UNIVERSIDAD DE LOS ANDES
DEPARTMENT OF ECONOMICS

Predicting criminal behavior with Lévy flights using real data from Bogota

Supervised by
Professor Álvaro José Riascos Villegas

Presented by
Mateo Dulce Rubio
201224184

October 29, 2018

Predicting criminal behavior with Lévy flights using real data from Bogota*

Mateo Dulce Rubio†

Abstract

I use residential burglary data from Bogota, Colombia, to fit an agent-based model following truncated Lévy flights (Pan et al., 2018) elucidating criminal rational behavior and validating repeat/near-repeat victimization and broken windows effects. The estimated parameters suggest that if an average house or its neighbors have never been attacked, and it is suddenly burglarized, the probability of a new attack the next day increases, due to the crime event, in 79 percentage points. Moreover, the following day its neighbors will also face an increment in the probability of crime of 79 percentage points. This effect persists for a long time span. The model presents an area under the Cumulative Accuracy Profile (CAP) curve, of 0.8 performing similarly or better than state-of-the-art crime prediction models. Public policies seeking to reduce criminal activity and its negative consequences must take into account these mechanisms and the self-exciting nature of crime to effectively make criminal hotspots safer.

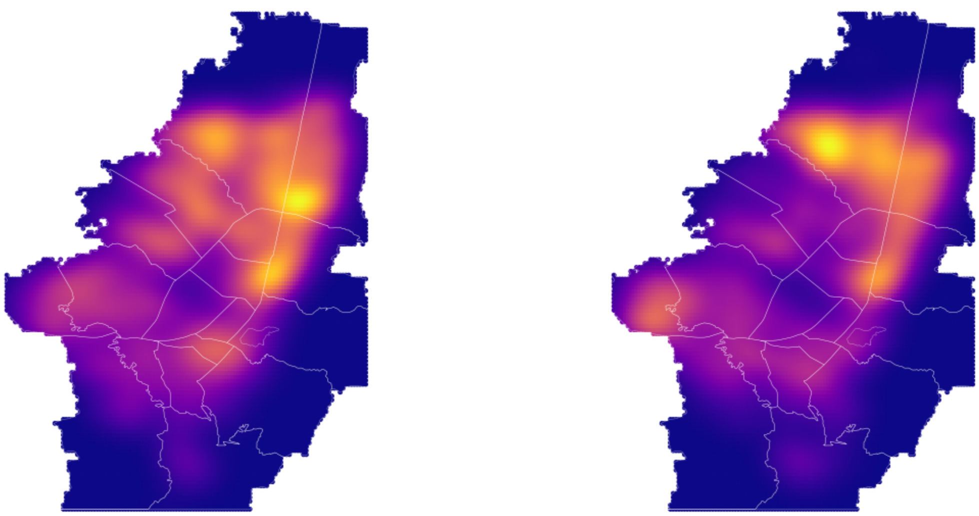
Keywords: Criminal behavior, Crime prediction model, Machine learning, Agent-based model.
JEL Classification: K42, H39, C53, C63.

*Thesis for the Masters in Economics, Universidad de los Andes.
†Universidad de los Andes & Quantil | Matemáticas Aplicadas. m.dulce10@uniandes.edu.co.

1 Introduction

Crime is a persistent problem that most modern cities face and Bogota is no exception. However, crime is not uniformly distributed over the whole city but presents spatio-temporal clustering patterns. For instance, some neighborhoods deal with higher criminal rates than others, and certain times of the day and days of the week are far more dangerous than others.

This is particularly true in the case of residential burglaries in Bogota, even though the location of criminal targets (houses) remains constant over time. As shown in Figure 1, residential burglary is not ubiquitous but is concentrated in specific zones of the city, with those areas preferred by criminals changing over time. Specifically, between the first and second semester of 2012 the burglary events displace, leading to the dissipation of old hotspots, the formation of new ones, and the consolidation of some others.



(a) January - June, 2012. (b) July - December, 2012.

Figure 1: Residential burglary hotspots - Bogota. KDE fitted to burglary data, bw=0.01.

Furthermore, between 2012 and 2015, two percent of Bogota's streets were accounted for all homicides and a quarter of all crimes reported in the city, but they received less than 10% of police time and narrow public services (Blattman, Green, Ortega, & Tobón, 2017). Understanding the emergence, diffusion and dissipation of these aggregates of criminal occurrences, called *hotspots*, is crucial for the efficient assignment of scarce police budget in order to prevent crime and its negative consequences.

With this in mind, *Predictive Policing* arises as the use of statistical inference and machine learning techniques to identify vulnerable crime areas, i.e., regions with a high estimated probability of crime occurring.[1] Nevertheless, all these models assume the crime as a random event and try to estimate the underlying process of crime generation, but fail giving insights on the rational behavior of offenders.

In this work, I use residential burglary data from Bogota to fit an agent-based model (Pan et al., 2018) elucidating criminal rational behavior, and validating repeat/near-repeat victimization and broken windows effects. In particular, using a likelihood function derived by Lloyd, Santitissadeekorn, and Short (2016), I found statistical evidence to support that crime is not a random event evenly distributed over the city, but presents spatio-temporal clustering patters, due to its self-exciting nature.

Specifically, the estimated parameters of the model suggest that if a house with average attractiveness index or its neighbors have never been attacked, and it is suddenly burglarized, the probability of a new attack the following day increases, due to the crime event in 79 percentage points. Moreover, the next day, its neighbors will also face an increment in crime probability of 79 percentage points. This effect persists for a long period of time. Public policies and policing strategies that seek to reduce criminal activity and its negative consequences must take into account these mechanisms and the self-exciting nature of crime, to effectively make criminal hotspots safer.

The model presents an area under the Cumulative Accuracy Profile (CAP) curve of 0.8, performing similarly or better than state-of-the-art crime prediction models (Mohler, Short, Brantingham, Schoenberg, & Tita, 2011), and particular well-known cases of the model (Chaturapruek, Breslau, Yazdi, Kolokolnikov, & McCalla, 2013; Short et al., 2008). The CAP was constructed by computing the Hit Rate (portion of the crimes correctly predicted by the model) for different percentages of the city area marked as hotspots.

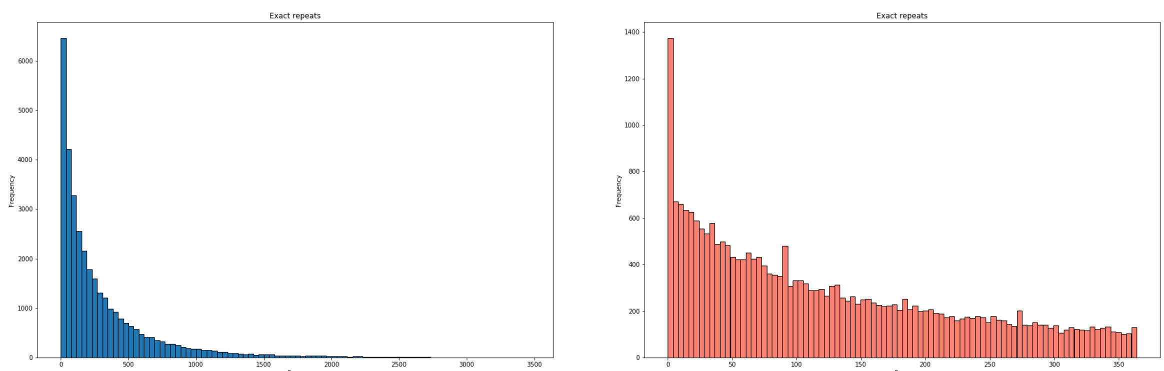
The methodology used relies on Becker (1968) and Ehrlich (1973) economic theory of crime, which suggests that criminals behave rationally, weighing the expected utility of perpetrating a crime, the expected (des)utility of being caught, the probability of arrest, prosecution and sentencing, and the opportunity cost of doing other, perhaps illegal, activities.

For instance, evidence suggests that burglars often prefer to return to a previously burglarized neighborhood in part because is where they already have good information about the type of prop-

[1] For a review and comparison of different crime prediction models applied to Bogota data, see Barreras, Diaz, Riascos, and Ribero (2016).

erties that could be stolen and the routines of the police and its inhabitants (Wright & Decker, 1994). This pattern is known as repeat and near-repeat victimization. Moreover, the past occurrence of crimes in a given area creates an atmosphere of lawlessness and crime-tolerant region that encourages the occurrence of even more crimes, documented as the *broken windows* effect (Wilson & Kelling, 1982).

Figure 2 plots histograms of the days between burglary events that took place in the exact same 180 by 320 meters cell in Bogota, for all reported residential burglaries from 2004 to 2013 in the city. The histograms present a spike in short periods of days, suggesting a criminal rationality pattern to return to the exact or near locations of past events, leading to an increase likelihood of repeat/near-repeat victimization in the days ater a residential burglary event. How this micro-scale perceived facilities and benefits of perpetrating a crime leads to the formation of macro-scale aggregates of criminal activity is still a matter of study.



(a) Histogram between burglary events - 10 years (b) Histogram between burglary events - 1 year

Figure 2: Repeat victimization evidence

Such criminal behavior was mathematically modeled by Short et al. (2008), with a dynamic expected utility of burglarizing a house that varies in response to mechanisms of repeat/near-repeat victimization and broken windows theory. In this model, criminals move through the city following a random walk biased towards high attractiveness regions. In each time step, burglars decide to either burglarize the house where they are located or to move to an adjacent place, taking into account the expected utility of the criminal act and its opportunity cost, as proposed by Becker (1968) and Ehrlich (1973).

However, burglars can search more efficiently for houses with high attractiveness by doing larger jumps, i.e. modeling the criminals locomotion using Lévy flights, with the distribution of step lengths obeying a power law (Chaturapruek et al., 2013; Pan et al., 2018). Evidence of this behavior was found by Hesseling (1992) using solved crime data for residential burglaries in Utrecht, the Netherlands. In this work, she founds that crimes perpetrated in the inner-city were mostly committed by offenders living outside the inner-city, suggesting that criminals are willing to travel farther distances for more attractive targets, and use different means of transportation to achieve their goal.

With this in mind, Chaturapruek et al. (2013) generalized the model proposed by Short et al. (2008) modeling criminal locomotion with Lévy flights under which criminals are allowed to move to farther sites in the city, while Levajkovic and Zarfl (2016) extend the Lévy flights model to a 2-dimensional Lattice. The model was then generalized by Pan et al. (2018) allowing criminals to move according to a truncated Lévy flight with a limited jump range reflecting a limited traveling distance, biased by the perceived attractiveness of houses.

Fitting these agent-based models on real crime datasets and testing its predictive accuracy, shed light on the motivation and incentives of criminal offenders, leading to a better design of public policies and assignment of police resources. To date, this family of models have never been tested on real criminal data.

2 Methodology

Becker (1968) and Ehrlich (1973) economic framework suggests that offenders act rationally, perpetrating a crime whenever their expected benefits are higher than their expected costs. That is, if the following inequality holds:

$$(1-p)U_c - pS > U_l, \tag{2.1}$$

where p is the perceived probability of arrest, prosecution and sentencing, U_c the utility of crime, S the (des)utility of punishment, and U_l the opportunity cost of crime (Gomez-Cardona, Mejia, & Tobon, 2017).

In the short run, punishments do not vary a lot as they are the result of long legislative processes, and the opportunity cost of doing other legal activities also remains quite constant due to labor market rigidities. Thus, under this framework, more crimes will be perpetrated at a certain place if its expected utility increases or its perceived probability of apprehension falls, outweighing the opportunity cost of robbing somewhere else.

Such criminal behavior was mathematically modeled by Short et al. (2008), with a dynamic perceived utility of crime at different locations in the city varying through repeat/near-repeat victimization and broken windows mechanisms. The model was generalized by Pan et al. (2018) to account for a more realistic form of human locomotion, allowing criminals to explore the city in search of attractive houses according to a truncated Lévy flight.

Specifically, the statistical model of criminal behaviour proposed by Short et al. (2008) is based on a two dimensional lattice where each vertex $s = (x_s,\ y_s)$ represents a house with an *attractiveness* $A_s(t)$ assigned. This attractiveness index displays the expected benefit of burglarizing the house perceived by the criminal, and is composed of two quantities:

$$A_s(t) = A_s^0 + B_s(t). \tag{2.2}$$

A_s^0 is a static, but possibly spatially heterogeneous, component, while $B_s(t)$ varies with interactions of house s with burglars, capturing repeat/near-repeat victimization patterns and the broken windows effect discussed previously.

Concretely, the dynamic component $B_s(t)$ of the attractiveness index $A_s(t)$ depends upon past burglary events at house s, increasing a constant value θ each time the house is burglarized. This increment only affects the attractiveness level for a finite time, such that

$$B_s(t+\Delta t) = B_s(t)(1-\omega\Delta t) + \theta E_s(t), \tag{2.3}$$

where $E_s(t)$ is the total number of burglaries at site s in the time interval starting at t, and ω accounts for the time span repeat victimization is more likely to occur. Furthermore, $B_s(t)$ spreads spatially to adjacent sites modeling near-repeat victimization and broken windows effect:

$$B_s(t+\Delta t) = \left[(1-\eta)B_s(t) + \frac{\eta}{4}\sum_{s'\sim s} B_{s'}(t)\right](1-\omega\Delta t) + \theta E_s(t), \tag{2.4}$$

with $\eta \in [0,1]$ a parameter measuring such spreading effect: higher values of η lead to higher attractiveness levels due to near burglary events. In equation (2.4), s' refers to adjacent sites to house s, four in the two dimensional grid.

Criminal agents are also modeled over the lattice, and in each time interval they can either burglarize the house where they are located, or move to another place. The probability of a criminal burglarizing a house s in a time interval of size Δt follows a standard Poisson process, a stochastic process widely used in queueing theory to model random events distributed in time, such as the arrival of customers, phone calls or, in this case, crime events:

$$p_s(t) = 1 - e^{-A_s(t)\Delta t}. \tag{2.5}$$

The expect number of crimes per burglar at a house s in the interval from t to $t+\Delta t$ is $A_s(t)\Delta t$; as expected, more attractive houses are burglarized more. Criminals that commit burglary are removed from the lattice displaying the tendency of fleeing the site of a crime after committing it. Moreover, each time interval burglars are generated with rate Γ at each grid vertex.

In the Pan et al. (2018) model, when a criminal does not perpetrate a crime, he moves to another house following a truncated Lévy flight biased toward regions with high attractiveness. The probability of moving from a house s to a house r is

$$q_{s\to r} = \frac{w_{s\to r}}{\sum_{s'\in\mathbb{Z}^2, s'\neq s} w_{s\to s'}}, \tag{2.6}$$

with

$$w_{s\to r} = \begin{cases} \frac{A_r}{||s-r||^{\mu}}, & ||s-r|| < L \\ 0, & ||s-r|| \geq L \end{cases}, \tag{2.7}$$

for some given L, larger jump allowed, and μ, the exponent of the underlying Lévy flight.

Note that criminals at each site of the lattice must leave after a time interval, either by burglarizing the house were they are located and being removed, or by moving to another place. Then, the number of criminals at a site s after one time interval, $n_s(t+\Delta t)$, consists of burglars coming from other houses that did not perpetrate a crime, and criminals generated at rate Γ:

$$n_s(t+\Delta t) = \sum_{\substack{r\in\mathbb{Z}^2 \\ ||s-r||<L}} [n_r(t) - E_r(t)]q_{r\to s} + \Gamma\Delta t. \tag{2.8}$$

Together, equations (2.4) and (2.8) describe the dynamics of the mean-field discrete model of criminal behavior following biased truncated Lévy flights. The former gives the dynamics of the attractiveness index while the later the distribution of criminals. The attractiveness propagates in space while simultaneously decaying in time and interacting with burglars to generate even more attractiveness. On the other hand, criminals are removed from the lattice due to burglary events as a reaction to the attractiveness, move following a truncated Lévy flights in search of attractive houses, and are generated at a constant rate.

2.1 Effect of L

As a particular case, when $L = 1$ Short et al. (2008) model is recovered,[2] in which criminals move according to a random walk biased to high attractiveness regions, but are only allowed to move to adjacent houses. The model remains the same except for the transition probability of a criminal moving from site s to site r, which reduces to:

$$q_{s\to r}(t) = \frac{A_r(t)}{\sum_{s'\sim s} A_{s'}(t)}, \tag{2.9}$$

where $s' \sim s$ indicates all adjacent sites to s, four in the two dimensional grid.

However, this model implicitly assumes that burglars have access just to one (slow) mean of transportation and limited knowledge only of the attractiveness level of neighboring sites. This could led to a high number of small hotspots due to the impossibility of criminals to visit farther sites of the city.

On the other hand, taking $L = \infty$ leads to a pure Lévy flights model of criminal behavior as developed by Chaturapruek et al. (2013). In this setting, burglars are able to search for houses with high attractiveness by doing larger jumps modeled using Lévy flights, with the distribution of step lengths obeying a power law. This leads to the following transition probability of a criminal moving from site s to house r:

$$q_{s\to r} = \frac{w_{s\to r}}{\sum_{s'\in\mathbb{Z}^2, s'\neq s} w_{s\to s'}}, \tag{2.10}$$

with

$$w_{s\to r} = \frac{A_r}{||s-r||^{\mu}}, \tag{2.11}$$

and μ the exponent of the underlying Lévy flight.

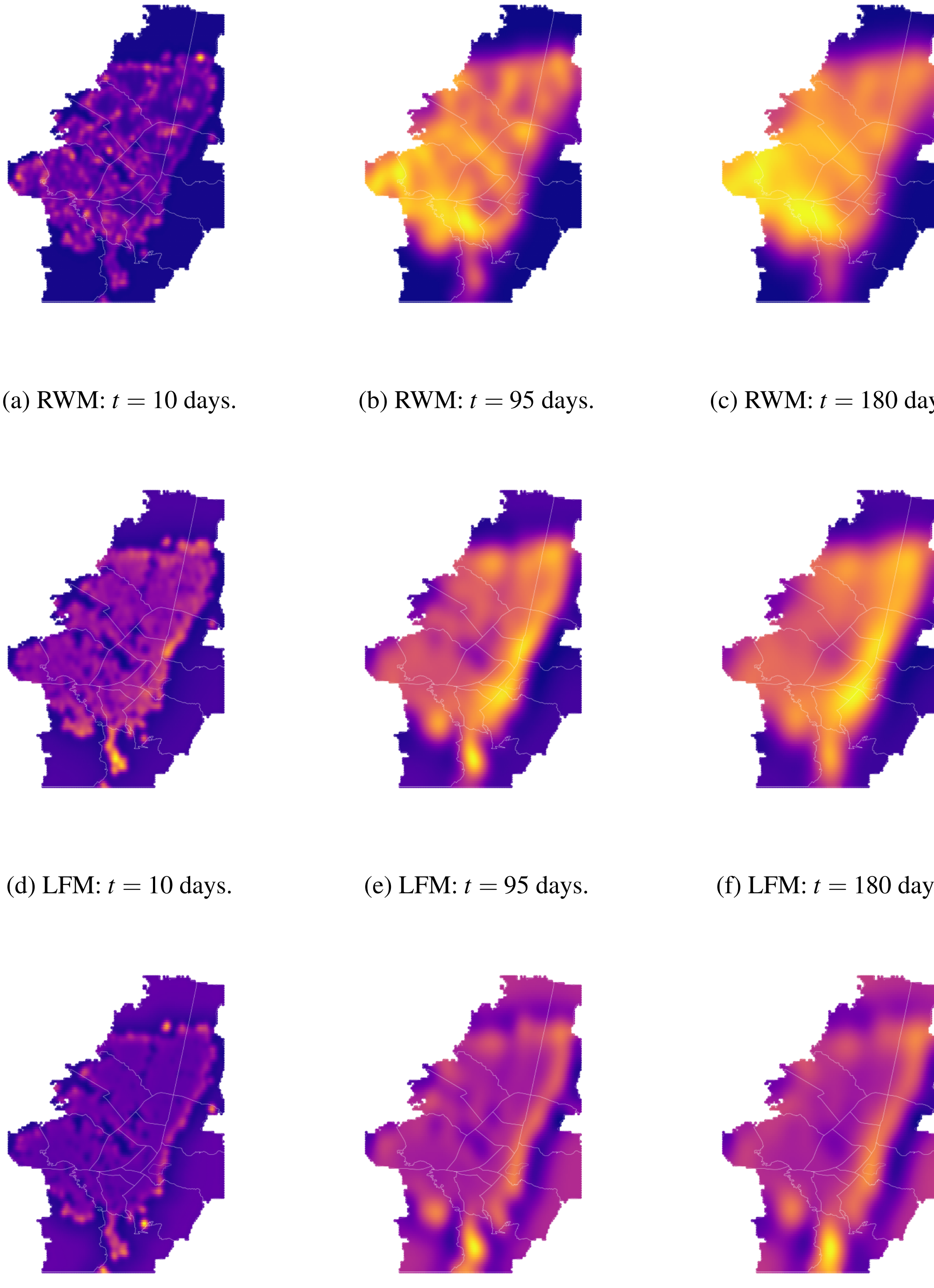
Nonetheless, this model could led to an opposite regime of aggregation of criminal activity: due to the possibility of traveling long distances, criminals may displace to the most attractive neighborhoods generating few static hotspots.

Figure 3 presents simulations of the criminal hotspots obtained with the model for different larger jumps allowed. Specifically, the criminal regimes of aggregation when $L = 1$ (RWM), $L = \infty$ (LFM) and $L = 2$ km (TLFM) are compared, using in each case the respective transition probability of moving from one house to another. All of the simulations started from the same initial attractiveness index, computed fitting a Kernel Density Estimation (KDE) to the spatial locations

[2] Here, $L = 1$ means that the larger jump allowed for criminals is equal to one grid space, such that burglars are able to move only to adjacent houses.

of residential burglaries from January to June 2012, with a bandwidth of 0.001. Furthermore, all of them began with one criminal located at each grid vertex.

The simulations used $A^0 = 0.5$, $\Gamma = 5$, $\theta = 5$, $\eta = 0.5$, $\omega = 0.035$, and evolve according to equations (2.4) and (2.8), with the number of crimes $E_s(t)$ replaced by the probabilistic number of crimes $p_s(t)n_s(t)$. The three cases considered were simulated for 180 days and the criminal hotspots formed at 10, 95 and 180 days are presented.



(a) RWM: $t = 10$ days.
(b) RWM: $t = 95$ days.
(c) RWM: $t = 180$ days.

(d) LFM: $t = 10$ days.
(e) LFM: $t = 95$ days.
(f) LFM: $t = 180$ days.

(g) TLFM: $t = 10$ days.
(h) TLFM: $t = 95$ days.
(i) TLFM: $t = 180$ days.

Figure 3: Criminal hotspots formation for different values of L. RWM: $L = 1$; LFM: $L = \infty$; TLFM: $L = 2$. All simulations used $A^0 = 0.5, \Gamma = 5, \theta = 5, \eta = 0.5, \omega = 0.035$. Initial state from KDE.

The top panels show the evolution of the Random Walk Model (Short et al., 2008) obtained when $L = 1$. Under this setting, only two type of areas are found: one with a very low crime rate and other with a very high crime intensity. The expected number of crimes is very uniform with one hotspots over most of the city.

The middle panels, on the contrary, present one big hotspot that keeps increasing its expect number of crimes during the simulation. The rest of the city presents different levels of predicted criminal activities with areas of low, middle and high crime rates.

Finally, the bottom panels consist of the truncated Lévy flights model of crime with dynamic hotspots emerging and dissipating at different times of the simulation. For instance, the hotspot at the south of the city changes its form during the hole simulation with no apparently steady state. These simulations suggest that $1 < L < \infty$ leads to the more realistic crime hotspots formation, that is, criminals moving according to a truncated Lévy flight.

3 Data assimilation

Given a dataset of times and locations of burglary events, the goal is to find the different parameters of the model that, together, best describe the observed crimes, seeking to elucidate criminal rationality and forecast future crime occurrences. Precisely, estimating the parameters θ, that measures repeat victimization, ω, which accounts for the time span repeat victimization is more likely to occur, and η, capturing near-repeat victimization and broken windows effect, might shed light on criminal behavior leading to a better assignment of police resources and the design of more effective public policies seeking to reduce crime and its consequences.

However, this is a tricky challenge as for a crime to be perpetrated in the model it is necessary not only to estimate the attractiveness index correctly but to have a motivated offender situated at the correct house. Moreover, the dynamic quantities of interest, burglars' location and house attractiveness, are unobserved. Instead, only data on the times and locations of actual crimes are available, which are a function of these two quantities under Becker (1968) and Ehrlich (1973) framework.

To address this issues, Lloyd et al. (2016) developed a data assimilation technique to fit dynamical agent-based models to crime data using Maximum Likelihood Estimation. The procedure is based on fitting the expected number of crimes given by the model at each location, to the actual times and locations of crime events.

Note that, in the model terminology, $n_s(t)A_s(t)$ gives the expected number of crimes at house

s in the time interval beginning at t. Then, given a dataset of N known crime events composed of its times and locations, $\{(s_k,\ t_k)\}_{k=1}^{N}$, a data assimilation procedure must seek that $n_{s_k}(t_k)A_{s_k}(t_k)$ be large for all $k = 1,\ldots,N$, with simultaneously an expected crime rate at other locations and times being low.

Assuming that the probability of crime at house $s \in (x_s - \Delta x,\ x_s + \Delta x) \times (y_s - \Delta y,\ y_s + \Delta y)$ in the time interval $(t_k,\ t_k + \Delta t)$ is governed by a Poisson process with rate

$$\lambda_{s,k} = n_s(t_k)A_s(t_k)\Delta x\Delta y\Delta t, \tag{3.1}$$

the probability of no attacks at site s in the time interval $(t_k,\ t_k + \Delta t)$, given a set of parameters Θ, is (Lloyd et al., 2016):

$$\mathbb{P}(\text{no crime at } s \text{ in } (t_k,\ t_k + \Delta t) \mid \Theta) = e^{-\lambda_{s,k}}. \tag{3.2}$$

Similarly,

$$\mathbb{P}(\text{one crime at } s \text{ in } (t_k,\ t_k + \Delta t) \mid \Theta) = \lambda_{s,k}e^{-\lambda_{s,k}}. \tag{3.3}$$

At each time interval, a burglar decides to engage in a criminal activity considering its location and the attractiveness index of houses at that time. Such attractiveness index already takes into account past burglary events, capturing repeat/near-repeat victimization and broken windows effect. Therefore, the events in each time step can be considered independent as they occur as a response to the information available at each time interval.

Now, suppose that no criminal event take place at a house s between t_k and t_{k+1}, with n time steps between them. Then, assuming that the events in each time interval are independent, the total probability of no attacks between t_k and t_{k+1} followed by a criminal event at t_{k+1} is given by

$$\begin{aligned}\mathbb{P}(\text{no crime at } s \text{ in } (t_k, t_{(k+1)}) \ \wedge \ \text{one crime at } s \text{ in } (t_{(k+1)},\ t_{(k+1)} + \Delta t) \mid \Theta)\\ = \lambda_{s,(k+1)}e^{-\lambda_{s,(k+1)}} * e^{-\sum_{i=0}^{n-1}\lambda_{s,i}}.\end{aligned} \tag{3.4}$$

Computing such probability for the whole training sample (N crimes and T periods), summing over the whole lattice and taking the continuum limit $\Delta x, \Delta y, \Delta t \to 0$, the likelihood function of observing the training data $\{(s_k, t_k)\}_{k=1}^{N}$ is obtained:

$$\mathbb{P}(\text{data} \mid \Theta) = \left(\prod_{k=1}^{N} n_{s_k}(t_k)A_{s_k}(t_k)\right) * \exp\left(-\int_0^T \int_Y \int_X nA\ dxdydt\right). \tag{3.5}$$

Finally, the log-likelihood function that will be used to fit the data is derived:

$$\mathbb{L}(\text{data} \mid \Theta) = \sum_{k=1}^{N} \log(n_{s_k}(t_k)A_{s_k}(t_k)) - \int_0^T \int_Y \int_X nA \; dxdydt. \tag{3.6}$$

The interpretation for the likelihood function is what one may look for: the expected number of crimes given by the product nA must be large where crimes actually occurred at the time they happened, but not with an arbitrarily large number of expected crime (spatio-temporal integral) over the lattice.

The dataset used in the present work corresponds to times and locations of reported residential burglaries in Bogota from 2012 (3,829 cases) and 2013 (3,369 cases) collected by the *Delinquential, Contraventional and Operative Statistical Information System* (SIEDCO in Spanish) from Metropolitan Police of Bogota. This data is currently available form the Department of Economics of the Universidad de los Andes.

3.1 Training procedure

To fit the truncated Lévy flights model for crime (Pan et al., 2018) to residential burglaries data from Bogota, the likelihood function derived in section 3 (Lloyd et al., 2016) was used. The training process is carried out to estimate the parameters of the model that, together, better replicate the known training dataset, seeking to elucidate criminal behavior.

For this procedure the urban area of Bogota was discretized by a lattice composed of $10,946$ homogeneous cells of 180 meters width (Δx) and 320 meters height (Δy), while the time step used Δt was 1 day. As in Pan et al. (2018) $\mu = 2.5$, and the larger jump allowed for a criminal at each time step was 10 km.

As discussed before, the data assimilation framework seeks that $n_{s_k}(t_k)A_{s_k}(t_k)$ be large for every $(s_k,\ t_k)$ in the training set. However, $n_s(t)A_s(t)$ is still a stochastic crime rate, and any observed crime is just one realization of this intensity. Thus, in the training process the model dynamics must evolve according to those realizations in the training dataset, rather than to the underlying expected crime rate.

With this in mind, the truncated Lévy flights model for residential burglary (Pan et al., 2018) was fitted with a training set composed of six consecutive months (182 days) of crime data, with the system evolving via

$$B_s(t+\Delta t) = \left[(1-\eta)B_s(t) + \frac{\eta}{4}\sum_{s'\sim s} B_{s'}(t)\right](1-\omega\Delta t) + \theta E_s(t), \tag{3.7}$$

$$n_s(t+\Delta t) = \sum_{\substack{r\in\mathbb{Z}^2 \\ ||s-r||<L}} [n_r(t) - E_r(t)]q_{r\to s} + \gamma\Delta t, \tag{3.8}$$

where $E_s(t)$ comes from the training data set. Note that $E_s(t)$ is precisely the quantity of interest, and its realization is what the model is trying to understand in the training procedure by maximizing the log-likelihood function, estimating the parameters that capture repeat/near-repeat victimization and broken windows mechanisms, shedding light to criminal rational behavior.

The six months training window was divided in two subgroups, each of 91 days. The first three months were used to rationalize the static attractiveness component and the burglars generation rate, $(A_s^0,\ \Gamma)$. Then, the following three months were used to estimate the parameters (θ,η,ω) concerning the dynamic attractiveness component, B_s, using the MLE technique obtained in section 3.

First, note that in a steady state in which the total number of criminals remains constant, the number of burglars removed from the system each time step due to burglary events must equals the number of criminals generated. Thus, the parameter Γ, accounting for the burglars generation rate, was set to the average daily number of crime events occurring in Bogota in the first three months of the training set.

On the other hand, the static attractiveness component must reflect the intrinsic attractive of each house. However, evidence, such as the presented in Figure 1, suggests that such intrinsic attractive is not homogeneous over the lattice. Thus, A_s^0 was rationalized using Kernel Density Estimation (bw$=0.001$) fitted to the locations of the crime events occurring during the first three months of the training dataset. The density obtained with the KDE was scaled between 0 and 1 such that each house has an static attractiveness proportional to its historical criminal rate.

Then, given Γ and A_s^0, the log-likelihood function was maximized with the second half of the training dataset to estimate the parameters θ, that measures repeat victimization, ω, which accounts for the time span repeat victimization is more likely to occur, and η, capturing near-repeat victimization and broken windows effect, seeking to elucidate criminal rational behavior.

3.2 Test procedure

Once training is complete and the parameters are estimated, the test procedure is carried out to measure the predictive power of the model over data it has never seen before. Concretely, to asses the predictive accuracy of the truncated Lévy flights model of residential burglaries, the system evolves with the estimated parameters starting from the attractiveness and criminal distribution after the last time interval of the training procedure, and the formed hotspots are compared with the known crimes of the test data.

As the goal is to measure the predictive accuracy of the model, it would be incorrect to use the crime events of the test set in its evolution.[3] Therefore, in equations (3.7) and (3.8) that describe the dynamics of the model, the number of crimes $E_s(t)$ is replaced by the probabilistic number of crimes $p_s(t)n_s(t)$:[4]

$$B_s(t+\Delta t) = \left[(1-\eta)B_s(t) + \frac{\eta}{4}\sum_{s'\sim s} B_{s'}(t)\right](1-\omega\Delta t) + \theta p_s(t)n_s(t), \tag{3.9}$$

$$n_s(t+\Delta t) = \sum_{\substack{r\in\mathbb{Z}^2 \\ ||s-r||<L}} [1-p_r(t)]n_r(t)q_{r\to s} + \Gamma\Delta t. \tag{3.10}$$

At each time interval of the test setting, the x% of the cells with higher expected numbers of crimes ($n_s(t)A_s(t)$) are marked as hotspots, for some given x. Then, the location of the known crime events in the test dataset are contrasted with the locations predicted by the model. To measure its predictive accuracy, I use the Hit Rate, a standard measure for crime prediction models defined as the portion of the known crimes in the test set correctly predicted by the model:

$$\text{Hit Rate} = \frac{\text{\# of crimes in predicted hotspots}}{\text{Total \# of crimes}}. \tag{3.11}$$

The test set corresponds to the burglary events that occurred in the month (31 days) immediately following the training set. Furthermore, by varying the percentage of cells flagged as hotspots in the test procedure a Cumulative Accuracy Profile (CAP) curve is constructed, such that the average Hit Rate obtained by the model in each time interval of the test window is a function of the percentage of cells marked as hotspots. Finally, the Area Under the CAP Curve (AUC) is used to compare the

[3]One can do the test procedure by incorporating the actual events every day and only predicting the next day. However, in practice data is not updated with such frequency, which motivates doing so in this way and, in addition, only to study the predictive power of the model for a future month.

[4]This could also be replaced by the expected number of crimes $n_s(t)A_s(t)$.

predictive accuracy of the TLFM, its particular cases RWM and LFM, and other state-of-the-art crime prediction models.

4 Results

The truncated Lévy flights model for residential burglaries was estimated six times using different training and test sets, to asses the stability of the estimated parameters and its predictive accuracy. Specifically, the first training set used was composed of residential burglaries occurring in Bogota from February to July 2012, with events in August 2012 corresponding to the test set. The second training set was from March to August 2012, testing with September 2012 data. Successively, the last training set used corresponds to burglary data from July to December 2012, predicting the events occurring in January 2013.

As explained in section 3.1, each six months training set was divided in two subgroups. The first three months were used to estimate the intrinsic attractiveness component A_s^0 and the criminal generation rate Γ. Then, the last three months were used to maximize the log-likelihood function (3.6) estimating the parameters θ, η and ω that capture repeat/near-repeat victimization and broken windows effect.

The particular cases, $L = 1$ (RWM) and $L = \infty$ (TLFM) were estimated and tested under the same procedure, using the respective transition probability in the dynamic evolution of the model. Furthermore, the predictive accuracy of the model is compared to other state-of-the-art crime prediction models, specifically Kernel Density Estimation (KDE) and Mohler et al. (2011) Self-Exciting Point Process modeling of crime (SEPP), over the whole six months of the training sets.

Appendix A presents the results of training the TLFM, LFM and RWM with the whole six months window, estimating A^0 (homogeneously over the lattice) and Γ together with the other parameters using MLE. Appendix B summarizes the SEPP model.

4.1 Predictive accuracy

Figure 4 presents the average Cumulative Accuracy Profile curve obtained in the six different training-test sets used to estimate the different models. In particular, the truncated Lévy flights model, pure Lévy flights model and Kernel Density Estimation have statistically the same predictive accuracy, being grater than the ones of the Random Walk and the Self-Exciting Point Process models for every percentage of cells flagged as hotspots.

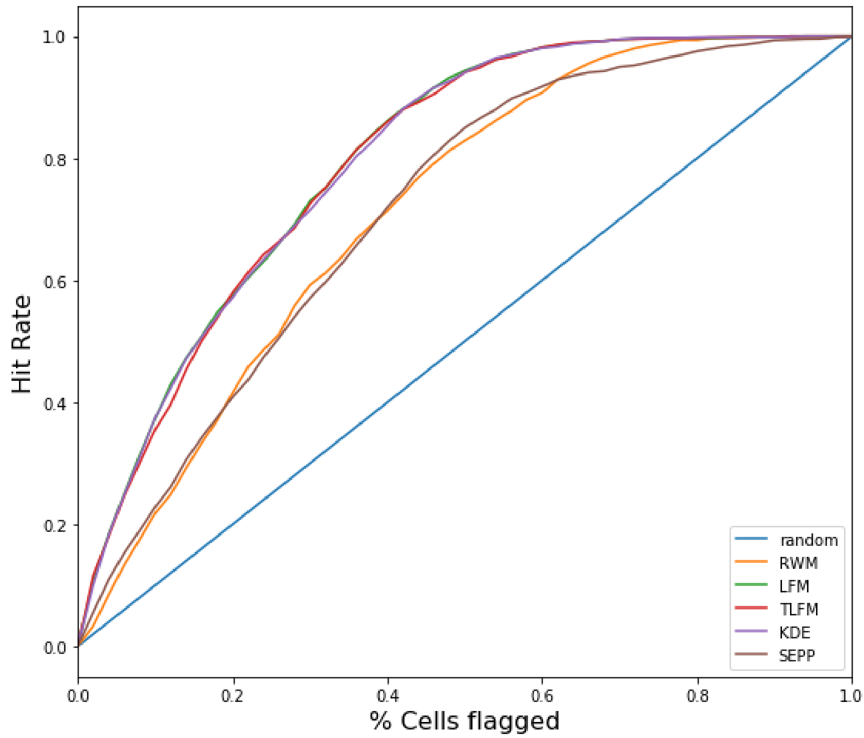


Figure 4: Average Cumulative Accuracy Profile curve for different crime prediction models.

The results are summarized in Table 1 in which the area under the average CAP curve (AUC) is presented for each of the models analyzed. TLFM and LFM present an AUC of almost 0.8, performing similarly or better than the state-of-the-art crime prediction models. The variance of the AUC obtained was, for each model, less than 0.0015, evidencing a independence of the predictive accuracy of the models from the election of the training and test sets.

	RWM	**LFM**	**TLFM**	**KDE**	SEPP
AUC	0.7152 (3.3e-4)	**0.7993** (5.3e-4)	**0.7969** (4.8e-4)	**0.7973** (5.8e-4)	0.7132 (0.0015)

Table 1: Area under average CAP curve for different crime prediction models. Standard deviation in parentheses.

The truncated and the pure Lévy flights models perform statistically equal. This could be explained, in part, by the fact that in a real world setting a pure Lévy flights model can not be simulated as cities are finite geographic spaces. The pure Lévy flights model fitted was, in fact, a truncated Lévy flight model with larger jump allowed equal to the size of Bogota ($< \infty$). However, the predictive accuracy of the KDE model evidence that, at least for residential burglaries in Bogota, hotspots present a very important persistent component over time, as the hotspots predicted by this model do not vary over time.

4.2 Estimated parameters

Table 2 presents the average value of the parameters estimated θ, that measures repeat victimization, ω, which accounts for the time span repeat victimization is more likely to occur, and η, capturing near-repeat victimization and broken windows effect, in the different training-test sets. The average rate of burglar generation Γ was of 11.6 ($\sigma^2 = 2.15$), while the average of the mean intrinsic attractive of houses $A^0 = 0.23$ ($\sigma^2 = 0.00026$). That is, on average for each simulation, that the static attractiveness component estimated over the lattice had a mean value of 0.23.

θ	ω	η
5.5133 (0.8974)	0.004 (1.4e-5)	0.9488 (7.4e-5)

Table 2: Estimated parameters Truncated Lévy flights model. Standard deviation in parentheses.

First of all, the decay rate $\omega = 0.004$ evidences a long time span when repeat/near-repeat victimization has an increased likelihood of occurrence. Concretely, the model shows that each time period (day) the dynamic attractiveness component decays 0.4%. This is consistent with the evidence presented in Figure 2, and with the high predictive accuracy of the KDE model as residential burglaries hotspots seem to present a very important persistent component over time: once a house is burglarized it becomes a high attractive target for offenders.

In the same fashion, the increase in attractiveness due to crimes parameter $\theta = 5.5133$ also exhibits a bias of criminals to return to previously burglarized houses. This could be explain, in part, because is where they already have good information of the types of properties that might be stolen and the routines of the police and their inhabitants (Wright & Decker, 1994).

For instance, consider a house s with an intrinsic attractiveness index equals to the average mean value estimated 0.23. If the house or its neighbors have never been attacked, and it is suddenly burglarized, the probability of a new attack the next day increases, due to the crime event, in

$$\begin{aligned} p(t)|_{E(t)=1} - p(t)|_{E(t)=0} &= (1 - e^{A^0+\theta}) - (1 - e^{A^0}) \\ &= e^{-0.23} - e^{-0.23-5.5133} \\ &= 0.7913. \end{aligned}$$

That is, an increase in 79.13 percentage points in the probability of being burglarized.

Moreover, the estimated parameter measuring triggering effects $\eta = 0.9488$ shows a strong and quickly spreading of attractiveness to neighboring sites. This validates the broken windows theory and the self-exciting nature of crime, with the past occurrence of crimes in a certain area creating an atmosphere of lawlessness and crime-tolerant region that encourages the occurrence of even more crimes.

Continuing with the previous example, a neighbor s' of the burglarized house s with the same static attractiveness component of 0.23 will face an increment in its probability of being burglarized, due to the crime event, of

$$\begin{aligned} p_{s'}(t+\Delta t) - p_{s'}(t) &= (1 - e^{A^0+\theta*\eta*(1-\omega)}) - (1 - e^{A^0}) \\ &= e^{-0.23} - e^{-0.23-5.5133*0.9488*(1-0.004)} \\ &= 0.7902. \end{aligned}$$

That is, an increase in 79.02 percentage points in the probability of being burglarized.

Thus, statistical evidence is found that support that crime is not a random event distributed evenly over the city, but presents spatio-temporal clustering patters. Furthermore, the found parameters elucidate criminals rational behavior, validating repeat/near-repeat victimization and broken windows effects. Public policies and policing strategies seeking to reduce criminal activity and its negative consequences must take into account these mechanisms and the self-exciting nature of crime, to effectively make criminal hotspots safer.

5 Incorporation of police control variables

Burglars do not only respond to past crime events but also to police control actions. As evidence suggests (see Figure 1), burglary hotspots are not static but vary through time in part due to a dynamic component in the attractiveness of houses and to responses of the police to criminal activity. Recent studies using crime data of cities in Colombia focus on the effect of public policies on criminal activity. Using impact evaluation methodologies, they evaluate the effect of police time and public resources in Bogota (Blattman et al., 2017), and public surveillance cameras in Medellin (Gomez-Cardona et al., 2017).

For instance, Blattman et al. (2017) studied if intensified state presence in criminal hotspots helps reducing crime in those areas. They randomly assigned nearly 2,000 Bogota's streets with high crime rate to eight months of doubled policing, increased municipal services, both, or neither, finding that in fact intensive policing makes high-crime streets safer. However, data from all streets suggest that this strategy of police control displace property crime to neighbor sites, with ambiguous impacts on violent crime.

On the other hand, Gomez-Cardona et al. (2017) used the installation of new surveillance cameras between 2013 and 2015 in Medellin, Colombia, as a quasi-experiment to investigate if this form of police control has any effect on crime. They found that total crime reports declined 23.5% and the level of arrests decreased in 31.5% in the coverage areas of the cameras after their installation, with no benefits to surrounding places. These results suggests that the effect of the surveillance cameras on crime occurrence is driven by a deterrent effect on criminals, as the monitoring capacity of the cameras is low and even decreased after the installation period, and there was not enough time to use camera images as proofs by the justice system.

These studies could be complemented by the model used in this work as, following the ideas of Jones, Brantingham, and Chayes (2010), police control variables could be introduced in the model to account for this criminal response to law enforcement actions, modifying the attractiveness perceived by criminals and, therefore, burglars' decisions to engage in criminal events. The attractiveness index is now given by

$$\tilde{A}_s(t) = e^{-\nu\cdot\psi_s(t)} A_s(t), \tag{5.1}$$

where $\psi_s(t)$ is a vector with police control variables such as police patrol, the presence of surveillance cameras, or lighting, and ν is a vector of positive constants.

The dynamics of police patrols could also be modeled as a truncated Lévy flight biased toward regions with high attractiveness index $A_s(t)$ in response to criminal activity. Doing so, the probability of a police moving from a house s to a house r is equal to

$$\hat{q}_{s\to r} = \frac{\hat{w}_{s\to r}}{\sum_{s'\in\mathbb{Z}^2, s'\neq s} \hat{w}_{s\to s'}}, \tag{5.2}$$

with

$$\hat{w}_{s\to r} = \begin{cases} \frac{A_r}{||s-r||^{\hat{\mu}}} & ||s-r|| < \hat{L} \\ 0 & ||s-r|| \geq \hat{L} \end{cases}, \tag{5.3}$$

for a given $\hat{L}$, larger police jump allowed, and $\hat{\mu}$, the exponent of the respective underlying Lévy flight.

Now the model is described by three equations:

$$B_s(t+\Delta t) = \left[(1-\eta)B_s(t) + \frac{\eta}{4}\sum_{s'\sim s} B_{s'}(t)\right](1-\omega\Delta t) + \theta E_s(t), \tag{5.4}$$

$$n_s(t+\Delta t) = \sum_{\substack{r\in\mathbb{Z}^2 \\ ||s-r||<L}} [n_r(t) - E_r(t)]\tilde{q}_{r\to s} + \Gamma\Delta t, \tag{5.5}$$

$$\kappa_s(t+\Delta t) = \sum_{\substack{r\in\mathbb{Z}^2 \\ ||s-r||<L}} \kappa_r(t)\hat{q}_{r\to s}, \tag{5.6}$$

with $\kappa_s(t)$ the number of police agents at house s in the time interval starting at t.

Fitting such model to real data of the locations of police, estimating the parameters ν, could shed light on interactions between criminals and law enforcement agents. Furthermore, using the right parameters, different policing strategies can be simulated and their effects on criminal activity compared in order to select the most effective one at reducing crime events. Nevertheless, this data is not publicly available for Bogota. Thus, an alternative could be to use already estimated effects of police control in Colombian cities, as found in Blattman et al. (2017); Gomez-Cardona et al. (2017), and compare the predictive accuracy of the model with police control variables with the initial one. This is subject of ongoing research.

6 Conclusions

Fitting the truncated Lévy flights agent-based model (Pan et al., 2018) to real crime data shed light on the rational behavior of criminals, without sacrificing predictive accuracy respect to other state-of-the-art crime prediction models. Using a MLE framework (Lloyd et al., 2016), I found statistical evidence to support that crime is not a random event evenly distributed over the city, but presents spatio-temporal clustering patters, due to its self-exciting nature. The estimated parameters of the model validate repeat/near-repeat victimization and broken windows theory, evidencing a bias of criminals to return to previously burglarized neighborhoods. Public policies and policing strategies that seek to reduce criminal activity and its negative consequences must take into account these mechanisms and the self-exciting nature of crime, to effectively make criminal hotspots safer.

A model that incorporates police control variables is proposed following Jones et al. (2010); Pan et al. (2018) ideas. Further work may seek to develop a data assimilation framework that shed light on interactions between criminals and law enforcement agents.

References

Barreras, F., Diaz, C., Riascos, A., & Ribero, M. (2016). A comparison of different crime prediction models for bogotá. *Documentos CEDE No 34*.

Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime* (pp. 13–68). Springer.

Blattman, C., Green, D., Ortega, D., & Tobón, S. (2017). *Pushing crime around the corner? estimating experimental impacts of large-scale security interventions* (Tech. Rep.). National Bureau of Economic Research.

Chaturapruek, S., Breslau, J., Yazdi, D., Kolokolnikov, T., & McCalla, S. G. (2013). Crime modeling with lévy flights. *SIAM Journal on Applied Mathematics*, *73*(4), 1703–1720.

Ehrlich, I. (1973). Participation in illegitimate activities: A theoretical and empirical investigation. *Journal of political Economy*, *81*(3), 521–565.

Gomez-Cardona, S., Mejia, D., & Tobon, S. (2017, 01). The deterrent effect of public surveillance cameras on crime. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2912947

Hesseling, R. B. (1992). Using data on offender mobility in ecological research. *Journal of Quantitative Criminology*, *8*(1), 95–112.

Jones, P. A., Brantingham, P. J., & Chayes, L. R. (2010). Statistical models of criminal behavior: the effects of law enforcement actions. *Mathematical Models and Methods in Applied Sciences*, *20*(supp01), 1397–1423.

Levajkovic, T., & Zarfl, H. M. M. (2016). Lévy processes, subordinators and crime modelling. *Novi Sad J. Math*, *46*(2), 65–86.

Lloyd, D. J., Santitissadeekorn, N., & Short, M. B. (2016). Exploring data assimilation and forecasting issues for an urban crime model. *European Journal of Applied Mathematics*, *27*(3), 451–478.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, *106*(493), 100–108.

Pan, C., Li, B., Wang, C., Zhang, Y., Geldner, N., Wang, L., & Bertozzi, A. L. (2018). Crime modeling with truncated levy flights for residential burglary models. *Mathematical Models and Methods in Applied Sciences*.

Short, M. B., D'Orsogna, M. R., Pasour, V. B., Tita, G. E., Brantingham, P. J., Bertozzi, A. L., & Chayes, L. B. (2008). A statistical model of criminal behavior. *Mathematical Models and Methods in Applied Sciences*, *18*(supp01), 1249–1267.

Wilson, J. Q., & Kelling, G. L. (1982). Broken windows. *Atlantic monthly*, *249*(3), 29–38.

Wright, R. T., & Decker, S. H. (1994). *Burglars on the job: Streetlife and residential break-ins*.

UPNE.

Appendix

A

Training with the whole six months from March to August 2012, a similar predictive accuracy is obtained for the test set corresponding to September 2012. The SEPP model was improved by capturing patterns of criminal occurence in different days of the week.

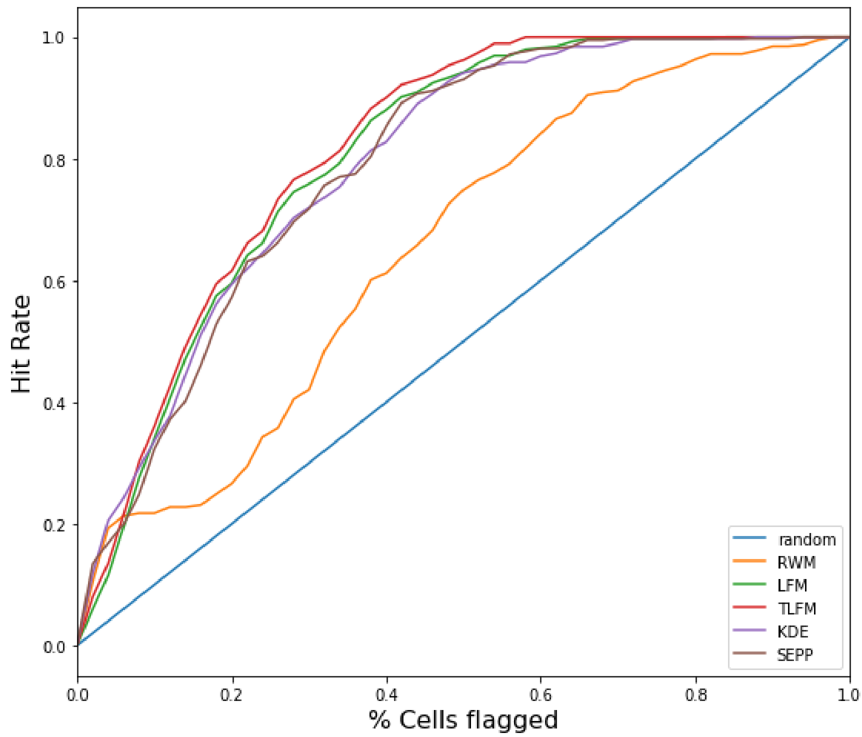


Figure 5: Cumulative Accuracy Profile curve for different crime prediction models. Training set: March - August 2012. Test set: September 2012.

	RWM	LFM	**TLFM**	KDE	SEPP
AUC	0.6545	0.8027	**0.8156**	0.7933	0.7906

Table 3: Area under CAP curve for different crime prediction models.

B

The spatio-temporal model proposed by Mohler et al. (2011) supposes that crime follows a self-exciting point process, in which past occurrence of criminal activity increment the likelihood of new crimes in a neighborhood of the first crime. The model estimates a function of spatio-temporal intensity capturing spatial patterns, temporal patterns, and triggering effects using Kernel Density Estimation and past occurrence of crimes. Similarly, this model captures repeat/near-repeat victimization and broken windows mechanisms but fails giving insights on criminal rationality.

First, crimes are classified between background and aftershock events, the former being those that arise independently given their spatio-temporal location, while the latter occur as triggering of past crimes nearby. Crime appearance is modeled as a self-exciting point process in which the past occurrence of crimes increases the probability of new crimes occurring in the future.

A spatio-temporal point process $N(x,y,t)$ is uniquely characterized by its conditional intensity $\lambda(x,y,t)$, which can be defined as the expected number of points falling in an arbitrarily small spatio-temporal region, given the points history $\mathscr{H}_t$ occurred until t (Mohler et al., 2011):

$$\lambda(x,y,t) = \lim_{\Delta x, \Delta y, \Delta t \downarrow 0} \frac{E[N\{(x, x+\Delta x) \times (y, y+\Delta y) \times (t, t+\Delta t)\} | \mathscr{H}_t]}{\Delta x \Delta y \Delta t}.$$

For the purpose of crime prediction and according to the initial assumptions on the behavior of crime, it is assumed that the conditional intensity takes the following functional form

$$\lambda(x,y,t) = \mu(x,y)\nu(t) + \sum_{\{k: t_k < t\}} g(x - x_k, y - y_k, t - t_k),$$

where $\mu(x,y)$ and $\nu(t)$ capture background crime occurence patterns according to their spatial and temporal locations, respectively. In a similar fashion, $g(x - x_k, y - y_k, t - t_k)$ captures how crime (x_k, y_k, t_k) propagates to other spatio-temporal locations.

To estimate the conditional intensity function, it is necessary to differentiate between background crimes and those triggered by past crimes, and use each of these families of data to estimate the functions μ and ν with background events and g with aftershock crimes. The training of the model is then based on *stochastic declustering* techniques and Kernel Density Estimation.

One way of improving this model is letting the kernel ν to capture temporal patterns, for example of criminal occurrence in different days of the weeks.