



Control points based on network analysis for hot spots by qualified theft, City of Aguascalientes, Mexico, 2011 – 2017.

Jonathan Steven Rivera Zambrano¹ and Victor Hugo Galván Noriega²

¹ CentroGeo Campus CDMX, CDMX, México. jsteven@centrogeo.edu.mx

² CentroGeo Campus CDMX, CDMX, México. victor_galvann@hotmail.com

Abstract. A methodology for the manipulation and analysis of georeferenced points as a result of the theft described in the City of Aguascalientes from 2011 to 2017, the detection of areas with high impact is generated, which end up becoming the inputs for the objective and strategic selection of points of control for timely intervention and constant monitoring of local authorities.

Keywords: Network Analysis, Density Estimation, Qualified theft.

1 Introduction

The amount of data that is currently handled can be a great potential for the identification of patterns or deeper reasoning to the behavior of human groups, however, its use and manipulation can be complicated. At present, a compendium of georeferenced data is analyzed due to the theft described in the City of Aguascalientes, from 2011 to 2017. The proposed methodology tries to capture from the possible sectorized distribution of the crimes, to the selection of control points that allow the intervention of the authorities in a timely manner in future criminal acts.

2 Network Analysis; Methodology

The analysis of networks takes its foundations, from postulates such as graph theory and topology, with its origin in discrete and applied mathematics, the need to problematize circumstances that allow identifying more efficient forms of communication between different places or knowing the relationship between Actors in a community is the most simplistic way of understanding the relationship between network analysis and graph theory, in the case of Geographic Information Sciences, one of its proposals in this field is the use of topology, which allows introduce to the stage the study of cities and social variables through roads [1].

The network analysis applied in this investigation, intends to relate the spatial grouping of denouncement of theft, georeferenced in the City of Aguascalientes, to the network of streets of the same, reducing the unit of analysis of polygons to lines, and given that the need lies in offering control points, to be taken by the authorities as strategic places in a timely intervention, allows a priori efficient selection of space.

2.1 Study Area and Data

Based on a set of data provided by the Secretary of Public Security of the State of Aguascalientes, an exploratory analysis of the vector data was carried out to identify the concentrations of crime points, regarding the type of theft and the date on which they occur in the streets of the city, once reviewed.

The study area was delimited to the Metropolitan Zone of the City of Aguascalientes that corresponds to the municipality of Aguascalientes, since this is the area in which there are a greater number of crimes, this is how a vector layer of criminal incidence was generated, selecting and cutting the georeferenced points that were within the polygon of the City of Aguascalientes, which was taken the INEGI Geostatistical Framework of 2017 [10]. The total points (recorded robberies) for the entire State of Aguascalientes was 61,117, for the present study in the capital, the study area was 50,492 crimes, 86% of the incidents.

Although it was initially filtered by three types of theft, so that the process was manageable by a standard computer, it can be mentioned that the process had a good result in terms of information management being possible to include all categories of theft type being burglary, business, passerby, carrier, auto parts, vehicle, public transport, individual transport, recovered objects and others. This is how the information in the attribute table that each crime account has was maintained. The road network that was used was extracted from Open Street Map, as already mentioned the shape was cut to the City of Aguascalientes.

2.3 Topology Review

The established methodology goes from obtaining the street network, to editing its topology, improving network connectivity. Topology rules are applied to detect faults and thus solve them, then the rules that were stipulated to find errors between the lines of the network are presented:

- It must not cross: a line of a layer should not overlap lines of the same layer, any line that overlaps is an error, no correction is made.

- It should not hang: a line of a layer must touch lines of the same layer at both ends, any end point where the line does not touch another line is an error, here it was applied as an exception since when relating to lines that had nodes it was highlighted that they were final streets, and being a road network that was cut with the City polygon, mostly the errors were marked in said cut.

- It should not overlap itself: a line characteristic of a layer must not cross or overlap, any line where the characteristic overlaps is an error, no error is presented therefore no corrections are made.

- It should not be intercepted in itself: a line characteristic of a layer must not cut itself, any line where the characteristic overlaps itself or any point where the characteristic intersects is an error, no observation is presented therefore no There are modifications in this section.

- It must be a line without divisions: a line characteristic of a layer must not have more than one part, any line characteristic with more than one part is an error, 184 errors were identified, because the lines that are connected With polylines they do not have any separation, the tool provides the possibility of making each line independent with the option called “explode”, separating the polylines in segmented lines.

- You must not intercept or touch the interior: a line that intersects with other lines must generate nodes or divisions where they are touched, otherwise it is considered an error, this option found

the greatest number of errors and that in previous processes were Those who had generated problems at the time of calculating a density kernel, so it is recommended that this rule be analyzed carefully, 17,645 errors were presented, being solved with the option “divide”.

2.4 Density Estimation

The density estimation is the guide that was chosen to synthesize the amount of information and georeferenced points that the prosecutor of the State of Aguascalientes delivered, mainly because it is a large volume but mainly to interpret clusters and possible patterns in the crime of qualified theft. Subsequently, a search was made of network analysis tools in both free and proprietary software, in QGIS version 2.18 and ArcMap v. 10.2.

V.Kernel.Vector. In QGIS, a network analysis tool called “v.kernel.vector” developed by Okabe and other authors [10] was identified, which, based on a layer of points (crimes) and a layer of a network (streets) generates a vector file of criminal density. Assign a value to each street according to the crime incidence that occurs in each one, high values mean a high incidence of robberies and low values a low incidence of crimes. The tool is still in development and presents errors in its algorithms since the output files obtained give values of “0”, so it is unlikely to determine the most relevant streets with respect to others. As for the criminal incidence, it was decided to rule it out. In the same way the alternative of generating a kernel in Qgis (called heat map) was studied, it delivers a raster format. The objective was to transform the raster into a vector and the values that had been set would be assigned to the street network. Below are the settings that had been used:

- Cell size no larger than 1/5 of the image scale so, 1: 50,000 scales 10 meters.
- Search radius of 1 km of 564, 19 meters
- Square kilometer Area Unit -Enviroments -Processing extent: same as layer floors or the study layer
- $A = \pi * r^2$ $r = \sqrt{A} / \pi$
- 1 km area, radius = 0.56km
- Area of 0.333 km radius = 0.33 km
- Kernel density is calculated in ArcMap with the specifications previously named mainly 10 scale and radius of 564.19 (1km) meters, but it was also done on a smaller scale at 325.57 meters (0.3 km)
- The kernel from raster density to polygon is passed, where the zones of the different concentrations of thefts can be bounded.
- It is intended to cross the areas with greater incidence or kernel, in order to give hierarchical values to the network or streets, this to then make calculations of the application “location assignment” of Arcgis with hierarchy values and have greater relationship between camera crime zone

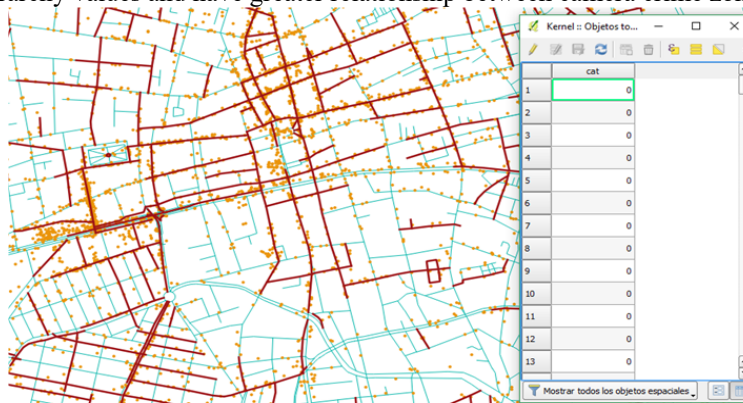


Illustration 1: Density kernel of the streets of the City of Aguascalientes, street network (blue), criminal occurrence (orange), kernel (red), density kernel attribute table shows values of 0.

Spatial Analysis Along Networks (SANET). On the other hand, using the ArcMap 10.2 software, a tool for network analysis called Spatial Analysis Along Networks (SANET) developed by Okabe and Sugihara [11] was identified, the installation file is necessary to download from the website (<http://sanet.csis.u-tokyo.ac.jp>). The tool developed for ArcGIS can also perform a kernel of density in networks from two vector layers (points and lines) so the same data was used. Therefore, once the files are loaded, the tool automatically detects the layers (streets and crimes), the estimation method is indicated: "equal continuous division to nodes" or "equal discontinuous division to nodes". The maximum distance distance of the points to the street and the distance of the street segments are defined as this will determine the value given to the street. The execution time of the algorithm will depend on the sample size and the assigned values of the bandwidth and cellwidth may take a few minutes or even several hours. Once the execution is finished, the output file is generated which shows the network of streets with different values defined by the number of crimes in that street.

SANET tool operation. The tool between different work functions with which it meets is the one used for this project, called "Kernel density estimation".

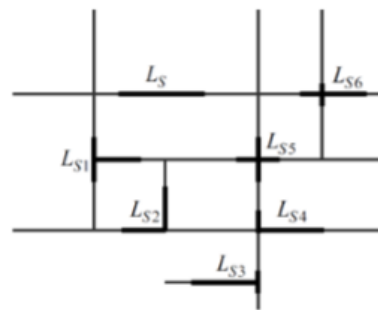
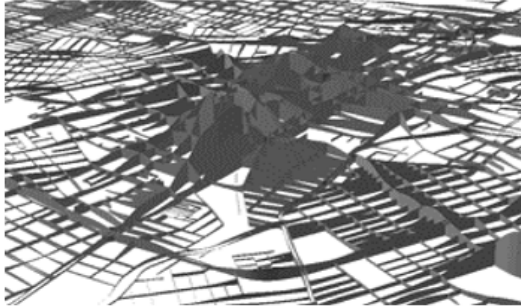


Illustration 2: 3D model generated in ArcScene for viewing areas with a high degree of crime. The authors Okabe, A. and K. Sugihara [11] describe the different ways of operating the application, among the most prominent parts is that: The calculation of a density probability is through a binomial point process, "the process of homogeneous point, the probability that a point is located in L_s is equal to that it is over L if for $i = 1, \dots, 6$ (thick line segments) because $ILSI = ILSiI$ being maintained for $i = 1, \dots, 6$." This is expressed by the equation:

$$\Pr[n(L_s) = k] = nC_k \left(\frac{L_s}{L}\right)^k \left(1 - \left(\frac{L_s}{L}\right)\right)^{n-k}$$

Equation 1: Okabe, A. y K. Sugihara [11].

The reference to the binomial point process due to the assumption of the same number of events or points with the same number of elements of k (the product being between the number of combinations C of k elements chosen from n elements, with the quotient of the length of the network line set and the length of the network raised ak by at least one of the quotient of the length of the network line set and the length of the network raised to nk). The process of the previous equation is called the homogeneous binomial point process. A binomial point process with the probability of non-given function is referenced as an inhomogeneous binomial point process.

The cell size (cellwidth) which is the measure used to develop the previous equation which aims to apply a hypothesis that the authors handle and under which they try to find the distribution of points on

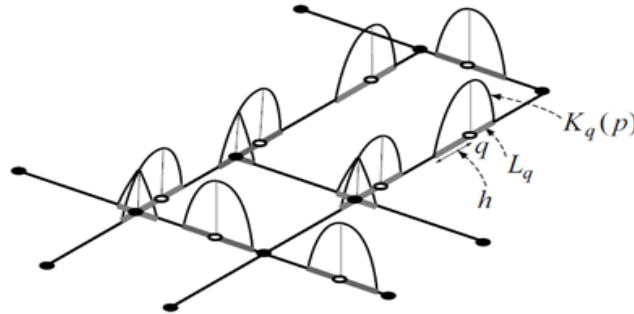
the network. Affirming that the points are generated by a homogeneous binomial point process corresponding to the hypothesis of complete spatial randomness (for its acronym in English CSR), otherwise the hypothesis is approved a non-homogeneous binomial process characterized by a function of non-uniform probability density. "For example, when we want to test the hypothesis that car accidents occur in proportion to the volume of traffic, we derive the probability density function of a traffic density function (a field function) over a road network." [11]

It is necessary to prove that a kernel density estimator is impartial, to which the author proposes to determine by measuring the probability density uniformly defined by the equation:

$$f(x) = \frac{1}{|\tilde{L}|} \text{ for } x \in \tilde{L}$$

Equation 2: Probability density; where \tilde{L} is the density probability function length called uniform distribution

The following shows the design of a network and the elements under which the SANET tool



calculates kernel density

Illustration 3: Density Kernel functions on a network established in a plane (the kernel used is indicated by the gray segments of thick lines) "Taken from Okabe, A. and K. Sugihara [11: p. 178].

$$K_q(p) \begin{cases} \geq 0 & \text{for } p \in L_q \\ = 0 & \text{for } p \in \tilde{L} \setminus L_q \end{cases}$$

$$\int_{p \in L_q} K_q(p) dp = 1$$

Equation 3: Function restrictions $K_q(p)$, taken from Okabe, A. and K. Sugihara [11: p.178]

In a non-directed network with a random point q a subnet L_q is constructed, this is how it is estimated that the shortest distance between any point in the subnet is less than or equal to h , which represents the width of the search buffer. All this is represented in the function $K_q(p)$. The function that builds the kernel density map is made using the following function.

Where \tilde{L} / L_q is the complement of L_q with respect to \tilde{L} , dp is the integration operator that symbolically indicates an infinitesimal line segment around p in the network \tilde{L} , and the integral over $p \in L_q$ is the integration of $K_q(p)dp$ along the line segments of L_q (ie a one-dimensional integral). Referring to $K_q(p)$ as a network kernel density function in q o KD network function, in short, q as the center of the kernel of $K_q(p)$, L_q as the kernel of $K_q(p)$ and h as the width of band [11: 178].

The process that analyzes the tool also takes into account the continuity of the networks or on the contrary that the calculation is carried out taking into account the interruption of the network as seen

appendix A. However, the process of each one is different and complex, so the process is invited by the authors[11].

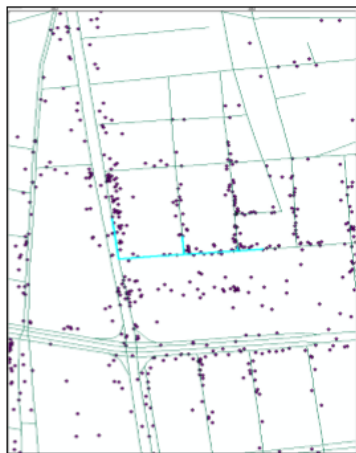
SANET test; Choice of method and configuration for density Kernel measurement.

Next, it can be seen that under the same conditions of point density and street network, the SANET tool estimates the density of points differently. Due to the type of configuration it can be performed continuously or discontinuously, the configuration is bandwidth 300 and cellwidth 30. The following observations stand out:

- The number of street divisions are the same, as well as the minimum and maximum possible density of crimes by segment, which is 0 and 350,711.
- The estimation of crimes by segment is different, in the configuration “kernel of density by equal discontinuous division” there is a tendency to classify more segments with a higher crime density, which generates more hotspot in the street network, taking as a point Starting with the fact that streets with a density of points greater than a standard deviation are chosen for case 1 of discontinuous kernel, it is estimated that the average of crimes with said deviation is 57.723, therefore segments that have equal or greater than that value and for that matter 2.
- With the continuous kernel it is presented that, with 1 standard deviation above the average, the number of crimes is 55.463, so that those that are equal to or greater than that value are chosen. For case 1 with a discontinuous kernel, a total number of 682 segments is presented, unlike the configuration of “kernel of density by equal continuous division” which has a total of 567 segments for the same crime example. Next, it can be distinguished that for case 1 there would be a greater number of streets chosen than for case 2, followed by the image an analysis of said difference will be presented.

The difference between the number of segments greater from case 1 to case 2 stands out that it is generated by a particular calculation for each case. Through the following image, it will be possible to identify that the discontinuous method for the same points generates a higher estimate, that is, for the same street they take into account a greater number of crimes that are clearly already calculated. Unlike the continuous method where the calculation of the crimes that are close to him is made. The calculation of discontinuous kernel, therefore, generates a bias that generates a count of points that have already been counted (See illustration 5).

Illustration
calculation
continuous



5: Density
for
kernel on

the left and discontinuous on the right

Faced with this characterization of the continuous and discontinuous types, it is taken that the best type of kernel configuration of the SANET application for the present study is the same continuous division density kernel.

Bandwidth and cell selection for SANET

- Case 1: bandwidth of 100 and a cellwidth of 10.
- Case 2: bandwidth of 200 and a cellwidth of 20.
- Case 3: bandwidth of 300 and a cellwidth of 30.

Standard deviation of the Average column (average points per street segment)

Case 1: Standard deviation 14.596, mean 7.239, a standard deviation above the average 21.835. Of 18,800 segments made by the SANET tool to cut the street network of the City of Aguascalientes. The attribute table was filtered, from the average column, choosing the segments that were above a standard deviation for which a total of 1,053 segment is presented.

Case 2: Standard deviation 25.276, mean 14.668, a standard deviation above the average 39.944. As for the second option of filtered segments above a standard deviation, a total of 697 segments are found.

Case 3: Standard deviation 33.962, average 21.500, a standard deviation above the average 55.463. For the latter case, 567 segments in total are obtained with the same standard deviation characteristics.

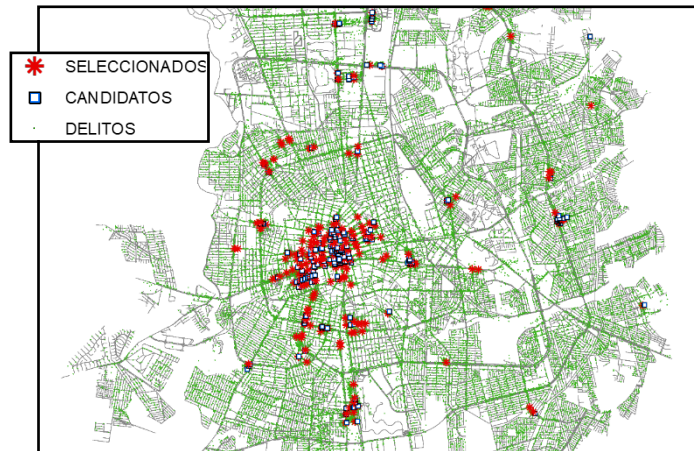
As you can see the density of points is variable depending on the bandwidth or cellwidth configuration clearly the increase is proportional, three configurations were taken one above that recommended by the author and the other by below. Therefore, a very high bandwidth and cell configuration will generate a wider count in the network, which implies that points already detected by another network are counted. Unlike if a very small bandwidth is established where points that are far from the network cannot be detected, it will clearly affect the result. It is considered appropriate to take the suggestion of the author of maintaining an intermediate point, with a bandwidth of 200 and a cell width of 20 where it does not take points very far from its place nor that it is not possible to detect those that are outside the network or street.

2.2 Control Point Selection (*location allocation*)

Those roads that were above a standard deviation of the average of theft qualified were extracted in a vector file, later processed for the assignment of nodes at the intersections, beginnings and ends of the network, that would be the input for the selection of points of strategic control in the opportune intervention of the authorities, which is regulated with the necessary number of cameras, the impedance in length of visualization and the greater allowed coverage of crime points, thus the ideal places for the installation of surveillance cameras would be designated or smart monitoring points [12].

The filter between the still important number of nodes is done by means of the Location Allocation extension (tool) of the Network Analyst compendium of the ArcGIS software. First, the location of the candidate nodes to control points that were chosen by SANET is presented, then the entire density of georeferenced points for theft, and through the configuration of facilities to plaintiffs (control points to theft points) Run the program in search of the 200 control points with the highest coverage capacity in a range of 200 meters (impedance). Below is part of the result:

Illustration 6: result of the network analysis methodology



3 Discussion

As a result, the Location Allocation tool mentions that with the established configuration of 200 cameras and 200 meters of impedance the total points found is 9,264 of the total of 50,492 (remember that these are the crimes concentrated by the prosecution from 2011 to 2017 of 10 different categories), this is how you have a total coverage of 18.35%. It is relevant to mention that this project is based on a historical analysis and that the objective is not necessarily to reach a high number of points, but to find the location of the densest places in crime rates.

4 Conclusion

The City of Aguascalientes, despite being an industrial city and generating sources of employment, presents a large number of criminal events. Grouping the events of crime in areas where there is a constant flow of people such as: the historic center and the esplanade of the San Marcos Fair as well as other areas where the same type of environment probably occurs. The proposed density kernel with the values that were determined with bandwidth and cellwidth (200-20) yielded the best model, however, the results are representative for the study area and for the crime of theft data. Therefore, the crime incidence can only be explained at this level with the study data, so the existence of crime incidence patterns must be determined from a temporary analysis that contemplates changes in population flows as can be the San Marcos Fair.

The location or location methods such as the location allocation tool certainly help to solve or generate alternatives for the possible location of candidate sites for the placement of surveillance cameras. Network analysis can solve a variety of urban problems in a city. To plan cities in a better way in which society is benefited and the impact generated is positive

5 Acknowledgment

We thank the CentroGeo and CONACYT center for opening the community doors in CIG, through the scholarship in the specialty in geomatics, as well as teachers Rodrigo Tapia and Ana Alegre for allowing our participation in the project carried out with the Government of Aguascalientes.

APPENDIX A

$$K_q(p) = \begin{cases} \frac{k(d_S(q,p))}{(n_{11}-1)(n_{12}-1) \cdots (n_{k-1}-1)} & \text{for } d_S(q, v_{k-1}) \leq d_S(q,p) < d_S(q, v_k), \\ 0 & \text{for } d_S(q,p) \geq h, \end{cases} \quad k = 1, \dots, m, \quad (9.8)$$

$$K_q(p) = \begin{cases} k(d_S(q,p)) & \text{for } 0 \leq d_S(q,p) < 2d_S(q, v_1) - h, \\ k(d_S(q,p)) - \frac{n-2}{n_1} k(2d_S(q, v_1) - d_S(q,p)) & \text{for } 2d_S(q, v_1) - h \leq d_S(q,p) < d_S(q, v_1), \\ \frac{2}{n_1} k(d_S(q,p)) & \text{for } d_S(q, v_1) \leq d_S(q,p) < h. \end{cases}$$

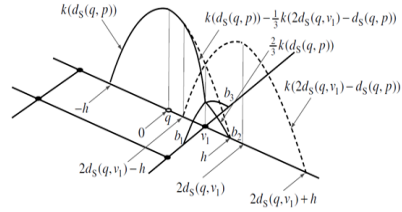
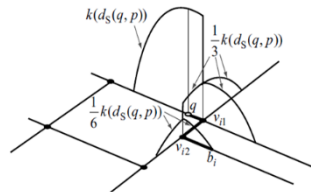


Illustration 4: Discontinuous and continuous analysis of the tool respectively

References

1. Curtin, K. (2007). Network Analysis in Geographic Information Science: Review, Assessment, and Projections. *Cartography and Geographic Information Science*, Vol. 34 (2) pp. 103-111 (2007) pp. 103-111. <https://pdfs.semanticscholar.org/45bb/1558bf4bdd0baf027c2d145ce97f96a33d81.pdf>
2. ESRI. (2017). Cómo funciona la densidad kernel. Recuperado de: <https://pro.arcgis.com/es/pro-app/tool-reference/spatial-analyst/how-kernel-density-works.htm>
3. ESRI. (2017). Comprender el análisis de densidad. Recuperado de: <http://desktop.arcgis.com/es/arcmap/10.3/tools/spatial-analyst-toolbox/understanding-density-analysis.htm>
4. ESRI. (2017). Densidad de Kernel. Recuperado de: <https://pro.arcgis.com/es/pro-app/tool-reference/spatial-analyst/kernel-density.htm>
5. ESRI. (2017). Entornos de análisis y Spatial Analyst. Recuperado de: <https://pro.arcgis.com/es/pro-app/tool-reference/spatial-analyst/analysis-environments-and-spatial-analyst.htm>
6. ESRI. (2017). Reglas de topología y arreglo de errores de topología. Recuperado de: <http://desktop.arcgis.com/es/arcmap/10.3/manage-data/editing-topology/geodatabase-topology-rules-and-topology-error-fixes.htm#>
7. Gobierno del Estado de Aguascalientes (2017) Plan Estatal de Desarrollo 2016 2022. México.
8. Instituto Nacional de Estadística y Geografía (INEGI) 2016. Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública. Recuperado de: <http://www.beta.inegi.org.mx/temas/incidencia/>
9. Jaitman, Laura y Ajzenman, Nicolas. (2014). Crime Concentration and Hot Spot Dynamics in Latin America IDB Working papers N° IDB-WP-699 Inter-American Development Bank. <https://publications.iadb.org/bitstream/handle/11319/7702/Crime-Concentration-and-Hot-Spot-Dynamics-in-Latin-America.pdf?sequence=1&isAllowed=y>
10. Okabe, A., Satoh, T., Sugihara, K. A kernel density estimation method for networks, its computational method and a GIS-based tool. *International Journal of Geographical Information Science* Vol. 23 (1), 7–32 (2009)
11. Okabe, A. and K. Sugihara (2012) *Spatial Analysis Along Networks: Statistical and Computational Methods*, Chichester: JonWiley, a volume in the Wiley series of Statistics in Practice.
12. Secretaría de Gobernación. (2016) Norma Técnica para Estandarizar las características técnicas y de interoperabilidad de los sistemas de ideovigilancia para la seguridad pública. Recuperado de: http://www.secretariadoejecutivo.gob.mx/docs/pdfs/consejo/Norma_tecnica_sistemas_video_vigilancia.pdf