Road Traffic Accident Black Spot Determination by using Kernel Density Estimation Algorithm and Cluster Statistical Significant Evaluation

Khanh Giang LE (Vietnam), Pei LIU and Liang-Tay LIN (Chinese Taipei)

Key words: Traffic Accident (TA), Black Spots (BS), Geographic Information System (GIS), Kernel Density Estimation (KDE), Local Moran's I.

SUMMARY

Determining road collision black spot locations plays an important role in reducing significantly the number of traffic accidents. The article presents a new procedure that identifies road traffic accident black spot locations by using GIS-based kernel density estimation algorithm and evaluates the statistical significance of resulting collision clusters. The results of the paper show that the approach was effective and exact in identifying road traffic accident black spot in Hanoi, Vietnam, simultaneously these hot spots were ranked according to their level of dangerousness. These outcomes will not only enable transport authorities to know comprehensively the reasons for each collision but also to help them manage and deal with hazardous areas according to the prior order in case of limited expense and allocate traffic safety sources suitably.

Road Traffic Accident Black Spot Determination by using Kernel Density Estimation Algorithm and Cluster Statistical Significant Evaluation

Khanh Giang LE (Vietnam), Pei LIU and Liang-Tay LIN (Chinese Taipei)

1. INTRODUCTION

Traffic accidents (TA) are one of the important issues over the world. According to the reports of World Health Organization (WHO), there are more than 1.24 million deaths and about 50 million people injured as results of TA every year in the world (WHO, 2013). To decrease significantly the number of crashes, it is crucial to understand where and when traffic accidents happen frequently. The locations, where are identified by a high accident occurrence compared with the other locations, are known as black spots. The past studies showed that the occurrences of TA are infrequently random in space and time. In fact, these locations identified by several key factors such as geometric design, traffic volume, or weather conditions, etc. (Chainey and Ratcliffe, 2013).

WHO reported that there were over a third of deaths owing to TA in low and middle-income nations among vehicles, cyclists, and pedestrians (WHO, 2013). Vietnam is a developing country thus TA issue also is one of the most concerns of transportation authorities. The annual social expenditure of TA in Hanoi, is the capital of Vietnam, in term of medical treatment, deaths, and property damage occupy 2.9% GDP (5-12 billion USD) (Mai, 2018).

In 2017, there were 20,000 traffic crashes, about 8,200 deaths and 17,000 injured on Vietnam's road networks (Giang, 2018). Currently, non-spatial modelling has been used in Vietnam to identify TA hot spots, namely: Accident Frequency Method (AFM) (classification by level of injury) over one year period (MOT, 2012). This is the oldest and simplest method to identify dangerous locations. However, this method has many limitations such as lacking of visualizing, connecting between space and time, ranking of hot spot's priority, does not take into consideration traffic volume, which has a direct relationship with crash frequency. Therefore, the results have bias toward high-volume locations and suffers from the RTM bias (Li, 2006). Currently, there has not any study dealing with collision mapping in Vietnam.

Geographic Information System (GIS) is a very powerful tool for analyzing traffic safety. GIS can visualize the locations of accidents and store its attributes. Thus, it is easy to find the reasons behind each collision. Spatial data usage plays an important role on traffic safety analysis. GIS enables us to collect, store, manipulate, query, analyze, and visualize the spatial data (Lloyd, 2010; Satria and Castro, 2016).

Spatial analysis of TA has been popularly applied to explore hot spots (Anderson, 2009). GIS has been applied as a management system for accident analysis by combination of spatial statistical methods (Shafabakhsh et al., 2017). In the recent years, the combination of GIS and statistical analysis is increasingly more used by many researchers for assessing the road accidents (Yalcin and Duzgun, 2015; Benedek et al., 2016). Kernel density estimation (KDE)

is one of the most popular density-based methods and has been widely used for detecting dangerous road segments (Xie and Yan, 2013). However, KDE method has a drawback is that the uncertainty about the exact location of the traffic collision is showed by the search bandwidth of the kernel (Anderson, 2009). Thus, KDE only is better for visualization purpose than for determining TA black spot locations (Plug et al., 2011). The same issue was showed by (Xie and Yan, 2008), KDE method lacks an investigation of the statistical significance of the high-density locations. Recently, there are very few researches that investigate comprehensively statistical significance of KDE method. Thus, how to identify which clusters is statistical significance is really necessary.
Therefore, in our study, firstly, TA black spot locations was determined by using GIS-based kernel density estimation algorithm, after that statistical significance of resulting collision clusters was evaluated, and then their order was arranged in accordance with their significance. Finally, to validate this approach, we compare the results with traffic accident reports during three years (2015-2017) in Hanoi, Vietnam. The purpose of this paper is to present an improved procedure of identifying TA black spots. The remainders of the article are arranged as follows. Section two depicts data and methods. Section three illustrates analysis results and discussions. Section four presents the check and validation of the results. Finally, conclusions are presented in section five.

2. MATERIALS AND METHODS

2.1 Data and Research Region

This study was carried out in Hanoi, Vietnam. Two different databases were used for this study. First, a road network map was provided in a shape file format, which includes specifications of roads such as road length, road width, road type, and speed limits. Second, a traffic accident database in three years (2015-2017) was provided by the Transport Police Department in Hanoi (Fig. 1). Such a time span is sufficient because there are many records and the characteristics of the TA remains unchanged relatively (Elvik, 2008). There are 1,132 crashes were recorded on Hanoi's roads. The collision database was provided in an Excel file and contained significant accident parameters such as the date and time of a crash, crash location, accident types, age and sex of drivers, the number of the injured, etc.

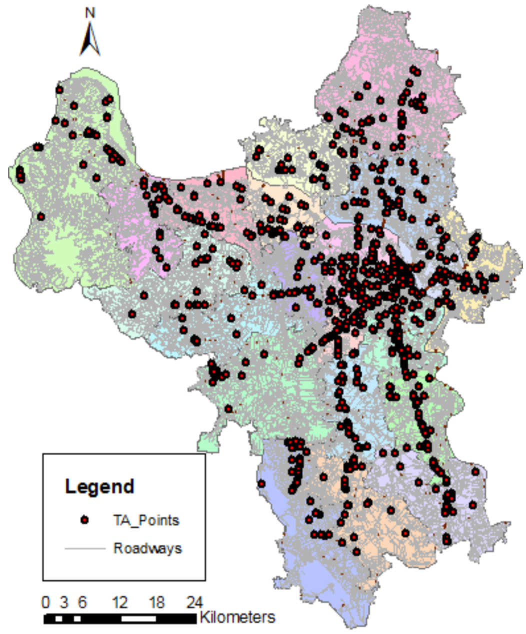

Fig. 1. TA locations in Hanoi (2015-2017).

2.2 Kernel Density Estimation (KDE)

KDE is one of the most effective methods to determine the spatial models of TA (Blazquez and Celis, 2013; Satria and Castro, 2016). The density of events is calculated within a definite research radius in the study areas to create a smoothed surface. A kernel function is utilized to assign a weight to the area surrounding the events proportional to its distance to the point event. From there, the value is highest at the point event location centre and decrease smoothly to a value of zero at the radius of the research circle (see Fig. 2). At the end, a smoothed continuous density surface is generated by adding the individual kernels in the research area (Anderson, 2009; Rahimi and Shad, 2017). The intensity at a specific location is calculated by Eq. (1):

$$f(s) = \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{d_i}{h}\right) \quad (1)$$

where f(s) is the density estimate at the location s, n is the number of observations, h is the bandwidth or kernel size, K is the kernel function, and d_i is the distance between the location s and the location of the i^{th} observation.

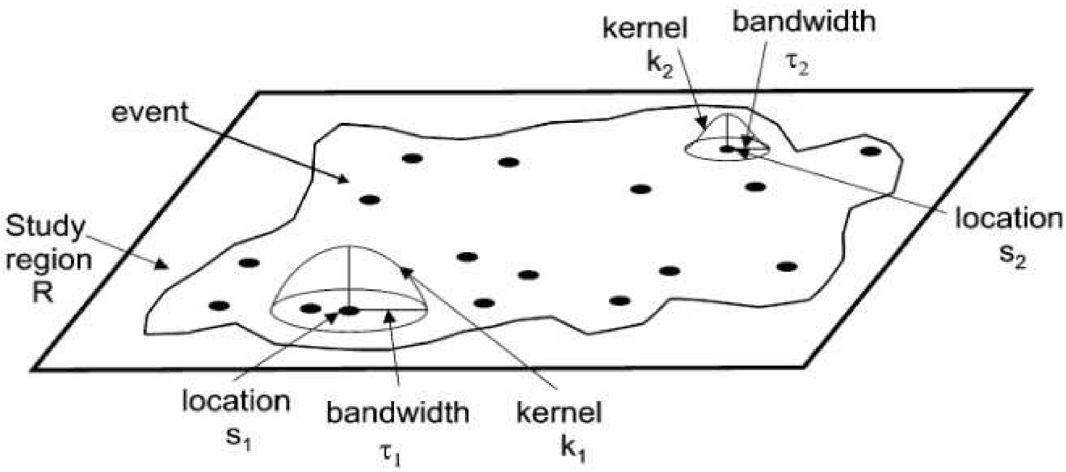

Fig. 2. Diagram of how the quadratic kernel density method works and is the basis for the density method used for this study (source: Bailey and Gatrell, 1995).

2.3 Anselin Local Moran's I

The Cluster and Outlier Analysis tool identifies spatial clusters of features with high or low values. The tool also identifies spatial outliers. To do this, the tool calculates a local Moran's I value, a z-score, a pseudo p-value, and a code representing the cluster type for each statistically significant feature. The z-scores and pseudo p-values represent the statistical significance of the computed index values.

The local Moran's I (Anselin, 1995) is one of the most widely used Local Indicators of Spatial Association (LISA) statistics that is applied to evaluate the statistical significance (Satria and Castro, 2016). It measures the statistical correlation between attributes at each location in a study area and the values (usually the statistic means) in the neighboring locations. It also tests the significance of this similarity. Formally, the local Moran's I can be expressed as Eq. (2):

$$I_i = z_i \sum_{j=1}^{n} w_{i,j} z_j \qquad (2)$$

where z_i and z_j represent the deviations from the mean, and $w_{i,j}$ is the spatial weight between features i and j.

In general, there are four types of correlation among neighbouring values: high-high (H-H), low-low (L-L), high-low (H-L), and low-high (L-H). (H-H) and (L-L) indicate that there is a positive autocorrelation, while (H-L) and (L-H) show that there is a negative autocorrelation (O'Sullivan and Unwin, 2010). The (H-H) areas are relevant for hazardous location detection and show locations where a high number of crashes are surrounded by high values (Xie and Yan, 2013).

3. RESULTS AND DISCUSSIONS

3.1. Analysis Results

The output of KDE method is presented in a raster format consisting of a grid of cells. The two main parameters that influence the KDE are cell size and bandwidth. The choice of bandwidth is quite subjective (Anderson, 2009). In our research, we tried to practice it ten times including 100 m, 200 m, …, 1000 m in order to find the optimal bandwidth for our research. Finally, we considered 1000 m-bandwidth value because it enables us visualize TA back spot locations easily. However, it is not always a good idea to choose a large bandwidth, as the TA black spot locations will not be accurate. This is true as the mention of (Anderson, 2009) is that the uncertainty about the exact location of the traffic collision is showed by the search bandwidth of the kernel. Fig. 3. shows TA black spot locations in two different bandwidth values are 500 m and 1000 m.

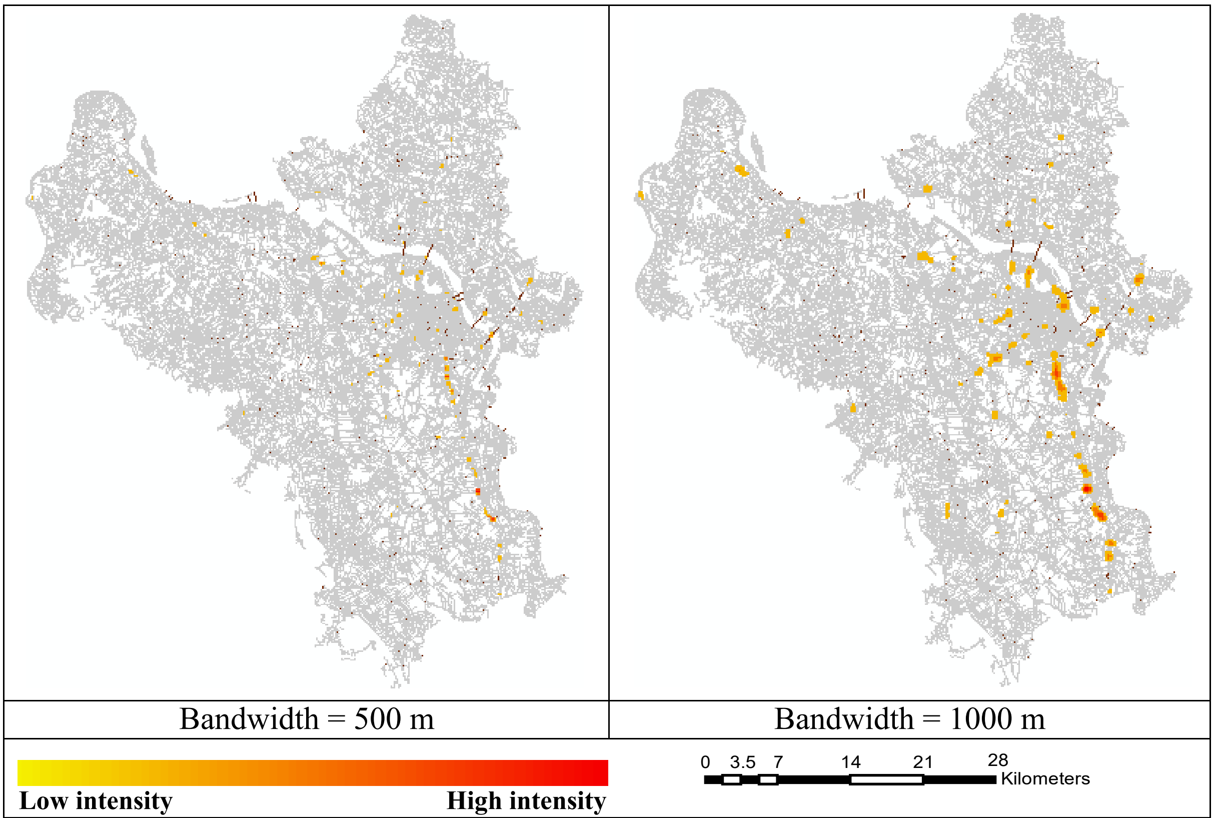

Fig. 3. Accident black spot positions in Hanoi (2015-2017).

With the positions of TA in Fig. 2, it is impossible to find out TA black spot locations. However, KDE method enables us visualize TA black spots easily. Fig. 3 shows that red colored areas are TA black spots in Hanoi (2015-2017), which mainly concentrate on NH-1A section such as Van Dien station, Cho Tia station, Quang Trung – Nguyen Trai intersection, Ha Dong, etc. However, the main advantage of the KDE method as opposed to classical statistic clustering methods is that the uncertainty about the exact position of the TA is showed by the bandwidth of the kernel – this means something like spreading the risk of an accident (Anderson, 2009). Therefore, it is necessary to investigate the importance of the resulting clusters of TA and find out the most hazardous location.

3.2. Assessing the Importance of TA Black Spots

The clusters which form the local maxima of the kernel function are determined. However, we tried to identify the statistical significance of a cluster more objectively. There is less than 1% likelihood that this clustered pattern could be the result of random chance (z-score of 10). As a result, the next step is to determine the bandwidth that maximizes clustering phenomenon.

The Local Moran's I spatial autocorrelation was run at a variety of distances and for each of those it got a z-score which is the level of statistical significance. It diverged from our randomization assumption, and when we compare these z-scores across different distance increments, we find that some are higher than others.

Next, we applied the cluster and outlier analysis to generate a map of hot spots and cold spots (Fig. 4). The red points show areas of clusters where we have features with similarly high values near each other so in this case what that means is that we have got TA point's high priority near each other, and those clusters are statistically significant meaning that they are very far from random arrangement.

In contrast, the green points represent areas where they have similarly low values clustered near each other, so those points are low priority. However, in this case, these points did not appear. Besides, the orange and the yellow points are outliers that are points in which case we find a high value or high TA surrounded by low TA and conversely for the yellow points. In this case, the red points were mainly occurred on NH-1A such as Van Dien station, Cho Tia station; Thang Long Boulevard - Me Tri intersection and Pham Van Dong Road.

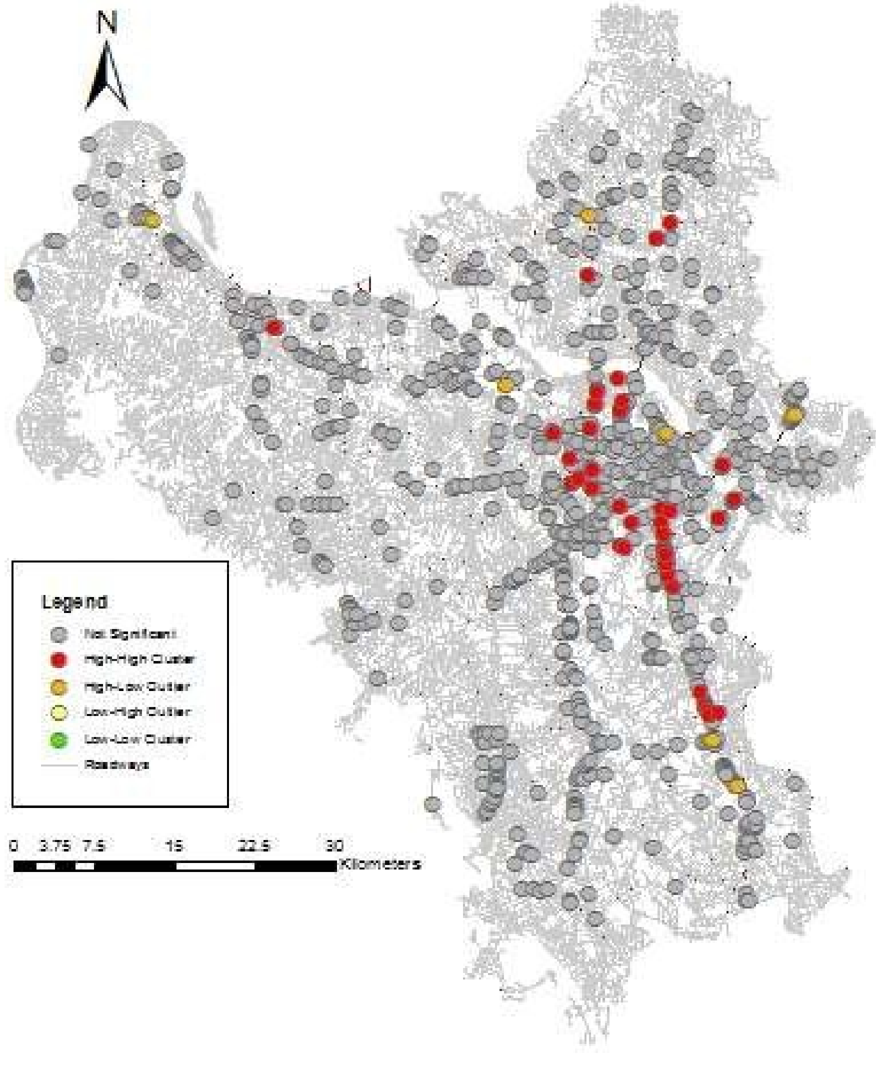

Fig. 4. TA Hot spots, Cold spots with statistical significance.

4. THE VALIDATION OF THE RESULTS

The results of the proposed methodology showed that this approach is appropriate for overcoming drawbacks of KDE method. Fig. 5 illustrates road traffic accident hotspot priority map. In general, there are four types of correlation among neighboring values: high-high (H-H), low-low (L-L), high-low (H-L), and low-high (L-H). (H-H) and (L-L) indicate that there is a positive autocorrelation, while (H-L) and (L-H) show that there is a negative autocorrelation. The (H-H) areas are relevant for hazardous location detection and show locations where a high number of crashes are surrounded by high values. In this case, the red points (circled in red) are (H-H) clusters where TC occurred frequently and these locations were mainly occurred on NH-1A such as Van Dien station, Cho Tia station; Thang Long Boulevard - Me Tri intersection and Pham Van Dong Road.

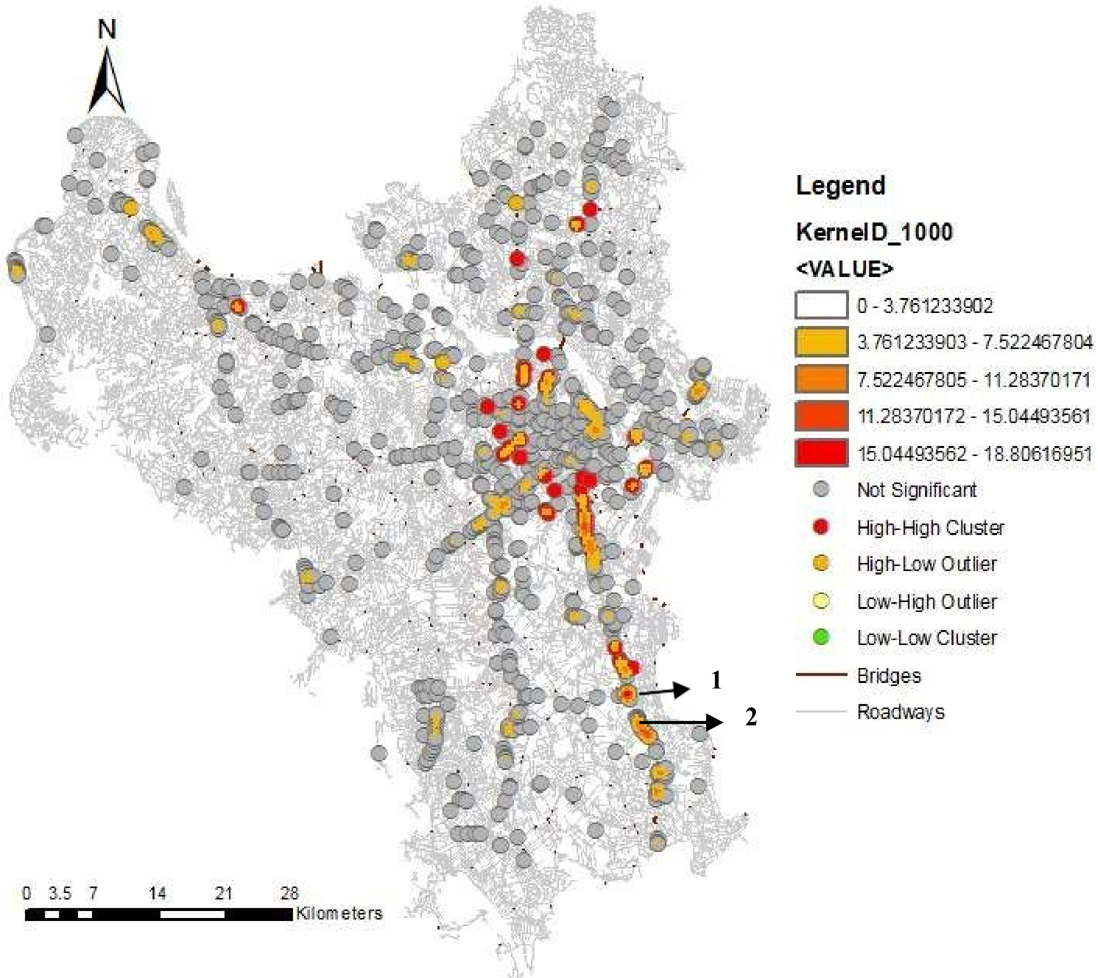

Fig. 5. TA black spot map with dangerous levels

In addition, this approach investigated the statistical significance of the high-density locations. For instance, location 1 (Fig. 5) was identified as a high density point of TA through applying KDE method. However, after investigating the statistical significance of the high-density locations, this location was identified as a (H-L) outlier point. It means this location is a high TA surrounded by low TA. Location 2 (Fig. 5) was identified as a highly dense zone of TA through applying KDE method. But, after investigating the statistical significance of the high-density locations, this location was determined as a not significant point (grey color) (Figure 5). The results of the proposed methodology are appropriate to the observations from the reality and the reference data. This proposed methodology enables traffic authorities understand the situations clearer and comprehensively.

5. CONCLUSION

The paper proposed a new procedure that determines road traffic accident black spot locations by using GIS-based kernel density estimation algorithm, evaluates the importance of resulting collision clusters, and then arranges them in accordance with their significance. The results of the paper show that the approach was effective and exact in identifying road traffic accident black spot in Hanoi, Vietnam. These outcomes will not only enable traffic authorities to understand comprehensively the causes behind each collision, but also to help them manage and deal with hazardous areas according to the prior order in case of limited budget and allocate traffic safety resources appropriately.

The integration of KDE method and statistical significance evaluation of the resulting clusters of TA help to overcome the drawbacks of KDE method. From there, the determination of TA black spot locations will be improved with high accuracy. The results of the paper show that TA black spots mainly occurred in NH-1A namely Van Dien station, Cho Tia Station, and at Nguyen Trai - Quang Trung intersection, Thang Long Boulevard - Me Tri intersection and Pham Van Dong road. This is also the first study about this issue in Vietnam, so the contribution of the article will help the traffic authorities easily solve this problem not only in Hanoi, but also can apply for other cities.

However, within the scope of the paper, there is a limitation is that does not take traffic volume in identifying TA hot spots. Therefore, in the forthcoming studies, the authors will solve this issue. In addition, the authors will deploy this application online, which not only helps the traffic authorities, police patrol to update emergence information easily but also provide the citizen a black spot map in an updated, accurate, and visual way.

REFERENCES

Anderson, T. K., 2009, Kernel density estimation and K-means clustering to profile road accident hotspots, Accident Analysis and Prevention, Elsevier.

Anselin, L, 1995, Local indicators of spatial association – LISA, Geographical Analysis, Vol. 27, No. 2, Ohio State University Press.

Bailey, T. C., Gatrell, A. C., 1995, Interactive Spatial Data Analysis. John Wiley and Sons, New York.

Benedek, J., Ciobanu, S. M., Man, T. C., 2016, Hotspots and social background of urban traffic crashes: a case study in Cluj-Napoca (Romania). Accident Analysis & Prevention, 87, 117-126.

Blazquez, C. A., Celis, M. S., 2013, A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile. Accident Analysis and Prevention, 50, 304–311, Elsevier.

Chainey, S., Ratcliffe, J., 2013, GIS and Crime Mapping, John Wiley and Sons, England.

O'Sullivan, D., Unwin, D., 2010, Geographic information analysis, John Wiley and Sons, New York.

Elvik, R., 2008, A survey of operationa definitions of hazardous road locations in some European countries, Accident Analysis and Prevention, 40, 1830-1835, Elsevier.

Erdogan, S., Yilmaz, I., Baybure, T., Gullu, M., 2008, Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar, Accident Analysis and Prevention, 40, 174-181, Elsevier.

Ha Mai, 2018, Việt nam mất khoảng 130 tỉ usd chi phí cho tai nạn giao thông trong 15 năm, https://thanhnien.vn/thoi-su/viet-nam-mat-khoang-130-ti-usd-chi-phi-cho-tai-nan-giao-thong-trong-15-nam-954438.html [accessed: 13:50, 25/09/2018].

Thu Giang, 2018, Ủy ban An toàn giao thông Quốc gia tổng kết công tác năm 2017, http://backantv.vn/tin-tuc-n17855/uy-ban-an-toan-giao-thong-quoc-gia-tong-ket-cong-tac-nam-2017.html [accessed: 12:09, 25/09/2018].

Li, L, 2006, A GIS-based Bayesian approach for analyzing spatial-temporal patterns of traffic crashes, Doctoral dissertation, Texas A&M University.

Lloyd, C. D., 2010, Spatial data analysis: an introduction for GIS user, Oxford University Press.

MOT, 2012, Thông tư 26/2012/TT-BGTVT, Quy định về việc xác định và xử lý vị trí nguy hiểm trên đường bộ đang khai thác, BGTVT, Vietnam.

Plug, C., Xia, J. C. and Caulfield, C., 2011, Spatial and temporal visualisation techniques for crash analysis, Accident Analysis and Prevention, Elsevier.

Rahimi, S., Shad, R., 2017, Identification of road crash black-sites using geographical information system, International Journal for Traffic and Transport Engineering (IJTTE) 7(3):368-380.

Satria, R., Castro, M., 2016, GIS tools for analyzing accidents and road design: a review, Transportation Research Procedia 18: 242 – 247.

Shafabakhsh, G. A., Famili, A., Bahadori, M. S., 2017, GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran, J. Traffic Transp. Eng. (Engl. Ed.) 4 (3): 290-299.

WHO, 2013, Global status report on road safety 2013. Supporting a decade of action, World Health Organization, Department of Violence and Injury Prevention and Disability, Geneva.

Xie, Z., Yan, J., 2013, Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach, J. Transp. Geogr, Elsevier.

Yalcin, G., Duzgun, H. S., 2015, Spatial analysis of two-wheeled vehicles traffic crashes: Osmaniye in Turkey. KSCE Journal of Civil Engineering, 19(7): 2225-2232.

BIOGRAPHICAL NOTES

Khanh Giang Le
Doctoral Candidate, PhD program of Civil and Hydraulic Engineering (2017 – Now).
Institution: College of Construction and Development, Feng Chia University, Taiwan.
Master Degree: Birmingham City University, United Kingdom (2013 – 2014)
Bachelor Degree: University of Transport and Communications, Hanoi, Vietnam (2001-2006)
He is a lecturer at Geodetic Division, Civil Engineering Faculty, the University of Transport and Communications, Hanoi, Vietnam (2006 – Now).
He is interested in applying GIS, GPS, spatial statistics, and geospatial analysis in transportation sector and urban studies.

Associate Professor **Pei LIU**
He is an Associate Professor of College of Construction and Development, Feng Chia University, Taichung, Taiwan.
He got his PhD degree in Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, America. He is an expert in Highway Engineering, Artificial Intelligence Methods, Numerical Methods, and Pavement Engineering.

Professor **Liang-Tay LIN**
He is Dean of College of Construction and Development, Feng Chia University, Taichung, Taiwan.
He got his PhD degree in Department of Civil Engineering, National Taiwan University, Taiwan. He is President of Chinese Institute of Transportation, Taiwan. He was Director General of Transportation Bureau, Taichung City Government. He is Director of innovation centre for Intelligent Transportation and Logistics. He is an expert in Traffic Engineering, Traffic Control, Traffic Flow Theory, and Urban Traffic Management.

CONTACTS

Khanh Giang LE
Doctoral Candidate, Ph.D program of Civil and Hydraulic Engineering.
College of Construction and Development, Feng Chia University.
Taichung
Taiwan
Tel. + 886 984267484
Email: khanhgiang298@gmail.com