Characterizing and analyzing urban dynamics in Bogota

Marco De Nadai
Emmanuel Letouzé
Marta C. González
Bruno Leprì

June 2017


Contact at AFD: Thomas Roca (rocat@afd.fr)

1 University of Trento, denadai@fbk.eu
2 Data-Pop Alliance, eletouze@datapopalliance.org
3 MIT, martag@mit.edu
4 FBK, lepri@fbk.eu

L'Agence Française de Développement (AFD), institution financière publique qui met en œuvre la politique définie par le gouvernement français, agit pour combattre la pauvreté et favoriser le développement durable. Présente sur quatre continents à travers un réseau de 72 bureaux, l’AFD finance et accompagne des projets qui améliorent les conditions de vie des populations, soutiennent la croissance économique et protègent la planète. En 2014, l’AFD a consacré 8,1 milliards d’euros au financement de projets dans les pays en développement et en faveur des Outre-mer.

Les opinions exprimées dans ce papier sont celles de son (ses) auteur(s) et ne reflètent pas nécessairement celles de l’AFD. Ce document est publié sous l’entière responsabilité de son (ses) auteur(s).

Les Papiers de Recherche sont téléchargeables sur : http://librairie.afd.fr/

AFD Research Papers

AFD Research Papers are intended to rapidly disseminate findings of ongoing work and mainly target researchers, students and the wider academic community. They cover the full range of AFD work, including: economic analysis, economic theory, policy analysis, engineering sciences, sociology, geography and anthropology. AFD Research Papers and other publications are not mutually exclusive.

Agence Française de Développement (AFD), a public financial institution that implements the policy defined by the French Government, works to combat poverty and promote sustainable development. AFD operates on four continents via a network of 72 offices and finances and supports projects that improve living conditions for populations, boost economic growth and protect the planet. In 2014, AFD earmarked EUR 8.1bn to finance projects in developing countries and for overseas France.

The opinions expressed in this paper are those of the author(s) and do not necessarily reflect the position of AFD. It is therefore published under the sole responsibility of its author(s).

AFD Research Papers can be downloaded from: http://librairie.afd.fr/en/
Characterizing and analyzing urban dynamics in Bogota

Marco De Nadai, University of Trento, denadai@fbk.eu
Emmanuel Letouzé, Data-Pop Alliance, eletouze@datapopalliance.org
Marta C. González, MIT, martag@mit.edu
Bruno Lepri, FBK, lepri@fbk.eu

Abstract

Containing crime without affecting the livability of the urban environment is a major challenge in our society. Traditionally, researchers relate crime to socio-economic disorganization and people’s routine activity, as it influences effective control and suitable targets. An important open question is what the role the urban fabric plays. Although empirical research has shown that the physical urban environment is an essential factor for urban vitality and health, we lack evidence of any clear relationship between the structural characteristics (e.g. roads and land use mix) of neighborhoods and crime. Here, by using open data and mobile phone records, we explore this link with a spatial regression model that analyzes the environmental and the social conditions to which each part of the city is exposed. We found that physical characteristics of the city connected to higher urban diversity better explain the emergence of crime than traditional socio-economic conditions and, together, physical characteristics and socio-economic conditions improve the performance of previous approaches. This result suggests that urban diversity and natural surveillance theories play an important role in the proliferation of crime, and the knowledge of this role can be exploited in urban planning to reduce crime.

Key words: Welfare programs, Well-Being, Poverty, Family Structure and Planning, Economics of Minorities, Races, Indigenous Peoples and Immigrants, Non-labor Discrimination, Demand and Supply of Labor, Urban and Transportation Analysis, Housing, Land Use in Urban Economies

JEL Classification: I38, J12, J13, J15, J18, J21, J23, J61, J64, J68, J71, N36, O18, R14, R21, R23, R28, R41, R52

Original version: English

Accepted: June 2017
Characterizing and analyzing urban dynamics in Bogota

Marco De Nadai\textsuperscript{1,2}, Andrés Clavijo\textsuperscript{4}, Marta C. González\textsuperscript{3}, Emmanuel Letouzé\textsuperscript{4} and Bruno Lepri\textsuperscript{*2}

\textsuperscript{1}University of Trento. \textsuperscript{2}FBK. \textsuperscript{3}MIT. \textsuperscript{4}Data–Pop Alliance

Containing crime without affecting the livability of the urban environment is a major challenge in our society. Traditionally, researchers relate crime to socio-economic disorganization and people’s routine activity, as it influences effective control and suitable targets. An important open question is what the role the urban fabric plays. Although empirical research has shown that the physical urban environment is an essential factor for urban vitality and health, we lack evidence of any clear relationship between the structural characteristics (e.g. roads and land use mix) of neighborhoods and crime. Here, by using open data and mobile phone records, we explore this link with a spatial regression model that analyzes the environmental and the social conditions to which each part of the city is exposed. We found that physical characteristics of the city connected to higher urban diversity better explain the emergence of crime than traditional socio-economic conditions and, together, physical characteristics and socio-economic conditions improve the performance of previous approaches. This result suggests that urban diversity and natural surveillance theories play an important role in the proliferation of crime, and the knowledge of this role can be exploited in urban planning to reduce crime.

1 Introduction

The rapid growth of cities and the increase of population mobility have challenged our ability to understand crime. The primary focus of criminology research has been on people that commit crimes, and the reason they are involved. For crime to happen three conditions are supposed to be present and interact: the presence of a motivated offender who is willing to commit a crime, a suitable target, and the absence of guardians that would otherwise prevent the crime from taking place \cite{6}. In this equation crime offenders, victims and guardians are all affected by socio-economic conditions, the social disorganization (e.g. unemployment) of community \cite{9} and the place where they intersect. Thus, place matters.

Environmental criminology suggests that place not only is logically required, but also influences the likelihood of becoming a crime hotspot through its physical characteristics. Accordingly, place is one of the five necessary and sufficient components that constitute a criminal incident, namely place, time, law, offender and victim \cite{3}. Thus, environmental criminologists are interested in land use, street design, traffic patterns and daily activities of people. However, scholars have virtually ignored other theories (e.g. social disorganization), and bounded their discussion to macro-areas of few cities.

Urban planners and sociologists argue that cities are not a mere artificial construction that group people. A city is a vital process of the people who compose it; and its neighborhoods are the elementary form of cohesion in urban life \cite{15}. One of the seminal books in city planning is Jane Jacobs’ The Death and Life of Great American Cities \cite{13}. In this book she introduced the concept of eyes-on-the-street, which suggests that safety can be maintained by citizens through urban surveillance. For this to work, some physical qualities need to be present in the neighborhoods (i.e. a mix of residential, commercial and recreational land uses) to guarantee the diversity and continuous presence

\*lepri@fbk.eu
of people throughout the day. It is thus clear the tight coupling of environmental criminology and urban planning theories.

Traditional approaches on describing crime have failed to provide a clear and broad description of the desirable characteristics the different parts of the city should possess to keep crime events low.

In the present study we seek to fill this gap, formalizing the hypothesis that physical characteristics of the city not only are related to better life conditions [17] and vitality [7], but also greatly influence crime. Thus, we create two types of models. One is focused on describing how physical characteristics influence crime in each part of the city. The other predicts crime events in a city from the structural features, and answers to the question "can we predict crime from the physical characteristics of the city?". Thanks to several new sources of data and a Negative Binomial model, we model crime through physical and socio-economic characteristics, but also spatial and routine-activity information.

We find that structural characteristics of the city, namely Jacobs' diversity conditions, are better predictors than socio-economic status, and that these results are robust across different spatial aggregations. Also, we find that the number of inhabitants' routine movements between the neighborhoods is closely related to crime, with highly-connected points of the city experiencing a higher number of crimes, as suggested by the routine-activity theory [6]. Finally, we observe that the combination of structural, routine and socio-economic information provides better estimates of crime than each on its own. Together, these observational results suggest that the city structure has a strong connection with crime, and that improving its qualities can discourage criminality.

Thus, our main contributions are: i) we focus on place in a new fashion and show how physical characteristics greatly influence crime events; ii) we built a now-casting model, which is portable from one city to many, able to predict crime counts in a city; iii) we employ new sources of data and combine multiple criminology theories in Bogota.

This paper is organized as follows: in Section 2 we review the literature in this field. In Section 4 we outline the proposed approach and the evaluation process. Finally, we show our results in Section 5, before discussing the implications and limitations in Section 6.

2 Related work

One of the most prolific place that established the hallmark of environmental criminology is Chicago. In the University of this city, sociologists and criminologists started to consider neighborhoods as unit of analysis, both from the social and political (or administrative) perspective.

Ernest Burgess developed a concentric-zone model, based price changes in housing values, to study crime patterns in Chicago [4]. He observed that the distribution of social problems and crime vary in respect to the distances to the center. On the basis of this model, Clifford Shaw extensively researched how young people, juvenile delinquency and adult offenders were distributed in space [8]. He introduced the spot maps, delinquency rate maps, radial maps and zone maps that established a landmark on crime mapping.

In the recent years, many empirical and predictive studies flocked thanks to new methods mainly coming from computer science, and the increasing availability of new sources of data. Graif and Sampson [11] examined the connection of immigration and socio-economic diversity to homicide. Thanks to a spatial model they found that immigrant concentration is either unrelated or inversely related to homicides, whereas language diversity is negatively correlated to homicides. Verifying how diversity of people influences crime in neighborhoods was the goal of the study of Traummueller et al. [19]. To observe people dynamics they used mobile phone records broken by age and gender. They observed that age-diversity and presence of non-residents are linked to lower criminality. Bogomolov et al. [2] used mobile phone data in a similar fashion, and predicted crime hotspots thanks to a Random Forest regression. By only using the ambient population characteristics extracted from mobile phone records they were able to predict with almost 70% of accuracy whether an area would have high or low crime levels in time. The assumption of observing ambient population ignoring the movements of people in the city was tackled by Graif et al. [9]. They argued that people are continuously exposed to different neighborhoods, and they proposed a network of neighborhoods to describe this. This approach is considered important for understanding how changes of activity spaces can influence crime.

Strikingly, very few scholars considered multiple sources of data and theories in their discussions. Moreover, the importance of place is limited to spatial-autocorrelation or the topology of street
patterns (e.g. [16]). Contrarily to this, Wang et al. [20] examined crime in Chicago by leveraging census counts as well as new sources of data available on the web, such as crowd-generated Points Of Interest (POI) and taxi flows. POIs were expected to be linked to higher crime incident as they represent suitable targets; taxi flows were considered as proxy for trips made by humans. The incremental now-casting model built by Wang et al. showed that crime rate can be estimated with a relative error of 30% by using demographic information, but it could be also reduced by 5% with the POI data. Finally, he showed that the complexity of the crime in Chicago could be further explained by the taxi flow network, which further improved the results by 5%.

3 Description of data

4 Methods

Crime is considered a rare occurring phenomenon: a small proportion of people are victimized, with also few unreported and undiscovered crime. Crime offenses are not distributed evenly in space; they tend to cluster in parts of the city that exhibits similar characteristics, favorable to the spread of crime. Rare events, expressed through count variables, have been widely modeled on Ordinary Least Squares (OLS) through a logarithmic and square root transformation of the response variable. However, this introduces extra estimation bias, homoscedasticity assumptions and impossible predictions [14]. Moreover, spatial auto-correlation is rarely accounted for. For this reason, we use a Negative Binomial regression (NB) for count variables, and an eigenvector spatial filter (ESF) to account for spatial auto-correlation.

In this paper we want to describe a city but also to be able to predict it. Differently from the common meaning, here we use the word "predict" to mean the estimation of a property not directly observed (nowcasting).

Thus, we first employ a descriptive model for Bogota where it is possible to understand the interactions of each component of the model to describe crime events. Then, we create a predictive model validated with a 5-fold Cross-validation with 1000 repetitions (to avoid overfit). This allows to answer to the question "can I predict crime events from the characteristics of the city?".

4.1 A regression model for rare events

The Poisson regression model, a class of Generalized linear models (GLM), is particularly attractive to model count response variable. However, its restrictive assumption to have identical mean and variance is violated with many real-world situations. The NB introduces an extra parameter $\kappa$ to the Poisson model with parameter $\lambda_i$, where the variability of $\lambda_i$ has a gamma distribution with mean $\mu_i$ and index $v$. It follows that $Y_i \sim NB(\kappa, \mu_i/\kappa)$ and:

$$E(Y_i) = \kappa \frac{\mu_i}{\kappa} = \mu_i$$

$$Var(Y_i) = \mu_i \kappa - \mu_i$$

with $\kappa$ accounting for the extra variability with a quadratic function on $\mu_i$, and as $\kappa \rightarrow \infty$ the distribution of $Y_i$ converges to a Poisson random variable. With a log-link function on $\mu_i$, the NB model can be written as:

$$Ln[E(Y_i)] = Ln[E(i)] + \beta_0 + \sum_{k=1}^{n} X_k B_k$$

where $X_0, X_1, \ldots, X_n$ are the covariates, $\beta_0, \beta_1, \ldots, \beta_n$ the regression parameters, and $Ln[E(i)]$ is the offset variable and its role is to control for size differences across the units.

4.1.1 Spatial auto-correlation

Models dealing with spatial data analysis require to test for spatial auto-correlation. Positive (negative) spatial auto-correlation refers to the attitude of nearby attributes to have similar (dissimilar) values. There are numerous quantitative methods to measure and deal with spatial auto-correlation. The eigenvector spatial filter (ESF) [12] introduces a set of independent variables that account for the spatial relationship of the variables. These variables are a subset of the eigenvectors extracted from the numerator of the Moran’s I coefficient [5]:

$$(I - \frac{11^T}{n})W(I - \frac{11^T}{n})$$

where $I$ is a $(n \times n)$ identity matrix, 1 is a $n \times 1$ vector of ones, and $W$ is a generic $(n \times n)$ distance matrix. This matrix can be either be defined with geographical distance or flow of people and freight.

To model the proximity of each region, we define the $W$ distance matrix as the inverse squared distance separating the observations:

$$w_{i,j} = \begin{cases} 0 & \text{if } i = j \\ \frac{1}{d_{i,j}^2} & \text{otherwise} \end{cases}$$
where \(d(i,j)\) is the Euclidean distance between \(i\) and \(j\), and \(\gamma\) is a penalization parameter. We set \(\gamma = 2\) to place a greater weight on close observations and leave a marginal role to distant observations. The matrix is variance stabilized.

The \(n\) eigenvectors describe the full-range of uncorrelated spatial patterns; but employing all \(n\) eigenvectors in a regression framework is not desirable for reasons of model parsimony. Thus, we first filter the eigenvectors with low (and opposite) Moran’s I (\(\frac{MI_E}{MI_{max}} \geq 0.25\)). Then, from this subset we select the smallest subset of eigenvectors \(\{E_1, E_2, \ldots, E_p\}\) each time adding, in a step-wise fashion, the eigenvector that reduces the most the Akaike Information Criterion (AIC) of the model. The process stops when the AIC does not decrease anymore. The step-wise process does not guarantee to select the best eigenvectors, but it is a very simple and fast method to select the orthogonal eigenvectors to add. For further details on eigenvector selection and implementation strategies see Tiefelsdorf et al. [18]. The final subset of candidate eigenvectors represents the spatial filter for the variable analyzed. Thus, the aspatial NB model defined in Equation (3) takes the form:

\[
Ln[E(Y_i)] = Ln[E(i)] + \beta_0 + \sum_{k=1}^{n} X_k B_k + \sum_{j=1}^{p} E_j B_{n+j}
\]  

4.2 Spatial aggregation

Bogota is regionally recognized for the important strides it has made in reducing violence related to organized crime in recent years. To understand the multitude of aspects that influence crime in a neighborhood, we first have to define it. A neighborhood is a geographical unit composed by people who usually interact with each-other, and sharing common goals. This spatial community has a loose definition [13] and it has to be "big enough (in population) to swing weight in the city as a whole, but small enough so that street neighbourhoods were not lost or ignored". From this description we selected the Unidades de Planeamiento Zonal (UPZ) as a valid spatial aggregation for neighborhoods. The function of a UPZ is to help in the planning in the development of urban norms in the city. There are 113 UPZ and their average population is 63,720, with average area of 3.7 in 2009.

4.3 Measure of crime

Crime counts represent a measure to map and understand, in absolute terms, where the crime takes place. Nonetheless, crime rates over residential population are usually preferred as they assess the risk of people to be victimized in a particular location [1]. In a NB model, a rate variable is defined with an offset, which is a variable that is forced to have a coefficient of 1 in the model. This is particular useful to create risk maps, but we think it is a too restrictive constraint. To achieve a greater flexibility we prefer to add population as covariate and estimate its coefficient in the model. A population coefficient (\(B_p\)) greater than one means that spatial units with more population have higher rate of criminality. On the contrary, units with less population have fewer crimes per inhabitants when \(B_p < 1\).

4.4 Daily routines

As aforementioned, daily routines are supposed to influence the presence of offenders, victims and guardians in a place. Moreover, frequent trips between two places are supposed to influence each other’s crime. Wang et al. [20] used the taxi flow network with
Crime rates (over residents)

Figure 2: Crime rate (\(|\text{crimes}|/(\beta_p|\text{population}|)\)) in Bogota for each UPZ unit.

4.5 Jane Jacobs’ diversity

As aforementioned, our hypothesis is that Jane Jacobs’ diversity conditions have an impact on crime measures. Thus, in this section we describe the metrics, priorly validated [7, 17], that operationalize the Jacobs’ theory.

Land use mix. A common way of quantifying the mixture of land uses is through average Shannon entropy. The average entropy, which we here call Land Use Mix (LUM), is defined as:

\[
\text{LUM}_{3L,j} = -\sum_{j \in 3L} P_{i,j} \log(P_{i,j}) / \log(|3L|)
\]

where \(P_{i,j}\) is the percentage of square meters having land use \(j\) in unit \(i\), and \(3L = \{\text{residential, commercial and institutional, park and recreational}\}\) represents the land uses considered. The LUM ranges between 0, wherein the unit is composed by only one land use (e.g. residential), and 1, wherein developed area is equally shared among the \(n\) land-uses. The problem with this index is that it depends on the way land uses are grouped together, and no distinction is made on the order of land uses. Thus, an entropy of 0.75 could either mean high land use mix with a major role of residential land use, or high land use mix with a major role of parks. To better represent the different outcome we also employ a second entropy measure, based on the distinction between residential and non-residential land uses:

\[
\text{LUM}_{rnr,i} = -\sum_{j \in rnr} P_{i,j} \log(P_{i,j}) / \log(|rnr|)
\]

where \(rnr = \{\text{residential, non-residential,}\}\).

Jacobs argued for mixing primary uses so that people are on the street at different times of the day. To characterize spatial use in terms of activities, we determine whether each place is used daily (e.g., convenience stores, restaurants, sport facilities) or not. Based on that, we define the average accessibility of the buildings in a spatial unit \(i\) as:

\[
\text{A}_i = \frac{1}{|B_i|} \sum_{j \in B_i} \text{dist}(j, \text{closest}(j, D))^{-1}
\]

where \(D\) is the set of places that are used on a daily bases (e.g. convenience and grocery stores), \(\text{dist}(a,b)\) is the Euclidian distance between \(a\) and \(b\), \(\text{closest}(a,C)\) is a function that finds the closest item in set \(C\) from point \(a\), and \(B_i\) is the set of buildings in unit \(i\).

Consistent with the methodology used by the website developers to calculate Walk Score\(^1\), we define the weighted walkability score as:

\[
\text{walk}_i = \frac{1}{|B_i|} \sum_{c \in C} w_c \sum_{b \in B_i} w \text{dist}(b, \text{closest}(b, \text{POI}_c))^{-1}
\]

where \(C = \{\text{Grocery, Food, NightLife, Shops, Cultural}\}\), \(\text{POI}_c\) is the set of POIs of category \(c\), and \(w_c\) is the weight of importance to POIs, which is 3 for Food, Nightlife and Grocery POIs, 2 for Shops and 1 for others. \(w \text{dist}(a,b)\) is a linear function from 1, when the \(\text{dist}(a,b) <= 500\) m, to 0 when the \(\text{dist}(a,b) >= 2500\) m.

\(^1\)http://www.walkscore.com
**Small blocks.** Small blocks are believed to support stationary activities and provide opportunities for short-term and low-intensity contacts, easing into interactions with other people in a relaxed and relatively undemanding way. We compute the average block area among the set $B_i$ of blocks in unit $i$ as:

$$\text{Blocks area}_i = \frac{1}{|B_i|} \sum_{b \in B_i} \text{area}(b)$$  \hfill (12)

In addition, we compute the average distance between each building and the nearest street, a proxy for the concept *eyes on the street*, which suggest that the safety of neighborhoods can be maintained through continued surveillance of their inhabitants. For each unit $i$ and the set $S$ of streets, it is defined as:

$$\text{Eyes on the street}_i = \frac{1}{|B_i|} \sum_{b \in B_i} \text{dist}(b, \text{closest}(b, S))^{-1}$$  \hfill (13)

**Buildings.** Jacobs stressed the importance of having diverse buildings in a district to create vitality. Diverse buildings allow the mix of different socio-economic groups, as well as the tendency of an easier accommodation of creative people and small enterprises.

Colombia has a fiscal policy, called stratum, that classifies buildings in different regimes of tax payments for utilities and rents. Stratum is based on the external physical characteristics of the building, and it reflects the quality of life of residents with a six-level classification from 1 (lower low) to 6 (high). For this reason, we computed the heterogeneity of a unit $i$ as:

$$\text{Strata}_i = \sqrt{\frac{1}{|H_i|} \sum_{b \in H_i} (\text{strata}_b - \text{Strata}_i)^2}$$  \hfill (14)

where $H_i$ is the set of houses belonging in district $i$.

**Concentration.** Jacobs’ fourth and final condition is about having concentration of both buildings and people. First, we determine population density measures by dividing the number of people by the unit’s net area.

$$\text{Population density}_i = \frac{|\text{Population}_i|}{\text{area}_i}$$  \hfill (15)

Then we compute the floor-area density per each unit $i$ as:

$$\text{Buildings density}_i = \frac{|\text{Buildings}_i|}{\text{area}_i}$$  \hfill (16)

**Border Vacuums.** Border vacuums are places that act as physical obstacles to pedestrian activity. For instance, parks can be a hub of pedestrian activity, if efficiently managed [13], but they could also be deplorable places in which criminality flourishes (especially at night). Thus, we compute the average closeness of each building from the nearest park as:

$$\text{Closeness to LP}_i = (\frac{1}{|B_i|} \sum_{j \in B_i} \text{dist}(j, \text{closest}(j, LP)))^{-1}$$  \hfill (17)

where dist$(j, \text{closest}(j, LP))$ is the distance between block $j$ and its closest large park.

### 4.6 Covariates

The number of committed crimes is mainly influenced by the number of residents, their social disorganization and routine activity. Social disorganization is the inability of the neighborhood to maintain effective social control. Social disorganization is higher in deprived areas, social heterogeneous units and in places with high unemployment rate, which is also a proxy for motivated offenders. We define the social heterogeneity through unemployment rate:

$$\text{unemployment}_i = \frac{|\text{unemployed residents}_i|}{|\text{residents}_i|}$$  \hfill (18)

and a proxy of income heterogeneity through the weighted standard deviation of property values:

$$\text{social heterogeneity}_i = \sqrt{\frac{\sum_{b=1}^{n} w_b^2 \sigma^2}{\left(\sum_{b=1}^{n} w_b\right)^2}}$$  \hfill (19)

where $x_b$ is the property value of a block in spatial unit $i$ and $w_b$ is the residential population count of block $b$.

The number of residents is calculated as:

$$\text{population}_i = |\text{residents}_i|$$  \hfill (20)
4.7 Evaluation

The $R^2$ statistic is an intuitive interpretation of the proportion total variation in outcome that is accounted for by a OLS model. Concerning GLMs there is no directly analogous $R^2$ measure. For this reason, GLMs models are usually evaluated through their AIC, Pseudo-$R^2$ and Root Mean Squared Error (RMSE). One of the most interesting Pseudo-$R^2$ measures is the McFadden Pseudo-$R^2$, which:

\[
\text{McFadden Pseudo-}R^2 = 1 - \frac{\log \hat{L}(M_{\text{full}})}{\log \hat{L}(M_{\text{intercept}})}
\]

where $\hat{L}(M_{\text{full}})$ is the log likelihood of the full model and $\hat{L}(M_{\text{intercept}})$ is the log likelihood of the null model. It is worth to remember that this is not a true measure of fit, because it only compares the log likelihood of the full model with the one of the null model.

In the predictive model, evaluated through the K-fold Cross-validation ($K = 5$), we create multiple models that use a subset of the features. Thus, the subsets are:

- Socio-economic: demographic and social disorganization variables;
- City: Jane Jacobs’ diversity variables;
- Dynamics: daily routines variables;

and the combinations of them. This allows to understand the contribution of each subset to the description of the crime events in a city.

5 Results

Our contribution in this paper is twofold. At first we focus on describing the relation of the various factors with crime. Then, we built a predictive model to understand how the results can be generalize in different cities.

5.1 Descriptive model

From the results (in Table 2) we can observe the $\beta$ coefficients of features to understand the importance of each variable, holding the others as constant. The most important variables to describe crime are building density, population density and closeness to daily-use buildings. Particularly, the concentration variables are very important in describing crime, with building density that is positively correlated with crime (0.498). By contrast,
<table>
<thead>
<tr>
<th></th>
<th>β coefficient</th>
<th>std. error</th>
<th>p-value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Land use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land use mix (LUMHL)(8)</td>
<td>-0.078</td>
<td>0.029</td>
<td>**</td>
<td>[-0.135, -0.021]</td>
</tr>
<tr>
<td>Land use mix (LUMHNL)(9)</td>
<td>-0.098</td>
<td>0.034</td>
<td>**</td>
<td>[-0.166, 0.032]</td>
</tr>
<tr>
<td>Closeness daily buildings(^\d)(10)</td>
<td>0.212</td>
<td>0.035</td>
<td>***</td>
<td>[0.144, 0.281]</td>
</tr>
<tr>
<td>Walkability score (11)</td>
<td>0.146</td>
<td>0.060</td>
<td>*</td>
<td>[0.029, 0.263]</td>
</tr>
<tr>
<td><strong>Small blocks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Block area (12)</td>
<td>-0.033</td>
<td>0.054</td>
<td></td>
<td>[-0.140, 0.073]</td>
</tr>
<tr>
<td>Eyes on the street (13)</td>
<td>0.270</td>
<td>0.055</td>
<td>***</td>
<td>[0.163, 0.378]</td>
</tr>
<tr>
<td><strong>Buildings</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strata diversity(^\d)(14)</td>
<td>0.070</td>
<td>0.030</td>
<td>*</td>
<td>[0.011, 0.129]</td>
</tr>
<tr>
<td><strong>Concentration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population density (15)</td>
<td>-0.363</td>
<td>0.045</td>
<td>***</td>
<td>[-0.451, -0.276]</td>
</tr>
<tr>
<td>Building density (16)</td>
<td>0.498</td>
<td>0.054</td>
<td>***</td>
<td>[0.393, 0.603]</td>
</tr>
<tr>
<td><strong>Vacuums</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closeness parks(^\d)(17)</td>
<td>0.031</td>
<td>0.039</td>
<td></td>
<td>[-0.045, 0.108]</td>
</tr>
<tr>
<td><strong>Covariates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployment (18)</td>
<td>0.110</td>
<td>0.036</td>
<td>**</td>
<td>[0.039, 0.181]</td>
</tr>
<tr>
<td>Population (20)</td>
<td>0.6647</td>
<td>0.058</td>
<td>***</td>
<td>[0.550, 0.779]</td>
</tr>
<tr>
<td>Social heterogeneity (19)</td>
<td>0.159</td>
<td>0.041</td>
<td>***</td>
<td>[0.078, 0.240]</td>
</tr>
<tr>
<td>OD eigenvectors (Sec. 4.4)</td>
<td>-0.184</td>
<td>0.042</td>
<td>***</td>
<td>[-0.267, -0.101]</td>
</tr>
</tbody>
</table>

Spatial eigenvectors 2  
McFadders Pseudo-R\(^2\) 0.145  
RMSE 92.81  
Moran’s I (p-value) 0.02 (0.42)

\(^\d\) This is not a true measure of fit, and not bounded to 1. It indicates the degree to which the model parameters improve upon the prediction of the null model.

Table 1: Negative Binomial regression model that describes the number of crimes in each spatial unit.

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>C</th>
<th>D</th>
<th>S+D</th>
<th>C+D</th>
<th>S+C</th>
<th>Full</th>
</tr>
</thead>
<tbody>
<tr>
<td>McFadders Pseudo-R(^2)</td>
<td>0.077</td>
<td>0.113</td>
<td>0.085</td>
<td>0.106</td>
<td>0.120</td>
<td>0.141</td>
<td>0.143</td>
</tr>
<tr>
<td>RMSE</td>
<td>231.93</td>
<td>145.04</td>
<td>312.70</td>
<td>181.76</td>
<td>133.36</td>
<td>143.35</td>
<td>127.76</td>
</tr>
</tbody>
</table>

\(^\d\) This is not a true measure of fit, and not bounded to 1. It indicates the degree to which the model parameters improve upon the prediction of the null model.

Table 2: Negative Binomial regression models that predict the number of crime in each spatial unit. The results are average across 1000 iterations of a 5-fold Cross-validation. S: demographic and social disorganization variables only; C: Jane Jacobs’ diversity variables only; D: daily routine variables only; S+D: demographic, social disorganization and daily routine variables; C+D: Jane Jacobs’ diversity and routine variables; S+C: demographic, social disorganization and Jane Jacobs’ diversity variables; Full: model with all the variables.
the higher the population density is, the less crime events there are.

Small blocks are also very related to crime, as the distance of buildings from the nearest street has a positive correlation with crime events (0.270). This is in accordance to the eyes on the street theory of Jane Jacobs, that generates a virtuous loop which, in turn, increases public safety.

Covariates coming from criminology literature are also significant. We found that Social heterogeneity and unemployment are positively correlated with crime. Thus, higher deprivation and disorganization might stop the mechanisms by which residents themselves achieve guardianship and public order.

Finally, we find that social isolation increase crime, as it supposedly limits neighborhood possibilities and social capital [10]. Thus, we see that highly-connected points of the city experiencing lower number of crimes.

5.2 Predictive model

Our preliminary findings (see Table 2) indicate that structural characteristics of the city, namely Jacobs’ diversity conditions, are a better predictor of the target variables (the number of homicides and robberies) than socio-economic conditions such as unemployment and deprivation. Mobility networks, and thus the routine activity theory, improve the prediction of the model by 15%. The combination of structural and socio-economic information provides better predictions than each on its own.

6 Discussion

In this paper we modeled, for the first time, multiple aspects of urban life to describe and predict crime. We have done so by operationalizing the Jane Jacobs theory to describe the urban fabric, the social disorganization and the routine activity variables from criminology. We can now discuss some implications of our work.

Descriptive maps. Maps are invaluable tools in criminology to understand where the problems are, and to evaluate initiatives for crime prevention. Thus, our framework allows policy makers to visually analyze crime rates (Figure 2), crime counts (Figure 1) and, most importantly, spatial auto-correlations (Figure 3) and heterogeneous effects (Figure 4).

Factors for crime. With our descriptive model we have shown that it is possible to have a static description of what happens in the city, and how the multitude of complex features of city life come together. Therefore, now more than ever, it is important to control for the many relations that come together, especially as a consequence of urban fabric and mobility. Urban diversity and natural surveillance theories play an important role in the proliferation of crime, and the knowledge of this role can be exploited by policy makers to reduce crime.

Generalization. Our predictive model that the combination of structural and socio-economic information, and mobility, provides better predictions than each on its own. However, it is striking that the urban fabric has such an important role in the prediction. Thus, we think it is critical to consider it in new models and analysis in different cities in the world.

This work is not without any limitation. First, we don’t analyze crime over time. Then, cities are not to be considered island unto themselves, as they are embedded in a country-wide complex system of social interactions. Routine of residents exposes them to different cities, conditions and possibilities on a daily basis. Thus, we think that in the next future it is important to consider also these factors.

Our findings are apt to crime control and prevention action plans for Colombian cities. This study provides valuable insights for local governments so that they can base urban management decisions on empirical evidence on the deterrents of crime.

References


List of recent AFD Research Papers

AFD Research Papers are available through AFD’s website at https://www.afd.fr/en/ressources-accueil


