

Article

Fuzzy-Based Spatiotemporal Hot Spot Intensity and Propagation—An Application in Crime Analysis

Barbara Cardone ¹ and Ferdinando Di Martino ^{1,2,*} 

¹ Dipartimento di Architettura, Università degli Studi di Napoli Federico II, Via Toledo 402, 80134 Napoli, Italy; b.cardone@unina.it

² Centro di Ricerca Interdipartimentale “Alberto Calza Bini”, Università degli Studi di Napoli Federico II, Via Toledo 402, 80134 Napoli, Italy

* Correspondence: fdimarti@unina.it; Tel.: +39-081-253-8907; Fax: +39-081-253-8905

Abstract: Cluster-based hot spot detection is applied in many disciplines to analyze the locations, concentrations, and evolution over time for a phenomenon occurring in an area of study. The hot spots consist of areas within which the phenomenon is most present; by detecting and monitoring the presence of hot spots in different time steps, it is possible to study their evolution over time. One of the most prominent problems in hot spot analysis occurs when measuring the intensity of a phenomenon in terms of the presence and impact on an area of study and evaluating its evolution over time. In this research, we propose a hot spot analysis method based on a fuzzy cluster hot spot detection algorithm, which allows us to measure the incidence of hot spots in the area of study. We analyze its variation over time, and in order to evaluate its reliability we use a well-known fuzzy entropy measure that was recently applied to measure the reliability of hot spots by executing fuzzy clustering algorithms. We apply this method in crime analysis of the urban area of the City of London, using a dataset of criminal events that have occurred since 2011, published by the City of London Police. The obtained results show a decrease in the frequency of all types of criminal events over the entire area of study in recent years.



Citation: Cardone, B.; Di Martino, F. Fuzzy-Based Spatiotemporal Hot Spot Intensity and Propagation—An Application in Crime Analysis. *Electronics* **2022**, *11*, 370. <https://doi.org/10.3390/electronics11030370>

Academic Editor: Miin-shen Yang

Received: 10 December 2021

Accepted: 20 January 2022

Published: 26 January 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: hot spot; spatiotemporal hot spot detection; fuzzy cluster; EFCM; reliability; fuzzy entropy; HR-EFCM

1. Introduction

Hot spot detection is a spatial analysis method aimed to detect regions on a map, called *hot spots*, within which a high concentration of events characterizing a specific phenomenon is localized. Each event is spatially referred and geometrically represented as a point on the map; cluster techniques are often applied on the dataset of events to detect cluster prototypes representing hot spots on the map.

Clustering methods are generally used to detect hot spots. The data points are made up of events assigned as elements with point geometry on the map. Clustering algorithms are used to locate and construct hot spots as elements with polygonal geometry on a map, corresponding to regions of the study area where the phenomenon is most insistent. Moreover, by analyzing the location and extension of hot spots detected in successive time frames, it is possible to study their evolution over time.

Many cluster-based hot spot detection methods are proposed in the literature. K-means [1] is applied to detect hot spots in crime analysis [2–5] and in fire analysis [6,7]. K-medoids [8] is used to detect hot spots in disease analysis [9] and crime analysis [10]. Fuzzy C-means (for short FCM) [11–13] is used by various authors to detect hot spots in crime analysis [14–17], road traffic crashes [18], and disease analysis [19].

Some researchers apply density-based clustering to detect also irregular shapes of hot spots on the map. Kernel density-based algorithms [20] are applied in crime analysis [21],

soil pollution [22], and traffic accident analysis [23]. The fast DBSCAN algorithm [24] is applied in [25] to detect hot spots with a high density of taxi passengers.

As a trade-off between the speed of execution of K-means and FCM and the accuracy in detecting the outline of hot spots obtained using density-based algorithms, in [26] a cluster-based hot spot detection method based on the extended FCM algorithm [27] (for short EFCM) is proposed; EFCM is an extension of the FCM algorithm in which cluster prototypes are hyperspheres in the feature space, rather than points, as in K-means and FCM. In [26], the hot spots are given by circles on the map; the authors show that these circles approximate the shape of the clusters detected by using density-based clustering. The EFCM hot spot detection algorithm is applied in [28] for disease analysis and in [29] for earthquake disaster analysis.

In particular, in [28] a method based on EFCM is proposed for spatiotemporal hot spot analysis. The dataset of events is partitioned in subsets, where each subset contains events occurred in a time frame. By analyzing the location and extent of the hot spots detected at each time step on the map, their displacement over time is traced; in addition, computing the spatial intersections between hot spots detected at consecutive time frames, it is possible to analyze in which geographical areas the phenomenon is persistent and in which geographical areas it has moved.

One of the major critical points in hot spot analysis is to define a measure of the intensity of the phenomenon on a specific area and to make an assessment of the reliability of this measure. The hot spot analysis algorithms proposed in the recent literature do not make use of a measure to evaluate the intensity of the phenomenon over an area and do not evaluate how reliable this measure is. In [30], an index evaluating the reliability of hot spots detected via EFCM clustering is proposed. This index measures the reliability of the hot spots measuring the fuzziness of clustering evaluated considering the De Luca and Termini fuzzy entropy of a fuzzy set [31,32]. In [33,34], the De Luca and Termini fuzzy entropy is used to measure the fuzziness of clustering detected by executing FCM; each fuzzy cluster constitutes a fuzzy set and its fuzzy entropy is measured by considering the membership degrees to it of the data points, so that the closer the cluster is to a crisp set, the less the measure of its fuzziness will be.

In this paper, we propose a new method to analyze the intensity and spatiotemporal evolution of hot spots detected using the EFCM spatiotemporal hot spot detection proposed in [28]. We assess the incidence over time of the phenomenon analyzed in a specific area by calculating an index called hot spot strength, which measures the percentage over time that the selected area is affected by hot spots. In addition, we measure the reliability of this evaluation, calculating a reliability index of the hot spot strength based on the hot spot reliability measure proposed in [30].

The main contributions of our research are summarized below:

- In addition to analyzing the evolution of the phenomenon in a selected area for each time frame, our method evaluates with what intensity this area has been affected by the phenomenon in a given period of time; we measure this intensity by calculating the *hot spot strength* index. This measure is essential in an application context to understand with what intensity a certain phenomenon is spreading over an area of investigation;
- A measure of the reliability of the hot spot strength is proposed, using the reliability index of the hot spot [30] to assess the reliability of the hot spot strength measured in a time frame; it is given by the weighted average of the reliability of the hot spots that insist on the selected area in this time frame, where the weight assigned to a hot spot is given by the extension of its spatial intersection with the selected area.

The EFCM hot spot detection algorithm and the fuzzy entropy hot spot reliability measure are summarized in Section 2. In Section 3, we present our method. Section 4 shows the results of its application in crime analysis on an area of study given by the City of London. Final considerations and future perspectives are included in Section 5.

2. Preliminaries

2.1. EFCM Hot Spot Detection

Let $X = \{x_1, \dots, x_N\} \subset \mathbb{R}^2$ a set of bi-dimensional data points extracted from a spatial event dataset. Each data point is a spatially referenced event given by its latitude and longitude coordinates.

EFCM returns cluster prototypes made of hyperspheres in the feature space. The $C^{(0)}$ clusters are assigned initially in EFCM; the optimal number of clusters C is found by dissolving, during each iteration, the two clusters most similar to each other if their similarity is greater than a prefixed threshold η .

Let $V = \{v_1, \dots, v_C\} \subset \mathbb{R}^n$ be the set of centers of the C clusters. Let U be the $C \times N$ partition matrix, where u_{ij} is the membership degree of the j th data point x_j to the i th cluster v_i . Let $r = \{r_1, \dots, r_C\}$ be the set of radii of the C clusters.

EFCM minimizes the following objective function:

$$J(\mathbf{U}, \mathbf{V}, \mathbf{r}) = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m \delta_{ij}^2 \tag{1}$$

where m is the fuzzifier parameter and δ_{ij} , interpreted as the distance between the i th cluster and the j th data point, is given by:

$$\delta_{ij} = \max(0, d_{ij} - r_i) \tag{2}$$

In (2), d_{ij} is the Euclidean distance between the center of the i th cluster and the j th data point and r_i is the radius of the i th cluster.

EFCM stops after t iterations if the difference $|\mathbf{U}^{(t)} - \mathbf{U}^{(t-1)}| = \max_{\substack{i=1, \dots, C \\ j=1, \dots, N}} |u_{ij}^{(t)} - u_{ij}^{(t-1)}|$ is

less than a prefixed stop iteration threshold ϵ .

The parameters to set before executing EFCM are:

- The fuzzifier parameter m ;
- The stop iteration threshold ϵ ;
- The threshold assigned to dissolve the most similar clusters η ;
- The initial number of clusters $C^{(0)}$.

EFCM returns the centers of the final clusters, their radii, and the $C \times N$ partition matrix.

EFCM is applied in [26] to detect hot spots, given by spatial regions where such events are localized with higher density. In hot spot detection, the features consist of the two geographical coordinates locating the events, and a hot spot is approximated by a circular area on the map. The prototype of the i th cluster detected by EFCM is given by a circle with center coordinates $v_i = (x_i, y_i)$ and a radius r_i . EFCM returns the centers of the C clusters V , their radii r , and the partition matrix U . The couple (v_i, r_i) identifies a circle on the map.

In [26], the authors show that EFCM can approximate the shapes of hot spots on the map and is robust with respect to the presence of noise and outliers.

2.2. Fuzzy-Entropy-Based Hot Spots Reliability Evaluation

In [33,34], a measure of the reliability of hot spots detected via EFCM based on the De Luca and Termini fuzzy entropy [31,32] is proposed.

The De Luca and Termini fuzzy entropy measures the fuzziness of a fuzzy set.

Let $F(X) = \{A: X \rightarrow [0, 1]\}$ be the family of fuzzy sets defined on a universe of discourse X .

Let $h: [0,1] \rightarrow [0,1]$ be a continuous function called the *fuzzy entropy function*, where:

1. $h(1) = 0$;
2. $h(u) = h(1 - u)$;
3. h is monotonically increasing in $[0, 1/2]$;
4. h is monotonically decreasing in $[1/2, 1]$.

The fuzzy entropy function h has a minimum (0) when u is 0 or 1 and a maximum when $u = 1/2$.

De Luca and Termini in [26,27] propose the following fuzzy entropy function:

$$h(u) = \begin{cases} 0 & \text{if } u = 0 \\ -u \cdot \log_2(u) - (1-u) \cdot \log_2(1-u) & \text{if } 0 < u < 1 \\ 0 & \text{if } u = 1 \end{cases} \quad (3)$$

which has the maximum value 1 when $u = \frac{1}{2}$; this is called Shannon's function.

If $X = \{x_1, x_2, \dots, x_N\}$ is a discrete set, the entropy measure of fuzziness of the fuzzy set A is given by:

$$H(A) = K \sum_{j=1}^N h(A(x_j)) \quad (4)$$

where K is a multiplicative constant. If $H(A) = 0$, then for each element $x_j, j = 1, \dots, N$ $A(x_j) = 0$ or $A(x_j) = 1$ and A coincides with a subset of the set X ; if for each element x_j $A(x_j) = 1/2$, then the fuzziness of A is maximal.

If A is a crisp set, its fuzziness is null and $H(A) = 0$. The higher the fuzziness of a fuzzy set, the closer the mean membership degree to the fuzzy set of X 's elements approaches $1/2$.

In [33,34], the fuzziness measure (4) is used to construct a new validity index applied to evaluate the optimal number of clusters in FCM. If A_i is the i th fuzzy cluster where $i = 1, \dots, C$, C is considered as a fuzzy set and u_{ij} is the membership degree of the j th data point to the i th cluster, the authors use the following fuzzy entropy measure of A_i :

$$H(A_i) = \frac{1}{N} \sum_{j=1}^N h(u_{ij}) \quad i = 1, \dots, C \quad (5)$$

where N is the number of data points and the De Luca and Termini fuzzy entropy function (4) is used.

In [30], the reliability of the i th detected hot spot is measured by calculating the fuzziness of the detected clusters. The reliability index of a detected hot spot is given by the formula:

$$R(A_i) = 1 - H(A_i) \quad i = 1, \dots, C \quad (6)$$

The reliability of each hot spot is evaluated by calculating its reliability index by (6); this is a value in the range [0,1]. Finally, the reliability thematic map is produced.

In [30], the authors propose an EFCM-based hot spot detection algorithm in which the reliability of each hot spot is calculated by (6).

Below we show this algorithm, abbreviated as the HR-EFCM (Algorithm 1).

Algorithm 1 HR-EFCM

1. Extract the event dataset $X = \{x_1, \dots, x_N\} \subset R^2$
 2. Execute EFCM ($X, m, \epsilon, \eta, C^{(0)}$) obtaining the partition matrix U with components u_{ij} , the cluster centers v_i , their radius r_i where $i = 1, \dots, C$ and $j = 1, \dots, N$
 3. **For** $l = 1$ to C // for all the clusters detected by EFCM
 4. $H \leftarrow 0$
 5. **For** $j = 1$ to N
 6. $H \leftarrow H + h(u_{ij})$ // where the Equation (5) is applied for the function $h(u)$
 7. $H \leftarrow H/N$
 8. $R_i \leftarrow 1 - H$
 9. **Next** i
 10. **Return** the partition matrix U , the cluster centers v_i , their radius r_i , and their reliability R_i $i = 1, \dots, C$
-

In [30], the HR-EFCM algorithm is applied to detect hot spots in disease analysis; the results show that the reliability of a hot spot is linearly dependent on the standard deviation

of the values of the membership degrees of the data points to the corresponding fuzzy cluster. Furthermore, comparative tests show that the reliability values calculated using the hot spot reliability evaluation algorithm are correlated to the reliability values assigned by the pool of experts.

3. The Proposed Framework

We propose a novel method based on the HR-EFCM algorithm that evaluates the reliability of the results of the spatiotemporal evolution of hot spots.

Let X be a dataset of georeferenced events partitioned in T subsets X_1, X_2, \dots, X_T , where X_t $t = 1, 2, \dots, T$ is a subset containing all events that occurred in a given time frame t .

For each subset, HR-EFCM is executed to detect the hot spots as circles on the map. For each hot spot, the reliability is calculated as in (6).

In order to analyze the localization and evolution of hot spots in a selected area on the map, the zones are defined in this area covered by hot spots and detected in each time frame; each of these zones consists of the extent of the spatial intersection between a hot spot and the selected area.

In Figure 1, an example is shown of this process, in which the dataset X is partitioned into three subsets, corresponding to three consecutive time frames.

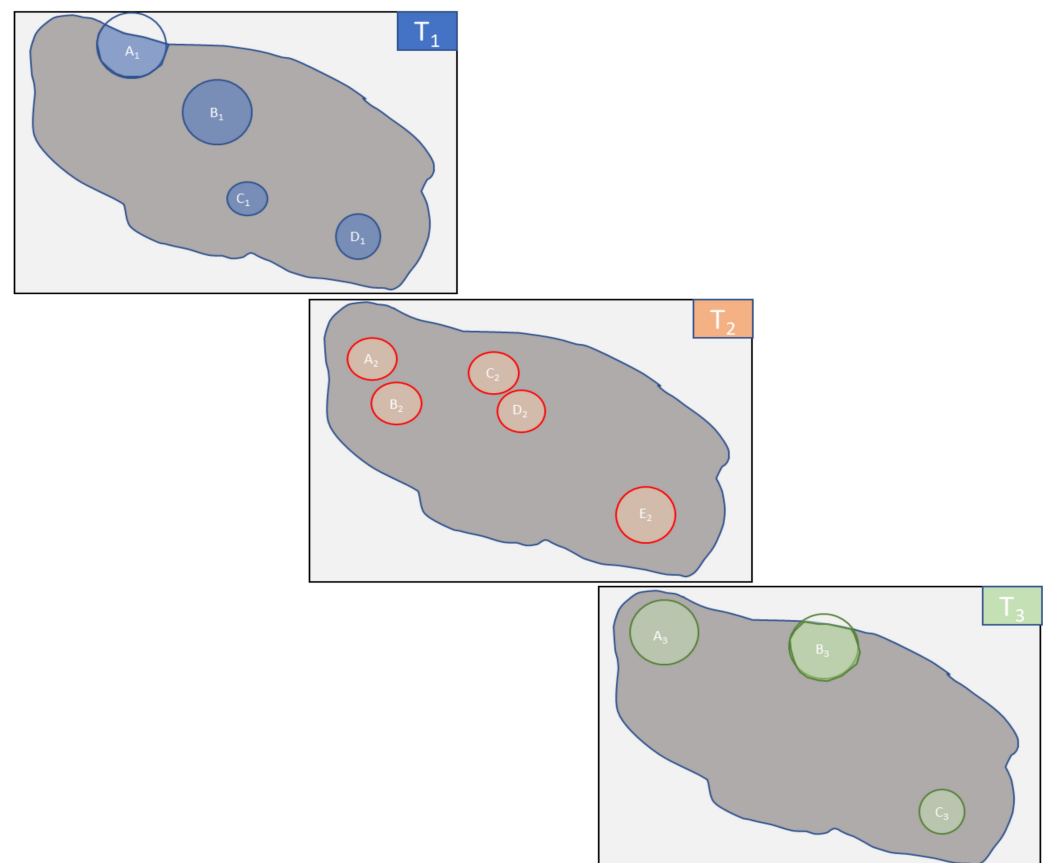


Figure 1. Example of hot spots detected and plotted in an area.

Figure 1 shows the hot spots detected in each of the time frames T_1 , T_2 , and T_3 . The original hot spots A_1 and B_3 in the figure are cut on the boundary of the selected area on the map. Therefore, an evaluation of the intensity of the phenomenon can be made by measuring the extension of the hot spots included in the selected area in relation to the extent of the entire selected area.

For each time frame an index is calculated, called the *hot spot strength*, which measures the percentage of the extent of the selected area covered by hot spots; moreover, an assess-

ment of the reliability of this measure is calculated, given by the weighted average of the reliability of each hot spot covering the selected area, in which the weight is constituted by the extent of the spatial intersection between the hot spot and the selected area.

The hot spot strength measured in the t th time frame is given by:

$$S_t = \frac{1}{D} \sum_{i=1}^{C_t} D_{i,t} \quad t = 1, \dots, T \quad (7)$$

where C_t is the number of hot spots detected in the t th time frame, $D_{i,t}$ is the extent of the spatial intersection between the i th hot spot detected in the t th time frame and the selected area, and D is the extent of the selected area.

If the i th hot spot does not intersect with the selected area, $D_{i,t}$ is null and this hot spot does not contribute to the calculation of the hot spot strength index S_t .

S_t takes on a value between 0 and 1; it is equal to 0 if no hot spot detected at the t th time frame covers the selected area; conversely, it is equal to 1 if the extent of the selected area is covered by hot spots detected at the t th time frame.

The reliability of the hot spot strength S_t is evaluated by the formula:

$$RS_t = \frac{\sum_{i=1}^{C_t} D_{i,t} \cdot R_{i,t}}{\sum_{i=1}^{C_t} D_{i,t}} \quad t = 1, \dots, T \quad (8)$$

where $R_{i,t}$ is the reliability of the i th hot spot detected in the t th time frame, varying in the range of [0,1].

RS_t varies between 0 and 1; it is equal to 0 if the reliability of all hot spots covering the selected area detected in the t th time frame is zero; conversely, it is equal to 1 if the reliability of all these hot spots is equal to 1.

In synthesis, the hot spot strength of the phenomenon in the selected area at the t th time frame is measured as the ratio between the sum of the extents of the spatial intersection between the hot spots detected in this time frame and the selected area and the extent of the selected area; its reliability is given by the weighted average of the reliability of each hot spot measured by (6), where the weight is the extent of the spatial intersection between this hot spot and the selected area.

In the preprocessing phase, HR-EFCM is executed for each subset in order to detect the C_t hot spots, $t = 1, \dots, T$, and calculate their reliability (Algorithm 2).

Algorithm 2 Spatiotemporal hot spots detection

1. Extract the event dataset X
 2. Partition the dataset into T subsets X_1, X_2, \dots, X_T
 3. **For** $t = 1$ to T //for each subset of events occurred in t th the time frame
 4. Execute HR-EFCM ($X_t, m, \varepsilon, \eta, C^{(0)}$)
 5. **Next** t
 6. **Return** the C_t hot spots detected and their reliability $t = 1, \dots, T$
-

After selecting an area on the map, for each time frame the hot spot strength and its reliability are calculated, respectively, via Equations (7) and (8) (Algorithm 3).

Algorithm 3 Spatiotemporal hot spots Strength Evaluation

1. Select a zone on the map
2. **For** $t = 1$ to T //for each subset of events occurred in t th the time frame
3. **For** $I = 1$ to C_t //for all the hot spots detected executing HR-EFCM on the t th subset
4. $S_t \leftarrow 0$
5. $RSN \leftarrow 0$
6. $RSD \leftarrow 0$
7. $D_{i,t} \leftarrow$ area of the part of the i th hot spot intersecting the selected region
8. $S_t \leftarrow S_t + D_{i,t}$
9. $RSN \leftarrow RSN + D_{i,t} * S_{i,t}$
10. $RSD \leftarrow RSD + D_{i,t}$
11. **Next** i
12. $RS_t \leftarrow RSN/RSD$
13. **Next** t
14. **Return** the hot spot strength S_t and its reliability $t = 1, \dots, T$

By analyzing the trend of the hot spot strengths calculated in each time frame, it is possible to evaluate how the diffusion of the phenomenon analyzed in the selected area has varied over time. Furthermore, the assessment of the reliability of the hot spot strength values allows the overall reliability of the results of the analysis to be evaluated.

To test the proposed method in an application context, we took into consideration various types of criminal events (robberies, shoplifting, car thefts, acts of sexual violence, etc.) in urban agglomerations in the City of London. The tests were carried out considering the Lower Super Output Areas (for short LSOAs) in the City of London as the study area and analyzing the spread of different types of criminal events that occurred from September 2011 to July 2021. In Section 4, the results of all experiments are shown and discussed.

4. Experimental Results

The LSOA in the City of London is shown in the map in Figure 2. We applied our method using a dataset of crime events that occurred in the LSOAs in the City of London from September 2011.



Figure 2. LSOAs of the City of London, shown within Greater London.

Following the neighborhood policing model known as sector policing (<https://www.cityoflondon.police.uk>), examined on 1 July 2021, the city of London is split into two sectors, east and west, with a senior leader responsible for each sector. Each sector is broken down further into three regions called *clusters*, including adjoint wards. Every cluster is guarded by a group of police officers, the Dedicated Ward Officers (DWO), responsible for maintaining order in that region.

The west sector includes the Fleet Street, Bank, and Barbican clusters, while the east sector includes the Monument, Liverpool Street, and Fenchurch Street clusters. A map of the six clusters is shown in Figure 3.

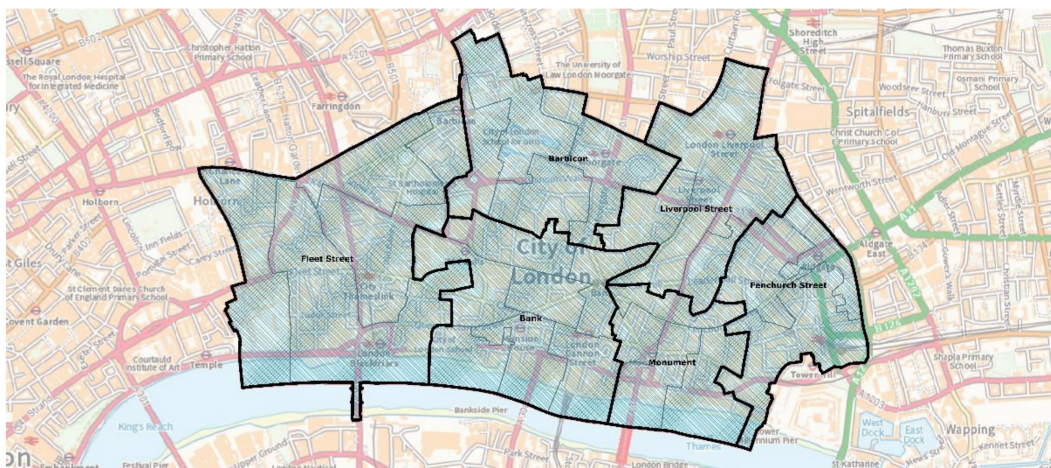


Figure 3. Map of the clusters in the City of London.

In these experimental tests, we applied the proposed method for analysis by type of criminal event, the incidence of the phenomenon on each of the six clusters in which the City of London is partitioned, and its evolution over time. This analysis will allow us to assess how effective the DWO's cluster surveillance may have been, monitoring how this effectiveness changes over time.

The database used in these experiments is composed of 22,310 georeferenced crime events that occurred in the City of London from September 2011 to July 2021. It is partitioned into 14 datasets corresponding to the 14 crime types recorded by the police in England and extracted from the website <https://data.police.uk/> (accessed on 1 July 2021).

We implemented our method using GIS ArcGIS Desktop 10.8. The geographical coordinate system used in our experiments was the projected Universal Transverse Mercator British National Grid coordinate system.

In the preprocessing phase, all the recorded events without geolocation information were discarded, while the other events were georeferenced and divided by type of crime and year of occurrence.

For each dataset, we executed our method by partitioning it in eleven subsets, corresponding to the crime events occurred in each time frame, where a time frame is given by a year.

Table 1 shows the number of recorded and georeferenced crime events belonging to each crime type that occurred in this period in the City of London.

Criminal events belonging to the typology "theft from the person" were recorded only starting from 2019; the "bicycle theft" and "possession of weapons" events started from 2013.

For each subset corresponding to a crime type, the *spatiotemporal hot spot detection* algorithm is executed and results are obtained for each year the hot spots are detected, as given by circular areas on the map.

Then, the *spatiotemporal hot spot strength evaluation* algorithm is executed for each crime type on the six clusters, measuring for each year the hot spot strength of the crime type on the cluster and its reliability.

For each cluster in the City of London, it is possible to analyze the locations and variation over time of the areas covered by hot spots detected by events belonging to a specific crime type.

Table 1. Number of crime events that occurred in the City of London from September 2011 to July 2021.

Crime Type	Number of Crime Events										
	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Anti-social behaviour	670	1710	1007	1056	786	1066	1064	1443	1328	751	336
Bicycle theft			260	392	259	359	290	460	453	455	63
Burglary	77	347	262	230	220	241	174	328	365	201	37
Criminal damage and arson	77	240	250	220	214	243	181	238	306	218	52
Drugs	183	536	415	429	348	340	242	410	791	636	255
Other crime	98	5987	6887	9459	5995	181	119	126	244	229	63
Other theft	859	2653	2397	2128	1567	1507	1119	1740	2953	935	106
Possession of weapons			17	25	15	43	37	71	85	53	17
Public order	82	204	172	201	164	250	175	430	515	297	115
Robbery	11	52	49	44	38	29	33	105	183	91	25
Shoplifting	205	636	590	566	628	656	572	915	1011	690	177
Theft from the person									944	494	131
Vehicle crime	49	172	214	218	91	175	121	252	219	165	37
Violence and sexual offences	203	624	660	745	833	931	693	1307	1514	710	224
Tot	2514	13,161	13,180	15,713	11,168	6021	4820	7825	10,911	5925	1638

For the sake of brevity, the results obtained for two types of crimes are detailed below for “drugs” and “shoplifting” in the Bank and Liverpool Street clusters.

The map in Figure 4 respectively shows the areas covered by hot spots detected for drugs crime events that occurred in the years 2017 (marked in blue) and 2018 (marked in red) in the Bank cluster.

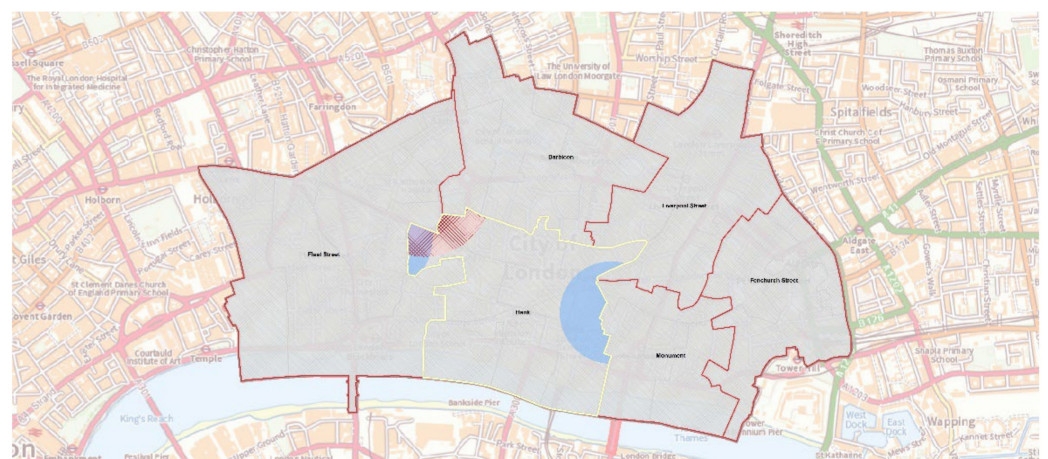


Figure 4. Drug crime hot spots in the Bank cluster for the years 2017 (marked in blue) and 2018 (marked in red).

Figures 5–7 respectively show the areas covered by hot spots detected for drug crime events in the years 2018–2019, 2019–2020, and 2020–2021.

One can observe that a large zone covered by a drug crime hot spot is present in the eastern area of the Bank cluster in 2017, which is no longer present in subsequent years. On the other hand, an area covered by hot spots persists in the northwest of the cluster.

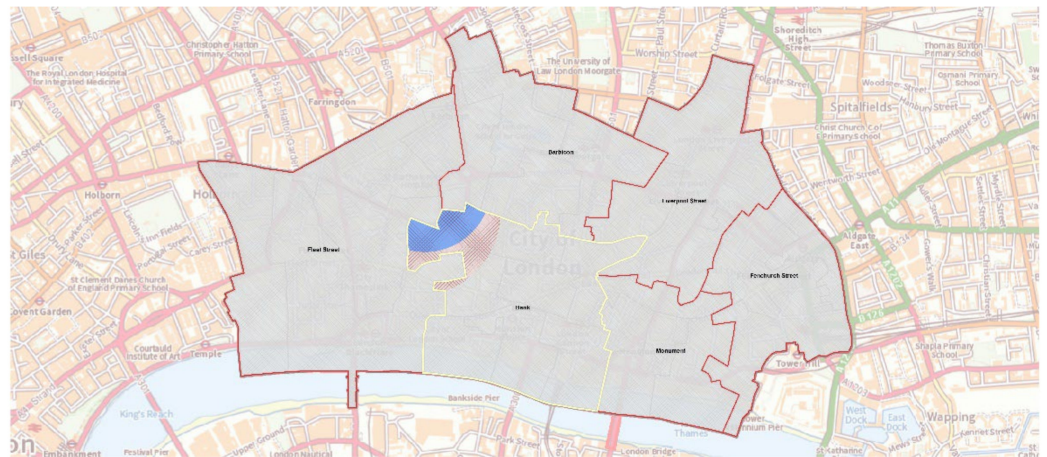


Figure 5. Drug crime hot spots in the Bank cluster for the years 2018 (marked in blue) and 2019 (marked in red).

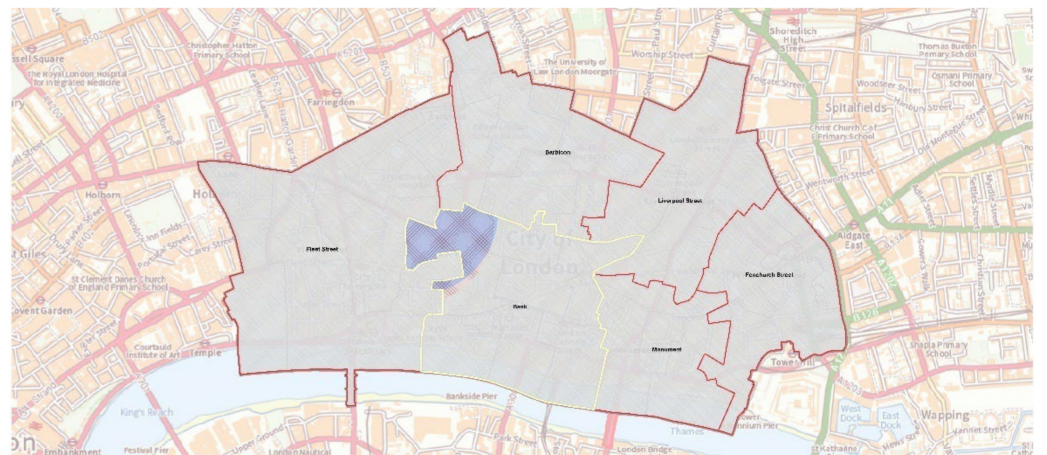


Figure 6. Drug crime hot spots in the Bank cluster for the years 2019 (marked in blue) and 2020 (marked in red).

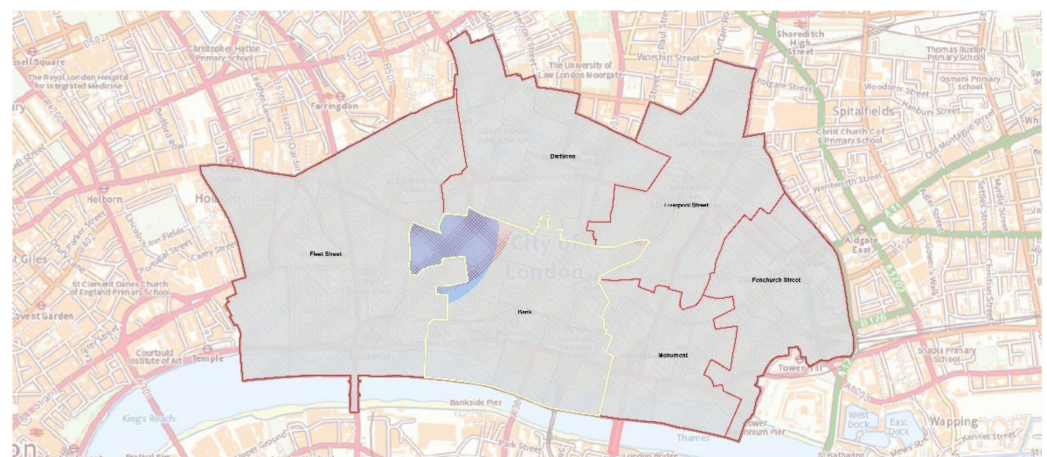


Figure 7. Drug crime hot spots in the Bank cluster for the years 2020 (marked in blue) and 2021 (marked in red).

Figure 8 plots a trend for the hot spot strength detected in the Bank cluster for drug crimes from 2011 to 2021.

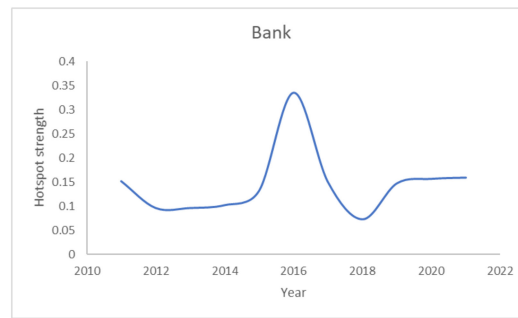


Figure 8. Trend of the hot spot strength for drug crimes in the Bank cluster.

The hot spot strength reaches a maximum in 2016 of 35%, then decreases to a minimum of less than 10% in 2018 and stabilizes at a value of around 15% from 2019.

Figure 9 shows the trend of the reliability of the hot spot strength calculated by (8). The reliability fluctuates between a minimum value of 0.75, reached in 2016, and a maximum value of 0.9. Since 2019, its trend has been constant, being approximately almost equal to 0.9.

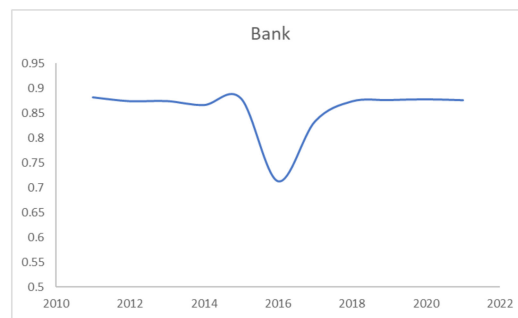


Figure 9. Trend of the reliability of the hot spot strength for drug crimes in the Bank cluster.

Figure 10 plots the trends of the hot spot strengths obtained for all the six clusters for drug crimes.

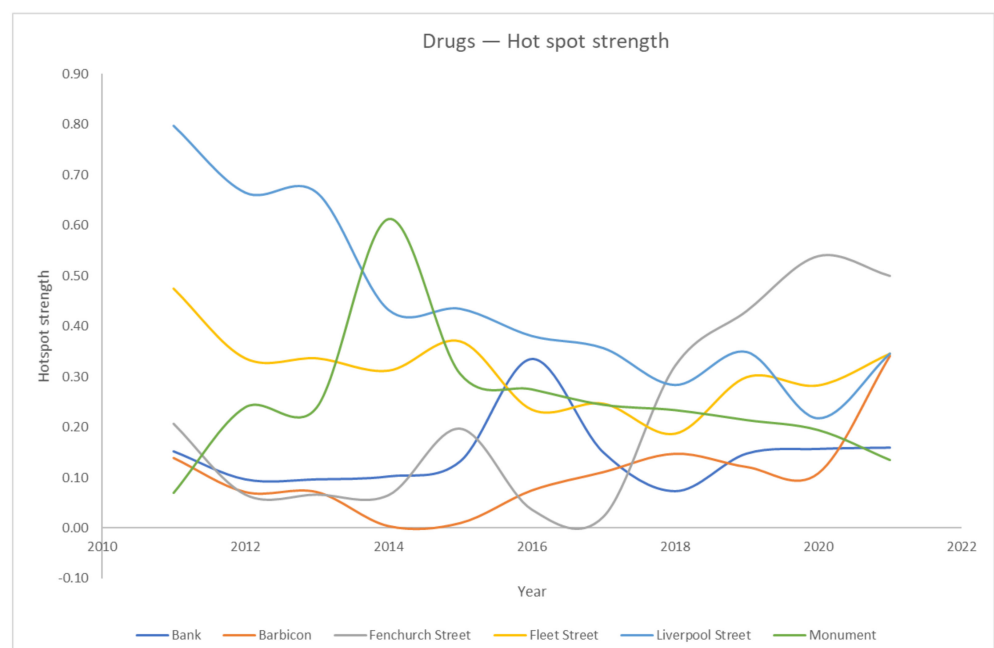


Figure 10. Trends of the hot spot strengths for drug crimes in all six clusters of the City of London.

Of particular importance in Figure 10 are the decrease over time of the hot spot strength in Bank (which halves from 80% in 2011 to 40% in 2021) and the increase in Fenchurch Street (which is zero in 2017 and reaches 50% in 2020).

The map in Figure 11 shows the areas covered by hot spots detected for drug crime events in the years 2017 (marked in blue) and 2018 (marked in red) in the Bank cluster.

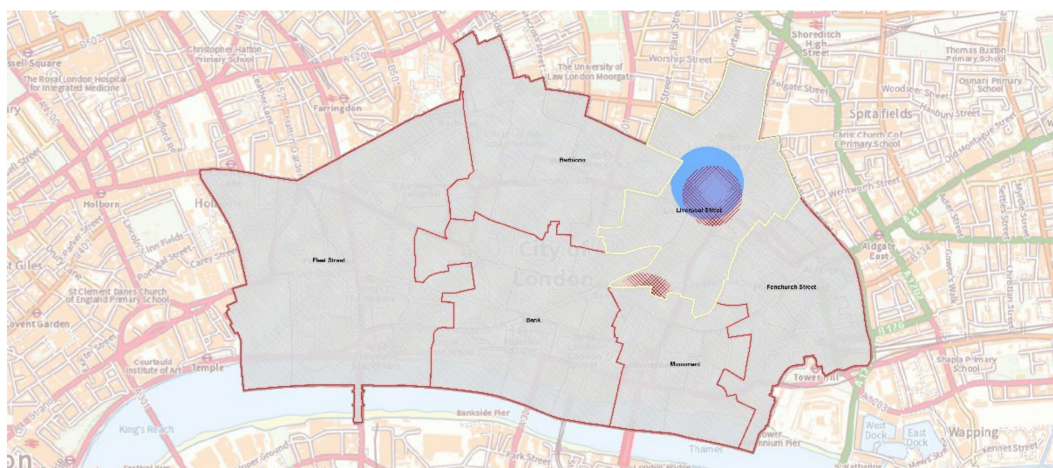


Figure 11. Shoplifting hot spots in the Liverpool Street cluster for the years 2017 (marked in blue) and 2018 (marked in red).

Figures 12–14 respectively show the areas covered by hot spots detected for drug crime events in the years 2018–2019, 2019–2020, and 2020–2021.

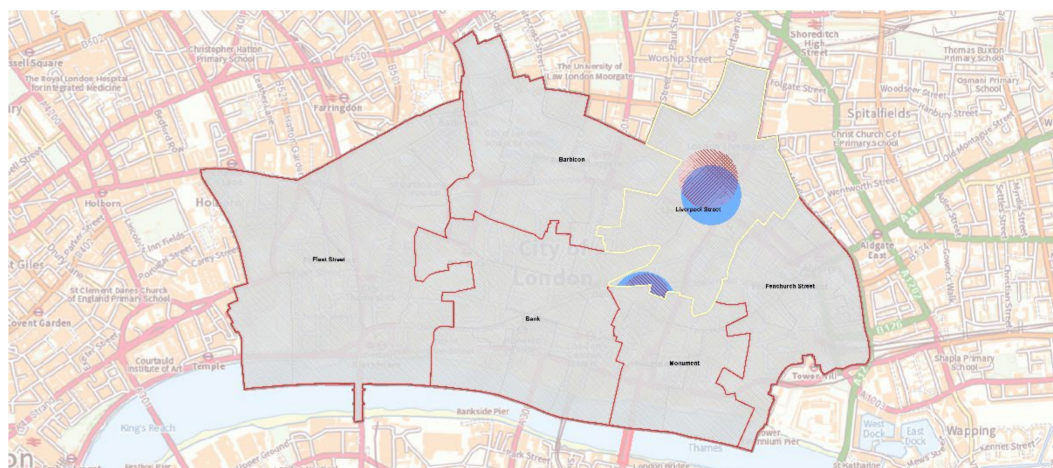


Figure 12. Shoplifting hot spots in the Liverpool Street cluster for the years 2018 (marked in blue) and 2019 (marked in red).

In the cluster there are two hot spots, one central and the other in the border area with the Monument cluster; the latter appears starting from 2018. Of particular significance is a reduction in the central hot spot from 2017 to 2021.

Figure 15 plots the trend of the hot spot strengths detected in the London Street cluster for shoplifting from 2011 to 2021.

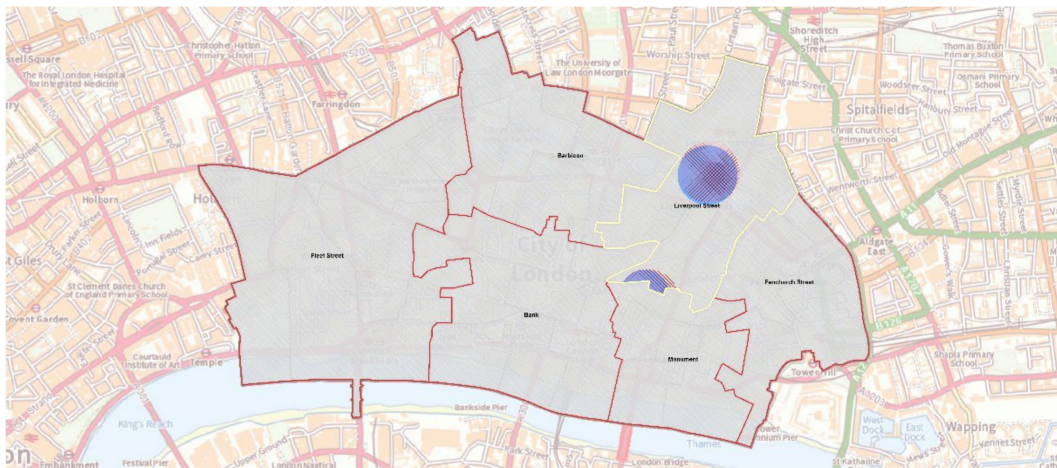


Figure 13. Shoplifting hot spots in the Liverpool Street cluster for the years 2019 (marked in blue) and 2020 (marked in red).

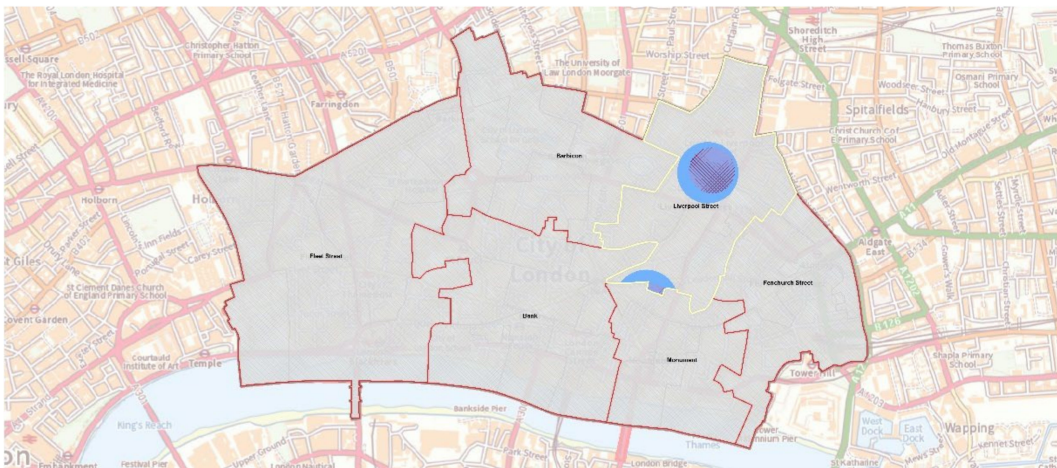


Figure 14. Shoplifting hot spots in the Liverpool Street cluster for the years 2020 (marked in blue) and 2021 (marked in red).

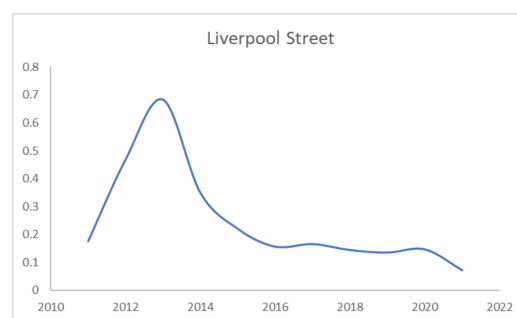


Figure 15. Trend of the hot spot strength for shoplifting in the Liverpool Street cluster.

This trend is similar to the one shown in Figure 8 for drug crimes in the Bank cluster. The hot spot strength reaches a maximum in 2013 of 70% and then decreases to reach a minimum of 10% in 2021.

Figure 16 shows the trend of the reliability of the hot spot strength calculated by (8).



Figure 16. Trend of the reliability of the hot spot strength for shoplifting in the Liverpool Street cluster.

The reliability fluctuates between a minimum value of 81%, reached in 2013, and a maximum value of 88%. Since 2016, it has remained approximately equal to this value.

To analyze the presence of correlations between the hot spot strength and the extension of areas with high data point density, the ratio between the extension of areas with annual data point density per square kilometer greater than the threshold and the extension of the Liverpool Street DWO cluster is calculated in each year. Here, we use three threshold values, 200, 300, and 400 data points per square kilometer. Table 2 shows for each year both the hot spot strength and the values of these ratios obtained, with the thresholds set at 200 (D200), 300 (D300) and 400 (D400) data points per square kilometer.

Table 2. Annual values for the percentages of areas with numbers of data points greater than 200, 300, and 400 data points per square kilometer.

Year	Hot Spot Strength	D200	D300	D400
2011	0.18	0.18	0.07	0.07
2012	0.47	0.15	0.10	0.20
2013	0.68	0.04	0.04	0.16
2014	0.35	0.22	0.13	0.00
2015	0.22	0.24	0.10	0.04
2016	0.16	0.19	0.05	0.02
2017	0.17	0.08	0.02	0.01
2018	0.15	0.07	0.01	0.04
2019	0.14	0.06	0.02	0.05
2020	0.15	0.03	0.02	0.08
2021	0.07	0.06	0.01	0.05

Figures 17–19 show the trends for the D200, D300, and D400 indices.

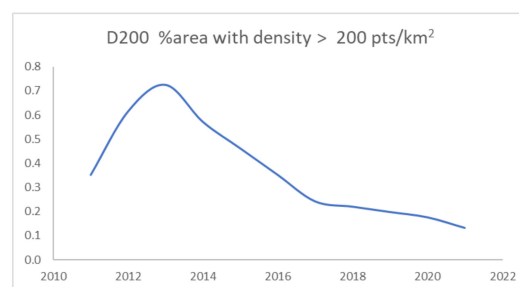


Figure 17. Trend of the percent of areas with density points greater than 200 pts/km² for shoplifting in the Liverpool Street cluster.

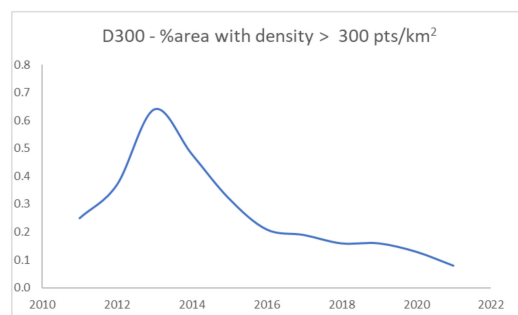


Figure 18. Trend of the percent of areas with density points greater than 300 pts/km² for shoplifting in the Liverpool Street cluster.

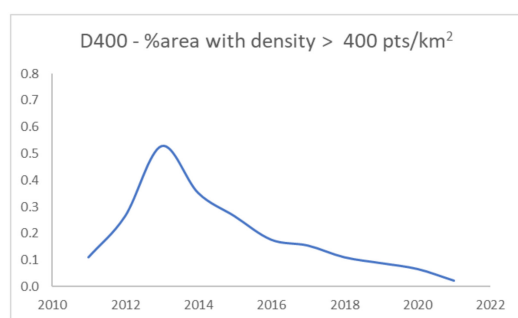


Figure 19. Trend of the percent of areas with density points greater than 400 pts/km² for shoplifting in the Liverpool Street cluster.

The three trends are similar to the hot spot strength trend. In particular, the trend for the D300 index is most similar to the trend for the hot spot strength, with a mean absolute difference of 5% (12% for D200 and 7% for D400) and a Pearson's linear correlation coefficient value of 0.935 (0.922 for D200 and 0.923 for D400). Similar trends are obtained for other crime types and in all DWO clusters.

Although the use of a simple density-based statistical analysis provides results approximately similar to those obtained by measuring the hot spot strength, this analysis, in addition to the disadvantage of depending on the choice of spatial density threshold, can only provide approximate results. In order to obtain more precise and reliable results, it is necessary to use density-based cluster algorithms; however, such algorithms have high computational complexity. The use of the hot spot strength method, on the one hand, involves linear computational complexity, as it executes an FCM-based cluster algorithm to detect the hot spots. Furthermore, it is able to provide reliable results and to evaluate this reliability by measuring the reliability of the hot spot strength.

Figure 20 plots the trends for the hot spot strengths obtained for all the six clusters for shoplifting in the City of London.

The trends in Figure 20 show a strong decrease in hot spot strength in Liverpool Street from 2013, which reaches a maximum value of 70%; in all other clusters, the strength never reaches a maximum value greater than 30%. For all clusters, the hot spot strengths for shoplifting are below 20% in 2021.

Trends for the hot spot strengths of other types of crimes also show decreases in recent years in all clusters of the City of London. These results show that in recent years, the control of the entire City of London by the police has been further intensified and improved. Furthermore, since the reliability of hot spot strength measures is always greater than 70%, and starting from 2016 is always greater than 80%, all of the results obtained can be considered reliable.

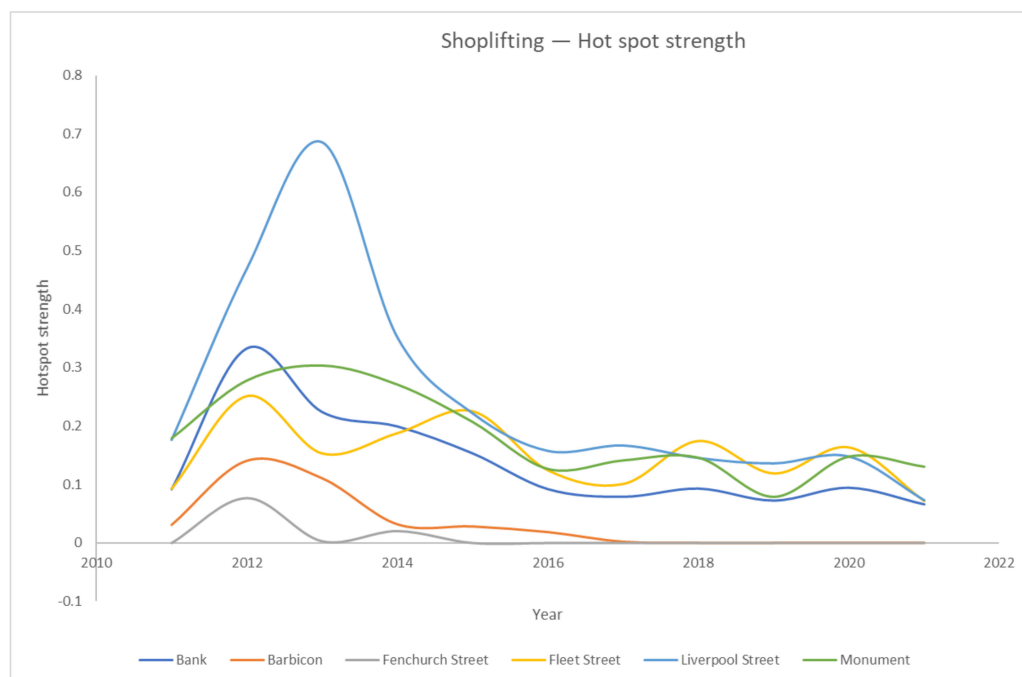


Figure 20. Trends of the hot spot strengths for shoplifting in all the six clusters of the City of London.

5. Conclusions

This paper presents a novel method aimed to analyze the spatiotemporal evolution of hot spots. The incidence of hot spots in a selected area is measured by calculating an index called the *hot spot strength*, and by evaluating its reliability using a method based on the De Luca and Termini fuzzy entropy measure.

This method is applied in crime analysis of the study area of the City of London, by acquiring the datasets published by the UK police relating to the various types of criminal events that occurred every year in the City of London from 2011 to 2021. The *hot spot strength* values measured on each cluster into which the city is partitioned and their reliability are calculated for each year and for each type of crime. The results show a general decrease in *hot spot strength* starting from 2016 in all clusters, which suggests an improvement in the control of the City of London by the wards in recent years. The *hot spot strength* measures can be considered reliable, as their reliability values are always greater than 0.7.

In the future, we intend to apply our method in different contexts and for the different problems, and to test it in GIS-based platforms to monitor the location, intensity, and temporal evolution of natural, anthropogenic, or climatic events in areas affected by a certain phenomenon.

Author Contributions: Conceptualization, B.C. and F.D.M.; methodology, B.C. and F.D.M.; software, B.C. and F.D.M.; validation, B.C. and F.D.M.; formal analysis, B.C. and F.D.M.; investigation, B.C. and F.D.M.; resources, B.C. and F.D.M.; data curation, B.C. and F.D.M.; writing—original draft preparation, B.C. and F.D.M.; writing—review and editing, B.C. and F.D.M.; visualization, B.C. and F.D.M.; supervision, B.C. and F.D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. MacQueen, J.B. Some Methods for Classification and Analysis of Multivariate Observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 21 June–28 July 1965 and 27 December 1965–7 January 1966; Le Cam, L.M., Neyman, J., Eds.; University of California Press: Oakland, CA, USA, 1967; Volume 1, pp. 281–297.
2. Levine, N. CrimeStat: A Spatial Statistical Program for the Analysis of Crime Incidents. In *Encyclopedia of GIS*; Shekhar, S., Xiong, H., Zhou, X., Eds.; Springer: Berlin/Heidelberg, Germany, 2010.
3. Agarwal, J.; Nagpal, R.; Sehgal, R. Crime Analysis Using K-Means Clustering. *Int. J. Comput. Appl.* **2013**, *83*, 4.
4. Sing, A.K.; Manimannan, G. Detecting Hot Spots on Crime Data Using Data Mining and Geographical Information System. *Int. J. Stat. Math.* **2013**, *8*, 5–9.
5. Hajela, G.; Chawla, M.; Rasool, A. A Clustering Based Hot Spot Identification Approach for Crime Prediction. *Procedia Comput. Sci.* **2020**, *167*, 1462–1470. [[CrossRef](#)]
6. Vadrevu, K.V.; Csiszar, I.; Ellicott, E.; Giglio, L.; Badarinath, K.V.S.; Vermote, E.; Justice, C. Hot Spot Analysis of Vegetation Fires and Intensity in the Indian Region. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 224–228. [[CrossRef](#)]
7. Khairani, N.A.; Sutoyo, E. Application of K-Means Clustering Algorithm for Determination of Fire-Prone Areas Utilizing Hot Spots in West Kalimantan Province. *Int. J. Adv. Data Inf. Syst.* **2020**, *1*, 9–16. [[CrossRef](#)]
8. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2005; p. 342. ISBN 978-0471735786.
9. Tabarej, M.S.; Minz, S. Rough-Set Based Hot Spot Detection in Spatial Data. In *Advances in Computing and Data Sciences. ICACDS 2019. Communications in Computer and Information Science*; Singh, M., Gupta, P., Tyagi, V., Flusser, J., Ören, T., Kashyap, R., Eds.; Springer: Singapore, 2019; Volume 1046, pp. 356–368. [[CrossRef](#)]
10. Hardika, E.; Atmaja, S. Implementation of k-Medoids Clustering Algorithm to Cluster Crime Patterns in Yogyakarta. *Int. J. Appl. Sci. Smart Technol.* **2019**, *1*, 38–48.
11. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Plenum Press: New York, NY, USA, 1981; p. 272, ISBN 978-0306406713.
12. Bezdek, J.C.; Ehrlich, R.; Full, W. The fuzzy C-means Clustering Algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [[CrossRef](#)]
13. Bezdek, J.C.; Pal, S.K. *Fuzzy Models for Pattern Recognition: Methods that Search for Structure in Data*; IEEE Press: New York, NY, USA, 1992; p. 544, ISBN 978-0780304222.
14. Grubestic, T.H. On the Application of Fuzzy Clustering for Crime Hot Spot Detection. *J. Quant. Criminol.* **2006**, *22*, 77–105. [[CrossRef](#)]
15. Kaur, R.; Sehera, S.S. Analyzing and Displaying of Crime Hot Spots Using Fuzzy Mapping Method. *Int. J. Comput. Appl.* **2014**, *103*, 25–28.
16. Ansari, M.Y.; Prakash, A. Application of Spatio-Temporal Fuzzy C-Means Clustering for Crime Spot Detection. *Def. Sci. J.* **2018**, *68*, 374–380. [[CrossRef](#)]
17. Win, K.N.; Chen, J.; Chen, Y.; Fournier-Viger, P. PCPD: A Parallel Crime Pattern Discovery System for Large-Scale Spatio-temporal Data Based on Fuzzy Clustering. *Int. J. Fuzzy Syst.* **2019**, *21*, 1961–1974. [[CrossRef](#)]
18. Bandyopadhyaya, R.; Mitra, S. Fuzzy Cluster-Based Method of Hot Spot Detection with Limited Information. *J. Transp. Saf. Secur.* **2015**, *7*, 307–323.
19. Besag, J.; Newell, J. The detection of clusters in rare diseases. *J. R. Stat. Soc. A* **1991**, *154*, 143–155. [[CrossRef](#)]
20. Devroye, L.; Rugosi, G. *Combinatorial Methods in Density Estimation*; Springer: Berlin/Heidelberg, Germany, 2001; p. 208, ISBN 978-0387951171.
21. Chaney, S.; Ratcliffe, J. *GIS and Crime Mapping Chap. 6, Identifying Crime Hot Spots*; John Wiley & Sons: Hoboken, NJ, USA, 2013; p. 402, ISBN 978-0-470-86099-1.
22. Lin, Y.-P.; Chu, H.-J.; Wu, C.-F.; Chang, T.-K.; Chen, C.-Y. Hotspot Analysis of Spatial Environmental Pollutants Using Kernel Density Estimation and Geostatistical Techniques. *Int. J. Environ. Res. Public Health* **2011**, *8*, 75–88. [[CrossRef](#)] [[PubMed](#)]
23. Hariforush, H.; Bellalite, L. A New Integrated GIS-based Analysis to Detect hot spots: A Case Study of the City of Sherbrooke. *Accid. Anal. Prev.* **2019**, *130*, 62–74. [[CrossRef](#)]
24. Kumar, K.M.; Reddy, A.R.M. A fast DBSCAN clustering algorithm by accelerating neighbor searching using groups method. *Pattern Recognit.* **2016**, *58*, 39–48. [[CrossRef](#)]
25. Huang, Z.; Gao, S.; Cai, C.; Zheng, H.; Pan, Z.; Li, W. A rapid density method for taxi passengers hot spot recognition and visualization based on DBSCAN. *Sci. Rep.* **2021**, *11*, 9420. [[CrossRef](#)]
26. Di Martino, F.; Sessa, S. The Extended Fuzzy C-means Algorithm for hot spots in Spatio-temporal GIS. *Expert Syst. Appl.* **2011**, *38*, 11829–11836. [[CrossRef](#)]
27. Kaymak, U.; Setnes, M. Fuzzy Clustering with Volume Prototype and Adaptive Cluster Merging. *IEEE Trans. Fuzzy Syst.* **2002**, *10*, 705–712. [[CrossRef](#)]
28. Di Martino, F.; Sessa, S.; Barillari, E.S.; Barillari, M.S. Spatio-temporal hot spots and Application on a Disease Analysis Case via GIS. *Soft Comput.* **2014**, *18*, 2377–2384. [[CrossRef](#)]
29. Di Martino, F.; Pedrycz, W.; Sessa, S. Hierarchical Granular hot spots Detection. *Soft Comput.* **2020**, *24*, 1357–1376. [[CrossRef](#)]
30. Di Martino, F.; Sessa, S. Fuzzy Entropy-Based Spatial hot spot Reliability. *Entropy* **2021**, *23*, 531. [[CrossRef](#)] [[PubMed](#)]

31. De Luca, A.; Termini, S. Entropy and Energy Measures of Fuzzy Sets. In *Advances in Fuzzy Set Theory and Applications*; Gupta, M.M., Ragade, R.K., Yager, R.R., Eds.; North-Holland: Amsterdam, The Netherlands, 1979; pp. 321–338.
32. De Luca, A.; Termini, S. A Definition of Non-probabilistic Entropy in the Setting of Fuzzy Sets Theory. *Inf. Control.* **1972**, *20*, 301–312. [[CrossRef](#)]
33. Di Martino, F.; Sessa, S. A New Validity Index Based on Fuzzy Energy and Fuzzy Entropy Measures in Fuzzy Clustering Problems. *Entropy* **2020**, *22*, 1200. [[CrossRef](#)]
34. Cardone, B.; Di Martino, F. A Novel Fuzzy Entropy-Based Method to Improve the Performance of the Fuzzy C-Means Algorithm. *Electronics* **2020**, *9*, 554. [[CrossRef](#)]