. (2017). Crime prediction by data-driven green's function method. *arXiv preprint arXiv:1704.00240*.
- Kennedy, L. W. and Caplan, J. M. (2010). *Risk Terrain Modeling Manual: Theoretical Framework and Technical Steps of Spatial Risk Assessment for Crime Analysis*.
- Kennedy, L. W., Caplan, J. M., and Piza, E. L. (2012). *A primer on the spatial dynamics of crime emergence and persistence*.
- Lee, Y., Eck, J. E., SooHyun, O., and Martinez, N. N. (2017). How concentrated is crime at places? a systematic review from 1970 to 2015. *Crime Science*, 6(1):6.
- Lewis, E., Mohler, G., Brantingham, P. J., and Bertozzi, A. L. (2012). Self-exciting point process models of civilian deaths in iraq. *Security Journal*, 25(3):244–264.
- Meyer, S., Elias, J., and Hohle, M. (2012). A space time conditional intensity model for invasive meningococcal disease occurrence. *Biometrics*, 68(2):607–616.
- Meyer, S., Held, L., and Hohle, M. (2017). Spatio-temporal analysis of epidemic phenomena using the R package surveillance. *Journal of Statistical Software*, 77(11):1–55.
- Mohler, G. (2014). Marked point process hotspot maps for homicide and gun crime prediction in chicago. *International Journal of Forecasting*, 30(3):491–497.
- Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., and Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108.
- Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., and Brantingham, P. J. (2015). Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*, 110(512):1399–1411.
- Munyo, I. and Rossi, M. A. (2015). First-day criminal recidivism. *Journal of Public Economics*, 124:81–90.

- Munyo, I., Rossi, M. A., et al. (2016). Is it displacement? evidence on the impact of police monitoring on crime. Technical report.
- Observatorio Nacional Sobre Violencia y Criminalidad (2019). *Denuncias de rapiña y hurto. 1 de Enero 2018 al 31 de Diciembre 2018*.
- Openshaw, S. (1983). *The Modifiable Areal Unit Problem*, volume 38.
- OpenStreetMap Foundation contributors (2019). Openstreetmap foundation.
- Padgham, M., Rudis, B., Lovelace, R., and Salmon, M. (2017). osmdata. *The Journal of Open Source Software*, 2(14).
- Pebesma, E. J. and Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5(2):9–13.
- Perry, W. L. (2013). *Predictive policing: The role of crime forecasting in law enforcement operations*. Rand Corporation.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reinhart, A. (2018). A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, 33(3):299–318.
- Reinhart, A. and Greenhouse, J. (2018). Self-exciting point processes with spatial covariates: modelling the dynamics of crime. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(5):1305–1329.
- Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(2):172–192.
- RStudio Team (2016). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.
- Veneri, F. (2016). Risk terrain modeling para el delito de rapiña en montevideo.
- Veneri, F. and Regueira, C. (2018). Incidencia de los factores socio demográficos en los patrones espaciales de los delitos contra la propiedad. análisis de micro territorios en montevideo.
- Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Springer.

- Wang, X. and Brown, D. E. (2012). The spatio-temporal modeling for criminal incidents. *Security Informatics*, 1(1):2.
- Warnes, G. R., Bolker, B., Gorjanc, G., Grothendieck, G., Korosec, A., Lumley, T., MacQueen, D., Magnusson, A., Rogers, J., and others (2017). *gdata: Various R Programming Tools for Data Manipulation*. R package version 2.18.0.
- Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology*, 53(2):133–157.
- Weisburd, D. and Telep, C. W. (2014). Hot spots policing: What we know and what we need to know. *Journal of Contemporary Criminal Justice*, 30(2):200–220.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H. (2018). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.3.1.
- Wickham, H., François, R., Henry, L., and Müller, K. (2018). *dplyr: A Grammar of Data Manipulation*. R package version 0.7.5.
- Wickham, H., Hester, J., and François, R. (2017). *readr: Read Rectangular Text Data*. R package version 1.1.1.
- Wickham, H. and Miller, E. (2017). *haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. R package version 1.1.0.
- Zeileis, A. (2014). *ineq: Measuring Inequality, Concentration, and Poverty*. R package version 0.2-13.

APÉNDICES

Apéndice 1

Librerías y software utilizado

Este trabajo fue realizado utilizando el lenguaje de programación R y la interfaz de usuario R studio ([R Core Team, 2017](#), [RStudio Team, 2016](#)). El código completo se presenta en el apéndice 4. Se utilizaron las siguientes librerías¹:

Utilitarios y manipulación de datos:

- dplyr: [Wickham et al. \(2018\)](#)
- gdata: [Warnes et al. \(2017\)](#)
- ggplot2: [Wickham \(2016\)](#)
- heaven: [Wickham and Miller \(2017\)](#)
- ineq: [Zeileis \(2014\)](#)
- lubridate: [Grolemund and Wickham \(2011\)](#)
- readr: [Wickham et al. \(2017\)](#)
- Rmisc: [Hope \(2013\)](#)
- stringer: [Wickham \(2018\)](#)

Mapas interactivos y datos Open Street Map

- mapview [Appelhans et al. \(2018\)](#)
- osmdata [Padgham et al. \(2017\)](#)

¹La siguiente clasificación responde al uso que se le dio a cada paquete, no al total de funcionalidades que proveen.

Manipulación de información geográfica

- gistool: [Brunsdon and Chen \(2014\)](#)
- maptools: [Bivand and Lewin-Koh \(2017\)](#)
- raster: [Hijmans \(2017\)](#)
- rgeos: [Bivand and Rundel \(2018\)](#)
- sp : [Pebesma and Bivand \(2005\)](#)

Conteo y estimación de densidad

- Gistool: [Brunsdon and Chen \(2014\)](#)
- spatialEco [Evans \(2017\)](#)
- spatialkernel: [Gómez-Rubio et al. \(2017\)](#)

Procesos puntuales y análisis exploratorios

- spatstat: [Baddeley and Turner \(2005\)](#)
- stpp: [Gabriel et al. \(2018\)](#)

Modelo endémico epidémico

- Surveillance: [Meyer et al. \(2017\)](#)

Apéndice 2

Curvas de concentración del delito

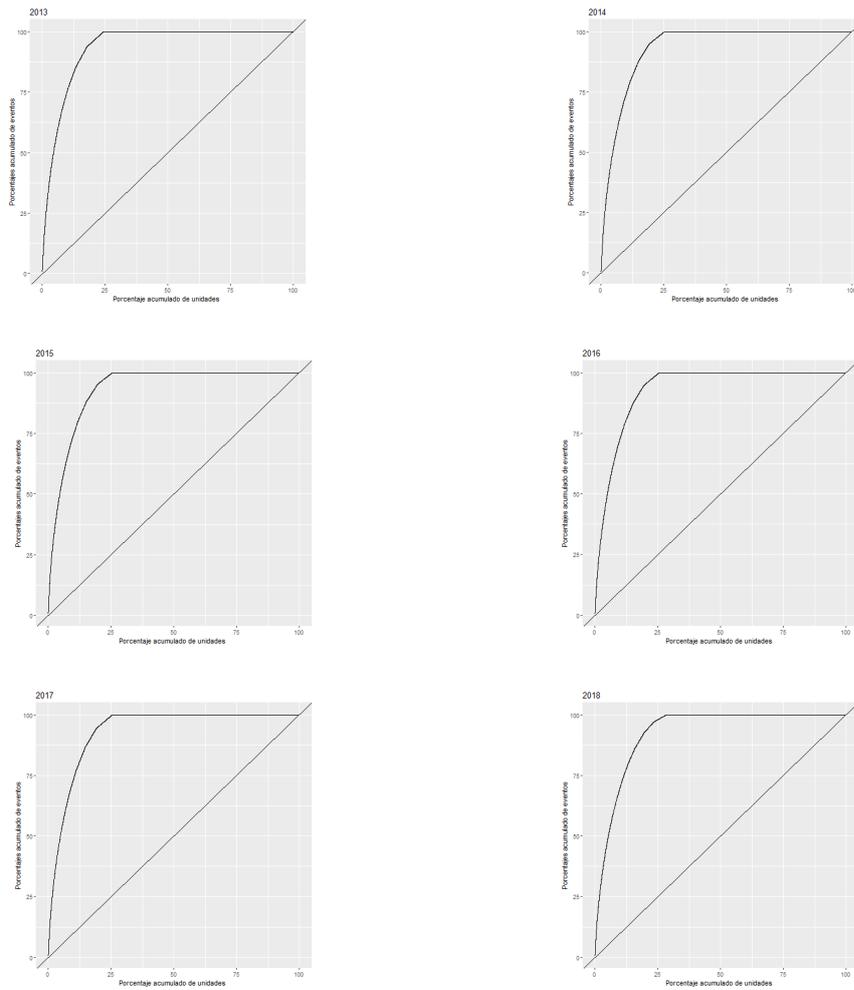
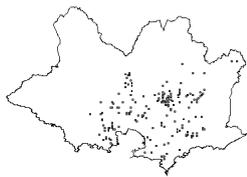


Figura 2.1: Estimación de curvas de concentración de Rapiña

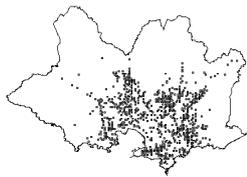
Las curvas de concentración son el resultado de ordenar las celdas de acuerdo al número de delitos. El eje de las ordenadas presenta el porcentaje acumulado de delitos y el de las abscisas el porcentaje acumulado de celdas, permitiendo establecer el nivel de concentración delictual.

Apéndice 3

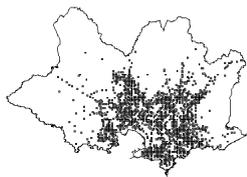
Huellas de patrullaje



(a)

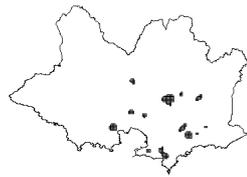


(b)



(c)

Figura 3.1: Conteo



(a)



(b)

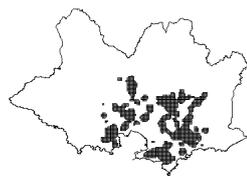
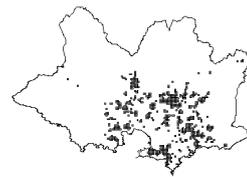
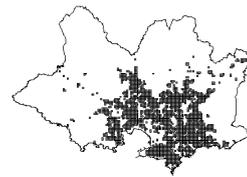


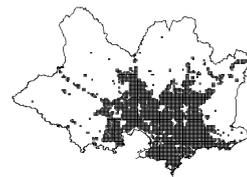
Figura 3.2: KDE



(a)

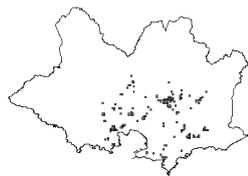


(b)



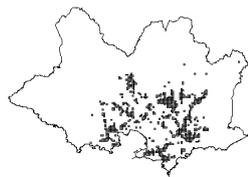
(c)

Figura 3.3: PHM

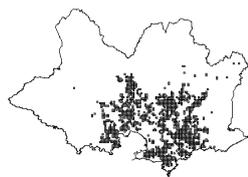


(a)

Huella MB gauss exp
patrullando 500 cuadrantes



(b)

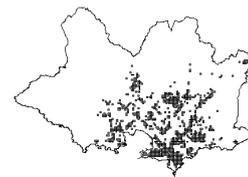


(c)

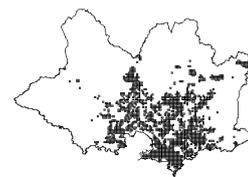
Figura 3.4: EP



(a)



(b)



(c)

Figura 3.5: EP cov

Las huellas de patrullaje permiten visualizar el número de celdas seleccionadas por los métodos durante los 20 días seleccionados al azar. (a) Indica que se seleccionaron 100 celdas, (b) 500 celdas, (c) 1000 celdas

Apéndice 4

Código en R

4.1. Preparación de base de datos

```
1 ##Inicio Proceso
2 rm(list=ls())
3 Base_final <- read_sav("2013-2018.sav")
4
5
6 ###Genero un var sobre tipo de delito
7 Base_final$RAPINA<-0
8 Base_final$RAPINA[ str_detect(Base_final$T[U+FFFD], "RAP[U+FFFD]")]<-1
9
10 ##Selecciono solo lo eventos de rapina
11 Base_final=subset(Base_final, RAPINA==1)
12 table(Base_final$Complemento, Base_final$[U+FFFD])
13
14 ###Re estructuro Lat y long
15 Base_final$X_num=as.numeric(str_replace(Base_final$X, ",", "."))
16 Base_final$Y_num=as.numeric(str_replace(Base_final$Y, ",", "."))
17
18 ##### Hora y fecha
19 ##Genero info de hecho en [U+FFFD]y fecha.
20 Base_final$Date <- Base_final$Hecho
21 Base_final$year=lubridate::year(Base_final$Date)
22 Base_final$day=lubridate::day(Base_final$Date)
23 Base_final$month=lubridate::month(Base_final$Date, label = TRUE)
24 Base_final$month_num=lubridate::month(Base_final$Date)
25 Base_final$hour=lubridate::hour(Base_final$Date)
26 Base_final$wday=lubridate::wday(Base_final$Date, label = TRUE)
27 Base_final$wday_num=lubridate::wday(Base_final$Date)
28
29 ##Genero una fecha simple y un indice correlativo desde 2013 enero 01
30 Base_final$fecha_simple<-lubridate::ymd(paste(Base_final$year, "-", Base_final$month_num, "-"
, Base_final$day))
31 Base_final$fecha_index<-difftime(Base_final$fecha_simple, lubridate::ymd(paste(2013, "-", 1,
"-", 1)), units="days")+1
32
33
34 ##Levanto el perimetro de montevideo
35 shp <- shapefile("./SHAPES/MVD.INE")
36 #Remuevo los delitos NA
37 BASE_COMPLETE<-Base_final[complete.cases(Base_final$X),]
38 BASE_COMPLETE<-as.data.frame(BASE_COMPLETE)
39 ##Remuevo lo que este por fuera de la ventana.Pierdo menos de un 1 por ciento.
40 coordinates(BASE_COMPLETE)<-c("X_num", "Y_num")
41 proj4string(BASE_COMPLETE) <- CRS("+init=epsg:32721")
42 W <- as.owin(shp)
43
44 #Todos
```

```

45 DELITOS_PPP<-as(BASE_COMPLETE,"ppp")
46 DELITOS_PPP<-DELITOS_PPP[W]
47 BASE_DELITOS_L<-DELITOS_PPP$marks
48 ##Guardo Numericos y redefino proy
49 BASE_DELITOS_L$X_num<-DELITOS_PPP$x
50 BASE_DELITOS_L$Y_num<-DELITOS_PPP$y
51 coordinates(BASE_DELITOS_L)<-c("X_num", "Y_num")
52 proj4string(BASE_DELITOS_L) <- CRS("+init=epsg:32721")
53
54 ##Preparo una grilla Espacial resolucio 200mts
55 grid <- raster(extent(shp))
56 res(grid) <- 200
57 proj4string(grid)<-proj4string(shp)
58 gridpolygon <- rasterToPolygons(grid)
59
60 # Intersecto grilla con contorno de MVD
61 MVD.grid <- raster::intersect(shp, gridpolygon)
62 ##Genero una ID correlativo y lo guardo en grila 2
63 MVD.grid@data$ID=seq(1,nrow(MVD.grid@data),1)
64 MVD.grid2<-subset(MVD.grid,select="ID")
65
66
67 ##Hago una rev de de proyecciones al mismo sistema.
68 keep(BASE_DELITOS_L,shp,MVD.grid2,sure=TRUE)
69 CRS.new<-CRS(proj4string(BASE_DELITOS_L))
70 shp <- spTransform(shp,CRS.new)
71 MVD.grid2 <- spTransform(MVD.grid2,CRS.new)
72
73 ##Save intermedio
74 save.image("PRUEBA.RData")
75 ##Genero centroides de las grilla
76 MVD.grid2_CENT<-gCentroid(MVD.grid2,byid=TRUE,id=MVD.grid2$ID)
77
78 #Extraigo covariables de open street map
79
80 AMENITY <- opq ("Montevideo Uruguay") %%
81 add_osm_feature(key = "amenity", value = "pub")
82 AMENITY <- osmdata_sp(AMENITY)
83 AMENITY_pts<-(AMENITY$osm.points)
84 AMENITY_pol<-(AMENITY$osm.polygons)
85 AMENITY_pol <- spTransform(AMENITY_pol,CRS.new)
86 AMENITY_pts <- spTransform(AMENITY_pts,CRS.new)
87
88 INDUSTRIAL <- opq ("Montevideo Uruguay") %%
89 add_osm_feature(key = "landuse", value = "industrial")
90 INDUSTRIAL <- osmdata_sp(INDUSTRIAL)
91 INDUSTRIAL<-INDUSTRIAL$osm.polygons
92 INDUSTRIAL <- spTransform(INDUSTRIAL,CRS.new)
93
94 ## DE LA IMM
95 Asentamientos <- shapefile("./Covariadas/asentamientos2007_2010")
96 Asentamientos <- spTransform(Asentamientos,CRS.new)
97
98 ##FINANCIAL & COMERCE
99 ATM<-subset(AMENITY_pts,amenity=="atm")
100 ATM_2<-subset(AMENITY_pol,amenity=="atm")
101 bank<-subset(AMENITY_pts,amenity=="bank")
102 bank_2<-subset(AMENITY_pol,amenity=="bank")
103 exchange<-subset(AMENITY_pts,amenity=="bureau_de_change")
104 exchange_2<-subset(AMENITY_pol,amenity=="bureau_de_change")
105 fuel<-subset(AMENITY_pts,amenity=="fuel")
106
107 ##MEDICAL
108 HOSP<-subset(AMENITY_pts,amenity=="hospital")
109 HOSP_2<-subset(AMENITY_pol,amenity=="hospital")
110 clinic<-subset(AMENITY_pts,amenity=="clinic")
111 clinic_2<-subset(AMENITY_pol,amenity=="clinic")
112
113
114 ##PUBLICS
115 table(AMENITY_pts$amenity)
116 BAR<-subset(AMENITY_pts,amenity=="bar")
117 BAR_2<-subset(AMENITY_pol,amenity=="bar")

```

```

118 pub<-subset (AMENITY_pts , amenity=="pub")
119 pub_2<-subset (AMENITY_pol , amenity=="pub")
120 restaurant<-subset (AMENITY_pts , amenity=="restaurant")
121 restaurant_2<-subset (AMENITY_pol , amenity=="restaurant")
122
123 nightclub<-subset (AMENITY_pts , amenity=="nightclub")
124 nightclub<-subset (AMENITY_pol , amenity=="nightclub")
125
126
127 ##Educ
128 school<-subset (AMENITY_pts , amenity=="school")
129 school_2<-subset (AMENITY_pol , amenity=="school")
130
131 college<-subset (AMENITY_pts , amenity=="college")
132 college_2<-subset (AMENITY_pol , amenity=="college")
133
134 university<-subset (AMENITY_pts , amenity=="university")
135 university_2<-subset (AMENITY_pol , amenity=="university")
136
137
138 ##Buffer de covariables 300 metros
139 ## INICIO CONTADOR en 0
140 MVD.grid2$ATM<-0
141 MVD.grid2$bank<-0
142 MVD.grid2$fuel<-0
143 MVD.grid2$exchange<-0
144 MVD.grid2$HOSP<-0
145 MVD.grid2$clinic<-0
146 MVD.grid2$Asentamientos<-0
147 MVD.grid2$SALUD_PRIVADO<-0
148 MVD.grid2$SALUD_PUBLICA<-0
149 MVD.grid2$SALUD_POLICLI<-0
150 MVD.grid2$edu_terciaria_PR<-0
151 MVD.grid2$edu_terciaria_PU<-0
152 MVD.grid2$edu_terciaria_NOU<-0
153 MVD.grid2$Liceo_Publico<-0
154 MVD.grid2$Liceo_Privado<-0
155 MVD.grid2$UTU<-0
156 MVD.grid2$INDUSTRIAL<-0
157 MVD.grid2$bar<-0
158 MVD.grid2$pub<-0
159 MVD.grid2$restaurant<-0
160 MVD.grid2$school<-0
161 MVD.grid2$college<-0
162 MVD.grid2$university<-0
163 ##Comienzo el contador
164 for ( i in 1:dim(MVD.grid2) [1] ) {
165 auxi_buff_0<-MVD.grid2_CENT[i]
166 auxi_buff_1<-gBuffer (auxi_buff_0,width =300 )
167 mapview (auxi_buff_1)
168 ##COMERCIO Y FIN
169 MVD.grid2@data [ i , ] $ATM<-nrow (ATM [ ! is.na (over (ATM, auxi_buff_1) ) , ])
170 MVD.grid2@data [ i , ] $bank<-nrow (bank [ ! is.na (over (bank, auxi_buff_1) ) , ]) + nrow (bank_2 [ ! is.na (
over (bank_2, auxi_buff_1) ) , ])
171 MVD.grid2@data [ i , ] $fuel<-nrow (fuel [ ! is.na (over (fuel, auxi_buff_1) ) , ])
172 MVD.grid2@data [ i , ] $exchange<-nrow (exchange [ ! is.na (over (exchange, auxi_buff_1) ) , ]) + nrow (
exchange_2 [ ! is.na (over (exchange_2, auxi_buff_1) ) , ])
173
174 MVD.grid2@data [ i , ] $bar<-nrow (BAR [ ! is.na (over (BAR, auxi_buff_1) ) , ]) + nrow (BAR_2 [ ! is.na (over (
BAR_2, auxi_buff_1) ) , ])
175 MVD.grid2@data [ i , ] $pub<-nrow (pub [ ! is.na (over (pub, auxi_buff_1) ) , ]) + nrow (pub_2 [ ! is.na (over (
pub_2, auxi_buff_1) ) , ])
176 MVD.grid2@data [ i , ] $restaurant<-nrow (restaurant [ ! is.na (over (restaurant, auxi_buff_1) ) , ]) +
nrow (restaurant_2 [ ! is.na (over (restaurant_2, auxi_buff_1) ) , ])
177
178 #LAND USE
179 MVD.grid2@data [ i , ] $Asentamientos<- ifelse (nrow (Asentamientos [ ! is.na (over (Asentamientos, auxi
_buff_1) ) , ]) > 1, 1, 0)
180 MVD.grid2@data [ i , ] $INDUSTRIAL<- ifelse (nrow (INDUSTRIAL [ ! is.na (over (INDUSTRIAL, auxi_buff_1) )
, ]) > 1, 1, 0)
181
182 MVD.grid2@data [ i , ] $HOSP<-nrow (HOSP [ ! is.na (over (HOSP, auxi_buff_1) ) , ]) + nrow (HOSP_2 [ ! is.na (
over (HOSP_2, auxi_buff_1) ) , ])

```

```

183 MVD.grid2@data[i,]$clinic<-nrow(clinic[!is.na(over(clinic,aux_i_buff_1)),])+nrow(clinic_2[!
    is.na(over(clinic_2,aux_i_buff_1)),])
184
185 MVD.grid2@data[i,]$$SALUD_PRIVADO<-nrow(SALUD_PRIVADO[!is.na(over(SALUD_PRIVADO,aux_i_buff_
    1)),])
186 MVD.grid2@data[i,]$$SALUD_PUBLICA<-nrow(SALUD_PUBLICA[!is.na(over(SALUD_PUBLICA,aux_i_buff_
    1)),])
187 MVD.grid2@data[i,]$$SALUD_POLICLI<-nrow(SALUD_POLICLI[!is.na(over(SALUD_POLICLI,aux_i_buff_
    1)),])
188 MVD.grid2@data[i,]$$sanatorio<-nrow(sanatorio[!is.na(over(sanatorio,aux_i_buff_1)),])
189
190 MVD.grid2@data[i,]$$edu_terciaria_PR<-ifelse(nrow(edu_terciaria_PR[!is.na(over(edu_
    terciaria_PR,aux_i_buff_1)),])>1,1,0)
191 MVD.grid2@data[i,]$$edu_terciaria_PU<-ifelse(nrow(edu_terciaria_PU[!is.na(over(edu_
    terciaria_PU,aux_i_buff_1)),])>1,1,0)
192 MVD.grid2@data[i,]$$edu_terciaria_NOU<-ifelse(nrow(edu_terciaria_NOU[!is.na(over(edu_
    terciaria_NOU,aux_i_buff_1)),])>1,1,0)
193 MVD.grid2@data[i,]$$Liceo_Publico<-ifelse(nrow(Liceo_Publico[!is.na(over(Liceo_Publico,aux_i
    _buff_1)),])>1,1,0)
194 MVD.grid2@data[i,]$$Liceo_Privado<-ifelse(nrow(Liceo_Privado[!is.na(over(Liceo_Privado,aux_i
    _buff_1)),])>1,1,0)
195 MVD.grid2@data[i,]$$UTU<-ifelse(nrow(UTU[!is.na(over(UTU,aux_i_buff_1)),])>1,1,0)
196
197 ##EDUC
198 MVD.grid2@data[i,]$$school<-nrow(school[!is.na(over(school,aux_i_buff_1)),])+nrow(school_2[!
    is.na(over(school_2,aux_i_buff_1)),])
199 MVD.grid2@data[i,]$$college<-nrow(college[!is.na(over(college,aux_i_buff_1)),])+nrow(college
    _2[!is.na(over(college_2,aux_i_buff_1)),])
200 MVD.grid2@data[i,]$$university<-nrow(university[!is.na(over(university,aux_i_buff_1)),])+
    nrow(university_2[!is.na(over(university_2,aux_i_buff_1)),])
201 }
202
203 ##Cierra el proceso y propongo dicotomizar.
204 ##Reestructuro en factores
205 MVD.grid2$ATM<-ifelse(MVD.grid2$ATM==0,0,1)
206 MVD.grid2$bank<-ifelse(MVD.grid2$bank==0,0,1)
207 MVD.grid2$fuel<-ifelse(MVD.grid2$fuel==0,0,1)
208 MVD.grid2$exchange<-ifelse(MVD.grid2$exchange==0,0,1)
209 MVD.grid2$SHOSP<-ifelse(MVD.grid2$SHOSP==0,0,1)
210 MVD.grid2$clinic<-ifelse(MVD.grid2$clinic==0,0,1)
211 MVD.grid2$Asentamientos<-ifelse(MVD.grid2$Asentamientos==0,0,1)
212 MVD.grid2$INDUSTRIAL<-ifelse(MVD.grid2$INDUSTRIAL==0,0,1)
213 MVD.grid2$bar<-ifelse(MVD.grid2$bar==0,0,1)
214 MVD.grid2$pub<-ifelse(MVD.grid2$pub==0,0,1)
215 MVD.grid2$restaurant<-ifelse(MVD.grid2$restaurant==0,0,1)
216 MVD.grid2$school<-ifelse(MVD.grid2$school==0,0,1)
217 MVD.grid2$college<-ifelse(MVD.grid2$college==0,0,1)
218 MVD.grid2$university<-ifelse(MVD.grid2$university==0,0,1)
219
220
221
222 ##Luego de este proceso le pongo a cada delito, que grid pertenece.
223 BASE.DELITOS.L <- point.in.poly(BASE.DELITOS.L, MVD.grid2)
224
225
226 ##Hago una preparacion de grilla espacio temporal para modelos Epi
227 ##Hago un grid para el surveillance.
228 tiles2stgrid <- function(tiles, start, T){
229 start <- sort.int(unique.default(start))
230 stgrid <- expand.grid(tile = row.names(tiles),
231 start = start,KEEP.OUT.ATTRS = FALSE, stringsAsFactors = TRUE)
232 cbind(stgrid,stop = rep(c(start[-1L], T), each = length(tiles)),
233 area = rep(rgeos::gArea(tiles, byid = TRUE), length(start)))}
234
235 stgrid_MVD <- tiles2stgrid(MVD.grid2, start = min(BASE.DELITOS.L$fecha_index),
236 T =max(BASE.DELITOS.L$fecha_index))
237
238 stgrid_MVD<-merge(stgrid_MVD,MVD.grid2,by.x="tile",by.y="ID",all.x=TRUE)
239 #Replico
240
241 stgrid_MVD2<-bind.rows(replicate(max(stgrid_MVD$stop)
242 , stgrid_MVD, simplify = FALSE))
243 stgrid_MVD2<-stgrid_MVD2 %% group_by(tile) %%

```

```

244 mutate(start = row_number())
245
246 stgrid_MVD2<-stgrid_MVD2 %>%
247 mutate(stop = start+1)
248
249
250 ##Creo un Bounding box.
251 lin <- as(shp, "SpatialLinesDataFrame")
252 pts <- as.data.frame(as(lin, "SpatialPointsDataFrame"))
253 BOUND<-as.data.frame(cbind(x=pts$coords.x1,
254 y=pts$coords.x2))
255
256 keep(BASE_DELITOS.L,BOUND,shp,MVD.grid2, stgrid_MVD2,sure=TRUE)
257 save.image("PRUEBA.RData")

```

4.2. Análisis descriptivo

4.2.1. Funciones para descriptivos y utilitarias

```

1 TABLA_CONCETRACION<-function(vector_r,anio){
2   aa<-sort(vector_r,TRUE)
3   seq_a<-1:length(aa)
4   acum_places <- 100*seq_a/length(seq_a)
5   acum_eventos <- 100*cumsum(aa)/sum(aa)
6   data_f<-as.data.frame(cbind(acum_places,acum_eventos))
7
8   return(data_f)
9 }
10 CURVA_CONCETRACION<-function(vector_r,anio){
11   aa<-sort(vector_r,TRUE)
12   seq_a<-1:length(aa)
13   acum_places <- 100*seq_a/length(seq_a)
14   acum_eventos <- 100*cumsum(aa)/sum(aa)
15   data_f<-as.data.frame(cbind(acum_places,acum_eventos))
16   ggplot(data_f, aes(acum_places,acum_eventos)) +
17     geom_line()+
18     # coord_cartesian(xlim = c(0,100))+
19     labs(title = paste(anio))+
20     xlab("Porcentaje acumulado de unidades ") + ylab("Porcentajes acumulado de eventos")+
21     geom_abline(slope=1, intercept=0)
22 }

```

4.2.2. Análisis descriptivos

```

1 rm(list=ls())
2 load("PRUEBA.RData")
3 rm(stgrid_MVD2)
4
5
6 ##Visualizo los datos
7 dates <- as.Date("2013-01-01")+BASE_DELITOS.L$fecha_index-1
8 dates<-as.POSIXlt(dates)
9 BASE_DELITOS.L$time<-as.numeric(dates)
10
11 ##Preparo un analisis Grafico y descriptivos simples.
12 CRS.new<-CRS(proj4string(BASE_DELITOS.L))
13 shp <- spTransform(shp,CRS.new)
14 MVD.grid2 <- spTransform(MVD.grid2,CRS.new)
15
16 serie<-as.data.frame(BASE_DELITOS.L) %>%
17 group_by(fecha_simple) %>%
18 summarize(RAPINA=sum(RAPINA, na.rm = TRUE))
19
20 plot(shp)#,main="Representación puntos")
21 plot(BASE_DELITOS.L,pch=".",add=T)
22
23 plot(serie$RAPINA,type="l",ylab="Denuncias",xlab="Dias desde 1 de enero 2013")

```

```

24
25 ## Concentracion delito, Panel
26 MVD.grid2$R_2013<-GISTools::poly.counts(subset(BASE.DELITOS.L,year==2013), MVD.grid2)
27 MVD.grid2$R_2014<-GISTools::poly.counts(subset(BASE.DELITOS.L,year==2014), MVD.grid2)
28 MVD.grid2$R_2015<-GISTools::poly.counts(subset(BASE.DELITOS.L,year==2015), MVD.grid2)
29 MVD.grid2$R_2016<-GISTools::poly.counts(subset(BASE.DELITOS.L,year==2016), MVD.grid2)
30 MVD.grid2$R_2017<-GISTools::poly.counts(subset(BASE.DELITOS.L,year==2017), MVD.grid2)
31 MVD.grid2$R_2018<-GISTools::poly.counts(subset(BASE.DELITOS.L,year==2018), MVD.grid2)
32
33
34 TABLA_MEDIAS<-rbind(
35 cbind(sum(MVD.grid2$R_2013),sum(MVD.grid2$R_2013>0)/nrow(MVD.grid2),mean(MVD.grid2$R_2013)
, sd(MVD.grid2$R_2013),Gini(MVD.grid2$R_2013)),
36 cbind(sum(MVD.grid2$R_2014),sum(MVD.grid2$R_2014>0)/nrow(MVD.grid2),mean(MVD.grid2$R_2014)
, sd(MVD.grid2$R_2014),Gini(MVD.grid2$R_2014)),
37 cbind(sum(MVD.grid2$R_2015),sum(MVD.grid2$R_2015>0)/nrow(MVD.grid2),mean(MVD.grid2$R_2015)
, sd(MVD.grid2$R_2015),Gini(MVD.grid2$R_2015)),
38 cbind(sum(MVD.grid2$R_2016),sum(MVD.grid2$R_2016>0)/nrow(MVD.grid2),mean(MVD.grid2$R_2016)
, sd(MVD.grid2$R_2016),Gini(MVD.grid2$R_2016)),
39 cbind(sum(MVD.grid2$R_2017),sum(MVD.grid2$R_2017>0)/nrow(MVD.grid2),mean(MVD.grid2$R_2017)
, sd(MVD.grid2$R_2017),Gini(MVD.grid2$R_2017)),
40 cbind(sum(MVD.grid2$R_2018),sum(MVD.grid2$R_2018>0)/nrow(MVD.grid2),mean(MVD.grid2$R_2018)
, sd(MVD.grid2$R_2018),Gini(MVD.grid2$R_2018)))
41 colnames(TABLA_MEDIAS)<-c("Total","Celdas Afectadas","Promedio","Desvio estandar","Gini")
42 rownames(TABLA_MEDIAS)<-c("2013","2014","2015","2016","2017","2018")
43
44 ##Preparp curva.
45
46 R13<-CURVA_CONCETRACION(MVD.grid2$R_2013,2013)
47 R14<-CURVA_CONCETRACION(MVD.grid2$R_2014,2014)
48 R15<-CURVA_CONCETRACION(MVD.grid2$R_2015,2015)
49 R16<-CURVA_CONCETRACION(MVD.grid2$R_2016,2016)
50 R17<-CURVA_CONCETRACION(MVD.grid2$R_2017,2017)
51 R18<-CURVA_CONCETRACION(MVD.grid2$R_2018,2018)
52
53 R13
54 R14
55 R15
56 R16
57 R17
58 R18
59
60 TR13<-TABLA_CONCETRACION(MVD.grid2$R_2013,2013)
61 TR14<-TABLA_CONCETRACION(MVD.grid2$R_2014,2014)
62 TR15<-TABLA_CONCETRACION(MVD.grid2$R_2015,2015)
63 TR16<-TABLA_CONCETRACION(MVD.grid2$R_2016,2016)
64 TR17<-TABLA_CONCETRACION(MVD.grid2$R_2017,2017)
65 TR18<-TABLA_CONCETRACION(MVD.grid2$R_2018,2018)
66
67 TABLA_CONCETRACION2<-rbind(
68 cbind(TR13[TR13$acum_eventos >=25,][1,1],TR13[TR13$acum_eventos >=50,][1,1]),
69 cbind(TR14[TR14$acum_eventos >=25,][1,1],TR14[TR14$acum_eventos >=50,][1,1]),
70 cbind(TR15[TR15$acum_eventos >=25,][1,1],TR15[TR15$acum_eventos >=50,][1,1]),
71 cbind(TR16[TR16$acum_eventos >=25,][1,1],TR16[TR16$acum_eventos >=50,][1,1]),
72 cbind(TR17[TR17$acum_eventos >=25,][1,1],TR17[TR17$acum_eventos >=50,][1,1]),
73 cbind(TR18[TR18$acum_eventos >=25,][1,1],TR18[TR18$acum_eventos >=50,][1,1]))
74 colnames(TABLA_CONCETRACION2)<-c("25 %","50 %")
75 rownames(TABLA_CONCETRACION2)<-c("2013","2014","2015","2016","2017","2018")
76 rm(R13,R14,R15,R16,R17,R18)
77 rm(TR13,TR14,TR15,TR16,TR17,TR18)
78
79
80 ##Intensidad 2d
81 DELITOS_PPP<-as(BASE.DELITOS.L,"ppp")
82 DELITOS_PPP$n
83
84 den_opti <- density(subset(DELITOS_PPP,year==2018)[W], sigma=300,diggle=T)
85 plot(den_opti,main="")
86
87 ##!= meses
88 den_opti_200_1 <- density(subset(DELITOS_PPP,year==2018 & month_num==1)[W], sigma=300,
diggle=T)
89 den_opti_200_6 <- density(subset(DELITOS_PPP,year==2018 & month_num==6)[W], sigma=300,

```

```

    diggle=T)
90 den_opti_200_12 <- density(subset(DELITOS_PPP, year==2018 & month_num==12 ) [W], sigma=300,
    diggle=T)
91
92 plot(den_opti_200_1, main="")
93 plot(den_opti_200_6, main="")
94 plot(den_opti_200_12, main="")
95
96 ##funcion k y L Solo para 2018
97 DELITOS_PPP2<-subset(DELITOS_PPP, year==2018) [W]
98 K <- Kest(DELITOS_PPP2)
99 L <- Lest(DELITOS_PPP2)
100 ##Funcion radial
101 pcf <- pcf(K)
102
103 plot(K, main="", xlim=c(0,500))
104 plot(L, main="", xlim=c(0,500))
105 plot(pcf, main="", xlim=c(0,500))
106
107 ##Version espacio temporal
108 DELITOS_PPP3<-subset(DELITOS_PPP2, month_num<=6)
109 DELITOS_PPP4<-subset(DELITOS_PPP2, month_num>6)
110
111
112 ##Trabajo solo con 180 dias
113 DATASET<-as.3dpoints(DELITOS_PPP3$x,
114 DELITOS_PPP3$y,
115 DELITOS_PPP3$marks$fecha_index)
116
117 u <- seq(0, 500, by = 50)
118 v <- seq(0, 60, by = 1) #
119
120 stik2 <- STIKhat(xyt = DATASET,
121 s.region = cbind(BOUND$x, BOUND$y),
122 dist = u, times = v, infectious = FALSE)
123
124 g <- PCFhat(xyt = DATASET,
125 dist = u,
126 times = v, s.region = cbind(BOUND$x, BOUND$y))
127
128
129 plotK(stik2, L=T, type="image", theta = -65, phi = 35)
130 plotK(stik2, L=T, type="persp", theta = -65, phi = 35)
131
132 plotPCF(g, type="persp", theta = 5, phi = 35)
133 plotPCF(g, type="image", theta = 5, phi = 35)
134
135 ## En covariables
136 rbind(table(DELITOS_PPP2$marks$ATM>0)[2]/DELITOS_PPP2$n*100,
137 table(DELITOS_PPP2$marks$bank>0)[2]/DELITOS_PPP2$n*100,
138 table(DELITOS_PPP2$marks$fuel>0)[2]/DELITOS_PPP2$n*100,
139 table(DELITOS_PPP2$marks$exchange>0)[2]/DELITOS_PPP2$n*100)
140
141 rbind(table(DELITOS_PPP2$marks$HOSP>0)[2]/DELITOS_PPP2$n*100,
142 table(DELITOS_PPP2$marks$clinic>0)[2]/DELITOS_PPP2$n*100)
143
144 rbind(table(DELITOS_PPP2$marks$bar>0)[2]/DELITOS_PPP2$n*100,
145 table(DELITOS_PPP2$marks$pub>0)[2]/DELITOS_PPP2$n*100,
146 table(DELITOS_PPP2$marks$restaurant>0)[2]/DELITOS_PPP2$n*100)
147
148 rbind(table(DELITOS_PPP2$marks$school>0)[2]/DELITOS_PPP2$n*100,
149 table(DELITOS_PPP2$marks$college>0)[2]/DELITOS_PPP2$n*100,
150 table(DELITOS_PPP2$marks$university>0)[2]/DELITOS_PPP2$n*100)
151
152 rbind(table(DELITOS_PPP2$marks$INDUSTRIAL>0)[2]/DELITOS_PPP2$n*100,
153 table(DELITOS_PPP2$marks$Asentamientos>0)[2]/DELITOS_PPP2$n*100)

```

4.3. Comparación de métodos.

4.3.1. Funciones para estimación y utilitarias

```
1 ##Funcion de predici[U+FFFD]n a un paso, t+1
2 PREDICT_1step <- function( fecha_pred, cte_tt, cte_esp, eps_t ){
3   ##Defino la fecha del indice para el cual voy a predecir
4   print( paste( " Realizando predici[U+FFFD]n para", fecha_pred ) )
5   INDICE_pred <- as.numeric( min( BASE_DELITOS_L[BASE_DELITOS_L$fecha_simple==fecha_pred,] $ fecha
6     _index ) )
7
8   ##Preparo contraste con realidad
9   COPIA <- as.data.frame( subset( MVD.grid2, select="ID" ) )
10  print( " Estableciendo constraste" )
11  CONTRASTE <- subset( BASE_DELITOS_L, fecha_index==INDICE_pred )
12
13  poly.counts( CONTRASTE, MVD.grid2 ) -> GRID_REAL_count
14  COPIA$REAL <- GRID_REAL_count
15
16  ##Genero muestra de entrenamiento
17  BASE_DELITOS_L <- subset( BASE_DELITOS_L, fecha_index < INDICE_pred & fecha_index >= INDICE_ini )
18
19  ##Hago conteo
20  print( " Estimaci[U+FFFD]n conteo" )
21  poly.counts( BASE_DELITOS_L, MVD.grid2 ) -> GRID_AUXI_count
22  COPIA$CONTEO_PASADO <- GRID_AUXI_count
23
24
25  ##Preparo centroide
26  MVD.grid2_CENT <- gCentroid( MVD.grid2, byid=TRUE, id=MVD.grid2$ID )
27  print( " Estimaci[U+FFFD]n KDE" )
28
29  ##realizo KDE en centroides.
30  kernel_est <- lambdahat( coordinates( BASE_DELITOS_L ), h=cte_esp,
31    gpts = coordinates( MVD.grid2_CENT ),
32    poly = NULL, edge = FALSE )
33
34  COPIA$kernel_est <- kernel_est$lambda
35  rm( kernel_est )
36  ##Realizo PHM
37  print( " Estimaci[U+FFFD]n PHM" )
38  PHM_est <- rep( 0, dim( COPIA ) [1] )
39  for ( i in 1:dim( COPIA ) [1] ) {
40    auxi_buff_0 <- MVD.grid2_CENT[ i ]
41    auxi_buff_1 <- gBuffer( auxi_buff_0, width = cte_esp )
42    pts_in <- BASE_DELITOS_L[ ! is.na( over( BASE_DELITOS_L, auxi_buff_1 ) ), ]
43    ##Si hay un punto en la zona de interes hago algo
44    if ( dim( pts_in@data ) [1] > 0 ) {
45      dist_x <- gDistance( pts_in, auxi_buff_0, byi=TRUE )
46      dist_t <- INDICE_pred - pts_in$fecha_index
47      flag_m60 <- ifelse( dist_t <= eps_t, 1, 0 )
48      PHM_est[ i ] <- sum( flag_m60 / ( ( 1 + 2 * dist_x / 200 ) * ( 1 + as.numeric( dist_t ) / 7 * 1 ) ) )
49    } }
50
51  COPIA$PHM_est <- PHM_est
52  rm( PHM_est )
53
54  ##Realizo EP
55  print( " Estimaci[U+FFFD]n En-Ep" )
56  print( " Estimaci[U+FFFD]n En-Ep: Base" )
57  MODELO_BASICO_0 <- twinstim( endemic = ~1+ATM+bank+fuel+exchange+
58    HOSP+clinic+Asentamientos+INDUSTRIAL+
59    bar+pub+restaurant+school+college+university,
60    epidemic = ~0,
61    data = myimdepi_MVD,
62    model=TRUE, verbose=FALSE, t0=INDICE_ini, T=INDICE_pred, finetune = TRUE )
63
64
65  ##Modelo ep con covariadas
66  MODELO_BASICO_1 <- update( MODELO_BASICO_0,
```

```

67 epidemic = ~1,
68 siaf=siaf.gaussian(), tiaf=tiaf.exponential(),
69 data = myimdepi.MVD,
70 model=TRUE, t0=INDICE_ini, T=INDICE_pred, finetune = FALSE, verbose=FALSE)
71 aa<-intensityplot(MODELO_BASICO_1, which = c("total intensity"),
72 aggregate = c("space"), tiles =MVD.grid2, plot = FALSE)
73
74 COPIA$Mod.Gauss_exp<-aa(coordinates(MVD.grid2))
75 COPIA$Mod.Gauss_exp<-ifelse(is.na(COPIA$Mod.Gauss_exp), 0,
76 COPIA$Mod.Gauss_exp)
77
78 rm(MODELO_BASICO_1, aa)
79
80 ### Modelo solo epidemico
81 MODELO_BASICO_2<-twinstim(endemic = ~1,
82 epidemic = ~1,
83 siaf=siaf.gaussian(), tiaf=tiaf.exponential(),
84 data = myimdepi.MVD,
85 model=TRUE, t0=INDICE_ini, T=INDICE_pred)
86 aa<-intensityplot(MODELO_BASICO_2, which = c("total intensity"),
87 aggregate = c("space"), tiles =MVD.grid2, plot = FALSE, sgrid=30)
88
89 COPIA$MB_gauss_exp<-aa(coordinates(MVD.grid2))
90 COPIA$MB_gauss_exp<-ifelse(is.na(COPIA$MB_gauss_exp), 0,
91 COPIA$MB_gauss_exp)
92 rm(MODELO_BASICO_2, aa)
93 print("Pronto")
94 return(COPIA)
95
96 }
97
98 ##Calculo el PAI
99 PAI_STEP<-function(SALIDA, COL, percentil_corte){
100 print(paste("Calculando para:", names(SALIDA)[COL]))
101 PTO<-quantile(SALIDA[, COL], seq(0, 1, .01))[percentil_corte+1]
102 PRED<-SALIDA[SALIDA[, COL]>PTO,]
103 HIT<-sum(PRED$REAL)/sum(SALIDA$REAL)*100
104 AREA<-nrow(PRED)/nrow(SALIDA)*100
105 PAI<-HIT/AREA
106 K<-nrow(PRED)
107 REPORTE<-cbind(HIT, AREA, PAI, K)
108 rownames(REPORTE)<-names(SALIDA)[COL]
109 return(REPORTE)
110 }
111 ## Curva
112 HIT_K_FIX<-function(SALIDA, max_size, step){
113 for(k in 3:dim(SALIDA)[2]) {
114 x = SALIDA[order(SALIDA[, k], decreasing= T),]
115 A<-seq(0, max_size, step)
116 HIT<-rep(0, length(A))
117 for(j in 2:length(A)){
118 HIT[j]<-sum(x[1:A[j],]$REAL)/sum(SALIDA$REAL)*100
119 }
120 HIT<-as.data.frame(HIT)
121 HIT$Lugares<-A
122 HIT$metodo<-names(SALIDA)[k]
123
124 if(k==3){HIT_DF<-HIT} else {HIT_DF<-rbind(HIT_DF, HIT)}
125 }
126 return(HIT_DF)
127 }
128
129 ##Calculo de ENTROPIA
130 ENTROPY_LISTA<-function(bind_modelo, step){
131 countflag<-0
132 for(k in 1:length(step)){
133 for(j in 2:dim(bind_modelo)[2]) {
134 x = bind_modelo[order(bind_modelo[, j], decreasing= T),]
135 LISTA<-x[1:step[k], 1]
136 flag<-bind_modelo$ID %n% LISTA
137 countflag<-countflag+flag
138 }
139 bind_modelo$countflag<-countflag

```

```

140 pi<-bind_modelo$countflag/sum(bind_modelo$countflag)
141 pi2<-pi [ pi > 0 ]
142 p3<-sum(pi2*log(pi2,base=2))
143 if (k==1){H<-p3} else {H<-cbind(H,p3)}
144 return (H)
145 }
146
147 ##Estudio de Mapas
148 Plot_cluster_esp<-function ( Salida_merge,max_size , fecha_pred ){
149 auxi<-subset (BASE_DELITOS_L,BASE_DELITOS_L$fecha_simple==fecha_pred)
150 for ( k in 3:dim ( Salida_merge ) [ 2 ] ) {
151 x = Salida_merge@data[order ( Salida_merge@data[k] , decreasing= T ) ,]
152 Salida_merge$FLAG<-Salida_merge$ID %n %x [ 1 : max_size , ] $ID
153 Salida_FLAG<-subset ( Salida_merge , FLAG==1)
154 Salida_FLAG<-gBuffer ( Salida_FLAG,width=.5,byid = TRUE)
155 Salida_FLAG<-disaggregate ( aggregate ( Salida_FLAG,dissolve=TRUE) )
156 plot (shp ,main=paste (" Puntos calientes seleccionados" ,names ( Salida_merge ) [k] ,
157 "\n patrullando" ,max_size , " cuadrantes" ) )
158 plot ( Salida_FLAG ,add=TRUE)
159 plot (auxi ,col="red" ,pch="X" ,add=TRUE ,size=.5)
160 }}
161
162 Plot_cluster_print<-function ( Salida_merge,max_size , fecha_pred ){
163 auxi<-subset (BASE_DELITOS_L,BASE_DELITOS_L$fecha_simple==fecha_pred)
164 for ( k in 3:dim ( Salida_merge ) [ 2 ] ) {
165 x = Salida_merge@data[order ( Salida_merge@data[k] , decreasing= T ) ,]
166 Salida_merge$FLAG<-Salida_merge$ID %n %x [ 1 : max_size , ] $ID
167 Salida_FLAG<-subset ( Salida_merge , FLAG==1)
168 Salida_FLAG<-gBuffer ( Salida_FLAG,width=.5,byid = TRUE)
169 Salida_FLAG<-disaggregate ( aggregate ( Salida_FLAG,dissolve=TRUE) )
170 png (paste (" FIG / " , "PC" ,names ( Salida_merge ) [k] ,
171 "n" ,max_size , ".png" ,sep = " " ) )
172 plot (shp ,main=" " )
173 plot ( Salida_FLAG ,add=TRUE)
174 plot (auxi ,col="red" ,pch="X" ,add=TRUE ,size=.5)
175 dev.off ()
176 }}
177
178 ##Estudio de Huella de patrullaje
179 Plot_huella_esp<-function ( bind_modelo,max_size , nombre ){
180 Salida_merge<-subset (MVD.grid2 , select="ID" )
181 Salida_merge$FLAG<-Salida_merge$ID %n %unlist (match_LISTA_sample ( bind_modelo,max_size ) )
182
183 Salida_FLAG<-subset ( Salida_merge , FLAG==1)
184 plot (shp ,main=paste (" Huella" , nombre ,
185 "\n patrullando" ,max_size , " cuadrantes" ) )
186 plot ( Salida_FLAG ,add=TRUE)
187 }
188 Plot_huella_print<-function ( bind_modelo,max_size , nombre ){
189 Salida_merge<-subset (MVD.grid2 , select="ID" )
190 Salida_merge$FLAG<-Salida_merge$ID %n %unlist (match_LISTA_sample ( bind_modelo,max_size ) )
191
192 Salida_FLAG<-subset ( Salida_merge , FLAG==1)
193 png (paste (" FIG / " , "HUELLA" , nombre ,
194 "n" ,max_size , ".png" ,sep = " " ) )
195 plot (shp ,main=" " )
196 plot ( Salida_FLAG ,add=TRUE)
197 dev.off ()
198 }
199
200 ##Calculo de estructura de zonas de patrullaje
201 CALCULO_EST_TOUCH<-function ( Salida_merge,COL,max_size ){
202 x = Salida_merge@data[order ( Salida_merge@data [COL] , decreasing= T ) ,]
203 Salida_merge$FLAG<-Salida_merge$ID %n %x [ 1 : max_size , ] $ID
204 Salida_FLAG<-subset ( Salida_merge , FLAG==1)
205 Salida_FLAG<-gBuffer ( Salida_FLAG,width=.5,byid = TRUE)
206 Salida_FLAG<-disaggregate ( aggregate ( Salida_FLAG,dissolve=TRUE) )
207 POLY_FLAG<-polyPerimeter ( Salida_FLAG)
208 AREA_FLAG<-area ( Salida_FLAG)
209 CI<-(POLY_FLAG/sqrt (AREA_FLAG))
210 Mat_Distancias <- gDistance ( Salida_FLAG,byid = T)
211 PR_distancias<-sum (Mat_Distancias) / (dim (Mat_Distancias) [1] *(dim (Mat_Distancias) [1] -1))
212 SALIDA<-data.frame (UNIT=length (CI) ,

```

```

213 PR_Perimeter=mean(POLY_FLAG) ,
214 PR_AREA=mean(AREA_FLAG) ,
215 PR_CI=mean(CI) ,
216 PR_distancias)
217 }
218
219
220 ### Utilitarios , listas y mathcing
221 match_LISTA_simple<-function(bind_modelo,step){
222 x = bind_modelo[order(bind_modelo[,3], decreasing= T) ,]
223 LISTA_3<-x[1:step,1]
224 x = bind_modelo[order(bind_modelo[,4], decreasing= T) ,]
225 LISTA_4<-x[1:step,1]
226 x = bind_modelo[order(bind_modelo[,5], decreasing= T) ,]
227 LISTA_5<-x[1:step,1]
228 x = bind_modelo[order(bind_modelo[,6], decreasing= T) ,]
229 LISTA_6<-x[1:step,1]
230 x = bind_modelo[order(bind_modelo[,7], decreasing= T) ,]
231 LISTA_7<-x[1:step,1]
232
233 LISTA<-list(LISTA_3,LISTA_4,LISTA_5,LISTA_6,LISTA_7)
234 return(LISTA)
235 }
236 match_LISTA_MATRIX<-function(bind_modelo,step){
237 LISTA<-match_LISTA_simple(bind_modelo,step)
238 Matrix<-matrix(rep(0,5*5),5,5)
239 colnames(Matrix)<-names(bind_modelo[3:7])
240 rownames(Matrix)<-names(bind_modelo[3:7])
241
242 for(i in 1:5){
243 for(j in 1:5){
244 Matrix[i,j]<-length(intersect(unlist(LISTA[i]),unlist(LISTA[j])))/step*100
245 }}
246
247 return(Matrix)
248 }
249 match_LISTA_sample<-function(bind_modelo,step){
250 x = bind_modelo[order(bind_modelo[,2], decreasing= T) ,]
251 LISTA_2<-x[1:step,1]
252 x = bind_modelo[order(bind_modelo[,3], decreasing= T) ,]
253 LISTA_3<-x[1:step,1]
254 x = bind_modelo[order(bind_modelo[,4], decreasing= T) ,]
255 LISTA_4<-x[1:step,1]
256 x = bind_modelo[order(bind_modelo[,5], decreasing= T) ,]
257 LISTA_5<-x[1:step,1]
258 x = bind_modelo[order(bind_modelo[,6], decreasing= T) ,]
259 LISTA_6<-x[1:step,1]
260 x = bind_modelo[order(bind_modelo[,7], decreasing= T) ,]
261 LISTA_7<-x[1:step,1]
262 x = bind_modelo[order(bind_modelo[,8], decreasing= T) ,]
263 LISTA_8<-x[1:step,1]
264 x = bind_modelo[order(bind_modelo[,9], decreasing= T) ,]
265 LISTA_9<-x[1:step,1]
266 x = bind_modelo[order(bind_modelo[,10], decreasing= T) ,]
267 LISTA_10<-x[1:step,1]
268 x = bind_modelo[order(bind_modelo[,11], decreasing= T) ,]
269 LISTA_11<-x[1:step,1]
270
271 x = bind_modelo[order(bind_modelo[,12], decreasing= T) ,]
272 LISTA_12<-x[1:step,1]
273 x = bind_modelo[order(bind_modelo[,13], decreasing= T) ,]
274 LISTA_13<-x[1:step,1]
275 x = bind_modelo[order(bind_modelo[,14], decreasing= T) ,]
276 LISTA_14<-x[1:step,1]
277 x = bind_modelo[order(bind_modelo[,15], decreasing= T) ,]
278 LISTA_15<-x[1:step,1]
279 x = bind_modelo[order(bind_modelo[,16], decreasing= T) ,]
280 LISTA_16<-x[1:step,1]
281 x = bind_modelo[order(bind_modelo[,17], decreasing= T) ,]
282 LISTA_17<-x[1:step,1]
283 x = bind_modelo[order(bind_modelo[,18], decreasing= T) ,]
284 LISTA_18<-x[1:step,1]
285 x = bind_modelo[order(bind_modelo[,19], decreasing= T) ,]

```

```

286 LISTA_19<-x[1:step,1]
287 x = bind_modelo[order(bind_modelo[,20], decreasing= T),]
288 LISTA_20<-x[1:step,1]
289 LISTA<-list(LISTA_2,LISTA_3,LISTA_4,LISTA_5,LISTA_6,LISTA_7,LISTA_8,LISTA_9,LISTA_10,
290 LISTA_11,LISTA_12,LISTA_13,LISTA_14,LISTA_15,LISTA_16,LISTA_17,LISTA_18,LISTA_19,
291 LISTA_20)
292 return(LISTA)
293 }

```

4.3.2. Funciones para estimación y utilitarias

```

1
2
3 \subsection{Estimaci[U+FFFD]n 8 de junio }
4 \begin{lstlisting}[basicstyle=\tiny,language=Python]
5
6 ##Transformaci[U+FFFD]n espacial
7 CRS.new<-CRS(proj4string(BASE_DELITOS_L))
8 shp <- spTransform(shp,CRS.new)
9 MVD.grid2 <- spTransform(MVD.grid2,CRS.new)
10
11 ##Preparo base de modelos EP
12 Event_MVD<-subset(BASE_DELITOS_L,select=c("ID","fecha_index","wday_num","hour"))
13 Event_MVD$fecha_index<-as.numeric(Event_MVD$fecha_index)
14 colnames(Event_MVD@data)<-c("tile","time","wdayn","hour")
15 Event_MVD$eps.t<-eps_t
16 Event_MVD$eps.s<-ventana_espacial
17 Event_MVD$time<-Event_MVD$time+.5
18
19 myimdepi_MVD <- as.epidataCS(events = Event_MVD,
20 stgrid = as.data.frame(stgrid_MVD2),W = shp)
21
22 rm(Event_MVD,stgrid_MVD2)
23 rm(SALIDA_0)
24
25 ##PRUEBA FIJA
26 ventana_temporal<-365
27 ventana_espacial<-300
28 eps_t<-30
29 fecha_predict<-"2018-06-08"
30
31
32 SALIDA_0<-PREDICT_1step(fecha_predict,ventana_temporal,ventana_espacial
33 ,eps_t)
34
35
36 ##Calculos PAI
37 pp<-90
38 A90<-rbind(PAI_STEP(SALIDA_0,3,pp),
39 PAI_STEP(SALIDA_0,4,pp),
40 PAI_STEP(SALIDA_0,5,pp),
41 PAI_STEP(SALIDA_0,6,pp),
42 PAI_STEP(SALIDA_0,7,pp))
43
44 pp<-95
45 A95<-rbind(PAI_STEP(SALIDA_0,3,pp),
46 PAI_STEP(SALIDA_0,4,pp),
47 PAI_STEP(SALIDA_0,5,pp),
48 PAI_STEP(SALIDA_0,6,pp),
49 PAI_STEP(SALIDA_0,7,pp))
50
51
52 pp<-99
53 A99<-rbind(PAI_STEP(SALIDA_0,3,pp),
54 PAI_STEP(SALIDA_0,4,pp),
55 PAI_STEP(SALIDA_0,5,pp),
56 PAI_STEP(SALIDA_0,6,pp),
57 PAI_STEP(SALIDA_0,7,pp))
58
59
60 #Calculo de curva
61 summary(SALIDA_0)

```

```

62 CURVA_0<-rbind(HIT_K_FIX(SALIDA_0,1000,50))
63
64 CURVA_0$metodo<-ifelse(CURVA_0$metodo=="CONTEO_PASADO", "Conteo",
65 ifelse(CURVA_0$metodo=="PHM_est", "PHM",
66 ifelse(CURVA_0$metodo=="kernel_est", "KDE",
67 ifelse(CURVA_0$metodo=="MB.gauss_exp", "EP",
68 ifelse(CURVA_0$metodo=="Mod.Gauss_exp", "EPcov",
69 CURVA_0$metodo))))))
70
71 png("FIG/TRADE-OFF_0.png")
72 ggplot(CURVA_0, aes(x=Lugares, y=HIT, colour=metodo)) +
73 geom_line()+geom_point()+ylab("TE")+xlab("celdas")
74 dev.off()
75
76 ##Plot clusters
77 Salida_merge<-subset(MVD.grid2, select="ID")
78 Salida_merge<-merge(Salida_merge, SALIDA_0, all.x=TRUE)
79 Plot_cluster_esp(Salida_merge,100, fecha_predict)
80 Plot_cluster_esp(Salida_merge,500, fecha_predict)
81 Plot_cluster_esp(Salida_merge,1000, fecha_predict)
82
83 Plot_cluster_print(Salida_merge,100, fecha_predict)
84 Plot_cluster_print(Salida_merge,500, fecha_predict)
85 Plot_cluster_print(Salida_merge,1000, fecha_predict)
86
87 TAB0_CON<-rbind(
88 cbind(Metodo=CURVA_0[CURVA_0$metodo=="CONTEO_PASADO" & CURVA_0$Lugares=="100",3],
89 Lugares=CURVA_0[CURVA_0$metodo=="CONTEO_PASADO" & CURVA_0$Lugares=="100",2],
90 TE=CURVA_0[CURVA_0$metodo=="CONTEO_PASADO" & CURVA_0$Lugares=="100",1],
91 CALCULO_EST_TOUCH(Salida_merge,3,100)),
92
93 cbind(Metodo=CURVA_0[CURVA_0$metodo=="CONTEO_PASADO" & CURVA_0$Lugares=="500",3],
94 Lugares=CURVA_0[CURVA_0$metodo=="CONTEO_PASADO" & CURVA_0$Lugares=="500",2],
95 TE=CURVA_0[CURVA_0$metodo=="CONTEO_PASADO" & CURVA_0$Lugares=="500",1],
96 CALCULO_EST_TOUCH(Salida_merge,3,500)),
97
98 cbind(Metodo=CURVA_0[CURVA_0$metodo=="CONTEO_PASADO" & CURVA_0$Lugares=="1000",3],
99 Lugares=CURVA_0[CURVA_0$metodo=="CONTEO_PASADO" & CURVA_0$Lugares=="1000",2],
100 TE=CURVA_0[CURVA_0$metodo=="CONTEO_PASADO" & CURVA_0$Lugares=="1000",1],
101 CALCULO_EST_TOUCH(Salida_merge,3,1000))
102
103 TAB0_KDE<-rbind(
104 cbind(Metodo=CURVA_0[CURVA_0$metodo=="kernel_est" & CURVA_0$Lugares=="100",3],
105 Lugares=CURVA_0[CURVA_0$metodo=="kernel_est" & CURVA_0$Lugares=="100",2],
106 TE=CURVA_0[CURVA_0$metodo=="kernel_est" & CURVA_0$Lugares=="100",1],
107 CALCULO_EST_TOUCH(Salida_merge,4,100)),
108
109 cbind(Metodo=CURVA_0[CURVA_0$metodo=="kernel_est" & CURVA_0$Lugares=="500",3],
110 Lugares=CURVA_0[CURVA_0$metodo=="kernel_est" & CURVA_0$Lugares=="500",2],
111 TE=CURVA_0[CURVA_0$metodo=="kernel_est" & CURVA_0$Lugares=="500",1],
112 CALCULO_EST_TOUCH(Salida_merge,4,500)),
113
114 cbind(Metodo=CURVA_0[CURVA_0$metodo=="kernel_est" & CURVA_0$Lugares=="1000",3],
115 Lugares=CURVA_0[CURVA_0$metodo=="kernel_est" & CURVA_0$Lugares=="1000",2],
116 TE=CURVA_0[CURVA_0$metodo=="kernel_est" & CURVA_0$Lugares=="1000",1],
117 CALCULO_EST_TOUCH(Salida_merge,4,1000))
118
119
120 TAB0_PHM<-rbind(
121 cbind(Metodo=CURVA_0[CURVA_0$metodo=="PHM_est" & CURVA_0$Lugares=="100",3],
122 Lugares=CURVA_0[CURVA_0$metodo=="PHM_est" & CURVA_0$Lugares=="100",2],
123 TE=CURVA_0[CURVA_0$metodo=="PHM_est" & CURVA_0$Lugares=="100",1],
124 CALCULO_EST_TOUCH(Salida_merge,5,100)),
125
126 cbind(Metodo=CURVA_0[CURVA_0$metodo=="PHM_est" & CURVA_0$Lugares=="500",3],
127 Lugares=CURVA_0[CURVA_0$metodo=="PHM_est" & CURVA_0$Lugares=="500",2],
128 TE=CURVA_0[CURVA_0$metodo=="PHM_est" & CURVA_0$Lugares=="500",1],
129 CALCULO_EST_TOUCH(Salida_merge,5,500)),
130
131 cbind(Metodo=CURVA_0[CURVA_0$metodo=="PHM_est" & CURVA_0$Lugares=="1000",3],
132 Lugares=CURVA_0[CURVA_0$metodo=="PHM_est" & CURVA_0$Lugares=="1000",2],
133 TE=CURVA_0[CURVA_0$metodo=="PHM_est" & CURVA_0$Lugares=="1000",1],
134 CALCULO_EST_TOUCH(Salida_merge,5,1000))

```

```

135
136
137 TAB0_MOD<-rbind(
138 cbind(Metodo=CURVA_0[CURVA_0$metodo=="Mod_Gauss_exp" & CURVA_0$Lugares=="100",3],
139 Lugares=CURVA_0[CURVA_0$metodo=="Mod_Gauss_exp" & CURVA_0$Lugares=="100",2],
140 TE=CURVA_0[CURVA_0$metodo=="Mod_Gauss_exp" & CURVA_0$Lugares=="100",1],
141 CALCULO_EST_TOUCH(Salida_merge,6,100)),
142
143 cbind(Metodo=CURVA_0[CURVA_0$metodo=="Mod_Gauss_exp" & CURVA_0$Lugares=="500",3],
144 Lugares=CURVA_0[CURVA_0$metodo=="Mod_Gauss_exp" & CURVA_0$Lugares=="500",2],
145 TE=CURVA_0[CURVA_0$metodo=="Mod_Gauss_exp" & CURVA_0$Lugares=="500",1],
146 CALCULO_EST_TOUCH(Salida_merge,6,500)),
147
148 cbind(Metodo=CURVA_0[CURVA_0$metodo=="Mod_Gauss_exp" & CURVA_0$Lugares=="1000",3],
149 Lugares=CURVA_0[CURVA_0$metodo=="Mod_Gauss_exp" & CURVA_0$Lugares=="1000",2],
150 TE=CURVA_0[CURVA_0$metodo=="Mod_Gauss_exp" & CURVA_0$Lugares=="1000",1],
151 CALCULO_EST_TOUCH(Salida_merge,6,1000))
152
153 TAB0_MB<-rbind(
154 cbind(Metodo=CURVA_0[CURVA_0$metodo=="MB_gauss_exp" & CURVA_0$Lugares=="100",3],
155 Lugares=CURVA_0[CURVA_0$metodo=="MB_gauss_exp" & CURVA_0$Lugares=="100",2],
156 TE=CURVA_0[CURVA_0$metodo=="MB_gauss_exp" & CURVA_0$Lugares=="100",1],
157 CALCULO_EST_TOUCH(Salida_merge,7,100)),
158
159 cbind(Metodo=CURVA_0[CURVA_0$metodo=="MB_gauss_exp" & CURVA_0$Lugares=="500",3],
160 Lugares=CURVA_0[CURVA_0$metodo=="MB_gauss_exp" & CURVA_0$Lugares=="500",2],
161 TE=CURVA_0[CURVA_0$metodo=="MB_gauss_exp" & CURVA_0$Lugares=="500",1],
162 CALCULO_EST_TOUCH(Salida_merge,7,500)),
163
164 cbind(Metodo=CURVA_0[CURVA_0$metodo=="MB_gauss_exp" & CURVA_0$Lugares=="1000",3],
165 Lugares=CURVA_0[CURVA_0$metodo=="MB_gauss_exp" & CURVA_0$Lugares=="1000",2],
166 TE=CURVA_0[CURVA_0$metodo=="MB_gauss_exp" & CURVA_0$Lugares=="1000",1],
167 CALCULO_EST_TOUCH(Salida_merge,7,1000))
168
169
170 tabla_mix_0<-rbind(TAB0_CON,TAB0_KDE,
171 TAB0_PHM,TAB0_MOD,TAB0_MB)
172
173 #####NIVEL DE MATHCING
174 round(match_LISTA_MATRIX(SALIDA_0,100),2)
175 write.csv2(round(match_LISTA_MATRIX(SALIDA_0,100),2),
176 "FIG/Match_mat_100.csv")
177 round(match_LISTA_MATRIX(SALIDA_0,500),2)
178 write.csv2(round(match_LISTA_MATRIX(SALIDA_0,500),2),
179 "FIG/Match_mat_500.csv")
180 round(match_LISTA_MATRIX(SALIDA_0,1000),2)
181 write.csv2(round(match_LISTA_MATRIX(SALIDA_0,1000),2),
182 "FIG/Match_mat_1000.csv")

```

4.3.3. Predicciones sobre días aleatorios

```

1 ##Selecciono dias al azar
2 set.seed(222)
3 VECT <- sample(1:365, 20, replace=F)+as.Date("2017-12-31") ##predigo para 2018
4 Patr<-500
5 step<-50
6 table(month(VECT))
7 table(weekdays(VECT))
8 write.csv2(t(VECT),"FIG/Vector_days.csv")
9
10 SALIDA_1<-PREDICT_1step(VECT[1],ventana_temporal,ventana_espacial,eps_t)
11 Sys.sleep(60)
12 gc()
13 SALIDA_2<-PREDICT_1step(VECT[2],ventana_temporal,ventana_espacial,eps_t)
14 Sys.sleep(60)
15 gc()
16 SALIDA_3<-PREDICT_1step(VECT[3],ventana_temporal,ventana_espacial,eps_t)
17 Sys.sleep(60)
18 gc()
19 SALIDA_4<-PREDICT_1step(VECT[4],ventana_temporal,ventana_espacial,eps_t)
20 Sys.sleep(60)
21 gc()

```

```

22 SALIDA_5<-PREDICT_1step(VECT[5], ventana_temporal, ventana_espacial, eps_t)
23 Sys.sleep(60)
24 gc()
25
26 SALIDA_6<-PREDICT_1step(VECT[6], ventana_temporal, ventana_espacial, eps_t)
27 Sys.sleep(60)
28 gc()
29
30 SALIDA_7<-PREDICT_1step(VECT[7], ventana_temporal, ventana_espacial, eps_t)
31 Sys.sleep(60)
32 gc()
33
34 SALIDA_8<-PREDICT_1step(VECT[8], ventana_temporal, ventana_espacial, eps_t)
35 Sys.sleep(60)
36 gc()
37 SALIDA_9<-PREDICT_1step(VECT[9], ventana_temporal, ventana_espacial, eps_t)
38 Sys.sleep(60)
39 gc()
40 SALIDA_10<-PREDICT_1step(VECT[10], ventana_temporal, ventana_espacial, eps_t)
41 Sys.sleep(60)
42 gc()
43
44 SALIDA_11<-PREDICT_1step(VECT[11], ventana_temporal, ventana_espacial, eps_t)
45 Sys.sleep(60)
46 gc()
47
48 SALIDA_12<-PREDICT_1step(VECT[12], ventana_temporal, ventana_espacial, eps_t)
49 Sys.sleep(60)
50 gc()
51
52 SALIDA_13<-PREDICT_1step(VECT[13], ventana_temporal, ventana_espacial, eps_t)
53 Sys.sleep(60)
54 gc()
55
56 SALIDA_14<-PREDICT_1step(VECT[14], ventana_temporal, ventana_espacial, eps_t)
57 Sys.sleep(60)
58 gc()
59
60 SALIDA_15<-PREDICT_1step(VECT[15], ventana_temporal, ventana_espacial, eps_t)
61 Sys.sleep(60)
62 gc()
63
64
65 SALIDA_16<-PREDICT_1step(VECT[16], ventana_temporal, ventana_espacial, eps_t)
66 Sys.sleep(60)
67 gc()
68
69 SALIDA_17<-PREDICT_1step(VECT[17], ventana_temporal, ventana_espacial, eps_t)
70 Sys.sleep(60)
71 gc()
72
73 SALIDA_18<-PREDICT_1step(VECT[18], ventana_temporal, ventana_espacial, eps_t)
74 Sys.sleep(60)
75 gc()
76
77 SALIDA_19<-PREDICT_1step(VECT[19], ventana_temporal, ventana_espacial, eps_t)
78 Sys.sleep(60)
79 gc()
80 SALIDA_20<-PREDICT_1step(VECT[20], ventana_temporal, ventana_espacial, eps_t)
81 Sys.sleep(60)
82 gc()
83
84 ##Calculo curvas
85 Patr<-1000
86 step<-50
87 CURVA_1<-HIT_K_FIX(SALIDA_1, Patr, step)
88 CURVA_2<-HIT_K_FIX(SALIDA_2, Patr, step)
89 CURVA_3<-HIT_K_FIX(SALIDA_3, Patr, step)
90 CURVA_4<-HIT_K_FIX(SALIDA_4, Patr, step)
91 CURVA_5<-HIT_K_FIX(SALIDA_5, Patr, step)
92 CURVA_6<-HIT_K_FIX(SALIDA_6, Patr, step)
93 CURVA_7<-HIT_K_FIX(SALIDA_7, Patr, step)
94 CURVA_8<-HIT_K_FIX(SALIDA_8, Patr, step)

```

```

95 CURVA_9<-HIT.K.FIX(SALIDA_9,Patr,step)
96 CURVA_10<-HIT.K.FIX(SALIDA_10,Patr,step)
97 CURVA_11<-HIT.K.FIX(SALIDA_11,Patr,step)
98 CURVA_12<-HIT.K.FIX(SALIDA_12,Patr,step)
99 CURVA_13<-HIT.K.FIX(SALIDA_13,Patr,step)
100 CURVA_14<-HIT.K.FIX(SALIDA_14,Patr,step)
101 CURVA_15<-HIT.K.FIX(SALIDA_15,Patr,step)
102 CURVA_16<-HIT.K.FIX(SALIDA_16,Patr,step)
103 CURVA_17<-HIT.K.FIX(SALIDA_17,Patr,step)
104 CURVA_18<-HIT.K.FIX(SALIDA_18,Patr,step)
105 CURVA_19<-HIT.K.FIX(SALIDA_19,Patr,step)
106 CURVA_20<-HIT.K.FIX(SALIDA_20,Patr,step)
107
108 ##Preparo estadisticos sobre las curvas
109 #####PREPARO UN group plot###
110 RES_1<-rbind(CURVA_1,CURVA_2,CURVA_3,CURVA_4,
111 CURVA_5,CURVA_6,CURVA_7,CURVA_8,
112 CURVA_9,CURVA_10,CURVA_11,CURVA_12,
113 CURVA_13,CURVA_14,CURVA_15,CURVA_16,
114 CURVA_17,CURVA_18,CURVA_19,CURVA_20)
115
116 STAT_C <- summarySE(RES_1, measurevar="HIT", groupvars=c("Lugares","metodo"))
117 write.csv(STAT_C,"FIG/STAT_C.60.csv")
118
119 ##Preparo etiquetas y grafico
120 STAT_C$metodo<-ifelse(STAT_C$metodo=="CONTEO_PASADO","Conteo",
121 ifelse(STAT_C$metodo=="PHM_est","PHM",
122 ifelse(STAT_C$metodo=="kernel_est","KDE",
123 ifelse(STAT_C$metodo=="MB.gauss_exp","EP",
124 ifelse(STAT_C$metodo=="Mod.Gauss_exp","EPcov",
125 STAT_C$metodo))))))
126
127 dev.off()
128 png("FIG/CURVE_TRADE_simul.png")
129 ggplot(STAT_C, aes(x=Lugares, y=HIT, colour=metodo)) +
130 geom_errorbar(aes(ymin=HIT-se, ymax=HIT+se), width=.1) +
131 geom_line() +ylab("TE")+xlab("Celdas")+
132 geom_point()
133 dev.off()
134
135
136 #####TABLAS MANUALES DE RESUMEN#####
137 #####TABLA CONTEO PASADO#####
138 bind_modelo<-as.data.frame(cbind(ID=SALIDA_1$ID,SALIDA_1$CONTEO_PASADO,
139 SALIDA_2$CONTEO_PASADO,
140 SALIDA_3$CONTEO_PASADO,
141 SALIDA_4$CONTEO_PASADO,
142 SALIDA_5$CONTEO_PASADO,
143 SALIDA_6$CONTEO_PASADO,
144 SALIDA_7$CONTEO_PASADO,
145 SALIDA_8$CONTEO_PASADO,
146 SALIDA_9$CONTEO_PASADO,
147 SALIDA_10$CONTEO_PASADO,
148 SALIDA_11$CONTEO_PASADO,
149 SALIDA_12$CONTEO_PASADO,
150 SALIDA_13$CONTEO_PASADO,
151 SALIDA_14$CONTEO_PASADO,
152 SALIDA_15$CONTEO_PASADO,
153 SALIDA_16$CONTEO_PASADO,
154 SALIDA_17$CONTEO_PASADO,
155 SALIDA_18$CONTEO_PASADO,
156 SALIDA_19$CONTEO_PASADO,
157 SALIDA_20$CONTEO_PASADO))
158 str(bind_modelo)
159
160
161
162 TABLA_CONTEO_PASADO<-rbind(
163 cbind(100,subset(STAT_C,Lugares==100 & metodo=="CONTEO_PASADO")$HIT,length(unique(unlist(
164 match_LISTA_sample(bind_modelo,100))))),ENTROPY_LISTA(bind_modelo,100)),
165 cbind(500,subset(STAT_C,Lugares==500 & metodo=="CONTEO_PASADO")$HIT,length(unique(unlist(
166 match_LISTA_sample(bind_modelo,500))))),ENTROPY_LISTA(bind_modelo,500)),
167 cbind(1000,subset(STAT_C,Lugares==1000 & metodo=="CONTEO_PASADO")$HIT,length(unique(unlist

```

```

    (match_LISTA_sample(bind_modelo,1000))) ,ENTROPY_LISTA(bind_modelo,1000))
166 )
167
168 rownames(TABLA_CONTEO_PASADO)<-rep("Coteo pasado",3)
169 colnames(TABLA_CONTEO_PASADO)<-c("Lugares","HIT","Lugares patrullados","ENTROPY")
170
171
172 ##PLOT
173 Plot_huella_print(bind_modelo,100,"Conteo")
174 Plot_huella_print(bind_modelo,500,"Conteo")
175 Plot_huella_print(bind_modelo,1000,"Conteo")
176 rm(bind_modelo)
177
178 ###TABLA KERNEL###
179 bind_modelo<-as.data.frame(cbind(ID=SALIDA_1$ID,SALIDA_1$kernel_est,
180 SALIDA_2$kernel_est,
181 SALIDA_3$kernel_est,
182 SALIDA_4$kernel_est,
183 SALIDA_5$kernel_est,
184 SALIDA_6$kernel_est,
185 SALIDA_7$kernel_est,
186 SALIDA_8$kernel_est,
187 SALIDA_9$kernel_est,
188 SALIDA_10$kernel_est,
189 SALIDA_11$kernel_est,
190 SALIDA_12$kernel_est,
191 SALIDA_13$kernel_est,
192 SALIDA_14$kernel_est,
193 SALIDA_15$kernel_est,
194 SALIDA_16$kernel_est,
195 SALIDA_17$kernel_est,
196 SALIDA_18$kernel_est,
197 SALIDA_19$kernel_est,
198 SALIDA_20$kernel_est))
199
200 TABLA_kernel_est<-rbind(
201 cbind(100,subset(STAT_C,Lugares==100 & metodo=="kernel_est")$HIT,length(unique(unlist(
    match_LISTA_simple(bind_modelo,100)))) ,ENTROPY_LISTA(bind_modelo,100)),
202 cbind(500,subset(STAT_C,Lugares==500 & metodo=="kernel_est")$HIT,length(unique(unlist(
    match_LISTA_simple(bind_modelo,500)))) ,ENTROPY_LISTA(bind_modelo,500)),
203 cbind(1000,subset(STAT_C,Lugares==1000 & metodo=="kernel_est")$HIT,length(unique(unlist(
    match_LISTA_simple(bind_modelo,1000)))) ,ENTROPY_LISTA(bind_modelo,1000)))
204 rownames(TABLA_kernel_est)<-rep("Kernel",3)
205 colnames(TABLA_kernel_est)<-c("Lugares","HIT","Lugares patrullados","ENTROPY")
206
207 ##PLOT
208 Plot_huella_print(bind_modelo,100,"Kernel")
209 Plot_huella_print(bind_modelo,500,"Kernel")
210 Plot_huella_print(bind_modelo,1000,"Kernel")
211
212 rm(bind_modelo)
213
214
215 ###TABLA PHM###
216 bind_modelo<-as.data.frame(cbind(ID=SALIDA_1$ID,SALIDA_1$PHM_est,
217 SALIDA_2$PHM_est,
218 SALIDA_3$PHM_est,
219 SALIDA_4$PHM_est,
220 SALIDA_5$PHM_est,
221 SALIDA_6$PHM_est,
222 SALIDA_7$PHM_est,
223 SALIDA_8$PHM_est,
224 SALIDA_9$PHM_est,
225 SALIDA_10$PHM_est,
226 SALIDA_11$PHM_est,
227 SALIDA_12$PHM_est,
228 SALIDA_13$PHM_est,
229 SALIDA_14$PHM_est,
230 SALIDA_15$PHM_est,
231 SALIDA_16$PHM_est,
232 SALIDA_17$PHM_est,
233 SALIDA_18$PHM_est,
234 SALIDA_19$PHM_est,

```

```

235 SALIDA_20$PHM_est))
236 TABLA_PHM<-rbind(
237 cbind(100,subset(STAT_C,Lugares==100 & metodo=="PHM_est")$HIT,length(unique(unlist(match_
LISTA_simple(bind_modelo,100))))),ENTROPY_LISTA(bind_modelo,100)),
238 cbind(500,subset(STAT_C,Lugares==500 & metodo=="PHM_est")$HIT,length(unique(unlist(match_
LISTA_simple(bind_modelo,500))))),ENTROPY_LISTA(bind_modelo,500)),
239 cbind(1000,subset(STAT_C,Lugares==1000 & metodo=="PHM_est")$HIT,length(unique(unlist(match_
LISTA_simple(bind_modelo,1000))))),ENTROPY_LISTA(bind_modelo,1000))
240 rownames(TABLA_PHM)<-rep("PHM",3)
241 colnames(TABLA_PHM)<-c("Lugares","HIT","Lugares patrullados","ENTROPY")
242
243 ##Plot
244 Plot_huella_print(bind_modelo,100,"PHM")
245 Plot_huella_print(bind_modelo,500,"PHM")
246 Plot_huella_print(bind_modelo,1000,"PHM")
247
248 rm(bind_modelo)
249
250 ###TABLA Modelo gauss exp####
251 bind_modelo<-as.data.frame(cbind(ID=SALIDA_1$ID,SALIDA_1$Mod_Gauss_exp,
252 SALIDA_2$Mod_Gauss_exp,
253 SALIDA_3$Mod_Gauss_exp,
254 SALIDA_4$Mod_Gauss_exp,
255 SALIDA_5$Mod_Gauss_exp,
256 SALIDA_6$Mod_Gauss_exp,
257 SALIDA_7$Mod_Gauss_exp,
258 SALIDA_8$Mod_Gauss_exp,
259 SALIDA_9$Mod_Gauss_exp,
260 SALIDA_10$Mod_Gauss_exp,
261 SALIDA_11$Mod_Gauss_exp,
262 SALIDA_12$Mod_Gauss_exp,
263 SALIDA_13$Mod_Gauss_exp,
264 SALIDA_14$Mod_Gauss_exp,
265 SALIDA_15$Mod_Gauss_exp,
266 SALIDA_16$Mod_Gauss_exp,
267 SALIDA_17$Mod_Gauss_exp,
268 SALIDA_18$Mod_Gauss_exp,
269 SALIDA_19$Mod_Gauss_exp,
270 SALIDA_20$Mod_Gauss_exp))
271 TABLA_Mod_Gauss_exp<-rbind(
272 cbind(100,subset(STAT_C,Lugares==100 & metodo=="Mod_Gauss_exp")$HIT,length(unique(unlist(
match_LISTA_simple(bind_modelo,100))))),ENTROPY_LISTA(bind_modelo,100)),
273 cbind(500,subset(STAT_C,Lugares==500 & metodo=="Mod_Gauss_exp")$HIT,length(unique(unlist(
match_LISTA_simple(bind_modelo,500))))),ENTROPY_LISTA(bind_modelo,500)),
274 cbind(1000,subset(STAT_C,Lugares==1000 & metodo=="Mod_Gauss_exp")$HIT,length(unique(unlist(
match_LISTA_simple(bind_modelo,1000))))),ENTROPY_LISTA(bind_modelo,1000))
275 rownames(TABLA_Mod_Gauss_exp)<-rep("Mod Gauss exp",3)
276 colnames(TABLA_Mod_Gauss_exp)<-c("Lugares","HIT","Lugares patrullados","ENTROPY")
277
278 ##Plot
279 Plot_huella_print(bind_modelo,100,"ModGaussexp")
280 Plot_huella_print(bind_modelo,500,"ModGaussexp")
281 Plot_huella_print(bind_modelo,1000,"ModGaussexp")
282 rm(bind_modelo)
283
284
285 ###TABLA MB gauss exp####
286 bind_modelo<-as.data.frame(cbind(ID=SALIDA_1$ID,SALIDA_1$MB_gauss_exp,
287 SALIDA_2$MB_gauss_exp,
288 SALIDA_3$MB_gauss_exp,
289 SALIDA_4$MB_gauss_exp,
290 SALIDA_5$MB_gauss_exp,
291 SALIDA_6$MB_gauss_exp,
292 SALIDA_7$MB_gauss_exp,
293 SALIDA_8$MB_gauss_exp,
294 SALIDA_9$MB_gauss_exp,
295 SALIDA_10$MB_gauss_exp,
296 SALIDA_11$MB_gauss_exp,
297 SALIDA_12$MB_gauss_exp,
298 SALIDA_13$MB_gauss_exp,
299 SALIDA_14$MB_gauss_exp,
300 SALIDA_15$MB_gauss_exp,
301 SALIDA_16$MB_gauss_exp,

```

```

302 SALIDA_17$MB_gauss_exp,
303 SALIDA_18$MB_gauss_exp,
304 SALIDA_19$MB_gauss_exp,
305 SALIDA_20$MB_gauss_exp))
306 TABLA_MB_gauss_exp<-rbind(
307 cbind(100, subset(STAT_C, Lugares==100 & metodo=="MB_gauss_exp")$HIT, length(unique(unlist(
      match_LISTA_simple(bind_modelo,100))))), ENTROPY_LISTA(bind_modelo,100)),
308 cbind(500, subset(STAT_C, Lugares==500 & metodo=="MB_gauss_exp")$HIT, length(unique(unlist(
      match_LISTA_simple(bind_modelo,500))))), ENTROPY_LISTA(bind_modelo,500)),
309 cbind(1000, subset(STAT_C, Lugares==1000 & metodo=="MB_gauss_exp")$HIT, length(unique(unlist(
      match_LISTA_simple(bind_modelo,1000))))), ENTROPY_LISTA(bind_modelo,1000))
310 rownames(TABLA_MB_gauss_exp)<-rep("MB_gauss_exp",3)
311 colnames(TABLA_MB_gauss_exp)<-c("Lugares", "HIT", "Lugares patrullados", "ENTROPY")
312
313 ##Plot
314 Plot_huella_print(bind_modelo,100,"MBgaussexp")
315 Plot_huella_print(bind_modelo,500,"MB_gaussexp")
316 Plot_huella_print(bind_modelo,1000,"MBgaussexp")
317 rm(bind_modelo)

```

4.3.4. Código para ejemplos

```

1 ##Ejemplo curva retorno
2 BA<-data.frame(Metodo=c(rep("A",11),rep("B",11)),
3 x=c(seq(0,500,50),seq(0,500,50)),
4 y=c(0,3,8,10,13,15,16,16.5,17,17.5,17.8,
5 0,2,6,8,12,13,14,16,16.5,17,17.5))
6
7 ggplot(BA, aes(x,y),group(Metodo)) +
8 geom_line(aes(linetype = Metodo))+geom_point()+
9 xlab("Lugares")+ylab("TE")
10
11 ### Ejemplo proceso puntuales
12 plot(redwood, type="p", cex = .5, "")
13 plot(quadratcount(redwood, nx=4, ny=4), cex=.8, add=TRUE)
14
15 KR<-Kest(redwood, correction = "none")
16 pcfR<-pcf(KR)
17
18 plot(KR, main=" ")
19 plot(pcfR, main=" ")
20
21
22
23 ###Espacio temporales
24 data("northcumbria")
25 data("fmd")
26 fmd <- as.3dpoints(fmd)
27
28 plot(fmd, s.region=northcumbria, mark=TRUE)
29
30
31 FMD <- as.3dpoints(fmd[, 1] / 1000, fmd[, 2] / 1000, fmd[,3])
32 Northcumbria <- northcumbria / 1000
33 Mt <- density(FMD[, 3], n = 1000)
34 mut <- Mt$y[findInterval(FMD[, 3], Mt$x)] * dim(FMD)[1]
35 h <- mse2d(as.points(FMD[,1:2]), Northcumbria, nsmse = 50, range = 4)
36 h <- h$h[which.min(h$mse)]
37 Ms <- kernel2d(as.points(FMD[, 1:2]), Northcumbria, h = h, nx = 5000,
38 ny = 5000)
39 atx <- findInterval(x = FMD[, 1], vec = Ms$x)
40 aty <- findInterval(x = FMD[, 2], vec = Ms$y)
41 mhat <- NULL
42 for(i in 1:length(atx)) mhat <- c(mhat, Ms$z[atx[i], aty[i]])
43 u <- seq(0, 10, by = 1)
44 v <- seq(0, 15, by = 1)
45 stik <- STIKhat(xyt = FMD, s.region = Northcumbria, t.region = c(1, 200),
46 lambda = mhat * mut / dim(FMD)[1], dist = u, times = v, infectious = FALSE)
47 g <- PCFhat(xyt = FMD, lambda = mhat * mut / dim(FMD)[1], dist = 1:20,
48 times = 1:20, s.region = Northcumbria, t.region = c(1,200))
49
50

```

```
51 plotK(stik ,L=T, type = "image", theta = -65, phi = 35)
52 plotPCF(g, type = "persp", theta = 65, phi = 35)
53 plotPCF(g, type = "image", theta = 65, phi = 35)
```