

## SEARCHING POSSIBLE ACCIDENT BLACK SPOT LOCATIONS WITH ACCIDENT ANALYSIS AND GIS SOFTWARE BASED ON GPS COORDINATES

<sup>1</sup> Pál HEGYI, <sup>2</sup> Attila BORSOS, <sup>3</sup> Csaba KOREN

Department of Transport Infrastructure, Széchenyi István University  
Egyetem tér 1, Győr, Hungary, 9026  
e-mail: <sup>1</sup> hegyip@sze.hu, <sup>2</sup> borsosa@sze.hu, <sup>3</sup> koren@sze.hu

Received 31 December 2016; accepted 19 June 2017

**Abstract:** There are several methods for the analysis of road accidents in a road network. In Hungary from 2011 GPS coordinates are used to identify the location of personal injury accidents. This method significantly improves the display of locations of accidents on the map, which can be then analyzed using GIS tools. Accident black spots are the most dangerous places in road networks identified by the density of the accidents in the network. One of the analysis methods is the accident density searching. The methods and algorithms used in some software may show differences in relation to one another. The aim of this research is comparing two applications by investigating the local road network in Győr. The analysis was made using the WEB-BAL accident analysis software using the density-based spatial clustering of applications with noise procedure and the QGIS software using the kernel density estimation method. The former is the official accident database and online software used for accident investigations and the latter is an open source geographic information system. The results are visualized in accident density plots and black spot maps.

**Keywords:** Accident analysis, Geographic information system, Road network, Accident black spot

### 1. Introduction

Black spots are the most dangerous locations in road networks; therefore the identification of these locations is important. This paper deals with the visualization of the possible locations of the black spots with two different methods. One of the methods is the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) data mining procedure, and the other is the Kernel Density Estimation (KDE) method. The

purpose of this investigation is to compare the results provided by the two methods in an urban road network. The visualization shows the possible black spots and can be used to select the real ones for further investigations.

## 2. Features of black spots

There are locations in the rural and urban road network where the number of accidents is relatively higher than at other similar locations. They can be either junctions or road sections. Definitions of black spots vary from country to country [1]. The definition of black spot by the Hungarian guideline [2] is:

- minimum 4 personal injury accidents per 3 consecutive years;
- in case of junctions the area of intersection with a given radius; or
- in case of road sections at least 100 m long section in urban areas and 1000 m long section in rural areas).

In Hungary accidents are localized by the road number and the kilometer post on rural roads and based on the postal address in the urban road network. From 2011 GPS coordinates are added to the identification of the location of personal injury accidents in both cases.

## 3. Database

The investigation of accidents was conducted in the local road network of the city of Győr using the algorithms DBSCAN and KDE. The map data came from OpenStreetMap (OSM) [3]. This is currently the only one available free of charge and up to date road map. Accident data came from the accident analysis software 'WEB-BAL'. These data were exported into an Excel sheet, based on which a shape file in GIS was created [4]. The time period investigated was three years from the beginning of 2011 until the end of 2013. The number of accidents was 643 with 737 personal injuries. *Fig. 1* shows the location of the accidents in the inner part of Győr.

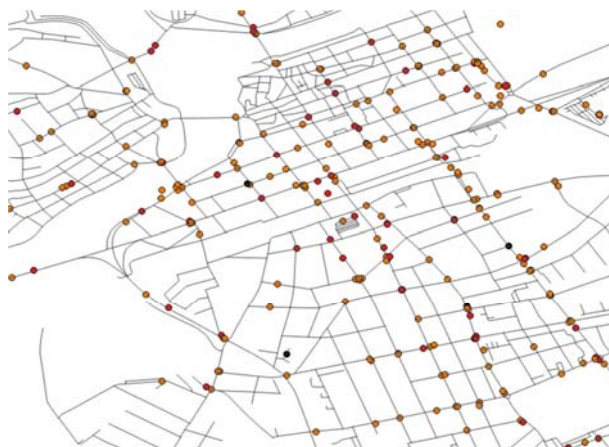
On the basis of the accident locations on the map, the two different methods can display locations where the density of the accidents is high; these are the so-called possible black spots. It is necessary to further investigate these locations to determine the real black spots, where the concentration of accidents is not a mere coincidence. The latter procedure is not subject of this paper.

## 4. Method of analysis

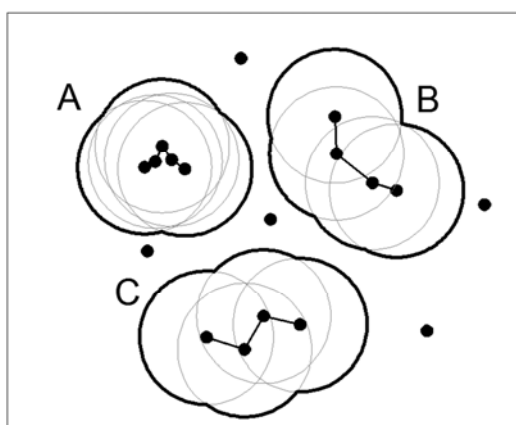
### 4.1. Density-based spatial clustering of applications with noise

The WEB-BAL accident analysis software uses the DBSCAN data mining procedure to find the black spots on the map where accidents are localized by their GPS coordinates [5], [6], [7]. This method checks whether other points can be found in a given distance around a certain point [8]. If a new point is found, the procedure is

repeated from the new point (*Fig. 2*). The end of the procedure is when there are no new dots in the enlargement distance [9], [10]. *Fig. 2* shows a theoretical example with three groups and five independent dots where the dots represent the accident locations and the groups are the possible black spots.



*Fig. 1.* Personal injury accidents in the local road network of Győr



*Fig. 2.* Spatial clustering with the method DBSCAN

The results are showed by the characteristics of the black spots. In this case the following parameters can be used:

- Enlargement distance;
- Minimum number of accidents (4);
- Minimum weighted accident density ( $0.001 \text{ w. acc./m}^2$ );
- Minimum calculated area ( $50 \text{ m}^2$ ).

These parameters are defined in the Hungarian guideline [2]. The used values are in parentheses.

#### 4.2. Kernel density estimation

Many GIS software use the kernel density estimation method to create density maps. In this case only one parameter was used, the radius. The formula of the KDE method is given as follows [11]:

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{d_i}{h}\right), \quad (1)$$

where  $f_n(x)$  is the density estimate at spatial  $x$ ;  $n$  is the number of observations (crashes) near location  $x$  within a radius of  $h$ ;  $h$  is the predefined bandwidth;  $K$  is the predefined kernel density function to measure the distance decay effect;  $d_i$  is the distance between the spatial unit  $x$  and the spatial unit where the  $i$ -th crash is located (Fig. 3).

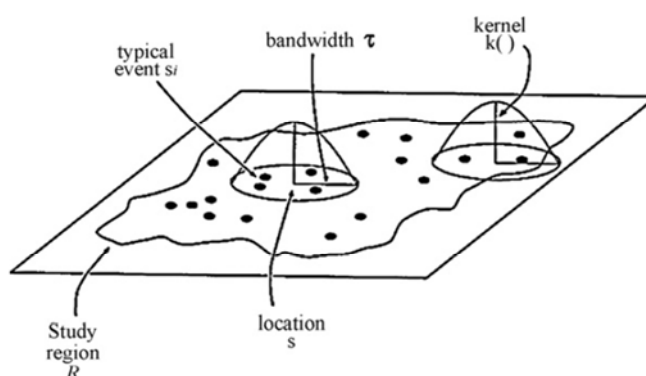


Fig. 3. Diagram of the kernel density estimation [12]

The results of this method are raster images. The heat-map appearance depends on the density of the dots [13]. The heat-map is a raster image, showing the density of the points. Each raster cell has a density value. The decay effect means that the value of the pixel will be zero in a given distance. 'The radius determines the distance around the point where the influence of the point will be felt' [14].

## 5. Results

In this chapter the results of the two methods are compared. In both cases several parameter values were applied, which influence the results of the visualizations. *Table I* shows the distances and the results of the number of the possible black spots, where the

distance is the enlargement distance in DBSCAN method, and the radius in KDE method.

Table I

The distances used and the number of black spots identified by the two methods

Distance [m]	Number of black spots found	
	DBSCAN	KDE
25	22	12
50	29	13
100	34	13
150	20	18
200	10	20

In case of DBSCAN method the results come from the WEB-BAL software [15] while the KDE method is based on the heat-map. The difference between the two methods is shown in *Fig. 4*.

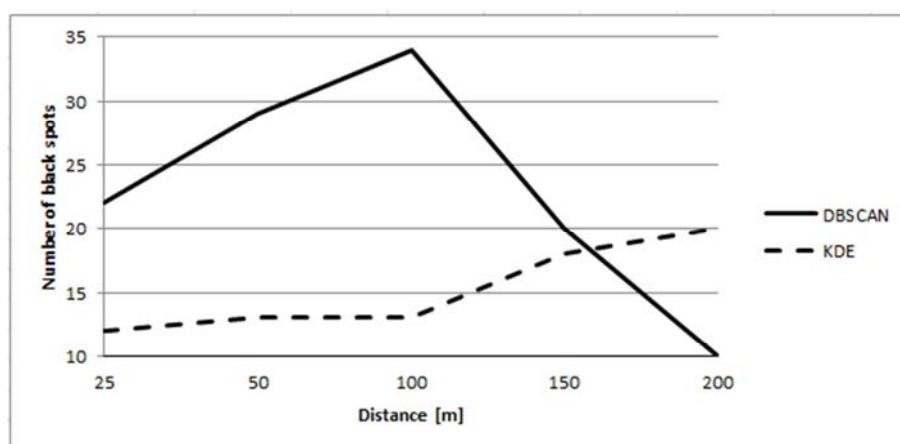


Fig. 4. Number of the black spots in the two methods

Contour lines produced on the basis of the heat-map have an attribute value called the 'height'. The QGIS software creates contour lines as a result of the values of the raster cells. In this case the number of black spots is the amount of the same height level lines. During the investigation various values were applied. *Table II* shows the attributes and the values used in all cases. In the first case the values used are the mean values of the height of the level lines with 1 meter level spacing. In the other cases different values for the 'height' attributes (being close to the mean value) and level spacing were used. In this paper results of black spot analysis made with the 'height' attribute 10 and level spacing 1 meter are presented.

Table II

The results of the number of black spots with different level spacing in QGIS software

Distance [m]	Number of black spots in level spacing [m]			Attribute of the contour lines (height) (Min-max value/ Used value)		
	1 meter	3 meter	4 meter	1 meter	3 meter	4 meter
25	12	13	13	1-21/10	1-21/9	1-20/8
50	13	14	15	1-21/10	1-21/9	1-20/8
100	13	18	21	1-21/10	1-21/9	1-20/8
150	18	18	23	1-21/10	1-21/9	1-20/8
200	20	25	27	1-21/10	1-21/9	1-20/8

Fig. 5 shows the additional values of the contour lines, which were used. There are different heights of the contour lines but the results are similar to each other. The maximum number of black spots is at the 100 m enlargement distance, while the lack spots continue to grow with distance in the KDE method.

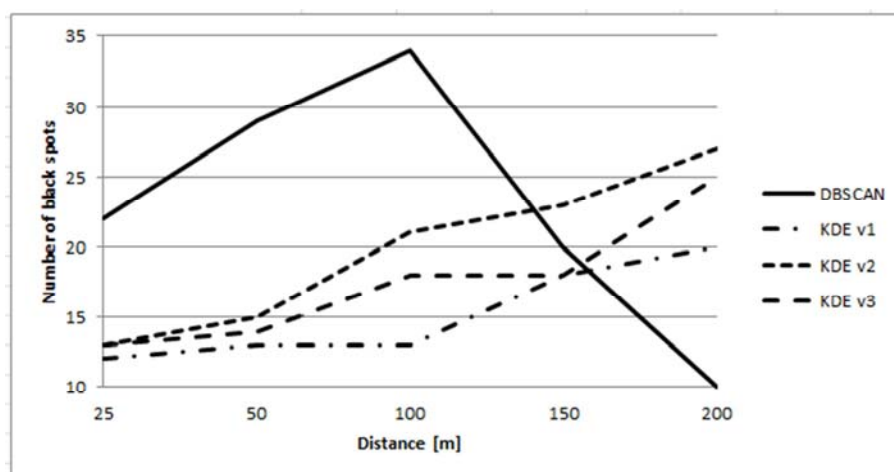


Fig. 5. Difference between the results of the two methods

The following figures (Fig. 6 - Fig. 7) show the comparison of the cartographic representations of black spots. In the first case the distance value is 50. The difference between the two methods is significant because the accident analysis software only identifies locations which meet the specified requirements of being a possible black spot, whereas the heat-map shows all the locations where there is an aggregation of accidents. The WEB-BAL software uses the OSM application to visualize the results, while the GIS software illustrates the heat-maps and the vector road network.

Using small distances the heat-map cannot illustrate black spots. In the left picture of Fig. 8 there are several possible black locations. They were found at junctions where there are more dots than in the sections of the streets. In this and in the following cases there are some well recognizable possible black spots in both maps. This is because

there are no high-density crash locations near the black spot. An example is shown in Fig. 8.

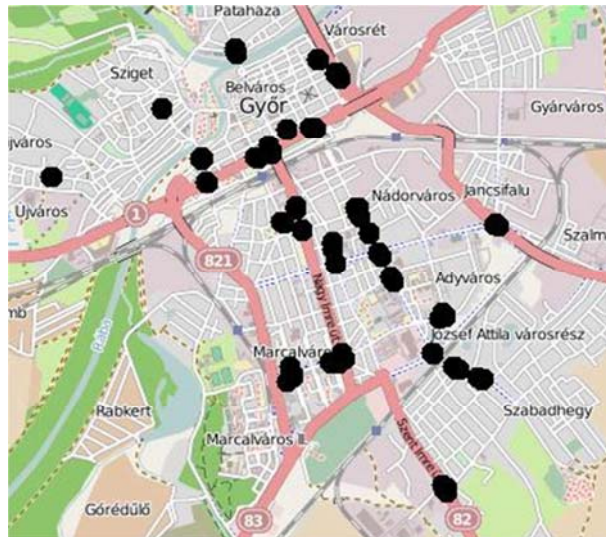


Fig. 6. Cartographic representations of the black spots (distance value 50 m)  
(WEB-BAL software)



Fig. 7. Cartographic representations of the high-density crash locations (distance value 50 m)  
(GIS software heat-map)

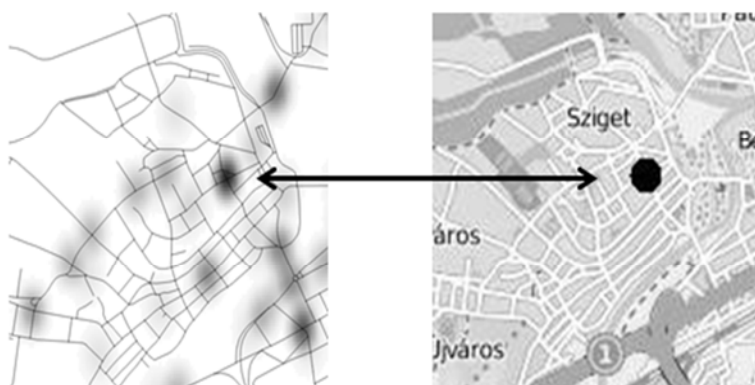


Fig. 8. Recognizable possible black spot on the maps  
(left side raster image, right side WEB-BAL software)

In the second case the distance value is 150 m. The value is used in the GIS software, and its result is the most similar to that of the DBSCAN method. Fig. 9 and Fig. 10 display several identical locations. The first map was made with the software WEB-BAL. It uses the method DBSCAN. On the other map high density crash locations are illustrated by the KDE method with GIS software.

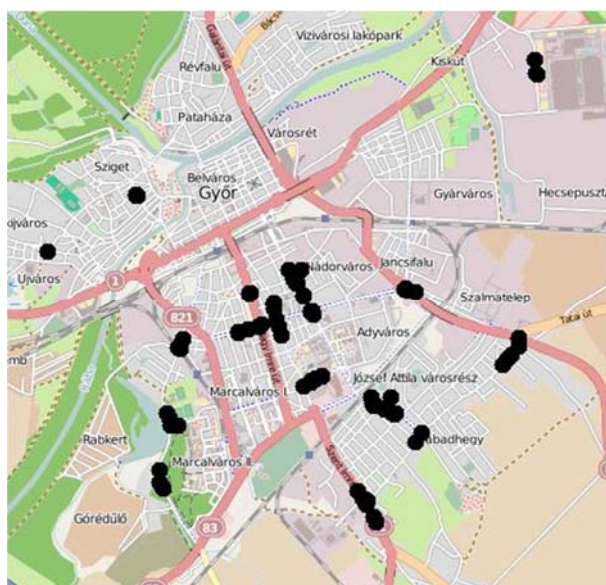
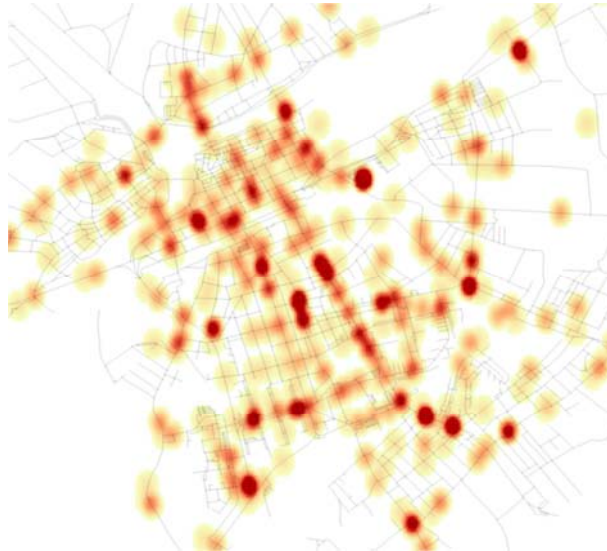


Fig. 9. Cartographic representations of black spots (distance value 150 m)  
(WEB-BAL software)

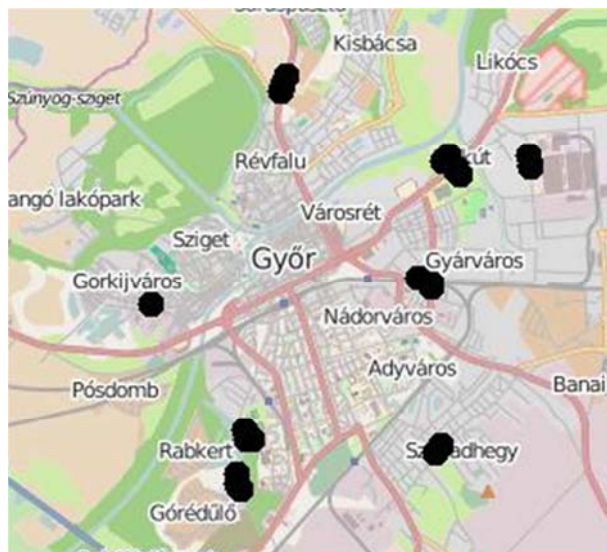
In the third example the distance value is 200. This distance is too large in the local road network, as the average distance of the junctions is about 150 meters in the



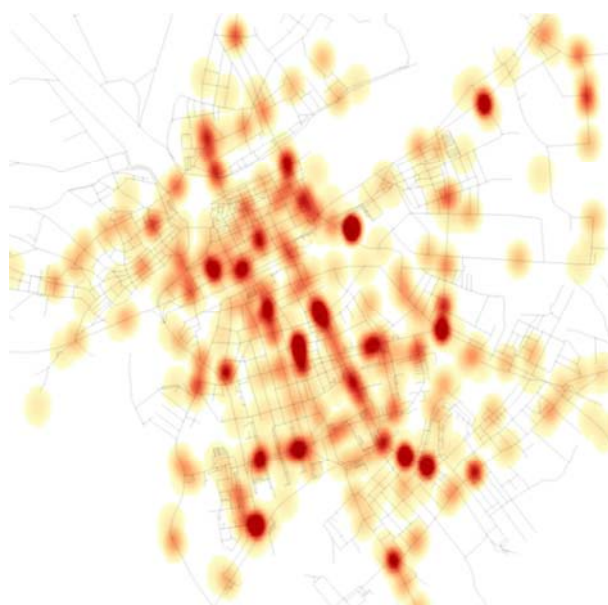
analyzed road network [16], [17]. This is the reason why *Fig. 11* and *Fig. 12* do not show too much similarity to each other.



*Fig. 10.* Cartographic representations of high-density crash locations (distance value 150 m)  
(GIS software heat-map)



*Fig. 11.* Cartographic representations of black spots (distance value 200 m)  
(WEB-BAL software)



*Fig. 12.* Cartographic representations of the high-density crash locations (distance value 200 m)  
(GIS software heat-map)

In this case the searching can falsely merge completely independent locations because the enlargement distance is greater than the average distance of the junctions.

Below a detailed analysis of the second case is given where the distance value is 150 m. This specific distance value is used because in this case the results (the number of the possible black spots) are similar in the two methods. The dots on the map were weighted in both cases depending on their severity (fatal - 10, serious - 3, minor - 1). During the analysis the QGIS software selected many places where there was a few number of accidents but they were fatal or serious. *Table III* shows the results of the main features.

*Table III*

The results of the main features of the possible black spots with different methods

	KDE	DBSCAN
Number of possible black spots	18	20
Number of accidents	58	124
Fatal accidents	10	3
Serious injury accidents	25	43
Minor injury accidents	23	78
Weighted number of accidents	198	237
Average number of accidents per possible black spots	3.4	6.2
Weighted number of accidents per possible black spots	11.6	11.9

In four cases, the KDE method also marked the places as possible black spots where only one fatal accident occurred. They are not possible black spots. The number of the accidents with minor injury is higher in the method DBSCAN. This is explained by the fact that the contour lines produced on the basis of the heat-map have a lower attribute value 'height' therefore the area has not been selected in method KDE. The average numbers of accidents per black spots are nearly half while the weighted numbers of accidents per black spots are about the same.

## 6. Conclusion

In this paper the results of the possible black spot identification were investigated with the WEB-BAL accident analysis software and QGIS software. The first software uses the DBSCAN method while the other one uses the Kernel Density Estimation method. There were several parameters used to identify the location of possible black spots. A conclusion of the study is that in urban road networks black spots cannot be clearly identified with heat-maps (raster images). The accidents are not separated sufficiently in the raster image. There are some identifiable locations of possible black spots on the heat-map, if they are away from their neighbors. These locations were recognized by both methods.

Heat-maps can well display the density of accidents in urban road networks, while possible black spots cannot be illustrated correctly. The number of possible black spots highly depends on the input parameters. In the KDE method fewer input parameters can be given, which reduces the subjective factors. The visualization is therefore more objective; however the numerical calculation already involves a great deal of subjectivity. The visual application of heat-maps can draw the attention to such accident concentrations that are ignored by usual possible black spot identification methods.

Heat-map investigation may also be used for the investigation of the influence of urban road patterns and junction types on the frequency of accidents. Further test areas can be used for the selection of the optimal radius.

## References

- [1] Elvik R. State-of-the-art approaches to road accident black spot management and safety analysis of road networks, *TØI Report*, Oslo, Norway, 2007, Paper 883.
- [2] *Hungarian design guideline for road planning*, e-ÚT 03.02.12.
- [3] [www.openstreetmap.org](http://www.openstreetmap.org) (last visited 1 December 2016).
- [4] Harirforoush H., Bellalite L. A new integrated GIS-based analysis to detect hotspots, A case study of the city of Sherbrooke, *Accident Analysis and Prevention*, (2016) In press <http://dx.doi.org/10.1016/j.aap.2016.08.015>
- [5] Szénási S. GPU implementation of DBSCAN algorithm for searching multiple accident black spots, *15th SGEM Geoconference on Informatics, Geoinformatics and Remote Sensing*, Albena, Bulgaria, 18-24 June 2015, Book 2, Vol. 1, pp. 647–652.
- [6] Szénási S. Parallel implementation of DBSCAN algorithm using multiple graphics accelerators, *16th SGEM Geoconference on Informatics, Geoinformatics and Remote Sensing*, Albena, Bulgaria, 28 June - 7 July 2016, Book 2, Vol. 1, pp. 327–334.

- [7] Mocsári T. Road safety conditions of municipal roads operation: identification of accident black-spots with GPS coordinates, (in Hungarian), *Thesis on Post Graduate Course about Transport Engineering Technology*, Széchenyi István University, 2016.
- [8] Dudík J. M., Kurosu A., Coyle J. L., Sejdic E. A comparative analysis of DBSCAN, K-means, and quadratic variation algorithms for automatic identification of swallows from swallowing accelerometry signals, *Computers in Biology and Medicine*, Vol. 59, 2015, pp. 10–18.
- [9] Sunita J., Parag K. Algorithm to determine  $\epsilon$ -distance parameter in density based clustering, *Expert Systems with Applications*, Vol. 41, No. 6, 2014, pp. 2939–2946.
- [10] Vallakati R., Mukherjee A., Ranganathan P. Situational awareness using DBSCAN in smart-grid, *Smart Grid and Renewable Energy*, Vol. 6, 2015, pp. 120–127.
- [11] Hao Y., Pan L., Jun C., Hao W. Comparative analysis of the spatial analysis methods for hot spot identification, *Accident Analysis and Prevention*, Vol. 66, 2014, pp. 80–88.
- [12] Bailey T. C., Gatrell A. C. *Interactive spatial data analysis*, John Wiley and Sons, New York, 1995.
- [13] Zhixiao X., Jun Y., Kernel density estimation of traffic accidents in a network space, *Computers, Environment and Urban Systems*, Vol. 32, 2008, pp. 396–406.
- [14] Documentation QGIS 2.14. [http://docs.qgis.org/2.14/en/docs/user\\_manual/plugins/plugins\\_heatmap.html?highlight=heatmap](http://docs.qgis.org/2.14/en/docs/user_manual/plugins/plugins_heatmap.html?highlight=heatmap) (last visited 6 December 2016).
- [15] WEB-BAL software homepage, [http://91.120.23.53/webbal\\_kkk/DBSCANBlackSpot.aspx](http://91.120.23.53/webbal_kkk/DBSCANBlackSpot.aspx), (last visited 1 December 2016).
- [16] Háznagy A., Fi I. Comparing the road networks of residential areas, *Pollack Periodica*, Vol. 11, No. 3, 2016, pp. 61–72.
- [17] Hegyi P. Road patterns of housing estates in Hungary, *Pollack Periodica*, Vol. 10, No. 1, 2015, pp. 83–92.