

Using Hotspots Algebra to analyze crime events and their urban context

Uso de un álgebra de puntos calientes para analizar eventos delictivos y su contexto urbano

Lenin Heredia G.
Universidad de Los Andes. Bogotá
Correo: hg.lenin10@uniandes.edu.co

Germán E. Bravo C.
Universidad de Los Andes. Bogotá
Correo: gbravo@uniandes.edu.co

Fecha de recibido: 02/11/2016 y Fecha de aprobación: 17/02/2017

Resumen

El análisis de puntos calientes es esencial en el campo de la criminología y bastante importante en la toma de decisiones para las agencias de policía, porque permite la mejora de la asignación de recursos policiales para las acciones oportunas y adecuadas. Existen diferentes técnicas para el análisis y la generación de puntos calientes, limitados por su dificultad para considerar otros factores urbanos y demográficos que podrían ser la causa de la aparición de estos puntos calientes o estos pueden influenciar sobre otros factores. Por otro lado, también es difícil e inusual analizar puntos calientes en un contexto temporal. Este trabajo busca a través de operaciones espacio-temporales basados en los puntos calientes, ir más allá del análisis del crimen clásico basado en puntos calientes, mediante la búsqueda de la influencia espacial de otros factores espaciales y analizando su relación en un contexto temporal.

El trabajo inicialmente presenta y analiza el rendimiento de diversas técnicas para la generación de punto calientes, llegando a la conclusión de que la técnica STAC soportado por *CrimeStat* es la más adecuada para el objetivo propuesto. A continuación, se define un álgebra de puntos calientes para soportar el estudio combinado del crimen y factores espaciales que lo afectan y/o son afectados por los eventos delictivos. El análisis temporal incluye la generación de puntos calientes para los días de la semana y/o meses en un año. De esta manera es posible estudiar las relaciones de causalidad y/o correlaciones entre los fenómenos estudiados y el entorno espacio-temporal donde se produce el crimen. Por lo tanto, es posible definir y aplicar medidas más adecuadas y oportunas, sobre todo en lo concerniente a la asignación de los recursos policiales.

Por último, se muestran algunos ejemplos de aplicación del análisis espacio temporal de álgebra de puntos calientes utilizando datos de delitos cometidos en la ciudad de Bogotá entre los años 2011 y 2013. Finalmente, se proponen algunos trabajos futuros en el tema.

Palabras Clave:

Álgebra de puntos calientes
Puntos calientes
Análisis del crimen
Elipses espaciales
STAC
CrimeStat
ArcGIS Desktop



‡Se concede autorización para copiar gratuitamente parte o todo el material publicado en la *Revista Colombiana de Computación* siempre y cuando las copias no sean usadas para fines comerciales, y que se especifique que la copia se realiza con el conocimiento de la Revista Colombiana de Computación.

Abstract

Hotspots analysis is essential in the criminology field and quite important in decisions making for police agencies because it permits the enhancement of allocation of police resources for timely and adequate actions. There exist different techniques for analysis and generation of hotspots, limited by its difficulty to consider other urban and demographic factors that could be the cause of the emergence of these hotspots or their influence over other factors. On the other hand, it is also difficult and unusual to analyze hotspots in a temporal context. This paper seeks, through spatial-temporal operations based in hotspots, to go beyond of classical crime analysis with hotspots, by looking for the spatial influence of other spatial factors over and analyzing also their relationship in a temporal context.

The paper initially presents and analyze the performance of various techniques for hotspot generation and determines that STAC technique from CrimeStat is the more suitable for the proposed objective. Then, it defines a hotspot algebra allowing the combined study of crime and spatial factors affecting it and/or been affected by crime events. Temporal analysis includes the hotspots generation for days of the week and/or months in a year. In this way it is possible to study causality relationships and/or correlations among the studied phenomena and the spatial-temporal environment where crime occurs. Therefore, it is possible to define and apply informed actions, primarily concerning the allocation of police resources.

Finally, it shows some application examples, thematic as well as temporal analysis, of hotspot algebra using crime data from Bogota for the years 2011 to 2013; finally, some future works in the subject are proposed.

Keywords:

Hotspots
Hotspots Algebra
Crime Analysis
Spatial Ellipses
STAC
CrimeStat
ArcGIS Desktop

1. Introduction

Crime mapping is employed by crime analysts to identify, visualize and analyze patterns in crime incidences. It is a key component in crime analysis and police strategy. For most analysts, the dominant paradigm for these activities has been the visual interpretation: just simply looking at the map and using common sense, judgment, knowledge of the jurisdiction and their dynamics; but in other cases, more analysis is needed and requires the use of statistical methods.

Crime is mapped for a number of reasons such as identify patterns in crime incidents and problems associated with them, identify hotspots, serving as a visual aid in the dissemination of information patterns and problems, show the relationship between geography and other factors, see the direction of patterns movement, check location data, create patrol districts, follow the crime changes, making maps for the deployment of general information to the police.

Crime rarely is distributed randomly or uniformly in the space [2], crime tends to be concentrated in specific locations, some areas have more crime than others, and some areas have different types of crime than others, meaning that some hotspots may be hotter than others. For strategies and purposes of troubleshooting, identifying high crime areas is useful for the development and evaluation of responses from the police.

There are several types of hotspots: 1) hot places (points or areas), where crime is concentrated in specific locations such as malls, stadiums or hospitals; 2) hot streets, which indicates that the crime, like illicit drug trafficking and pimping, is distributed similarly along an entire street segment; and 3) hot areas suggest equivalent crime risk in the whole area with a dramatic reduction or increase of the risk in the border, and are represented as thematic or choropleth maps. In all cases, a hotspot suggests that focusing the police activity in the area is adequate. Hot areas may be affected by natural, artificial and social boundaries, as for example gang territories [1] [2] [3].

There are different techniques for representing hotspots, like: 1) pin maps [1][2][4][16] to represent hot places; 2) thematic or choropleth maps [1][2] which, usually, refer the crime to administrative or jurisdictional areas; 3) network maps quadrants [2], that divide a region into a network of cells and the color gradient of each cell depends on the volume of events in the cell; 4) density maps [2] [3] [7], which are the most popular to represent hotspots for their aesthetic representation and are considered the best technique to represent the actual distribution of crime; and 5) standard deviation ellipses, based on statistical tests [3] [17], whose geometric representation is adequate and could easily be used in hotspots spatial algebra. To generate spatial standard deviation ellipses there are several methods

such as Nearest-Neighbor Hierarchical Spatial Clustering (NNHSC), Spatial and Temporal Analysis of Crime (STAC) [3] [17], K-Means Partitioning Clustering [3][7]. The selected tool and the method used in this work are CrimeStat [3] [17] and STAC, respectively.

In most cases the crime analysis ends with hotspots generation in the space, rarely considers the temporal context of phenomena under study, and rarely assesses causality relationships and/or correlations among other spatial factors. Furthermore, it is necessary to consider social variables like household density, socio-economic status, age risk, etc. can varies in space in the same way crime does; these variables are also correlated with geography and can explain geographical patterns in crime, as for example the identification street drug dealing patterns at city level as well as across neighborhoods [20]. Aiming the enhancement of the crime analyst's skills and the decision making process to mitigate crime, this works presents the development of a software artefact solving questions about the phenomenon under study and other spatial-temporal factors. Some of these questions are: Which and where are located the recreation and entertainment sites (museums, parks and theaters) in Bogota? Are the CAIS (Immediate Action Commands) well located to deal with people theft crime? How the people theft crimes affect the public transportation security? Which districts should be concerned by people theft in Bogotá? How changes the common murder crime along week days? Is it true that the liquor consumption in bars or nightclubs influence the common murder crime? How probable is that the motor vehicles theft (cars and motorcycles) could result in homicide? Which are the areas that embrace entertainment and recreational facilities (parks, museums and theaters) not affected by people theft crime? Where are located the sites where the consumption of liquor in bars or nightclubs may be triggering murders, but that are not influenced by people theft? Is it possible that threats have been materialized and ended in homicide? Where average theft is located? Which area concentrates the most of robberies? Which is the limit area for all robberies?.

The document describes the different types of hotspots, generation techniques, the hotspot algebra proposed, the development of the software artifact that supports spatial-temporal operations using the defined hotspots algebra, some obtained results for crime events in Bogotá for years 2011 to 2013, and finally some conclusions and future work.

2. Hotspots crime mapping

2.1. Techniques for generating Hotspots

2.1.1 Pin Maps

Pin maps are used to represent hot places, it is the most common and popular method for visualizing geographic patterns of crime [1] [2] [4] [16]. Incidents placed on the map are displayed through appropriate symbols to represent different categories of crime, also contain information such as the code that describes the type, date and time of the offense and other data related to the incidents. Hot places could be overlooked if lines or polygons are used to define hotspots at this level. The main disadvantage of pin maps is the difficulty to try to interpret spatial patterns and crime hotspots in large data sets.

3. Thematic or choropleth maps

A fairly popular technique to represent any spatial distribution of specific events are thematic or choropleth maps [1] [2] [3] [4] [16]. This kind of maps are divided into several polygons based, usually, on administrative or political areas such as census blocks, voting districts, neighborhoods or city limits, county, or a police section. Criminal events mapped as points are added to these geographic areas and counts over them can thematically show spatial patterns of crime in an area of interest. There are several classification schemes for generating such maps, such as: equal count, equal range, equal area, natural breaks, standard deviation, quantile, or custom ranges [2]; the classification system selected affect the outcome map and therefore the presence or absence of hotspots or high crime areas.

3.1. Network Map quadrants

This method compensates the limitations of thematic maps to overcome the problem of varying sizes and shapes of geographical administrative border areas. The unit for mapping could be a count of crimes per cell or a density value calculated from the count of cells and area [2]. This approach uses estimation techniques of surface and illustrates a geographical surface area. The process involves the calibration of two parameters, the cell size and radius search, so the results may vary significantly. The method involves placing a grid or network on top of a map (creating an array of cells), then evaluates a mathematical function from the center of each cell and performs a calculation on each cell, and within a predetermined search radius or bandwidth, producing a map with small grid cells thematically shaded. Some limitations of this technique are that the results fail to attract the user to any firm conclusions about the size or shape of hotspots. The outputs represented as geographical grids often result in loss of spatial detail within each quadrant and its quadrants, leading to inaccurate interpretation problems.

3.1.2. Interpolation methods and continuous surface smoothing

Interpolation and smoothing methods for analysis of density are the most popular for viewing and identifying hotspots [2] [3] [7] [17]. The method consists of the following steps: 1) a fine grid is generated on the distribution of points; 2) A three-dimensional function for each cell in the grid is calculated based on interpolation methods of kernel density (KDE), such as normal distribution, uniform distribution (flat), quartic distribution (spherical), triangular distribution (conical), negative exponential distribution. For each method a given search radius for each cell is set and weights for each point within the core radius is calculated, knowing that points closer to the center receive a higher weight and thus contribute more to the value of the total cell density. The score of a cell approaches the concentration of criminal activity. The value of cell density is an estimate and is influenced by the incidents that are within the search radius placed on top of the grid cell. 3) Finally, the values of final grid cells are calculated by adding all circle surfaces for each location. The variation in the input parameters could produce significantly different outputs, the choice of cell size or spatial resolution affecting the smoothness of the surface. There are no rules on setting these parameters and the choice of parameters is determined by the user depending on the type of crime, the size of the study area, etc.

3.1.3. Statistical Techniques

A repeatedly visual detection is sufficient to detect hotspots, but it is not enough to detect subtle changes in the geography of patterns over time or to calculate the correlations between two or more geographical variables, or to perform spatial predictions. The use of spatial statistics allows to measure the degree of clustering or dispersion within a group of incidents, to analyze spatiotemporal relationships, to analyze the movement of a serial offender, to detect geographic profiles, and to generate density maps [2] [3] [7]. In addition, there are statistical studies that correlate social disorganization factors with geographic patterns in crime, using for example, the negative binomial regression model [20].

3.1.3.2. Spatial Distribution Measures

The most basic statistical measures to understand overall patterns in crime data include the media center, the center of the average minimum distance, the median center, the geometric and harmonic mean. The main idea is to create polygons measuring the concentration of incidents, rather than determine a specific central point for the series. Some choices for these polygons include a rectangle showing the standard deviation of the X and Y coordinates, the circle of standard deviations and finally the standard deviation ellipse [3].

3.1.3.2. Spatial standard deviation ellipses

Spatial ellipses are a traditional method for detecting hotspots and an alternative for surface smoothing techniques based in kernel density analysis [3] [17]. Ellipses improve the distance standard deviation, avoid skewed distributions, and minimize the amount of extra space that appears in some circles of standard deviation. Ellipses are a common way of measuring the tendency of sets of points or areas, because the standard distance is calculated separately for X and Y.

These two measures define the axes of an ellipse covering the distribution of incidents, and is called standard deviation ellipse [3]. The ellipse allows the seeing of event distribution in an elongated way and with orientation, facilitating the interpretation of the studied phenomena. It is possible to specify the number of standard deviations to represent (1, 2 or 3, that cover approximately 68, 95 and 99 percent of all entry incidents respectively [3]. The comparison of hotspot spatial ellipses over time or other kind of hotspots help to determine whether there is a reduction or displacement of crime. The main limitation of spatial ellipses is that they represent an approximation of the actual crime distribution [3]. The grouping of events in elliptical clusters also excludes, in any visualization, the events that do not belong to a cluster.

3.1.3.3. Nearest-Neighbor Hierarchical Spatial Clustering

This method uses the technique of nearest neighbor analysis (NNA) to identify clusters of a minimum number of points defined by the user [3] [17]. The technique of nearest neighbor analysis identifies only those points that are closer than expected under spatial randomness. The first set of ellipses generated through this process is called first order cluster. Grouping these first-order clusters leads to second-order ellipses; grouping second order ellipses lead to third order ellipse, and so on, although in practice, the routine stops after identifying second or third level clusters.

3.1.3.4. Spatial and Temporal Analysis of Crime (STAC)

STAC [3][5] [17] explores the data by overlapping a grid over the study area and implementing a search circle to each cell. The routine counts the number of points in each circle in order to identify the denser clusters. Circles sharing the same points are added, so that hotspots may vary in size. STAC has some of the same parameters as NNHSC technique [3] [17], like the search radius, the minimum number of points per group and the number of standard deviations for the generated ellipses.

3.1.3.5. K-Means clustering

The K-Means technique creates a number of “K” ellipses defined by user, by partitioning crime data points in groups [3] [17]. The routine searches the best positioning of the K centers and then assigns each point to the its closest center.

3.2. Evaluating the fastest technique to generate hotspots

In order to choose the more efficient mechanism to generate crime hotspots, twelve scenarios and seven techniques are evaluated. The scenarios consist of five types of crime (crime in general, people theft, personal injury, mall theft, burglaries) in Bogota, and the seven weekdays to consider temporal analysis. The seven techniques were Getis Ord Gi*, Simple KDE, KDE generated with the Crime Analyst Tool, Thematic Maps, NNHSC, STAC and K-Means. For each sample, run time was measured and ranked as shown in Figure 1.

	Events	GI* Getis Ord	Crime Analyst	Thematic Maps	NNHSC	STAC	K-means	KDE
Total	97746	7	3	4	1	2	5	6
People theft	35314	6	3	4	7	1	1	5
People injury	17502	7	1	3	5	2	3	6
Mall theft	9180	7	3	4	5	1	1	6
Burglary	9141	7	2	3	5	1	3	6
Murder-Monday	275	7	4	6	1	1	1	5
Murder-Tuesday	263	7	4	6	1	1	5	1
Murder-Wednesday	253	7	4	6	1	1	5	1
Murder-Thursday	293	7	4	6	1	1	1	5
Murder-Friday	304	7	4	6	1	1	1	5
Murder-Saturday	449	7	3	6	4	1	1	4
Murder-Sunday	676	7	2	5	5	3	1	3

Figure 1. Ranking of performance for hotspots generation. 1 represents the fastest method.

It was found that the techniques that generate spatial hotspots represented by standard deviation ellipses are faster than the techniques that work with kernel density estimation and thematic maps. The techniques that generate spatial standard deviation ellipses do not present significant variation in runtime with each other, being STAC slightly faster than NNHSC (because it creates spatial ellipses in different order) and K-Means.

Having as criteria for selecting the technique the performance and the control in the outputs, STAC is selected as the technique to hotspots generation. In STAC the outputs are controlled through parameters, allowing the crime analysts to spend their time and effort not in the hotspots generation, but rather on hotspots algebra and in the discovering of new information, or on the study of relationships among hotspots or other geographic factors.

3.3. Temporary considerations in hotspots

Hotspots can change their shape and location over time, giving place to periods of appearance, persistence and decline of crime. Hence the time interval (time of day, day of the week, weekends, holidays and seasonal periods of high crime areas) selected in the analysis affect the size and shape of the hotspots. The hotspots analysis over time should help to determine if there is crime changes or reductions.

4. Hotspots algebra

4.1. Importance of hotspots algebra

In crime analysis there are different types of hotspots (hot places, hot streets and hot areas) referred to a phenomenon under study, that can be spatially related to other geo-referenced factors through spatial-temporal operations based on hotspots algebra. The spatial operations considered in the hotspots algebra include the union, intersection, difference, symmetric difference and complement, using as operands different types of hotspots related to a phenomenon under study.

The algebra is designed to enhance the analysis capabilities of criminologists by allowing the operation of hotspots of a given phenomenon with other geo-referenced factors, other hotspots and other spatial information. The analyst defines a hotspot expression using set operators as union, intersection, complement, difference and symmetric difference and data layers or hotspot layers as operands. A prior filtering of data can lead to spatial and/or temporal expressions of hotspot algebra. The evaluation of a hotspots algebraic expression generates new hotspots that could show causality relationships and/or correlations between hotspots and other factors involved in the expression. The information derived from hotspots algebra is valuable to enhance the crime analysis because the crime analyst could define more appropriate actions to combat, mitigate and prevent criminal phenomena.

Each hotspot type has a scope, hot places concentrate high incidence of events in specific places, hot streets group events along lines, while hot areas indicate uniform impact on entire areas with a drastic change (high or low) on the borders. Crime hotspots and other spatial factors are related through spatial operations; the union operation aims the generation of a new data layers from data layers having some similar characteristics and hence, by generalization, reduce the complexity of the analysis; the intersection is the most important and frequently used because it allows the discovery of new information by combining several hotspots; the operations of difference, symmetric difference and complement are used to exclude spatial factors in the analysis proposed by analysts.

4.2. General approach

Figure 2. Shows the process for generating new hotspots by using the hotspots algebra. Crime analysis handles different types of spatial information such as the phenomenon under study, sites of interest, administrative areas, and other geo-referenced information with different geometrical representation as points, lines or polygons; this spatial

information is handled by spatial layers, stored in a layer repository, from where it can be retrieved and managed when the analyst requires it.

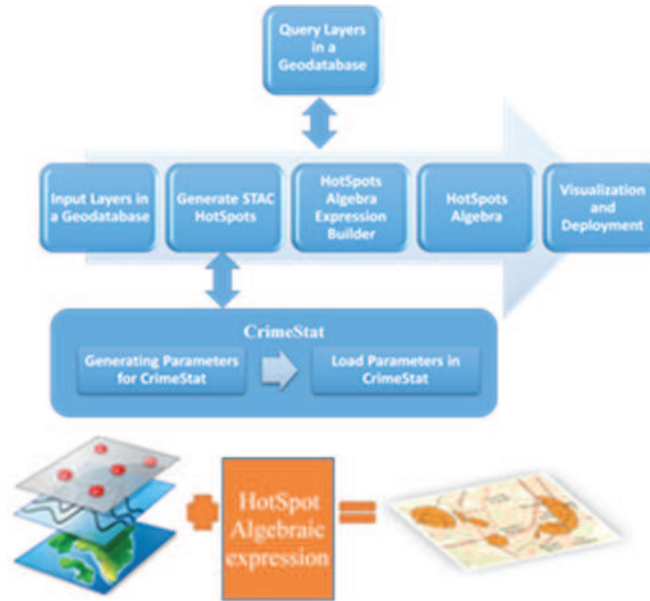


Figure 2. Hotspots algebra process.

The analysis is based on point layers associated with a specific type of information, from which hotspots are generated as spatial standard deviation ellipses whose boundaries contain high incidences of events under analysis. The generated hotspots are represented by elliptical polygonal layers and stored in a layer repository, to be retrieved and used in hotspots algebra expressions.

The layer repository has information associated with events, hotspots, and other geo-referenced factors describing the city, such as sites of interest, jurisdictional areas, districts, banks, malls, schools, colleges, public transportation, etc.

The expression generated by the hotspots algebra calculator (see an example in Figure 3) is parsed and solved by a geo-processing tool, which acts as a lexical analyzer, and generates new hotspots, which represent new information for possible relationships between the phenomenon in study and other spatial factors defined in the algebraic expression.

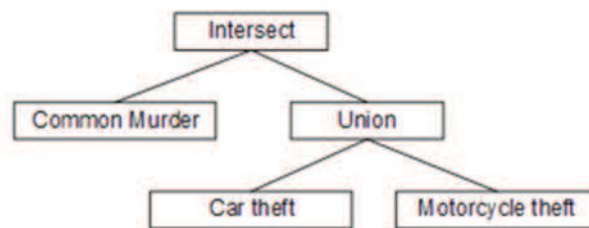


Figure 3. Generation of a hotspot algebra expression to analyze the correlation between vehicles theft and common murder.

The new generated hotspots are also stored in the layer repository, to be retrieved, visualized and deployed with their associated information whenever the analyst requires them.

Algebra Calculator hotspots allows the analyst to generate analysis scenarios, built through spatial operators, aiming to relate or correlate an event under study with other events or other spatial factors. The calculator generates a syntax tree where the nodes represent the spatial operators while the leaves represent the spatial layers. The syntax tree depends on the complexity of the study scenario proposed by the crime analyst, varying in both, depth and breadth.

Once proposed a study scenario represented by a hot spots algebraic expression, the generation of new hot spots is achieved by a geo-processing tool which parses the algebraic expression and generates new temporary layers, as the result of a spatial operation of two layers. The syntax tree is traversed in postorder, from bottom to top and from left to right. The latter is important for the non-commutative operations like difference, symmetric difference and for the unary operation of complement.

4.3. System architecture and implementation

Figure 4 shows the system architecture for the hotspots algebra calculator, and other geo-processing tools supporting the desired enhancement of crime analysis.

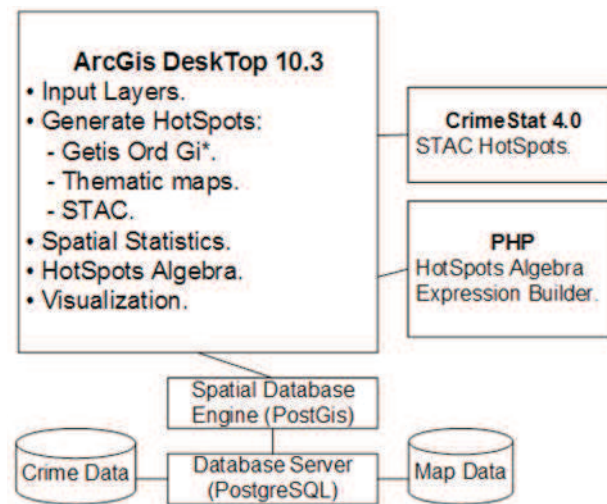


Figure 4. The system architecture supporting the crime analysis based on hotspots algebra.

A brief description of the components is: an ArcSDE geodatabase managed with PostGIS [10] to store the spatial layers, in shapefile format, and other alphanumeric information; CrimeStat is used to generate standard deviation ellipses hotspots using the STAC technique; a module to capture the hotspot algebraic expression, developed in PHP; and ArcGIS 10.3, that is used to generate the raster hotspots for thematic analysis and temporal analysis using the Getis Ord G_i^* technique, to solve the algebraic expression and to visualize the data and results.

The repository for storing and managing layers associated with different spatial data type is supported by an ArcSDE geodatabase, managed with the PostgreSQL relational database manager and is spatially enabled through PostGIS extension [10]; the recovery and management of layers is based on a descriptive data model also implemented in the PostgreSQL database.

Events point layers are retrieved from the geodatabase in order to generate spatial standard deviation ellipses hotspots with the STAC technique using CrimeStat 4.0 [17]. Communication between the system and CrimeStat is made through stored parameters files and retrieved from temporary directories. When a point layer associated to an event is selected, the system generates and saves a parameter file, which is then loaded to CrimeStat. These parameters specify the X and Y coordinates representing the study area and the output directory, and are then complemented with the spatial ellipses generation parameters, like the search radius, the search unit, the minimum amount of points that to constituting a cluster, the number of simulations, the scan type which can be rectangular or triangular and the number of standard deviations for the ellipses. Once the all parameters required are entered in STAC, the hotspots are generated and stored in a temporary directory as well as the new parameters specified in STAC are saved by the user to a temporary. Then, the system loads the hotspot shapefile and new parameters and stores them in the geodatabase in order to make them available for further visualization and processing.

The hotspot calculator has two main components: the expression editor and the expression evaluator. The expression editor gives access to all the information in the geodatabase, allowing the user to specify filters over the data in order to generate the ellipses hotspots focusing in a given phenomenon or in a given temporal analysis. The editor facilitates the writing of fully parenthesized expressions using the union, intersection, difference, symmetric difference and complement as operators and hotspot layers as operands. The editor also builds a syntax tree and allows the reusing of subexpressions, being in this way more powerful and easy to use. Due to integration issues between tools used, the editor is written as a web page, using PHP as development language.

The expression evaluator is implemented in ArcGis Desktop 10.3, using Python ToolBox, ArcObjects and the ArcPy library. The evaluation result is also stored in the geodatabase, as well as all the information concerning its generation, like date and time of execution, base data, hotspots, parameter for the hotspot generation and operations making part of the algebraic expression defined.

To visualize the results ArcGis Desktop 10.3 is also used. The information concerning how the results shown were generated is shown as a web link, using the hyperlink mechanism provided by ArcGIS.

5. Results

The case study works with crime events occurring in Bogota city, Colombia, during the period from 01/01/2011 to 02/04/2013. The data was provided by the Colombian National Police through the Study Center about Drugs and Security (CESED) of University of Los Andes.

The data contains 24 offense types, including people theft, murder, theft in commercial entities, residential theft, motor vehicle theft, among others.

Other data included in the project consist of other geo-referenced data representing theaters, museums, bars, restaurants, hotels, motels, supermarkets, malls, CAIs, banks, ATMs, citizen service centers, embassies, police stations, among others; there is also data related to administrative or jurisdictional areas such as neighborhoods, districts, strata and parks.

The examples shown are just a sample of all the questions that a crime analyst may have. These examples, with their corresponding syntax tree, aim to show how the crime analysis could be enhanced by posing questions in hotspot algebra that help to discover cause-effect relationships and/or correlations between the phenomena under study and other spatiotemporal factors.

The system helps to solves questions such as:

Are the CAIs well located to deal with people theft? (see Figure 5). The result can suggest to crime analyst how effective is the CAIs strategy and where the presence of police should be strengthened in order to deal with people theft crime. Is it true that the liquor consumption in bars and nightclubs could be influencing common murders? (see Figure 6). The result shows the intersection between bar and night clubs hotspots and murder hotspots. It can suggest or confirm to the analyst that the alcohol consumption in bars or nightclubs could be triggering the homicide crime and therefore, they could define actions in these places.

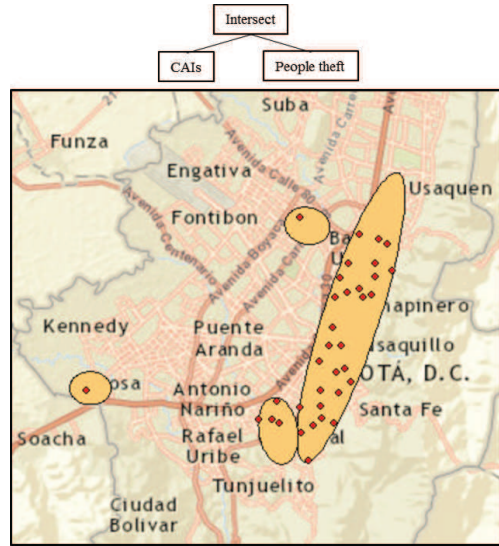


Figure 5. Personal thefts vs. CAIs. The ellipses are the hotspots and the points are the CAIs (Immediate attention centers).

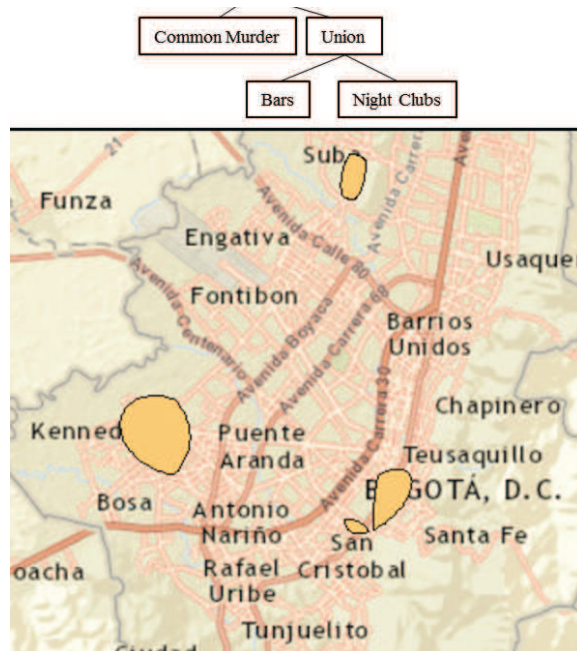


Figure 6. Alcohol Consumption vs murders.

How are related the vehicle theft with common murder? (see Figure 7). The result shows the intersection of murders hotspots with the union of car and motorcycles theft. It could help the crime analysts to assess whether or not cars or motorcycles theft could result in homicides, and therefore to define actions and places needing more control or police presence.

Which are the Transmilenio stations more susceptible to people theft? (see Figure 8). The result shows the intersection of Transmilenio stations (points) hot places with people theft hotspots. This can help crime analysts to have more security presence and take better actions in the stations; this result could also be useful for Transmilenio managers, who can make better security measures at these stations, by preventing the users, persuading the offenders and hence mitigate theft.

Which are the sections of the Transmilenio lines more affected for people theft? (see Figure 9). The result shows the intersection of Transmilenio lines (lines) hot streets with people theft hotspots. It helps the crime analysts to make better policing action measures, as well as more patrolling on those lines. It also can help to Transmilenio managers to provide more security to the units in these lines.

Which are the districts involved in areas that concentrate great amount of people theft? (see Figure 10). The result shows the intersection of districts with people theft hotspots. It can help the crime analysts to identify the districts having areas with higher concentration of theft. They can make better police actions in these areas. In addition, crime analysts may suggest to local governors to take more attention in these districts where people theft occurs. It may also suggest that criminals use districts borders to evade police checks and controls from local governments, and use them as natural protective barriers, because of jurisdictional problems.

Which are the areas that include entertainment and recreational places (parks, museums and theaters) not affected by people theft? (see Figure 11). The result shows the intersection of people theft hotspots with the union of entertainment and recreational facilities (result of the union of parks, museums and theaters). It could show to crime analysts, the areas with recreational and entertainment places where the presence of theft is almost zero, and possibly to reassign the police resources to other areas with more concentration of people theft.

Where are located the sites where the consumption of liquor in bars or night clubs may be triggering murders, but that are not influenced by people theft? (see Figure 12). The result shows the intersection between the union of bars and night clubs hotspots and murder hotspots, from the intersection result, people theft hotspots is removed. It can help the crime analysts or specialists in homicide unit, to identify areas where the consumption of liquor in bars or night clubs may be triggering murders, but excluding as a possible motive the people theft, and determine other possible causes for these murders. In addition, the security could be improved by taking greater control and police presence.

Is it possible that threats have been materialized and ended in homicide? (see Figure 13). It can help to crime analysts to determine if threats made quite possibly have been materialized and resulting in homicide. In this way, crime analysts can prevent and reduce homicides, paying more attention to the threats denounced.

How the common murder distribution changes over the weekdays? (see Figure 14). The answer can suggest the crime analysts in which days and places should be appropriate to take special actions to diminish murders.

These examples show that hotspots algebra is simple and could be very useful for generating new hotspots, which are important to enhance the crime analysis, based on the combination of different spatial operations such as union, intersection, difference, symmetric difference and complement, having as operands hotspots and other spatial factors.

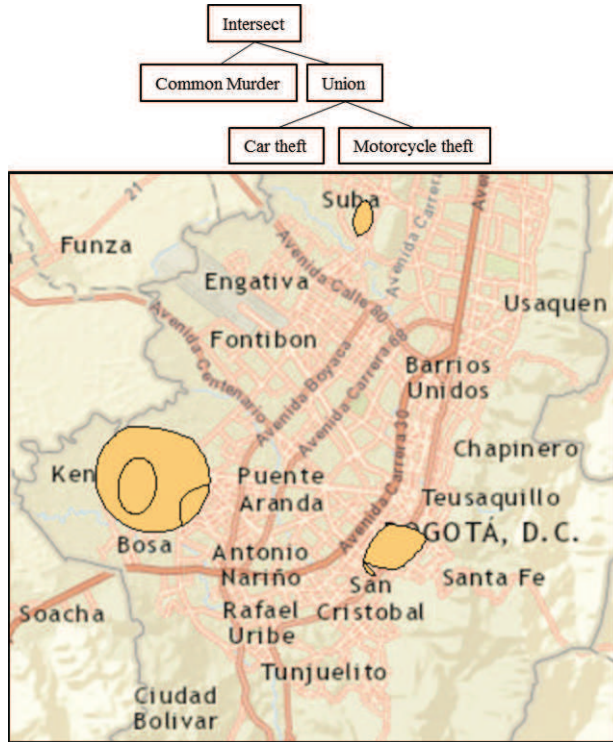


Figure 7. Motor vehicles theft vs. murders.

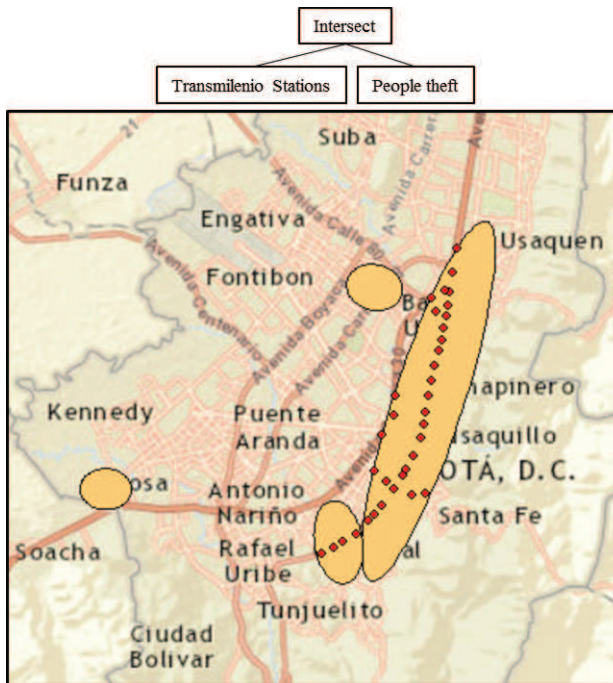


Figure 8. Transmilenio stations vs. people theft.

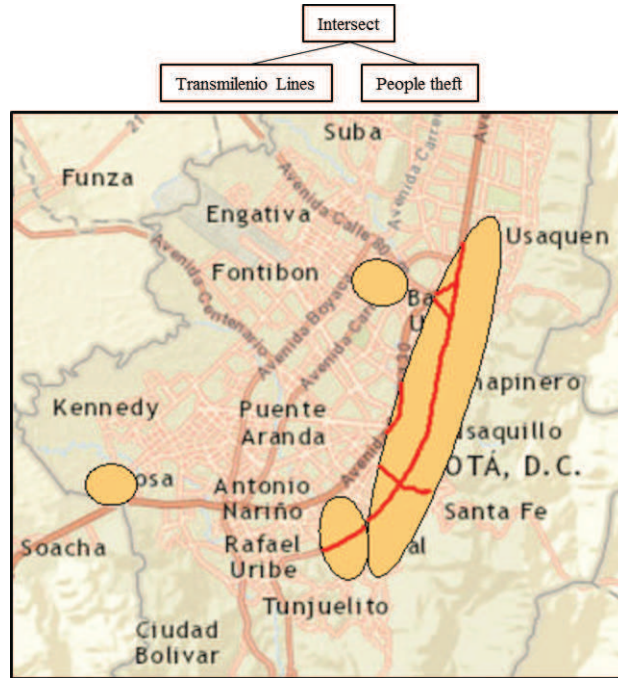


Figure 9. Transmilenio lines vs. people theft.

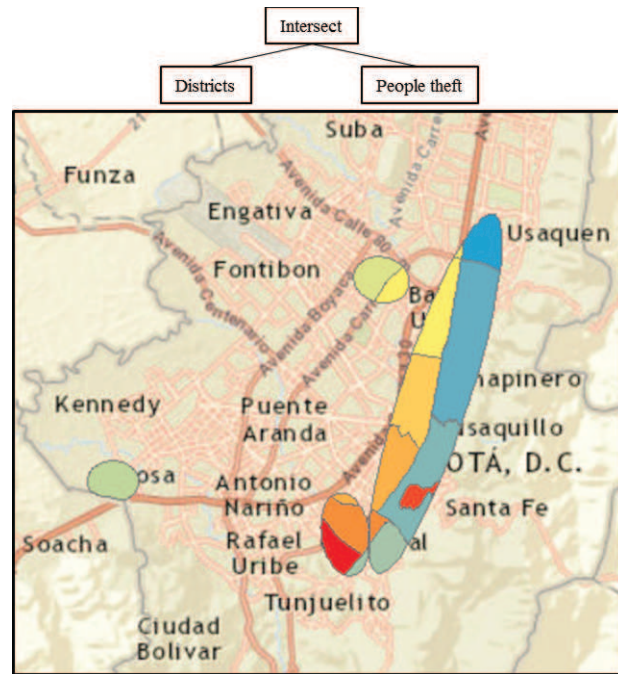


Figure 10. Districts vs. people theft.

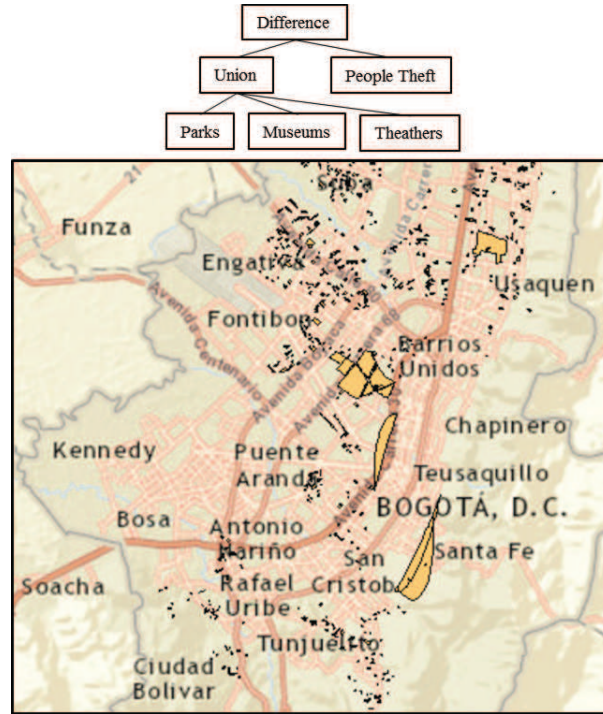


Figure 11. People theft vs. entertainment and recreational facilities.

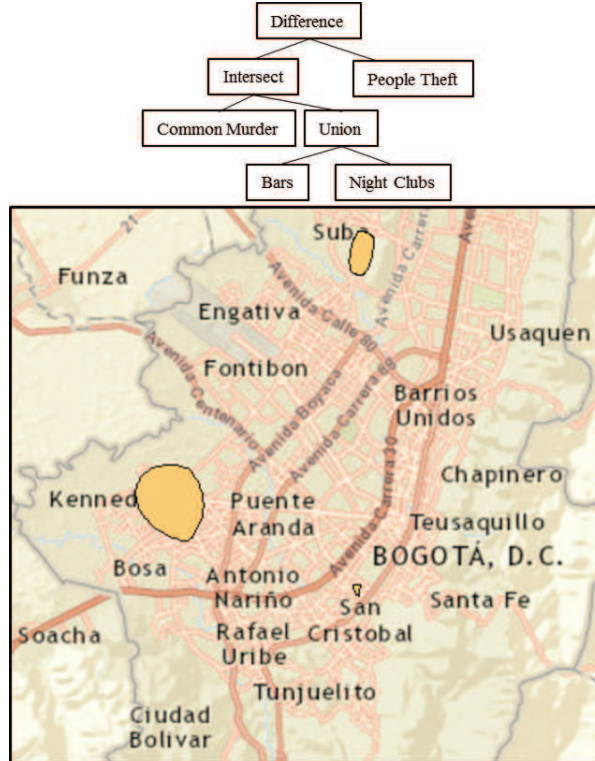


Figure 12. Alcohol Consumption in bars and night clubs vs murders and no people theft.

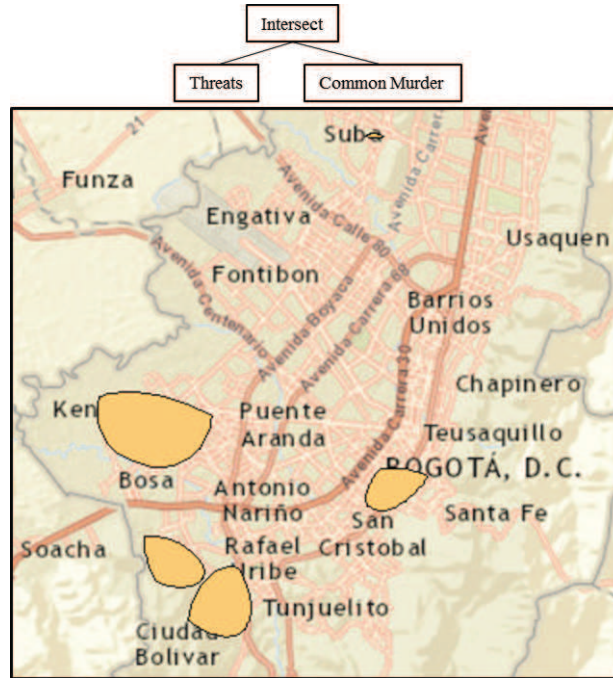


Figure 13. Threats vs murders. The result shows the intersection between threats hotspots and murder hotspots.

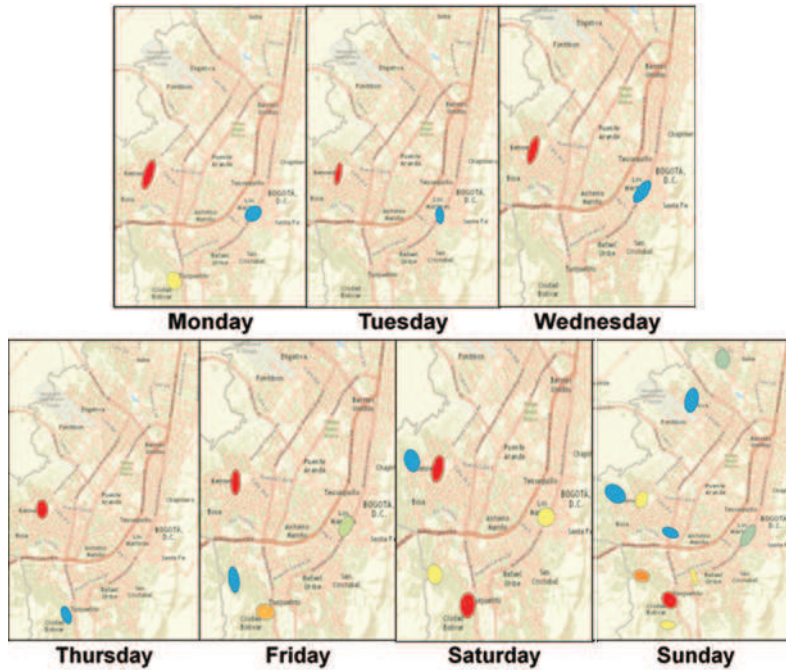


Figure 14. Hotspots for common murder in week days.

6. Conclusions

The main conclusions of this work are:

The hotspots generation represented by spatial standard deviation ellipses or convex polygons with the STAC technique, coherently represent criminal phenomenon under study.

The techniques that generate spatial standard deviation ellipses (NNHSC, STAC and K-Means) are faster than the techniques that work with kernel density estimation (KDE).

Of the seven techniques evaluated (Getis Ord G_i^* technique, Simple KDE, KDE generated with Crime Analyst Tool, Thematic Maps, NNHSC, STAC and K-Means), STAC turned out to be the best alternative, due to their speed and control of outputs, and in consequence is used in this work.

The use of ArcSDE geodatabase was appropriate because it allows an efficient storage and management of spatial layers, independently of their type of information and geometric representation, and also because it harness the power of relational databases systems.

The developed system is easy to use and intuitive, allowing the entry of crime layers, the generation of hotspots with techniques already presented, the generation of new hotspots using hotspots algebra, the visualization and deployment of different layers that have been stored and managed in ArcSDE geodatabase.

The generation of new hotspots by the hotspots algebra process is valuable to enhance the crime analysis for the decision makers concerned by crime, facilitating the discovering of new information, causality relationships and/or correlations between the phenomena under study and other spatial factors in a space-time context.

Finally, the developed system is applicable to study other domains, such as disease outbreaks, forest fires, traffic accidents, and others.

7. Future work

This work should lead to other developments like:

- Generate polygon hotspots from the results of raster hotspots generation, and used them as operands of the hotspot algebra.
- Define new operations on hotspots, such as forecasts through correlation analysis, time series and other methods, to enhance even more the crime analysis.
- Use other methods for hotspot generation as an alternative with possibly more precision. These methods could generate convex polygons hotspots with techniques already available in CrimeStat tool (NNHSC, K-Means).
- Implement the hotspot algebra application in free GIS platforms. In this way the platform would be available to many more law enforcement agencies, and have a positive impact on countries in development for leading with and preventing crime.
- Solve the compatibility issues between ArcGIS and CrimeStat, making the user interaction with the platform more uniform.

An alternative to quickly achieve the implementation of a more uniform platform using the statistical power of CrimeStat, is the use of the libraries provided by CrimeStat [21], that are free license for educational and research purposes. These libraries cover different statistical routines, being the hotspots analysis libraries the most important. The libraries should be put into operation on the .NET platform, making possible the generation of hotspots with the STAC technique

for point data. The hotspots algebra engine could also be handled on this platform, using ArcObjects ArcGis for the .NET platform or other GIS software for visualization.

References

- [1] Janet Reno, Daniel Marcus, Laurie Robinson, Noël Brennan, Jeremy Travis. "Mapping Crime: Principle and Practice". NCJ 178919, December 1999.
- [2] John E. Eck, Spencer Chainey, James G. Cameron, Michael Leitner, and Ronald E. Wilson. "Mapping Crime: Understanding Hotspots". NCJ 209393, 2005.
- [3] Susan C. Smith, Christopher W. Bruce. "CrimeStat III, User Workbook". The National Institute of Justice, June 2008.
- [4] Chainey and Ratcliffe, "GIS and Crime Mapping", 2005.
- [5] Chainey, S.P., Tompson, L., Uhlig, S. "The utility of hotspot mapping for predicting spatial patterns of crime". Security Journal, 2008.
- [6] Luc Anselin, Jacqueline Cohen, David Cook, Wilpen Gorr, and George Tita. "Spatial Analyses of Crime", Criminal Justice, 2000.
- [7] International Association of Crime Analysts (IACA). "Identifying High Crime Areas". Standards, Methods, & Technology (SMT) Committee White Paper 2013-02, October 2013.
- [8] International Association of Crime Analysts (IACA). "Definition and Types of Crime Analysis". Standards, Methods, & Technology (SMT) Committee White Paper 2014-02, October 2014.
- [9] Chao Ren, Rui Li*, Meng Li, Caihong Li. "A Projection-based Hotspot Analysis Method". 2013 IEEE.
- [10] Guiyun Zhou*, Jiayuan Lin, Wenfeng Zheng. "A Web-based Geographical Information System for Crime Mapping and Decision Support". 2012 IEEE.
- [11] Zhanhong Wang, Jianping Wu, Bailang Yu*. "Analyzing Spatio-temporal Distribution of Crime Hot-spots and Their Related Factors in Shanghai, China". 2011 IEEE.
- [12] Xiang Zhang*, Zhiang Hu, Rong Li and Zheng Zheng. "Detecting and Mapping Crime Hotspots Based on Improved Attribute Oriented Induce Clustering". National Key Technology R&D Program in the 11th Five-year Plan of china, task number: 2008BAH23B01.
- [13] Peng Chen, Tao Chen, Hongyong Yuan. "GIS Based Crime Risk Analysis and Management in Cities". 2010 IEEE.
- [14] MA Wei, CHEN Jianguo, CHEN Peng. "Illegal Activities Hotspot Analysis Based on GIS Methods". 2011 IEEE.
- [15] Li Li, Zhongbo Jiang, Ning Duan, WeiShan Dong, Ke Hu, Wei Sun. "Police Patrol Service Optimization Based on the Spatial Pattern of Hotspots". 2011 IEEE.
- [16] Wilpen L. Gorr, Kristen S. Kurland. "GIS Tutorial for Crime Analysis". Esri Press, 380 New York Street, Redlands, California. 2012.
- [17] Ned Levine (2015). CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 4.02). Ned Levine & Associates, Houston, Texas, and the National Institute of Justice, Washington, D.C. August.
- [18] Eric Pimpler. "Programming ArcGis 10.1 with Python Cookbook", Packt Publishing Ltd., Birmingham B3 2PB, UK. February 2013.

- [19] ESRI, "ArcGis Resource Center", <http://resources.arcgis.com/en/help/main/10.2/index.html#/00qn0000001p000000>, accessed in 2015.
- [20] Elenice Oliveira, Braulio Figueiredo Alves Silva and Marcos Oliveira Prates. "Street drug markets beyond favelas in Belo Horizonte, Brazil", 2015.
- [21] Ned Levine, CrimeStat Libraries (1.1). Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC, December 2012



Sobre los autores

Lenin Heredia G. Docente del Departamento de Ingeniería de Sistemas y Computación. Universidad de Los Andes, Bogotá.

Germán E. Bravo C. Docente del Departamento de Ingeniería de Sistemas y Computación. Universidad de Los Andes. Bogotá.



Este artículo se cita:

IEEE

L. Heredia G. and B. C. G. E., "Using Hotspots Algebra to analyze crime events and their urban context," *Rev. Colomb. Comput.*, vol. 18, no. 2, pp. 9–26, 2017.

APA

Heredia G., L., & E., B. C. G. (2017). Using Hotspots Algebra to analyze crime events and their urban context. *Revista Colombiana de Computación*, 18(2), 9–26.