Research Article

Hotspots Detection in Spatial Analysis via the Extended Gustafson-Kessel Algorithm

Ferdinando Di Martino[1] and Salvatore Sessa[1,2]

[1] *Dipartimento di Architettura, Università degli Studi di Napoli Federico II, Via Monteoliveto 3, 80134 Napoli, Italy*
[2] *Università degli Studi di Napoli Federico II, Centro Interdipartimentale per l'Analisi e la Progettazione Urbana Luigi Pisciotti, Via Toledo 402, 80134 Napoli, Italy*

Correspondence should be addressed to Ferdinando Di Martino; fdimarti@unina.it

We show a new approach for detecting hotspots in spatial analysis based on the extended Gustafson-Kessel clustering method encapsulated in a Geographic Information System (GIS) tool. This algorithm gives (in the bidimensional case) ellipses as cluster prototypes to be considered as hotspots on the geographic map and we study their spatiotemporal evolution. The data consist of georeferenced patterns corresponding to positions of Taliban's attacks against civilians and soldiers in Afghanistan that happened during the period 2004–2010. We analyze the formation through time of new hotspots, the movement of the related centroids, the variation of the surface covered, the inclination angle, and the eccentricity of each hotspot.

1. Introduction

Hotspot detection is a known spatial clustering process in which it is necessary to detect spatial areas on which specific events thicken [1]; the patterns are the events georeferenced as points on the map; the features are the geographical coordinates (latitude and longitude) of any event. Hotspot detection is used in many disciplines, as in crime analysis [2–4], for analyzing where crimes occur with a certain frequency, in fire analysis [5] for studying the phenomenon of forest fires, and in disease analysis [6–9] for studying the localization and the focuses of diseases. Generally speaking, for detecting more accurately the geometrical shapes of hotspot areas algorithms based on density [10, 11] are used and they measure the spatial distribution of patterns on the area of study, but these algorithms have a high computational complexity.

In [5, 12, 13] a new hotspot detection method based on the extended fuzzy C-means algorithm (EFCM) [14, 15] was proposed, which is a variation of the famous fuzzy C-means (FCM) algorithm that detects cluster prototypes as hyperspheres. With respect to the FCM algorithm, the EFCM algorithm has the advantages of determining recursively

the optimal number of clusters and being robust in the presence of noise and outliers. In [5, 12, 13] the EFCM is encapsulated in a GIS tool for detecting hotspots as circles displayed on the map. The pattern event dataset is partitioned according to the time of the event's detection, so each subset is corresponding to a specific time interval. The authors compare the hotspots obtained in two consecutive years by studying their intersection on the map. In this way it is possible to follow the evolution of a particular phenomenon studying how its incidence is shifting and spreading through time.

In this paper we present a new hotspot detection method based on the extended Gustafson-Kessel algorithm (EGK) [14, 15] for studying the spatiotemporal evolution of hotspots. Our aim is to improve the shape of the hotspots, maintaining a good computational complexity: indeed the EGK algorithm gives the cluster prototypes as hyperellipsoids and ellipses in the bidimensional case. The EGK algorithm is an extension of the Gustafson-Kessel (GK) algorithm [16] which we briefly present.

Let $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset R^n$ be a dataset composed of N patterns $\mathbf{x}_j = (x_{1j}, x_{2j}, \ldots, x_{nj})^T$, where x_{kj} is the kth
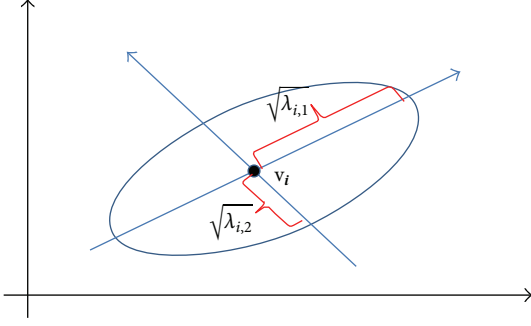
FIGURE 1: Example of ellipses cluster prototype using the GK algorithm.

component (feature) of the pattern $\mathbf{x}_j$. The GK algorithm minimizes the following objective function:

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

where C is the number of clusters fixed a priori, u_{ij} is the membership degree of the pattern $\mathbf{x}_j$ to the ith cluster ($i = 1, \ldots, C$), $\mathbf{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_C\} \subset R^n$ is the set of points given by the centroids of the C clusters, m is the fuzzifier parameter, and d_{ij} is the distance between $\mathbf{v}_i = (v_{1i}, v_{2i}, \ldots, v_{nj})^T$ and $\mathbf{x}_j$. The general form of this distance is given by

$$d_{ij} = \sqrt{\left(\mathbf{x}_j - \mathbf{v}_i\right)^T \mathbf{A}_i \left(\mathbf{x}_j - \mathbf{v}_j\right)}, \tag{2}$$

where $\mathbf{A}_i$ is the norm matrix, defined to be symmetric and positive. In the FCM algorithm $\mathbf{A}_i$ is equal to the identity matrix $\mathbf{I}$. In the GK algorithm the following Mahalanobis distance [17] is used:

$$d_{ij} = \sqrt{\left(\mathbf{x}_j - \mathbf{v}_i\right)^T \mathbf{P}_i^{-1} \left(\mathbf{x}_j - \mathbf{v}_j\right)}, \tag{3}$$

where $\mathbf{P}_i$ is the covariance matrix of the ith cluster given by

$$\mathbf{P}_i = \frac{\sum_{j=1}^{N} u_{ij}^m \left(\mathbf{x}_j - \mathbf{v}_i\right)\left(\mathbf{x}_j - \mathbf{v}_i\right)^T}{\sum_{j=1}^{N} u_{ij}^m}. \tag{4}$$

The covariance matrix $\mathbf{P}_i$ provides information about the shape and orientation of the cluster. The length of the kth axis of the hyperellipsoid is given by the root square of the kth eigenvalue λ_{ik} of $\mathbf{P}_i$. The directions of the axes of the hyperellipsoid are given by the directions of the eigenvectors of the matrix $\mathbf{P}_i$. In Figure 1 we show an example of ellipsoidal cluster prototype.

Using the Lagrange multipliers for minimizing objective function (1), we obtain the following solution for the centroids of each cluster prototype:

$$\mathbf{v}_i = \frac{\sum_{j=1}^{N} u_{ij}^m \mathbf{x}_j}{\sum_{j=1}^{N} u_{ij}^m}, \tag{5}$$
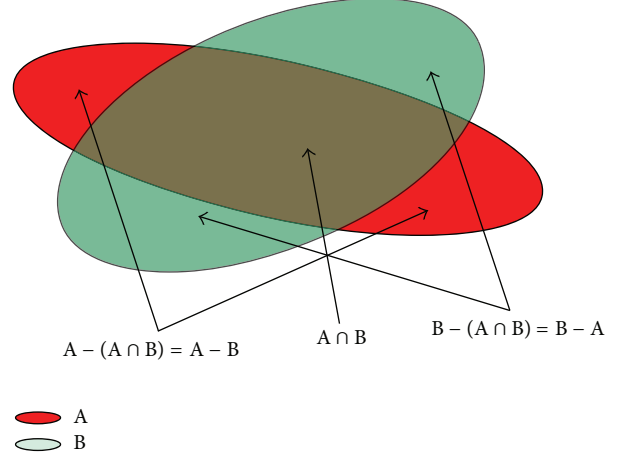

FIGURE 2: Intersection of two elliptical hotspots detected for events that happened in two consecutive periods.

where $i = 1, \ldots, C$ and u_{ij} are given by:

$$u_{ij} = \frac{1}{\left(\sum_{h=1}^{c} \left(d_{ij}^2 / d_{hj}^2\right)\right)^{2/(m-1)}}. \tag{6}$$

Initially the u_{ij}'s and the $\mathbf{v}_i$ are assigned randomly and updated in each iteration. If $U^{(l)} = (u_{ij}^{(l)})$ is the matrix $\mathbf{U}$ calculated at the lth step, the iterative process stops when

$$\left\|\mathbf{U}^{(l)} - \mathbf{U}^{(l-1)}\right\| = \max_{i,j} \left|u_{ij}^{(l)} - u_{ij}^{(l-1)}\right| < \varepsilon, \tag{7}$$

where $\varepsilon > 0$ is a prefixed parameter.

This algorithm is sensitive to the presence of outliers and noise and the number of cluster C is fixed a priori; as in the FCM algorithm, we need to use a validity index for determining an optimal value for the number of clusters C. In order to overcome these shortcomings, in [1, 16] the EGK algorithm is proposed which is a variation of the GK algorithm: there the optimal number of clusters is obtained during the iteration process. Furthermore, the EGK algorithm is robust with respect to the presence of noise and outliers.

In this paper we propose a new approach based on the EGK clustering method for detecting hotspots and studying their spatiotemporal evolution. Taking into consideration the bidimensional case, we obtain ellipses to be approximated as hotspot area better than the circular areas produced in the EFCM method.

Figure 2 shows an example of two intersecting elliptical hotspots, obtained as clusters detected by means of EGK method in two consecutive periods.

Figure 2 show three different regions:

(i) the area in which the hotspot A is not intersected by the hotspot B (corresponding to A − (A ∩ B) = A − B): this region can be considered as a set of geographical areas in which the prematurely detected event disappears successively;

(ii) the area of intersection A ∩ B: this area can be considered a geographical area in which the event persists in the course of time;

(iii) the area in which the hotspot B is not intersected by the hotspot A (corresponding to B − (A ∩ B) = B − A): this region can be considered as a set of geographical areas in which the prematurely not detected event propagates successively.

We can study the spatiotemporal evolution of the hotspots by analyzing the interactions between elliptical hotspots detected for consecutive periods, by verifying the presence of clusters in areas in which clusters have not yet been detected previously and the disappearance of clusters in areas previously covered by hotspots.

In this research we present a method for studing the spatiotemporal evolution of hotspots areas of war in Afghanistan; we apply the EGK algorithm for comparing consecutive years' event datasets corresponding to positions of Taliban's attacks against civilian and soldiers. Each event corresponds to the geolocalization of the site where Taliban's attack happened as well.

We study the spatiotemporal evolution of the hotspots by analyzing the intersections of hotspots corresponding to two consecutive years, the displacement of the centroids, the increase or reduction of the hotspots areas, and the emergence of new hotspots.

In Section 2 we give an overview of the EGK algorithm. In Section 3 we present our method for studying the spatiotemporal evolution of hotspots in spatial analysis. In Section 4 we present the results of the spatiotemporal evolution of hotspots. Our conclusions are contained in Section 5.

2. The EGK Algorithm

In the EGK algorithm we consider clustering prototypes given by hyperellipsoids in the n-dimensional feature's space. The ith hyperellipsoidal cluster prototype V_i is characterized by a centroid $\mathbf{v}_i = (v_{1i}, \ldots, v_{ni})$ and a mean radius r_i and we say that x_j belongs to V_i if $d_{ij} \leq r_i$.

The radius r_i is obtained considering the covariance matrix $\mathbf{P}_i$ of the ith cluster, defined by (4), whose determinant gives the volume of the ith cluster. Since $\mathbf{P}_i$ is symmetric and positive, it can be decomposed in the form

$$\mathbf{P}_i = \mathbf{Q}_i \mathbf{\Lambda}_i \mathbf{Q}_i^T, \tag{8}$$

where $\mathbf{Q}_i$ is an orthonormal matrix and $\lambda_i = (\lambda_{ij})$ is a diagonal matrix. The mean radius r_i is given as [15]

$$r_i = \frac{1}{n} \sqrt{\prod_{k=1}^{n} \lambda_{ik}^{1/n}} = \sqrt{(\det(\mathbf{P}_i))^{1/n}}. \tag{9}$$

In the EGK algorithm the objective function to be minimized is

$$J(\mathbf{X}, \mathbf{U}, \mathbf{V}) = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^m \left(d_{ij}^2 - r_i^2 \right), \tag{10}$$

where d_{ij} is given by the Mahalanobis distance (3). The solutions obtained for the centroids $\mathbf{v}_i$ are given by formula (5) and the functions u_{ij} are obtained by

$$u_{ij} = \frac{1}{\sum_{h=1}^{C} \left(d_{ij}^2 w_{ij} / d_{hj}^2 w_{hj} \right)^{1/(m-1)}}, \tag{11}$$

where $w_{kj} = \max(0, 1 - (r_k^2 / d_{kj}^2))$ for $k = 1, \ldots, C$. By setting $d_{kj}^2 w_{kj} = \max(0, d_{kj}^2 - r_k^2)$ and $\varphi_j = \text{card}\{k \in \{1, \ldots, C\} : d_{kj}^2 \leq r_k^2\}$ for any $j = 1, \ldots, N$, we have that formula (11) holds if $\varphi_j = 0$, while if $\varphi_j > 0$, one defines u_{ij} in the following way:

$$u_{ij} = \begin{cases} 0 & \text{if } d_{ij}^2 w_{ij} > 0 \\ \dfrac{1}{\varphi_j} & \text{if } d_{ij}^2 w_{ij} = 0. \end{cases} \tag{12}$$

Formula (12) produces the negative effect of diminishing the objective function (10) when a meaningful number of features are placed in a cluster; this effect can prevent the separation of the clusters. In order to solve this problem in [15], one starts with a small value r_i and by setting gradually $r_i := r_i + \beta^{(i)} / C^{(i)}$, where $C^{(i)}$ is the number of clusters at the ith iteration and $\beta^{(i)}$ is defined recursively as $\beta^{(0)} = 1$, $\beta^{(i)} = \min(C^{(i-1)}, \beta^{(i-1)})$ ($C^{(0)}$ is the initial number of clusters). By setting

$$I_{ik} = \frac{\sum_{j=1}^{N} \min \left(u_{ij}, u_{kj} \right)}{\sum_{j=1}^{N} u_{ij}}, \tag{13}$$

one defines the symmetric matrix $S = (S_{ik})$, where $S_{ik} = \max\{I_{ik}, I_{ki}\}$. If $S^{(l)}$ is the matrix S at the lth iteration ($l > 1$) and the threshold $\alpha^{(l)} = 1/(C^{(l)} - 1)$ is introduced, two indexes i^* and k^* are determined such that $S_{i^* k^*}^{(l)} = \max\{S_{ik}^{(l)} : i = 1, \ldots, N; k = 1, \ldots, C\} \geq \alpha^{(l)}$. Then these indices are merged by setting, for any $j = 1, \ldots, N$,

$$\begin{aligned} u_{i^* j}^{(l)} &:= u_{i^* j}^{(l)} + u_{k^* j}^{(l)}, \\ C^{(l)} &:= C^{(l-1)} + 1; \end{aligned} \tag{14}$$

thus, the k^*th row can be removed from the matrix $U^{(l)}$. In conclusion the following steps hold for the EGK algorithm.

(1) The user assigns initially $C^{(0)}$, $m > 0$ (usually $m = 2$), $\varepsilon > 0$, $S_{ik}^{(0)} = 0$, and $\beta^{(0)} = 1$.

(2) $u_{ij}^{(0)}$ are fixed randomly.

(3) v_i and r_i are calculated with formulae (5) and (9), respectively.

(4) u_{ij} are calculated with formulae (11) and (12).

(5) Determine i^* and k^* such that $S_{i^* k^*}^{(l)} = \max\{S_{ik}^{(l)} : i = 1, \ldots, N; k = 1, \ldots, C\}$.

(6) If $S_{i^* k^*}^{(l)} > \alpha^{(l)}$, then the i^*th and k^*th clusters are merged via formula (14) and the k^*th row is deleted from $U^{(l)}$.

(7) If formula (7) is satisfied, then the process stops otherwise go to (3) for the $(l+1)$th iteration.
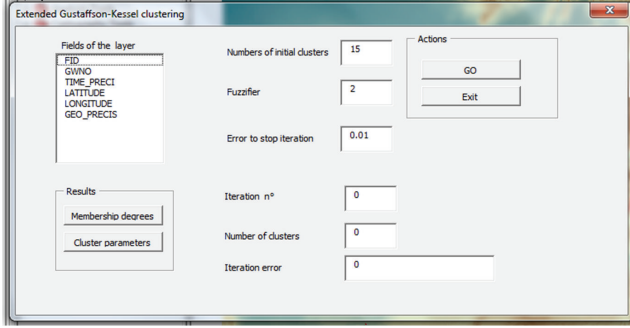
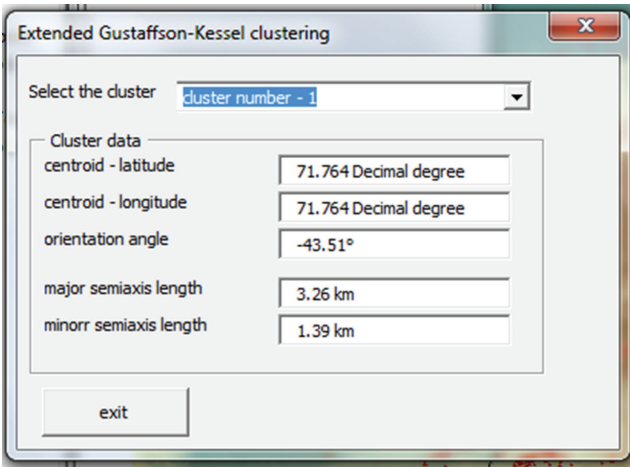Figure 3: Mask created in the ESRI/ArcGIS tool for managing the EGK process.

Figure 4: Mask created in the ESRI/ArcGIS tool for displaying the information of the detected cluster prototypes.

Figure 5: Spatiotemporal evolution of hotspots detected in three consecutive years.

Table 1: Results of the EGK applied to the event's subsets.

Year	Initial number of clusters	Final number of clusters	$|U^{(l)} - U^{(l-1)}|$	ε
2004–2006	15	7	0.79×10^{-2}	1×10^{-2}
2007	15	7	0.65×10^{-2}	1×10^{-2}
2008	15	7	0.68×10^{-2}	1×10^{-2}
2009	15	8	0.74×10^{-2}	1×10^{-2}
2010	15	8	0.81×10^{-2}	1×10^{-2}

3. Hotspots Detection and Evolution in Military War

Each pattern is given by the event corresponding to a place in which an attack has occurred. The two features of the pattern are the geographic coordinates of this place.

We divide the event dataset into subsets corresponding to the events that occurred in a specific year or set of years. For each subset of events we apply the EGK algorithm to detect the final cluster prototypes.

The dataset is extracted from the URL http://www.acleddata.com/data/asia/; the data are the geolocalizations of Taliban's attacks in Afghanistan during the period 2004–2010. The EGK algorithm is encapsulated in the ESRI/ArcGIS tool. Figure 3 shows the mask used for setting the parameters and running the EGK algorithm.

We can set other numerical fields for adding other features to the geographical coordinates. Initially we set the initial number of clusters, the fuzzifier m (equal to 2 by default), and the error threshold for stopping the iterations (equal to 0.01 by default). At the end of the process we displayed on the form of the number of iterations, the final number of clusters, and the error calculated at the last ite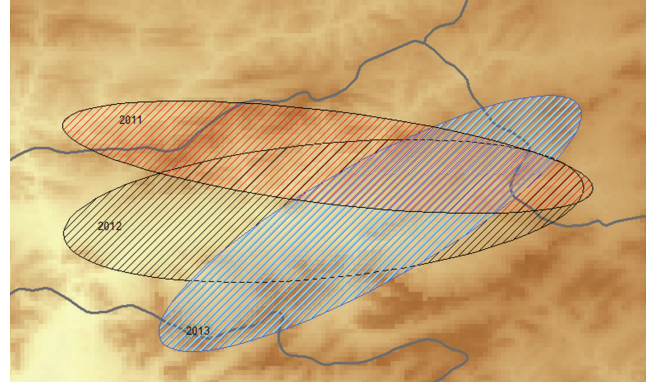ration. The resultant clusters are shown as ellipses on the geographical map and can be saved in a new geographical layer.

In Figure 4 we show the mask used for displaying the information of each elliptical prototype detected: centroid's coordinates, length of each semiaxis, and orientation of the ellipses with respect to the horizontal plane on the geographical map.

The final process concerns the comparative analysis of the hotspots obtained by the final clusters resulting for each subset of events. Figure 5 shows an example of the display of hotspots obtained as final clusters corresponding to three consecutive years.

In order to assess the expansion and the displacement of any hotspot, we measure the area covered by each hotspot, the distance between the centroids of two intersecting hotspots detected in consecutive periods, the variation of the inclination angle, the eccentricity, and the length of both semiaxis.

4. Test Results

After partitioning the dataset in the five periods 2004–2006, 2007, 2008, 2009, and 2010, respectively, we apply the EGK algorithm for detecting the sequences of elliptical cluster prototypes. We fix $m = 2$, $C^{(0)} = 15$, and $\varepsilon = 1 \times 10^{-2}$. Table 1 shows the results obtained for each period.

We present the details relating to the comparison of the hotspots by considering the event data that occurred in the five periods. In Figures 6, 7, 8, 9, and 10 we show the hotspots detected.
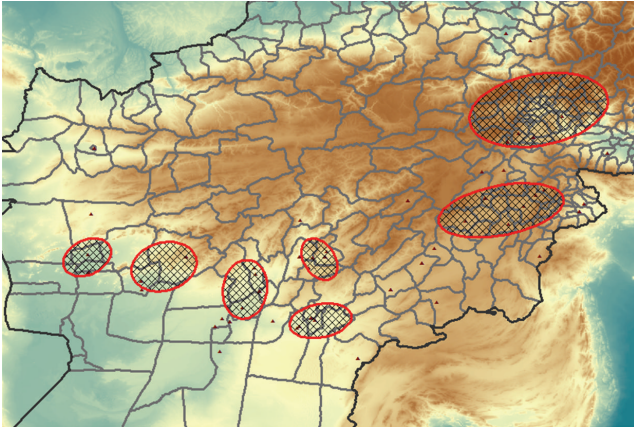
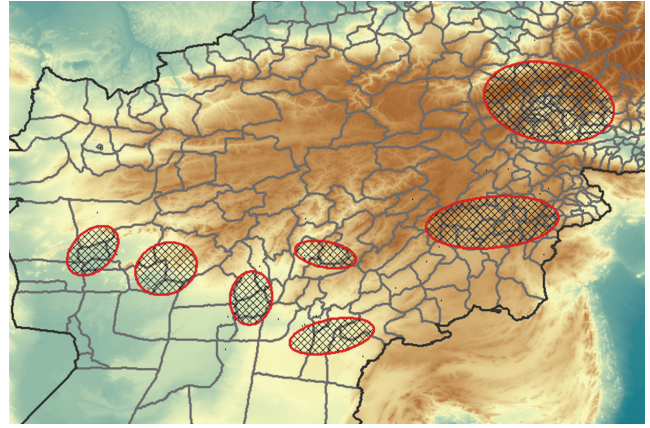FIGURE 6: Taliban's attack events: years 2004–2006.

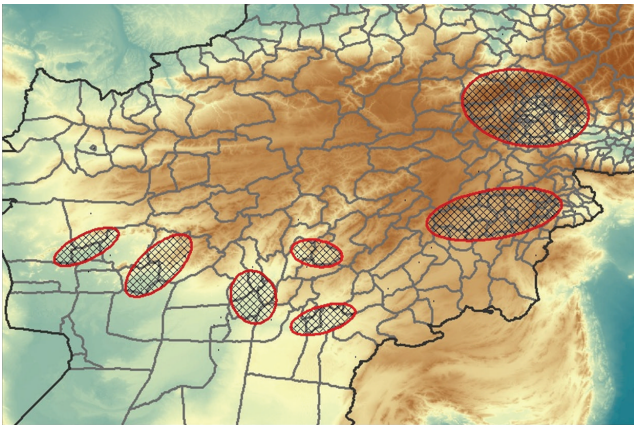FIGURE 8: Taliban's attack events: year 2008.
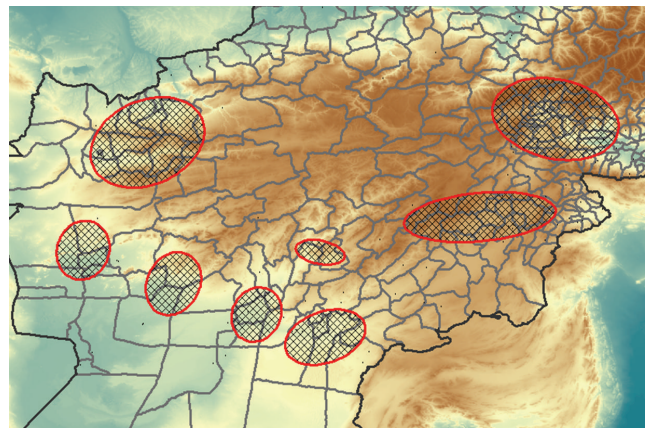

FIGURE 7: Taliban's attack events: year 2007.

FIGURE 9: Taliban's attack events: year 2009.

TABLE 2: Areas of hotspots detected in 2009 and 2010.

Hotspot ID	Area hotspot 2009	Area hotspot 2010	Intersection area	% of Intersection area
1	18804.11	18856.72	14936.25	79.43%
2	13580.19	13583.15	10632.51	78.29%
3	1516.89	1538.23	1011.87	66.71%
4	7889.86	6665.24	5174.83	65.59%
5	4987.08	6086.51	3706.52	74.32%
6	6361.20	5691.76	4490.73	70.60%
7	5761.71	7866.39	5175.92	89.83%
8	18460.84	23471.45	16255.32	88.05%

By analyzing Figures 6–8 we can deduce that in the period 2004–2008 seven hotspot areas approximated as ellipses are present; in these periods each hotspot modified only slightly its angle, width, and position of the centroid. In the years 2009 and 2010 a new hotspot is detected in a region neighboring with Turkmenistan. In Figure 10 the hotspots obtained for two consecutive years 2009 and 2010 are overlapped as well. In blue (resp., red) we enumerate the hotspots corresponding to the year 2009 (resp., 2010); see Figure 11. The hotspots are labeled, and the hotspot number 8 is the new hotspot detected and coming from overlap of the related hotspots.

In Table 2 the first column shows the labels of each hotspot; the second and third columns show the area, in km^2, of the hotspot detected in 2009 and 2010, respectively. The fourth column (resp., fifth) shows the intersection area of the two hotspots (resp., the percentage of area of the hotspot detected in 2009 covered by the corresponding hotspot detected in 2010, that is, the ratio "intersection area/area hotspot detected in 2009").

The results in Table 2 show that over 65% of the area of each hotspot detected in 2009 is also covered by the corresponding hotspot detected in 2010. Another significant result is the increase of the area of the hotspot 8, which exceeds 2×10^4 km^2 in 2010. In Table 3 we show the eccentricity of each hotspot and the distance between the centroids of each hotspot detected in 2009 and the corresponding one detected in 2010.

The results show that the eccentricity increases significantly in 2010 for hotspots 4 and 6, whereas it decreases for hotspot 3; the eccentricity remains almost unchanged for the remaining hotspots in 2010. Another significant result is the
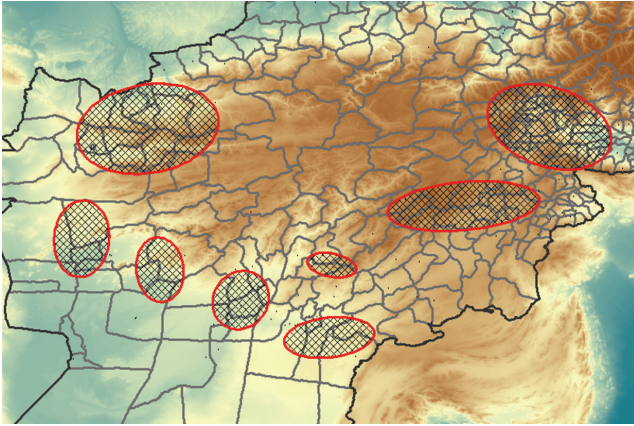
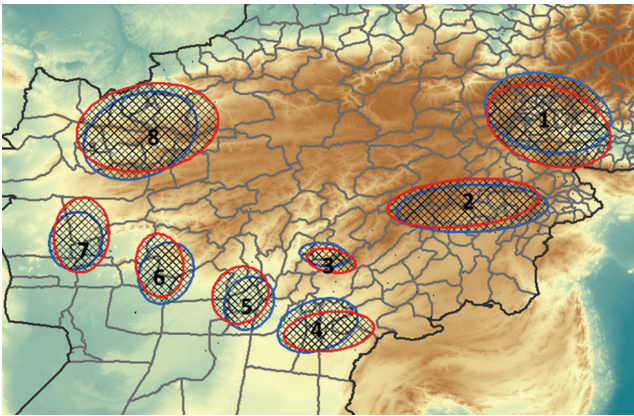Figure 10: Taliban's attack events: year 2010.

Figure 11: Hotspots detected for two consecutive periods: 2009 (blue) and 2010 (red).

Table 3: Eccentricity and centroid's distance between the hotspots detected in 2009 and 2010.

Hotspot ID	Eccentricity hotspot 2009	Eccentricity hotspot 2010	Centroids distance (km)
1	0.796	0.786	16.52
2	0.947	0.949	17.86
3	0.897	0.906	23.30
4	0.672	0.899	42.08
5	0.564	0.381	21.55
6	0.591	0.701	23.60
7	0.536	0.699	25.66
8	0.669	0.693	33.87

distance exceeding 40 km between the centroid of hotspot 4 detected in 2009 and the centroid of the corresponding hotspot detected in 2010.

5. Conclusions

We present a new approach for detecting hotspots in spatial analysis using the EGK clustering method encapsulated in a GIS tool. Similar to the EFCM algorithm, the EGK method is robust with respect to noise and outliers and we obtain the optimal number of the clusters iteratively during the process; furthermore, it has the advantage to detect hotspots of elongated shape. In our experiments we consider the site of Taliban's attacks in Afghanistan during the period 2004–2010. The spatial dataset is partitioned into subsets in order to study the evolution of the hotspots through time. We study the evolution of each hotspot in terms of movement of the centroids, surface covered, inclination, and eccentricity. The results show the formation, starting from 2009, of a new hotspot in the north-western zone neighboring with Turkmenistan. The results of the comparison of the hotspots detected in 2009 and 2010 show that this hotspot is increased with an extension of (about) $2 \times 10^4 \, km^2$.
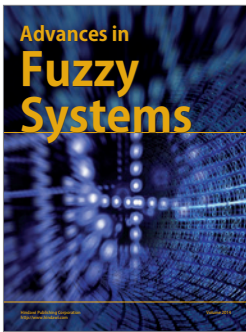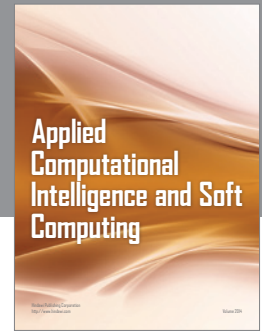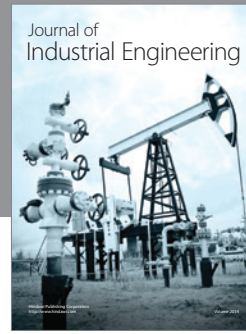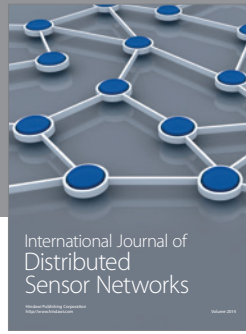
References

[1] T. H. Grubesic and A. T. Murray, "Detecting hotspots using cluster analysis and GIS," Annual Conference of CMRC, Dallas, 2001, http://www.ojp.usdoj.gov/ .

[2] S. P. Chainey, S. Reid, and N. Stuart, "When is a hotspot a hotspot? A procedure for creating statistically robust hotspot geographic maps of crime," in *Innovations in GIS 9: Socioeconomic Applications of Geographic Information Science*, D. Kidner, G. Higgs, and S. White, Eds., Taylor and Francis, London, UK, 2002.

[3] K. Harries, *Geographic Mapping Crime: Principle and Practice*, National Institute of Justice, Washington, DC, USA, 1999.

[4] A. T. Murray, I. McGuffog, J. S. Western, and P. Mullins, "Exploratory spatial data analysis techniques for examining urban crime," *British Journal of Criminology*, vol. 41, no. 2, pp. 309–329, 2001.

[5] F. Di Martino and S. Sessa, "The extended fuzzy C-means algorithm for hotspots in spatio-temporal GIS," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11829–11836, 2011.

[6] M. R. Barillari, U. E. Barillari, F. Di Martino, R. Mele, I. Perfilieva, and S. Senatore, "Spatio-temporal hotspot analysis for exploring evolution of diseases: an application to oto-laryngo-pharingeal diseases," *Advances in Fuzzy Systems*, vol. 2013, Article ID 385974, 7 pages, 2013.

[7] R. M. Mullner, K. Chung, K. G. Croke, and E. K. Mensah, "Geographic information systems in public health and medicine," *Journal of Medical Systems*, vol. 28, no. 3, pp. 215–221, 2004.

[8] K. Polat, "Application of attribute weighting method based on clustering centers to discrimination of linearly non-separable medical datasets," *Journal of Medical Systems*, vol. 36, no. 4, pp. 2657–2673, 2012.

[9] C.-K. Wei, S. Su, and M.-C. Yang, "Application of data mining on the development of a disease distribution map of screened community residents of taipei county in Taiwan," *Journal of Medical Systems*, vol. 36, no. 3, pp. 2021–2027, 2012.

[10] I. Gath and A. B. Geva, "Unsupervised optimal fuzzy cluster-ing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 773–780, 1989.

[11] R. Krishnapuram and J. Kim, "Clustering algorithms based on volume criteria," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 2, pp. 228–236, 2000.

[12] F. Di Martino, V. Loia, and S. Sessa, "Extended fuzzy c-means clustering algo-rithm for hotspot events in spatial analysis," *International Journal of Hybrid Intelligent Systems*, vol. 4, pp. 1–14, 2007.

[13] F. Di Martino and S. Sessa, "Implementation of the extended fuzzy C-means algorithm in geographic information systems," *Journal of Uncertain Systems*, vol. 3, no. 4, pp. 298–306, 2009.

[14] U. Kaymak, R. Babuska, M. Setnes, H. B. Verbruggen, and H. M. van Nauta Lemke, "Methods for simplification of fuzzy models," in *Intelligent Hybrid Systems*, D. Ruan, Ed., pp. 91–108, Kluwer Academic, Dordrecht, The Netherlands, 1997.

[15] U. Kaymak and M. Setnes, "Fuzzy clustering with volume pro-totypes and adaptive cluster merging," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 6, pp. 705–712, 2002.

[16] D. E. Gustafson and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in *Procedings of the 17th IEEE Conf Decis Control Incl Symp Adapt Processes*, pp. 761–766, San Diego, Calif, USA, January 1979.

[17] R. Gnanadesikan and J. R. Kettenring, "Robust estimates, resid-uals, and outlier detection with multiresponse data," *Biometrics*, vol. 28, pp. 81–124, 1972.