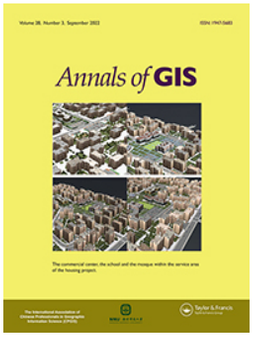

Annals of GIS

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tagi20

Crime beyond the edge: development of a tool to correct the edge effect on crime count

Daniel Salafranca Barreda, Diego J. Maldonado-Guzman & Patricia Saldaña-Taboada

To cite this article: Daniel Salafranca Barreda, Diego J. Maldonado-Guzman & Patricia Saldaña-Taboada (2022) Crime beyond the edge: development of a tool to correct the edge effect on crime count, Annals of GIS, 28:3, 279-292, DOI: 10.1080/19475683.2022.2052748

To link to this article: https://doi.org/10.1080/19475683.2022.2052748

Published online: 20 Mar 2022.

ANNALS OF GIS
2022, VOL. 28, NO. 3, 279–292
https://doi.org/10.1080/19475683.2022.2052748

Taylor & Francis
Taylor & Francis Group

Crime beyond the edge: development of a tool to correct the edge effect on crime count

Daniel Salafranca Barreda[a], Diego J. Maldonado-Guzman[b] and Patricia Saldaña-Taboada[c]

[a]Crímina Center, University Miguel Hernández, Elche, Spain; [b]Department of International Public, Criminal and Procedural Law, University of Cádiz, Cadiz, Spain; [c]Department of Criminal Law, University of Granada, Granada, Spain

ABSTRACT

The edge effect is a problem that can alter the results of some analyses, such as counting crime within a given geographic area. This article introduces a tool developed for *ArcGIS* toolbox, (*ArcGIS Geographic Information System)* to correct the border issues when using an aggregated crime data to artificially bounded space analytical units. It uses a method which considers those points located near the edge of the analysis unit, and avoids increasing the number of criminal points by assigning a value according to the distance of the edge. For this purpose, two functions based on decay with distance can be chosen: normal and linear. In order to show the performance of the tool, a sample of theft data occurred in 2016 in each census tract of Barcelona (Spain) district was used. These results show remarkable differences in the number of thefts in each census tract, before and after applying the edge correction. Some of the census tracts even went from experiencing no theft at all to having 5.5 or 4.5 incidents after correcting the edge effect. Finally, to demonstrate the benefits of the proposed tool, other strategies traditionally used as a solution for the edge effect were used. Then, the results are compared with those previously obtained.

ARTICLE HISTORY
Received 10 March 2021
Accepted 1 March 2022

KEYWORDS
Edge effect; *ArcGIS* toolbox; aggregated crime data; artificially bounded space; counting crime

1 Introduction

The disciplines responsible for the study of crime recognize that the criminal phenomenon can be understood and explained much more adequately if its geographical component is taken into consideration (Galdón Clavell and Pybus Oliveras 2011). Technological improvement and the development of Geographic Information Systems (GIS), from the 1990s, turned crime mapping and bounded space analysis into much simpler tasks that have improved over the years until today. In addition, the development of a set of integrated theories within environmental criminology, provided the crime mapping with a theoretical underpinning, explaining the results obtained from analysing the geographical and temporal component of crime.

In geographic crime analysis, point mapping is the most common way of depicting incidents. However, although this information is more precise regarding the exact location where the crime occurred, sometimes the crime analyst may have an interest or need to represent the data for an area in a summarized form (Harries 1999). In the latter case, the maps are obtained by aggregating the information to specific geographical units (municipalities, districts, neighbourhoods, census tracts, etc.).

For example, a quite usual type of cartographic representation of aggregated data are choropleth maps. Although their use for communicating the distribution of spatial phenomena has been extremely popular (Wei et al. 2017), it has also been the most misinterpreted and incorrectly produced type of map, of all the maps used most frequently (O'Sullivan and Unwin 2010). Though, applied to crime analysis, these maps represent, in a highly visual way, the amount of crime concentrated in each unit of analysis (a neighbourhood, for example); those units are not natural, as their boundaries are designed for administrative, policing, or political purposes.

Thus, a series of problems arise from using study areas whose boundaries are artificial, such as the so-called edge problem (Murray et al. 2001), which is one of the limitations to be considered in the ecological study of crime. The edge effect is present in those investigations that do not have the capacity to consider behaviours, objects or events that occur outside the studied geographic region, that is, beyond the limits of that region (Rengert and Lockwood 2009, 117). This problem would appear, according to Cruz Rot (2006), when points that are located outside the limits of the study area are not considered to estimate the characteristics of a point process. Therefore, it is suggested that

CONTACT Diego J. Maldonado-Guzman diego.maldonado@uca.es Avenida de la Universidad s/n. 11406. Jerez de la Frontera (Cádiz), Cadiz, Spain

this problem has to be addressed, since ignoring its effects means omitting that any geographic distribution or spatial interaction occurring within the unit of analysis, may extend beyond its boundaries (Gao et al. 2017).

In this sense, it is a serious mistake, in the geographic analysis of crime, to assume, for example, that crime hotspots (drug sales points, gas stations or nightclubs, for instance) located near the artificial boundary that separates one neighbourhood from another, will only have a criminogenic effect in the neighbourhood where they are located, without extending that effect beyond the boundary that separates that neighbourhood from the adjacent one.

Although the edge problem is present at all scales of spatial analysis, its consequences are accentuated when small geographic units, such as census tracts, are used, as the probability of points being located close to the axis is considerably higher. In the context of the geographical study of crime, these consequences can be classified into two types: the effects on some spatial crime analysis methods and the consequences on the estimates of incident numbers in each area.

Regarding the first type, the border issue can alter the results of methods that analyse crime concentration patterns. This is because if points located outside the boundaries of the studied area are excluded from the analysis, a series of events are ignored preventing the real distribution of the phenomenon from being recorded (Rengert and Lockwood 2009). The border issue also alters the calculations of the nearest neighbourhood index, so all statistical tests that employ such an index in their calculations will be equally sensitive to the edge effect. For example, in geographic profiling of crime, the CGT formula proposed by Rossmo (2000) is composed of the parameter B, which is calculated with half the average of the nearest neighbour index. Thus, the geoprofiling results in this case are sensitive to the edge effect.

Concerning the second type, the edge effect alters the results of operations to count the number of points located within each unit of analysis (e.g. a census tract). This may have negative repercussions in studies that use the number of crimes within an area as a variable, since the calculation of that number may be biased. Despite the fact that the polygon point counting technique is the most commonly used (Murray et al. 2001), such operation, available in several GIS such as ArcGIS or QGis, ignore the effect that a crime event occurring near the edge (or even on the edge), separating one unit of analysis from another, might have on the adjacent unit (Zhang, Suresh, and Qiu 2012). Thus, when a layer of crime points is aggregated to a layer, for example, of census tracts, such aggregation considers the point as an event occurring in a single location (the unit of analysis in question delimited by the edges). In this way, crime scenes occurring any closer to the border of one unit with another are considered to be exclusive to one unit. Figure 1 exemplifies this issue.

Aware of the consequences of the edge effect, several authors have proposed a number of possible solutions. Examples are: the use of Kernel density maps as an alternative to other edge-sensitive methods (Ratcliffe 2010), the use of buffer zones around the study areas (Rengert, Ratcliffe, and Chakravorty 2005; McCord and Ratcliffe 2007; Zhang, Suresh, and Qiu 2012), the use of straight lines with edge sections to know if points are located within the polygon (Olaya 2010), the use of simple methods based on omitting those points close to the boundary of one unit with another (Rengert and Lockwood 2009), or more complex solutions such as weighting points close to the edge (Ripley 1988). A more detailed explanation of some of these and other solutions can be found in Cruz Rot (2006) and in Goreaud and Pélissier (1999)

However, while attempts have been made to overcome the edge problem, the above proposals also have several limitations. For example, using kernel density as a solution does not allow to correct the counting incidents problems in choropleth maps, as kernel density does not give as an output a value to be aggregated to a spatial unity. For its part, the use of buffer zones is not a solution to the edge problem per se, since the real need is for the data to be spatially continuous (Rengert

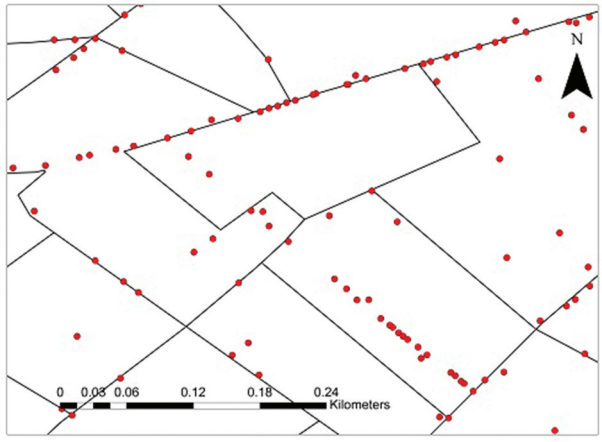

Figure 1. The image shows how the incidents are quite close to the border with the other neighbourhood, a border that does not really exist in the geographic space. For this reason, it is necessary to consider that each of these crimes has also occurred in some way in the adjacent neighbourhood, since the distance that separates them is minimal, and because such boundaries do not exist. As a consequence, the crime that occurs near the border will have some effect on the adjacent neighbourhood. Own elaboration.

and Lockwood 2009). In addition, this approach greatly increases the actual number of crimes that have occurred, as all incidents within the buffer are recounted in adjacent units. Figure 2 exemplifies this method to illustrate this limitation. On the other hand, the proposal to ignore incidents close to the edge produces a loss of information, which will be more serious when the number of points to be analysed is small. Finally, the use of straight lines cutting the axis is problematic when the ray traced touches the contour of the polygon or when the point is located right on the edge of the analytical unit.

Due to the consequences of the edge effect on geographical analyses in general and on spatial studies of crime in particular, and given the limitations of the suggested solutions, this article proposes a fresh method for correcting the edge effect that has been incorporated into the tool developed for the ArcGIS software. This tool makes it possible to automate the method presented. The process consists of duplicating the points close to the axis and weighting them using two possible decay functions with distance. This method overcomes the problems in obtaining the number of crimes in artificially delimited geographic units, since it allows computing the same crimes in two different adjacent units without increasing the real number of crimes. Likewise, the proposed method can be applied at any scale regardless of the analytical unit. It is also capable of detecting the limits of the study region, and its application is simplified by the creation of an ArcGIS tool that is available to the reader.

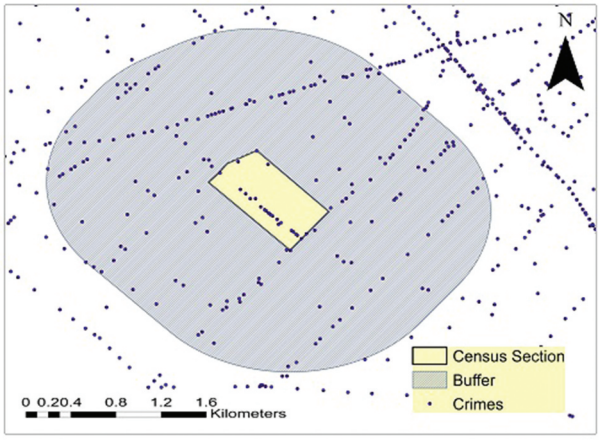

Figure 2. Example of the buffer zone method. The central polygon corresponds to the census tract in question. A buffer of 250 metres is created around the polygon (shaded area). The number of crimes occurring within the census tract is 37. However, in the adjacent area created with the buffer, 410 points are located. Therefore, these points at 250 metres from the axis are considered to have occurred within the census section, and so, instead of 37 crimes, 447 are considered to have occurred within the unit of analysis. Own elaboration.

The article is structured as follows: first, the proposed method to correct the edge problem is described, and the tool developed to automate all the steps described is introduced. Next, the results obtained before and after applying the proposed correction on theft crimes in the Ciutat Vella district of the city of Barcelona (Spain) are explained. Simultaneously, these results are contrasted with those derived from applying as a solution a Kernel Density Estimation (KDE) and the creation of buffer zones. The last section corresponds to a discussion of the method's contributions and a proposal for future research.

2. Method

2.1. Edge effect correction

The proposed method solves the edge effect by considering two factors. The first is the distance of the points to the axis, for which a distance in metres (to be chosen by the analyst) must be determined between each crime point and the edge of the analytical unit where that point is located. In order to select the best distance by which all points located below it are doubled, we suggest using the standard deviation (SD) of the distance between all the points in the sample and the nearest edge. This is similar to the nearest neighbourhood index, but instead of using the nearest average distance between all the points in the sample, the distance between all the points and their nearest edge is employed. The SD criteria is used to calculate the bandwidth size in a fixed kernel (Levine 2010). Thus, once the SD of the points-to-edge distances have been calculated, we suggest using this value as a distance criterion to select the points to be doubled. This distance-chose criteria allows to correct the border issue in any scale regardless of the size of the unit of analysis.

The criterions used to determine the best distance as a threshold when accounting the edge effect seem not to be based on any theoretical or statistic support. For example, Ewers and Didham (2006) shows different ways used to establish the critical percentage in order to determine the extension of the edge effect. The cited authors warn that no justification is given to select each of this criterions. Probably the way in which the optimal distance is determined is adapted to each solution purposed. As an example, Zhang, Suresh, and Qiu (2012) establish the buffer area size in two miles because that is the length of two inner city blocks, that is, the spatial unit of analysis they employ in their research. For its part,

the circumference-based Ripley's solution for the edge effect takes the distance between one point *i* and its nearest neighbour point to draw a circumference round the point *i*, using the proportion of the circumference within the study region as the weighting factor.

In our case, we choose the SD as a distance criterion, since it allows the analyst to take into account the average distance between all the points in the sample and their nearest edge. The SD has been selected over other possibilities (i.e. the mean) in order to deal with the extreme values.

Choosing an example of 100 metres (i.e. supposing that SD = 100), all crime events located within this distance will be duplicated to the adjacent unit of analysis (Figure 3). In case of one point is less than 100 metres from several unit of analysis, it would be doubled to the nearest one (in terms of distance).

Using only the first factor to fix the edge effect would result in an increase in the actual number of crime incidents, since each point near the axis is doubled according to the distance determined by the analyst. This means that if there are 1000 crimes in the original dataset and 300 of them are close to the edge (below the chosen distance), the total number of crimes would be 1300. Therefore, in order to correct this problem, the second factor indicated, i.e. the decay function for the duplicate point weighting, must also be taken into account. This procedure ensures that each duplicated point does not have a value of one, instead, this value is the result of a weighting estimated according to the distance of the point from the axis.

In this way, continuing with the previous example, if the original data set contains 1000 crimes and 300 are close to the edge (below the established distance), the total number of points is 1300, but the sum of all of them will be 1000 (the original number of criminal events in the sample). Despite Bailey and Gatrell (1995) claim that the choice of the weighting algorithm is not a crucial decision in obtaining the results, we have included the possibility of choosing between two functions for weighting the duplicate points that the method offers: the linear function and the normal function. The first is the simplest and implies that the value assigned to the point is constantly reduced as the distance from the point to the axis increases. Thus, the maximum value is located at the point and decreases steadily to a value of zero. On the other hand, the normal function is a curve (see Figure 4) in which the points obtain a smaller value as the distance to the edge decreases in an unequal manner.

Nevertheless, several adjustments have been made to the two weighting functions. The proposed method has inverted the two functions so that the value of the point decreases as the distance from the axis decreases (Figure 5). The reason for this inversion is that the effect of crime on the original analytical unit decreases as the point approaches the edge. It is then hypothesized that placing the point at the centre of the unit of analysis is not the same as placing it on the border with the adjacent unit. In the second case, the effect is distributed between both units.

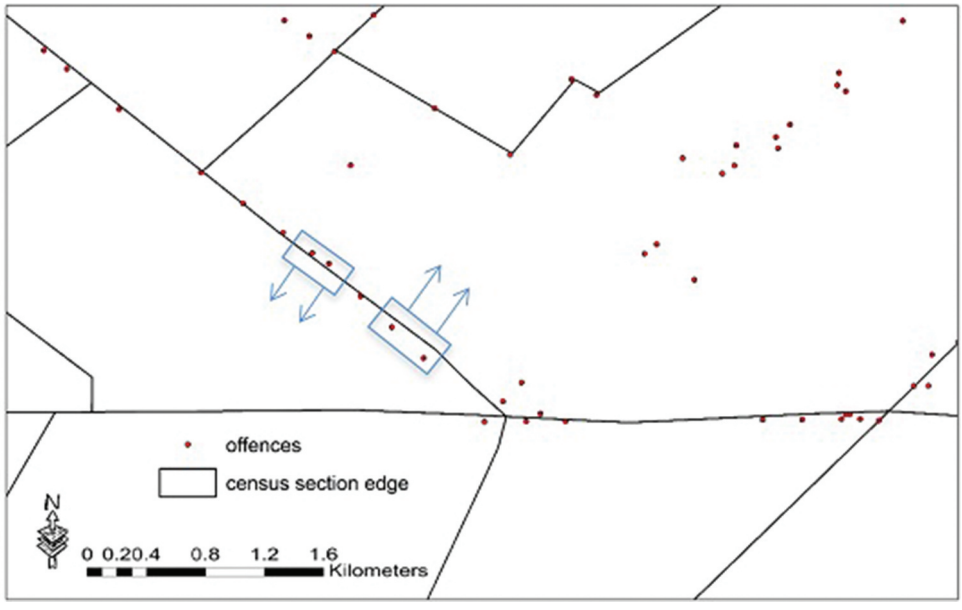

Figure 3. Example of point mirroring to the adjacent unit of analysis. The points within the box are less than 100 metres from the edge, so they are duplicated in the adjacent census tract. Own elaboration.

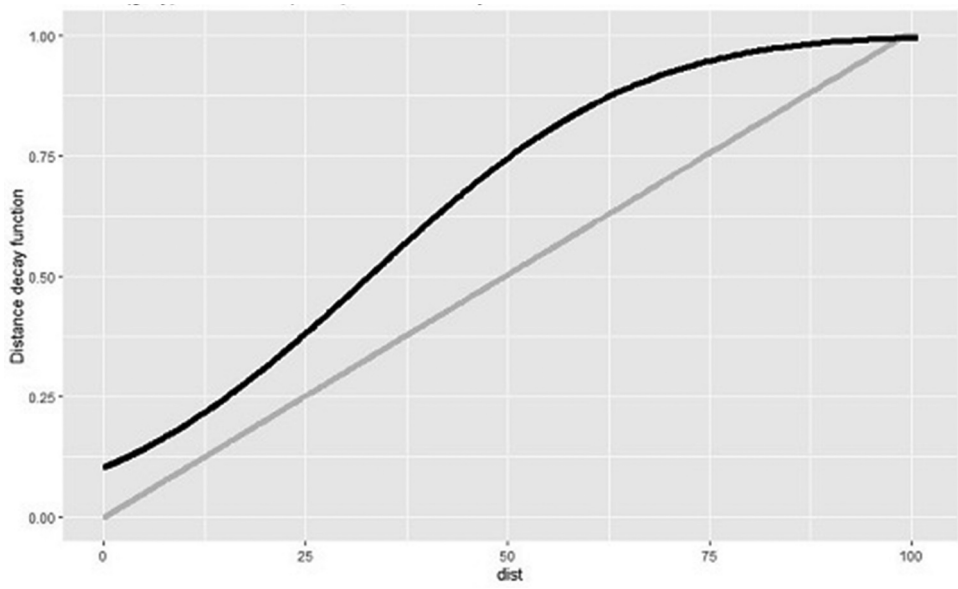

Figure 4. Linear (grey) and normal (black) distance decay functions. Own elaboration.

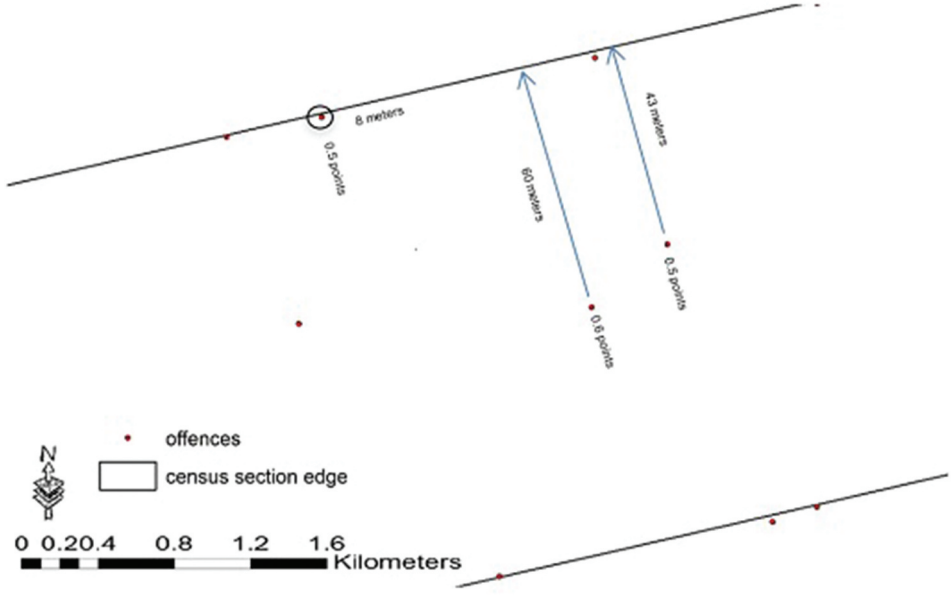

Figure 5. Point weighting using the linear function. With a set distance of 100 metres, all points within 50 metres of the edge will have a value of 0.5. Above 50 metres, the values of each point will depend on the distance to the axis. Own elaboration.

In addition to inverting the function, another adjustment has been made so that all points below half the selected distance have the same value (0.5) in both units of analysis (McCord and Ratcliffe 2009). This prevents the possibility of the original point having a lower value than its duplicate counterpart. For example, if a distance of 100 metres and a linear function have been entered, all points below 50 metres (half of 100) will have a value of 0.5 (both the original point and the duplicate point). In the case of points above half the set distance (but not exceeding it), the value of those is determined by the distance to the axis. For example, with a set distance of 100 metres, a point located 60 metres from the edge will have a value of 0.6 (and its duplicate counterpart will be worth 0.4). For a point located 75 metres from the edge will have a value of 0.75 (and its duplicate counterpart will be worth 0.25 points). Figure 6 shows an example with a distance set by the analyst of 100 metres.

As specified above, the weighting functions serve to solve the problem of multiplying the number of original crimes. However, there may be another limitation when, beyond the border of a unit of analysis (census tracts in

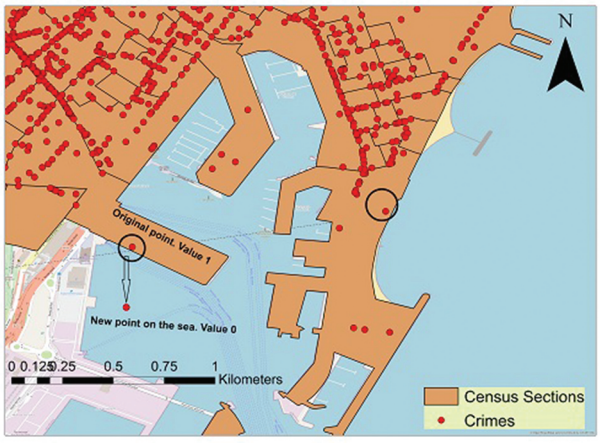

Figure 6. The points near the edges where there is no unit of analysis on the other side are duplicated, but do not lose their original value of one (1). Own elaboration.

this paper), there is no region under study. The method proposed here is designed to detect this type of situation. This is especially important when the sea is beyond the axis. Regarding the case of the study region in this article (see Figure 7), some census tracts in the Barceloneta quarter are surrounded by the sea. Here, the points close to the edge would lose their true value (value = 1), which would not be correct, since, although they are very close to the edge, there are no adjacent areas where they could deploy their effect. For this reason, even if the proposed method doubles the points, in the cases just described the duplicated point over the sea would have a value of zero, so the original point still retains its initial value, i.e. a value of one (Figure 7).

The tool detects the end of the studied region by noticing that beyond the edge there is not another polygon, i.e. that there is not a continuous surface. So, when beyond the edge there is the sea or another obstacle which interrupts the continuity (or the end) of the studied area, the points near this edge will not be considered for being doubled. As the tool proposed in this paper is designed to correct the border issue when counting points in aggregated spatial units (polygons), the tool has been programmed to detect the end of the studied area (artificial areas), not to detect discontinuities in the physical space (natural areas). This is because it is not important where the point is exactly located, but ensuring that it is actually located inside the given polygon.

It is also important to point that the tool always doubles the point to the nearest adjacent area, i.e. to the nearest neighbouring unity of analysis. This way of working make it necessary to think on two very strange but possible situations. Firstly, when analysing data aggregated to very small size areal unities, one point could be near to more than one contiguous area, but the point will be only doubled to the next one (these which is nearer the original area which contains the point close to the edge). Secondly, although highly difficult, one point could be located at the same distance (within

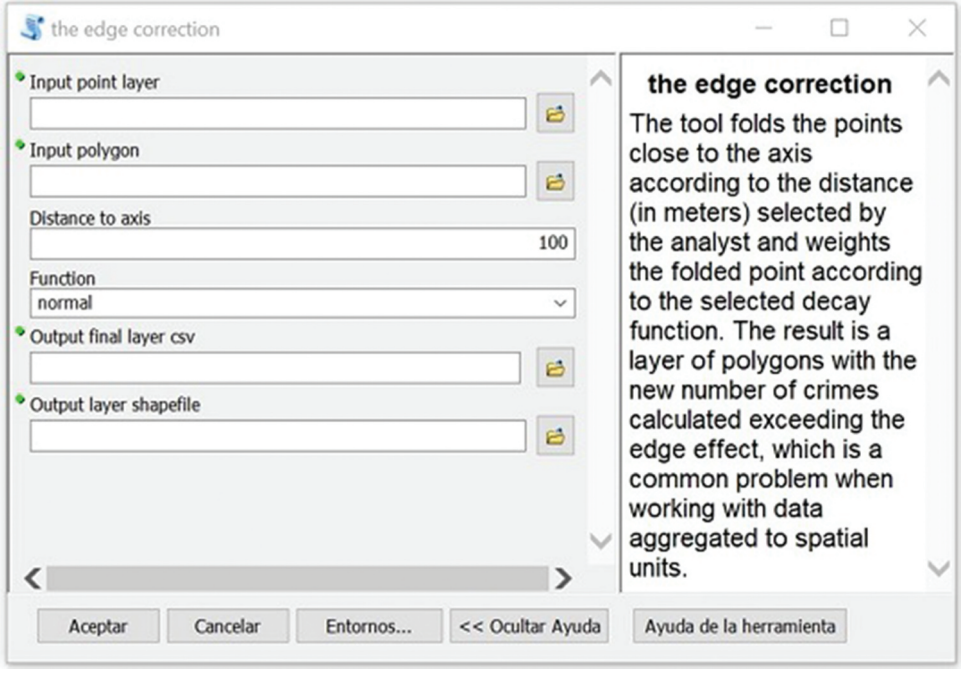

Figure 7. Interface of the tool "the edge correction" developed for the correction of the edge effect.

that specified by the analyst) from not only one but several neighbouring areas. In this case, the tool will not double this point to any of the four areas, but will give the original point the value of 0.5.

2.2. Development of an ArcGIS toolbox to correct the edge effects

The method described so far can be complex if performed manually. Therefore, we have developed a tool for ArcGIS that automates the whole procedure and allows us to obtain the number of offences in each unit with the correction of the edge effect (Figure 8). More specifically, a script has been developed in R language that will be incorporated into ArcGIS using R-Integration (Appendix 1). The final tool as well as the test data and the installation manual are freely available.[1] The obtained result is a file in shapefile format containing the number of offences in each unit after correction of the edge effect.

The data used to show the edge effect correction are the theft crimes occurred in 2016, in the census tracts that constitute the Ciutat Vella district of the city of Barcelona (Spain). The file with the georeferenced information of the thefts was obtained through the Department of Interior of the Generalitat de Catalunya,[2] and refer to all theft crimes occurred in the public streets of Barcelona that were recorded by Mossos d'Esquadra. The coordinates contained in the file were geographical (longitude and latitude).

In order to obtain the locations of the criminal events that occurred only in the Ciutat Vella district, a definition by query was made. As a result, the total number of thefts committed in 2016 was of 6558 crimes in Ciutat Vella (District 1). For its part, the layer of polygons corresponding to the census sections was obtained through the Cartographic and Geological Institute of Catalonia. To obtain the census sections belonging to the district under study, a definition was made by query. However, some census sections belonging to districts number 2, 3 and 10 were included. This operation was carried out so that the incidents located on the edge of the Ciutat Vella census sections, that had no more adjacent sections, were considered for the methodology and allowing the duplication of the point (Figure 9).

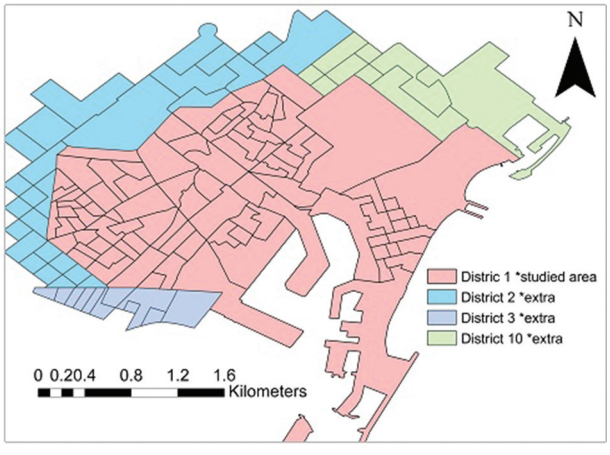

Figure 8. Census sections of the Ciutat Vella district together with some extra census sections of the adjacent districts. The sum of both layers constitutes the study area in this article. Own elaboration.

Before showing how the tool has been developed and the results of applying it, a guide with several steps is shown to facilitate using the tool and to remember the two decisions the analyst has to make. As is detailed in steps two and three in Figure 6, the analyst has to decide two parameters: (i) what points near the edge wants to be doubled (choosing the distance to the axis) and (ii) the distance-decay function to weight the value of the points (choosing between a normal o lineal function).

4. Results

In order to demonstrate the benefits of the proposed method, the results obtained before and after applying the edge effect correction are compared in Figures 10 - 12. The first one shows the number of theft crimes for each census tract using the traditional 'points-in-polygons' method, while Figures 11 and 12 show the results of correcting the edge effect using the linear and normal functions, respectively. ArcGIS Desktop 10.4.1 GIS was used for this purpose. As the SD of the average distance between all crimes and the edge of the studied region is 48, we double all points situated at or below 48 metres from the neighbourhood edges.

Figure 10 shows the results of applying the traditional operation that counts the number of points within each polygon. In this case, the number of crimes is integer and the census tract with the highest concentration of crime has a total of 1185 thefts. The following map (Figure 10) shows the results of applying the edge correction using the linear decay function.

It can be confirmed that, due to the weighting effect of distance, the number of crimes after applying the edge correction becomes decimal. In addition, the census tract that in Figure 6 had the highest concentration of crimes with a total of 1185 thefts now yield a result of 1207.26 crimes. Figure 11 shows the results of applying the edge correction using the normal type of decay function. Again, the results are decimal numbers. The census tract that in Figure 10 had a higher crime concentration with a total of 118 thefts, now yields for the normal function a result of 1211.80 crimes. This is a consequence of incorporating the points too close to the edge of the other adjacent census tracts.

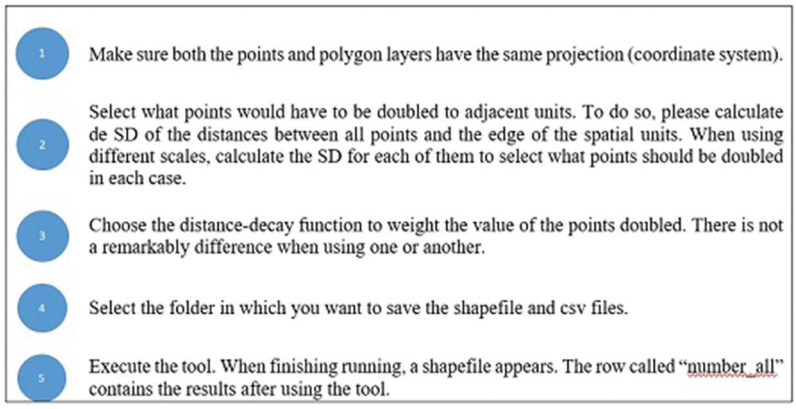

Figure 9. Steps to be followed to select the parameters of the tool and run it.

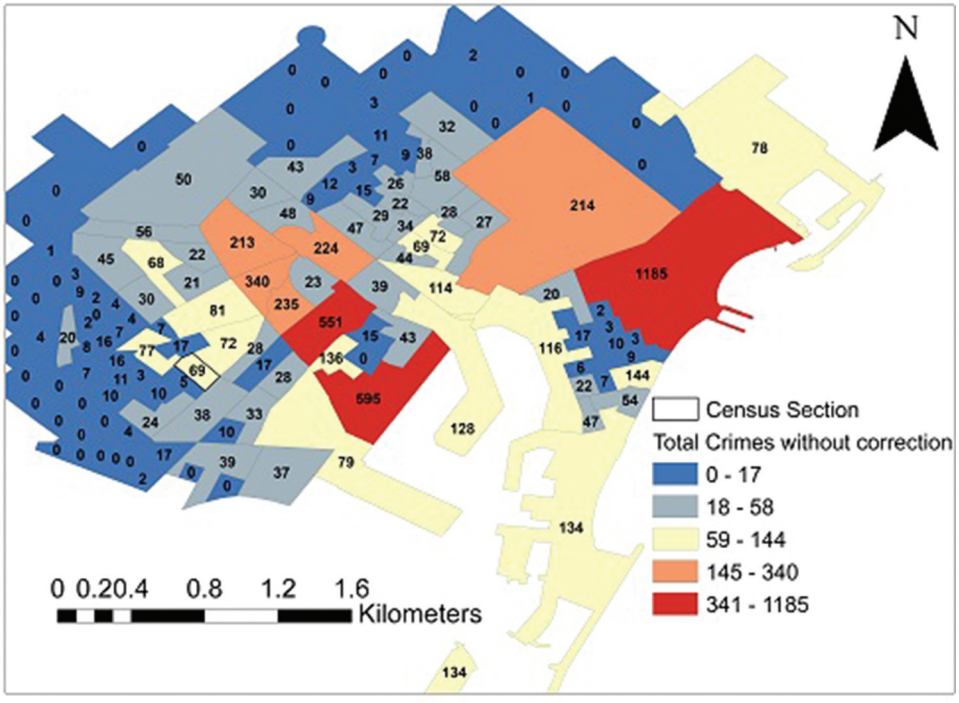

Figure 10. Number of crimes per census tract without edge effect correction. Own elaboration using ArcGIS 10.4.1.

To facilitate the comparison of the results, Table 1 shows the number of crimes for some of the census tracts in the study region. Those that represent the different possible outcomes have been chosen: census tracts in which the number of incidents increased after applying the correction, census tracts in which the number of incidents decreased after the correction, and census tracts that went from experiencing no crime to having several crimes.

The Table 1 shows how several census tracts experience a substantial change in the number of thefts after applying the edge correction. For example, census tract D had a total of 79 crimes after performing the polygon point count operation without correcting for the edge problem. After running the tool, the number of crimes for this census tract increased to 372.14 incidents for the linear function, and to 240.93 crimes for the normal function. The opposite case can be observed in census tract A, where the volume of crimes decreased considerably (222.9 less crimes for the linear function and 167.13 less crimes for the normal function). At the same time, section F, which initially contained no incidents, now has 4.5 crimes both for the linear and normal functions.

In order to compare the results obtained by the tool with those generated by using a KDE and the buffer-based solutions, Figures 13 -16 are shown together with Table 2. It should be recalled that the buffer-based

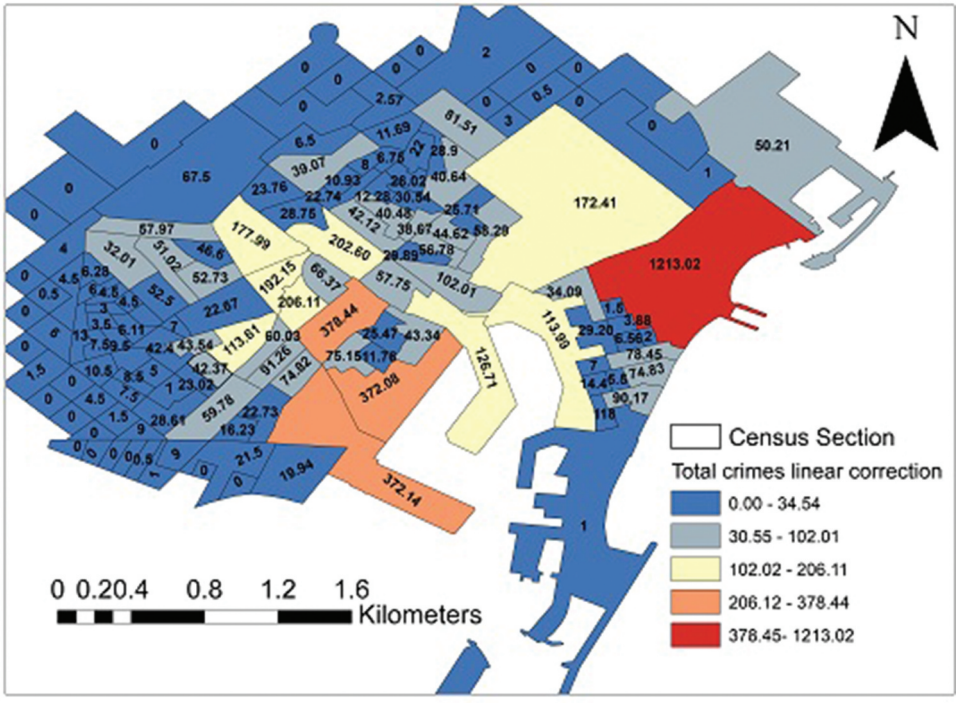

Figure 11. Number of crimes by census tract after applying the edge effect correction with the linear function. Own elaboration using ArcGIS 10.4.1.

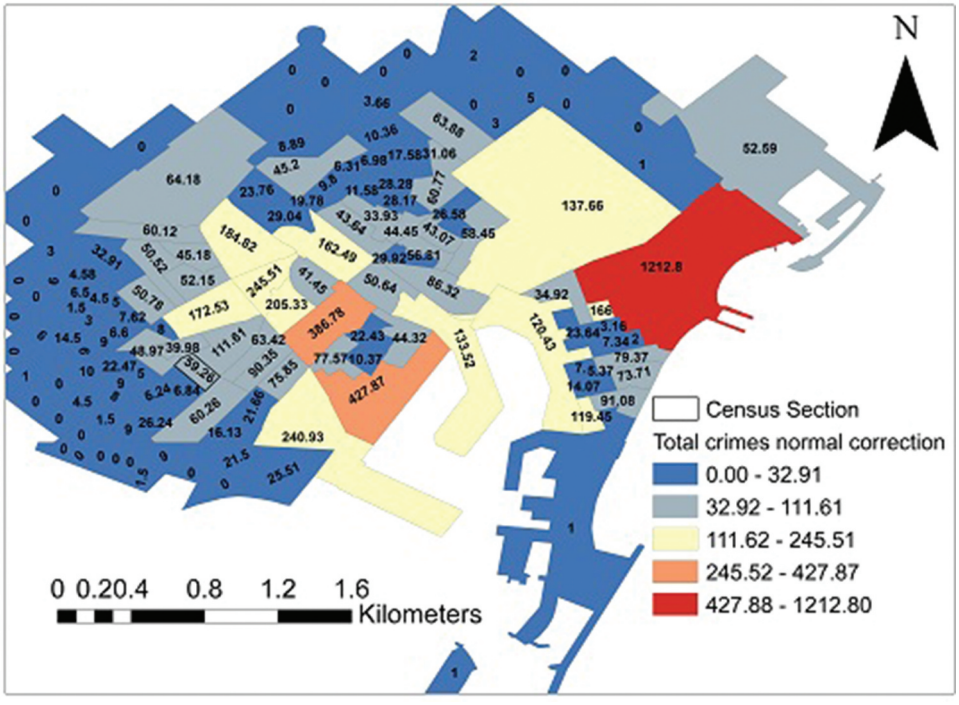

Figure 12. Number of crimes per census tract after correction of the edge effect with the normal function. Own elaboration using ArcGIS 10.4.1.

solution consists in creating an extra zone of a given size around the boundaries of each analysis unit (see Figure 2). The points located above this buffer zone are incorporated into the analysis unit in question to take them into account. This implies the need to create a buffer for each analytical unit and to count the number of points in each buffer. As the process is time-consuming, we compare the results for the census tracts

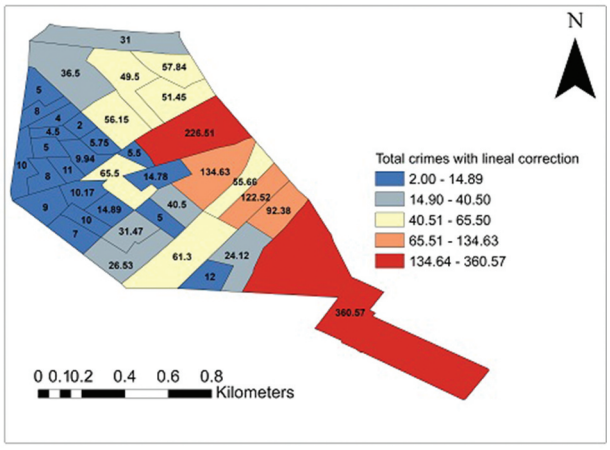

Figure 13. Results after applying the correction with the linear function.

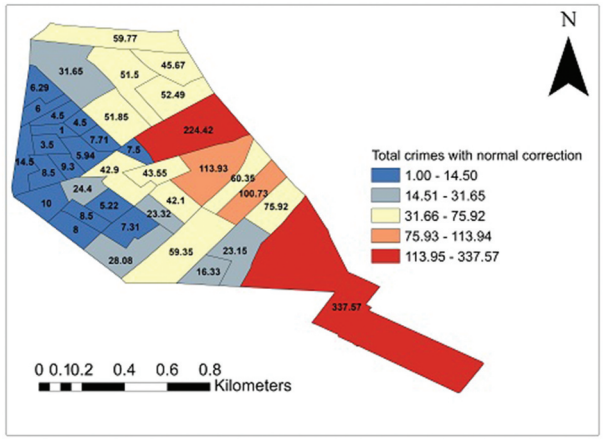

Figure 14. Results after applying the correction with the normal function.

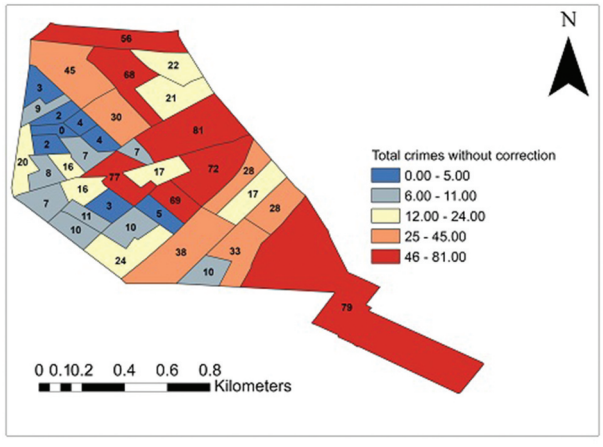

Figure 15. Results without applying any correction.

of one of the four neighbourhoods that make up the study region. In order to better compare the two solutions, the buffer size selected was 48 metres, the same distance that was introduced in the tool to correct the edge effect. Of the total 6558 thefts that occurred in the entire study region, 2091 were recorded in the Raval neighbourhood and on the edges of the immediately adjacent census tracts. The results of applying the different solutions to the Raval crimes can be seen in the following figures. Results of applying linear correction, normal correction, no correction and the buffer correction are shown in Figure 13, 14, 15 y 16, respectively.

Looking at the figures above, the number of crimes in each census tract after applying the buffer-based solution increases dramatically. This does not occur when applying the weighting solution that we propose in this article. With the tool presented here, only two of the census tracts have experienced a noticeable increase in the number of crimes because a high number of points were closely located to the axis in adjacent census tracts. However, the differences in the number of points when the buffer solution is applied are much more prominent. For example, one of the census tracts went from 79 thefts to 770 thefts after incorporating the buffered points. Therefore, the original number of points in the data has increased significantly. However, the application of the method we propose keeps the original number of crimes analysed accurate. Table 2 shows this. Note how the application of the buffer has almost quadrupled the number of incidents.

A strategy which is not sensible to the border issue is the KDE. When it is used, KDE produces a continuous surface considering all the points regardless of the edge of the spatial unit of analysis. However, the KDE solution is not useful when the analyst is interested in getting a value (for example, the number of crimes occurred) to be aggregated to a spatial unit (a census section, for example). This is because KDE only creates a heat map overcoming the artificial edges but does not give a value which could be used as a variable (for example, the amount of crime in each area).

Figure 17 shows a KDE applied to our crime data. It can be seen that we only are able to know where the hot spots are concentrated, but we do not get a crime count in each census section as a result. A possible alternative to get a value which could be aggregated to an area is using KDE by dividing the study area into cells. Each cell has a specific KDE value (z-value). Thus, by selecting the cells contained inside each spatial unit (polygon) we can calculate a value (the average z-values) that could be assigned to the specific unit (Xu, Pennington-Gray, and Kim 2018; Maldonado-Guzmán 2020). Figure 18 shows

Table 1. Comparison of the resulting number of crimes in each spatial unit.

Census Section	No edge correction	With linear correction	With normal correction
A	595	372.08	427.87
B	1185	1213	1212.8
C	50	67.5	64.18
D	79	372.14	240.93
E	214	172.41	137.66
F	0	4.5	4.5

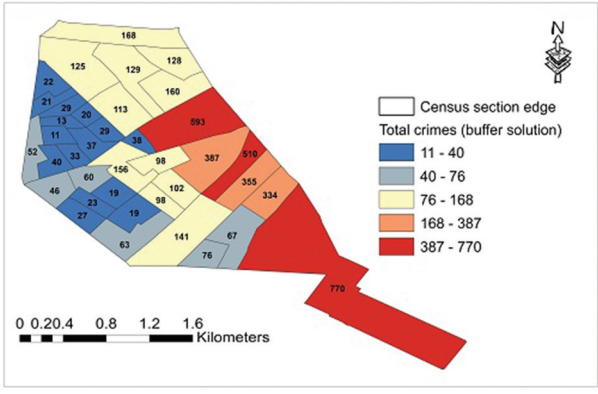

Figure 16. Results after applying the buffer-based correction.

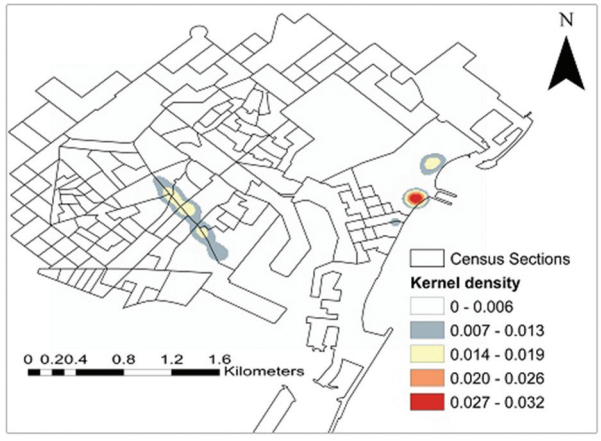

Figure 17. KDE of theft crimes occurred in 2016 in Ciutat Vella district. The bandwidth chosen has been 9,57732000000249E-05 (that suggested by default considering the data introduced).

Table 2. Total number of crimes resulting from each solution applied.

Number of crimes without correction	Number of crimes with normal function	Number of crimes with linear function	Number of crimes with buffer solution
2091	2091	2091	5112

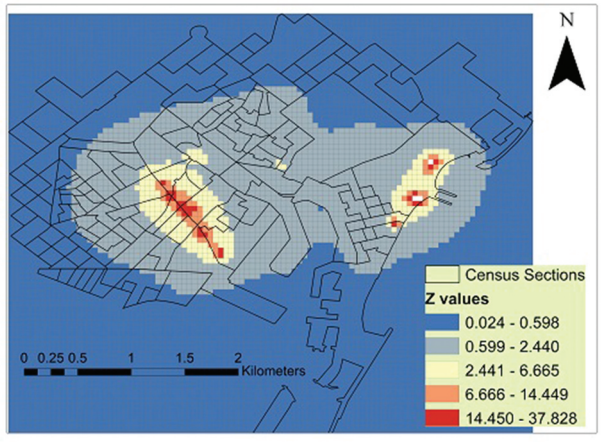

Figure 18. KDE using *Crimestat* software. With a fixed bandwidth of 9,5773E-05, the study region has been divided in 13,700 cells of 100 square metres each.

how the study area has been divided into 17,300 cells, each of them with a size of 100 square metres. Figure 19 shows the values of KDE in each cell. Thus, the value for the census section in the sample is 8.33.

However, doing the previous process manually is quite time-consuming. For example, to calculate the value for each spatial unit it is necessary to carefully draw the contours of the polygon in order to only select the cells inside it for the average calculation. Unless the analyst uses a software to compute this process automatically, repeating this method for each polygon manually can be tedious, especially when the number of spatial units is high. Moreover, the same cell can be contained by two or more polygons, and also some cells usually are not entire inside the polygon, so its whole value should not be counted when calculating the average value of KDE.

5. Discussion

This paper proposes a solution to the edge problem when counting crime incidents in artificial analysis units. For this purpose, a tool has been developed in ArcGIS that allows to apply such a solution automatically. By doubling the point to the adjacent unit in a weighted way, the results are more in line with the spatial reality of the criminal phenomenon. In this way, a fresh methodology is presented that overcomes the limitations of other approaches that have also been used to solve the edge effect when counting point data.

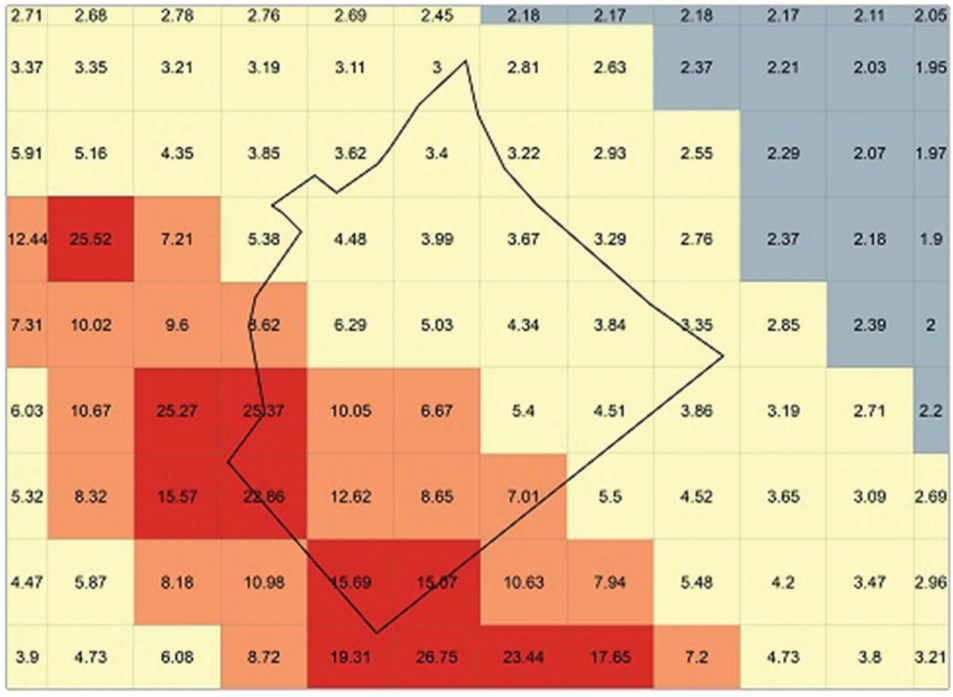

Figure 19. Example of calculating the average value of crime density to add it to a specific census tract. The numbers inside the polygon are considered to calculate the average value.

The proposed method for solving the border issue shows a number of advantages over other proposals. First, it makes it possible to compute the same crime in two different adjacent units, but without increasing the actual number of incidents occurring. Second, the procedure can be applied at any scale, regardless of whether the analytical unit is a country, city, neighbourhood or census tract. The only necessary thing is to adapt the distance for the new scale.

Third, this method can detect when the study region ends. Either because there are no other units on the other side of the boundary or because of the presence of the sea, the value of the point near the edge is automatically corrected. Lastly, this procedure achieves a sort of continuous surface in the territory, since the occurrence of a crime near the border of a census tract does not prevent that crime from being taken into account in the adjacent census tract.

Moreover, beyond the improvement in obtaining the number of crimes compared to the traditional point-in-polygon method, the tool can be useful for the analysis of fear of crime, since it considers the effect that a crime committed in one neighbourhood may have on the neighbourhood next door. The way by which crime incidents are counted after applying the tool makes it possible to consider those crimes that occurred in proximity to, but out of, the study region. For example, Maldonado-Guzmán et al. (2021) point out that in Barcelona the Besós quarter has high levels of fear of crime, although it has low levels of crime rates. The mentioned authors argue that this neighbourhood is adjacent to the Mina quarter, an area which belongs to another municipality different from Barcelona and that it has high levels of crime and social and physical disorder. If crime data of the Mina quarter is incorporated into the Barcelona crime data, the correction of the border issue by using the tool allows to consider crime incidents committed out of but near the limits of the study region, then improving the understanding of the relationship between crime concentration and fear of crime.

In order to weight the crime incident, depending on its distance from the axis, two decay functions with distance have been used. The results of applying one or the other are quite similar, being consistent with the results obtained by Ratcliffe and Taniguchi (2008) in their analysis. However, future research should compare the results generated by the tool as a function of the type of crime analysed and the spatial distribution pattern of the points. In this way, it could be revealed whether the use of one decay function or another is more appropriate based on the data to be analysed. Also, future modifications could be made to adapt the distance between the points and the edges depending on the area size of each spatial unit. Doing so, the tool could automatically adapt the best threshold distance to select which points are doubled, thus working as an adaptive bandwidth

Despite the improvements cited above, the tool purposed in this paper has some limitations which should be under consideration. Firstly, the current version of the tool only takes into account the chosen distance from the point to the nearest edge as a weighting factor. Nonetheless, also the size of both unities of analysis, that which contains the original point and that in which this point will be doubled, should be considered for the point weighting. Doing so it is important because when working with irregular boundaries, the influences on the adjacent spatial unit for the points are not equal in size, even though the two points on one side of the edge and the other are located exactly the same distance from the edge and within the selected distance by the analyst. Thus, futures developments of the tool need to incorporate the differences in the size areas together with the distance weighting criterion.

Secondly, the results obtained after using the tool are not integer numbers. This could limit the use of the results as dependent variable when using count regression models such as Poisson or Negative Binomial models, both typically used in the analysis of spatial crime data (Osgood 2000; Hilbe 2011). However, this limitation can be easily solved if the analyst rounds the number of crime. Operating by this way is not a problem, as the decimals numbers are so small that the rounding process does not increase the total amount of original points in the sample. Once the analyst has rounded the results after correcting the edge effect, an offset variable could be used to consider the population or areal unity size in which the data is observed (Hilbe 2017). This process allows to calculate the crime rates or densities instead of using the raw number of incidents. Several statistic packages such as SPSS or STATA include this option.

Thirdly, the analyst might be interested in choosing specifically which points within the distance specified have to be doubled to the neighbouring areas. The current version of the tool does not allow to do that automatically, but the analyst can use one of the csv files the tool generates to find which points have been doubled and, of all of them, which would not have wanted to be doubled. Then, the analyst can create a layer with those selected points and make a spatial join between the original layer and the layer containing the points which have been removed from the doubled incidents list.

In spite of the limitations above discussed, the tool makes improvement to the process of counting points in polygons skewed by the edge effect. Until the present, analysts made the traditional spatial join to get the number of points within polygons without be worried about such effect. The tool developed here allow the analyst to correct this problem in a more automatically and exactly way. Also, it is worth to note that the tool has been developed to obtain a general solution to the edge effect according to the needs that the authors have found in their research. It is expected that with the opportunity to offer this tool to other analysts, further improvements and modifications will be incorporated as other professionals have other needs arising from their own investigations. The authors are delighted to receive the suggestions of the analysts who use the tool, in order to adapt the tool to the real problems that these analysts encounter in practice.

Bibliographical note

Daniel Salafranca-Barreda is PhD candidate at the University of Miguel Hernández and a member of the Atypical Homicide Research Group. His research interests are in crime mapping and geographic profiling.

Diego Maldonado-Guzmán is a PhD Candidate and predoctoral fellow at the Department International Public Criminal and Procedural Law of the University of Cadiz, Spain and a member of Penal system, criminality, and security group (SEJ-378). His research interests are in geographical crime analysis, social disorganization and crime, geographic profiling, and online child sexual abuse.

Patricia Saldaña-Taboada is a PhD candidate and predoctoral Fellow at the Department of Criminal Law of the University of Granada, Spain. Her research interest are in cryptocurrencies and digital crime, cryptomarkets and crime and place.

Notes

1. The link to download the tool is: https://github.com/utm30web/Edge-effect.git
2. Download available on the website of the Cartographic and Geological Institute of Catalonia from this link http://bit.ly/2oCDRy3

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Diego J. Maldonado-Guzman http://orcid.org/0000-0002-5283-0634

References

Bailey, T. C., and A. C. Gatrell. 1995. *Interactive Spatial Data Analysis*. London: Longman.

Cruz Rot, M. 2006. "Introducción al análisis de datos mapeados o algunas de las (muchas) cosas que puedo hacer si tengo coordenadas." *Ecosistemas* 15 (3): 19–39.

Ewers, R. M., and R. K. Didham. 2006. "Continuous response functions for quantifying the strength of edge effects." *Journal of Applied Ecology* 43 (3): 527–536. doi:10.1111/j.1365-2664.2006.01151.x.

Galdón Clavell, G., and M. Pybus Oliveras. 2011."Crisis económica y gestión de la inseguridad ciudadana: Los mapas de delincuencia."*Revista Catalana de Seguretat Pública*24: https://www.raco.cat/index.php/RCSP/article/view/244856

Gao, F., W. Kihal, N. Le Meur, M. Souris, and S. Deguen. 2017. "Does the edge effect impact on the measure of spatial accessibility to healthcare providers?" *International Journal of Health Geographics* 16 (1): 46. doi:10.1186/s12942-017-0119-3.

Goreaud, F., and R. Pélissier. 1999. "On explicit formulas of edge effect correction for Ripley's K -function." *Journal of Vegetation Science* 10 (3): 433–438. doi:10.2307/3237072.

Harries, K. 1999. *Mapping crime: principle and practice*. Washington: National Institute of Justice. https://www.ojp.gov/pdffiles1/nij/178919.pdf

Hilbe, J. M. 2011. *Negative binomial regression*. United Kingdom: Cambridge.

Hilbe, J. M. 2017. "The statistical analysis of count data/El análisis estadístico de los datos de recuento." *Cultura y Educación* 29 (3): 409–460. doi:10.1080/11356405.2017.1368162.

Levine, N. 2010. "Kernel Density Interpolation." In *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.3)*. Houston, TX: National Institute of Justice. https://www.ojp.gov/pdffiles1/nij/grants/242960-242995.pdf

Maldonado-Guzmán, D. 2020. "Airbnb and crime in Barcelona (Spain): testing the relationship using a geographically weighted regression." *Annals of GIS* 1–14. doi:10.1080/19475683.2020.1831603.

Maldonado-Guzmán, D. J., P. Saldaña-Taboada, and A. Miguel-Alvaro. 2021. "Spatial analysis of perceived insecurity in Barcelona: The role of adjacent neighbourhoods and psychosocial factors involved." *Anuario de Psicología Jurídica* 31 (1): 63–73. doi:10.5093/apj2021a7.

McCord, E. S., and J. H. Ratcliffe. 2007. "A micro-spatial analysis of the demographic and criminogenic environment of drug markets in Philadelphia." *Australian and New Zealand Journal of Criminology* 40 (1): 43–63. doi:10.1375/acri.40.1.43.

McCord, E. S., and J. H. Ratcliffe. 2009. "Intensity value analysis and the criminogenic effects of land use features on local crime patterns." *Crime Patterns and Analysis* 2 (1): 17–30. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.364.6362&rep=rep1&type=pdf

Murray, A. T., I. McGuffog, S. Western, and P. Mullins. 2001. "Exploratory spatial data analysis techniques for examining Urban crime." *British Journal of Criminology* 41 (2): 309–329. doi:10.1093/bjc/41.2.309.

O'Sullivan, D., and D. J. Unwin. 2010. *Geographic information analysis*. New Jersey: John Wiley and Sons. doi:10.1002/9780470549094.

Olaya, V. (2010). *Sistemas de información geográfica*. https://volaya.github.io/libro-sig/

Osgood, D. W. 2000. "Poisson-based regression analysis of aggregate crime rates." *Journal of quantitative criminology* 16 (1): 21–41. doi:10.1023/A:1007521427059.

Ratcliffe, J. H. 2010. "The spatial dependency of crime increase dispersion." *Security Journal* 23 (1): 18–36. doi:10.1057/sj.2009.16.

Ratcliffe, J. H., and T. Taniguchi. 2008. "Is crime higher around drug-gangs streets corners? two spatial approaches to the relationship between gang set spaces and local crime levels." *Crime Patterns and Analysis* 1 (1): 23–45. http://ttaniguchi.net/wp-content/uploads/2014/08/Ratcliffe-Taniguchi-2008-Is-crime-higher-around-drug-gang-corners.pdf

Rengert, G. F., and B. Lockwood. 2009. "Geographical units of analysis and the analysis of crime." In *Putting Crime in Its Place. Units of Analysis in Geographic Criminology*, edited by D. Weisburd, W. Bernasco, and G. J. N. Bruinsma, 109–123. New York: Springerpp. doi:10.1007/978-0-387-09688-9_5.

Rengert, G. F., J. H. Ratcliffe, and S. Chakravorty. 2005. *Policing illegal drug markets: geographic approaches to crime reduction*. New York: Criminal Justice Press.

Ripley, B. D. 1988. *Statistical inference for spatial processes*. New York: Cambridge University Press. doi:10.1017/CBO9780511624131.

Rossmo, K. 2000. *Geographic Profiling*. Miami: CRC Press.

Wei, R., T. H. Grubesic, M. Manfredi, F. Mele, D. Amparore, G. Cattaneo, I. Morra, R. M. Scarpa, and F. Porpiglia. 2017. "An alternative classification scheme for uncertain attribute scheme." *Minerva Urologica E Nefrologica = the Italian Journal of Urology and Nephrology* 69 (4): 604–615. doi:10.1080/00330124.2017.1288573. *The professional geographer*.

Xu, Y., L. Pennington-Gray, and J. Kim. 2018. "The sharing economy: a geographically weighted regression to examine crime and the shared lodging sector." *Journal of Travel Research* 58 (7): 1–16.

Zhang, H., G. Suresh, and Y. Qiu. 2012. "Issues in the aggregation and spatial analysis of neighborhood crime." *Annals of GIS* 18 (3): 173–183. doi:10.1080/19475683.2012.691901.