

Integration of Probability and Clustering Based Approaches in the Field of Black Spot Identification

Maen Ghadi¹, Árpád Török^{1*}, Katalin Tánczos²

¹ Department of Automotive Technologies,
Faculty of Transport Engineering and Vehicle Engineering,
Budapest University of Technology and Economics
H-1521 Budapest, P.O.B. 91, Hungary

² Department of Transport Technology and Economics,
Faculty of Transport Engineering and Vehicle Engineering,
Budapest University of Technology and Economics
H-1521 Budapest, P.O.B. 91, Hungary

* Corresponding author, e-mail: torok.arpad@mail.bme.hu

Received: 27 November 2017, Accepted: 03 October 2018, Published online: 19 October 2018

Abstract

The objective of the paper is to define a complex methodology to analyze black spot locations of road infrastructure network combining the benefit of both; Empirical Bayes method and K-mean clustering approach. In the first step, K-mean algorithm is used to define homogeneous accident clusters. The homogeneity is described in three terms: traffic conditions, geometric design of the road and accident characteristics. Then, Empirical Bayes method is applied to define black spots based on the determined clusters. Due to the combination of the introduced methods, a powerful technique is provided that is able to identify high-risk locations and cluster dependent segment length as the output of the model.

Keywords

black spots, empirical Bayesian, k-mean algorithm, cluster analysis, road safety

1 Introduction

The identification of road sections characterized by high crash risk is considered the first step for any successful road safety management process. Based on Wen Cheng's and Simon's (2005) [1] researches, the objective of black spot (BS) identification process is "to identify the locations in a transportation system that have problems and its effects will be revealed through evaluated its crash frequency related to other similar locations". Until now, many different BS identification methods have been suggested. Empirical Bayes (EB) method is considered as a reasonably state-of-the-art methodology. Montella (2010) [2] compares seven well-known BS identification methodologies and concluded that EB is the most effective approach among them. The same result is obtained by Qu and Meng (2014) [3]. EB usually applies Poisson or negative binomial models to predict accident frequency for a given road segment or entity (i.e. intersection, ramp) in a given road type [4]. In order to predict accident frequency in road links more effectively, generally two types of data is applied to

provide input for EB methods: time series of accident data of each specific road segment and of a group of similar (or homogeneous) segments. During the identification of homogeneous accident clusters, road segments properties in accident locations, together with causes and types of accidents (e.g. road geometry, traffic characteristics, number of lane, traffic volume) should be considered as key attributes of the cluster generation process. Considering that traffic accident data is heterogeneous, in general, researchers often try to reduce heterogeneity by focusing on a specific type of accidents. For instance, Aditya and Grembek (2016) [5] have focused on pedestrian crashes at crossing and midblock areas in their analysis, Bédard et al. (2002) [6] have investigated single-vehicle crashes with fixed objects, and Greibe (2003) [7] built an accident prediction model for urban junctions. However, due to heterogeneity of data certain accident factors can remain hidden. Most often, classification of accidents is only based on researchers' experiences, methodological decisions or

wishes. Black and Thomas (1998) [8], have segmented the road arbitrarily into equal sections to examine the changing in Moran's index values of neighbor segments. Kwon et al. (2013) [9] have introduced further segmentation methodology based on the differences in road characteristics, i.e. number of lanes, average annual daily traffic (AADT). They concluded that different segmenting processes based on different set of infrastructure attributes can result consequently different estimation values regardless of which evaluation method is used. Others have used a moving fixed length window, considering that as a more efficient segmentation approach [10], [11]. However, this method has also some remarkable weakness especially considering the so called 'first come first serve' problem and the fixed analysis interval, which could result in either false positives or false negatives during the identification of BS locations [12]. Now it seems that, in order to efficiently apply EB method in case of homogeneous segments, it is useful to couple EB method to other approach that allows the aggregation of data in homogeneous segments or clusters. Clustering technique is one of the important data-mining approaches that could be used in finding hidden relationships and patterns of a large number of accidents [13, 14], and classify them according to the similarities in their attributes and spatial distributions. Depaire et al. (2008) [15] have used a latent class generation method to identify homogeneous traffic accident types. They suggest that applying clustering as a preliminary analysis can reveal hidden relationships and also can help in generating homogeneous road segments. Practically, homogeneity of accident classes could have many definitions considering the type of accidents, severity, spatial distribution, type and number of involved vehicles, driver characteristics and many other issues. Therefore, it is difficult to reach a pure definition of homogeneity. In our research paper, homogeneous segments have been defined based on spatial distribution patterns of crashes and traffic and geometric characteristics of the accident location. [5], [16]. Furthermore, it has to be mentioned that the combination of EB methods and clustering techniques has been successfully applied in other fields to identify critical locations (e.g. medicine) [17], [18]. Luca et al. (2012) [19] have used C-mean clustering to identify accident attributes from which the EB model was subsequently constructed.

Spatial clustering can be applied in case of many types of road database, characterized by different aggregation level. In many researches, during safety investigations and black-spot analysis traffic accidents are examined as the function

of speed, traffic volume or other variables (e.g. differentiating continuous road segments or intersections [16], [20], [21]). The emphasis tends to be on explaining how different road segments or elements influence crash likelihoods. On the other hand, many well-accepted, well-structured methods handle the segmentation process and even the black-spot analysis for entire roads components [5], [19] or even entire districts without the differentiation of intersections and continuous road sections. The reason for the application of a more general model, the relatively high aggregation level, and the skipping of a more detailed investigation is the spread of the application of more general factors in this scientific field that are related to the spatial and the temporal distribution of the accidents.

The initial research has been focused on the spatial variations of automobile accidents. Cluster analysis tool aims to assemble different accidents into groups based on their spatial distribution factor. In case of this approach the degree of proximity between two accidents is maximal if belonging to the same group and minimal otherwise. Thus, cluster analysis can be used to discover structures in data without providing an explanation or interpretation. [22], [23]. However spatial pattern can be used to provide possible explanations regarding the relationship, between different accidents based on their spatial locations, regardless of the road characteristic of the location (i.e. road segment, intersection). In other words, spatial clustering can classify similar accidents, based on their characteristics (causes of occurrence), into groups without having information on these characteristics initially. Since it has a significantly higher probability that neighboring accidents in space and time have strongly related causes of occurrence than distantly located accidents and thus their distribution regarding causes and types are more homogeneous [24], [25].

In accordance with the above introduced methodological results and accepted assumptions the objective of this research is to define an effective accident classification approach which results such clusters of accidents, which can be characterized by homogeneous infrastructure attributes and road locations. In this case, the accident classes define specific road sections as well, where the EB method can be applied.

2 Cluster analysis

Cluster analysis have several definitions, depending on where it is used, but generally its major goal is to organize a large data set (or set of objects) into constrained number of smaller homogeneous groups based on similarities or

dissimilarities between the components of the investigated sample. Thus generally, discovering the size and structure of the analyzed data can be considered as a fundamental goal of cluster analysis.

Partitioning methods are one of the most well-known clustering algorithms. This method is generally based on finding a clustering structure that minimizes a certain error criterion. The method of the Sum of Squared Error (SSE) is a reasonably commonly used technique to derive this error. A widely used algorithm that applies squared error is the K-mean algorithm. The analytical approach presented in this paper is based on the K-mean clustering algorithm.

The principle of this technique is as follows: each object (accident) in this method is represented by a geographical point, and each point has different attributes and coordinates. The algorithm starts with an initial set of cluster centers chosen randomly for a predetermined number of clusters (k). In the iteration process, each data object is assigned to its nearest center, according to Euclidean distance of the accidents. In the next step cluster centers are recalculated in accordance with equation (1) [26]. The iteration stops when no more cluster centers need to be relocated,

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} q_k, \quad (1)$$

where μ_k is the mean of cluster k and N_k is the number of objects (road crashes) belonging to cluster k .

In our case two input variables are applied to calculate distance of cluster elements from cluster centres; the spatial location of each accident and the AADT data within a homogeneous road geometry. The spatial location is a key variable in classifying accidents, since it has significantly higher probability that neighbouring accidents have strongly related causes than in case of distant located accidents. Considering that the accidents are located on the road network, a section-kilometre based localization seems to be more appropriate to represent spatial identification of accidents, than geographical coordinates (the longitude and latitude). It allows us to avoid classifying accidents in the same cluster, which are closely located but occurred on different roads. Accordingly in K-mean clustering method a cumulated distance parameter has been defined in one dimension space for each accident starting from the beginning of the road, as presented in Fig. 1. To consider the homogeneity of traffic during the segmentation process, AADT is also involved in the clustering method as a main input variable (See Fig. 1), so each resulted cluster - which expectedly equals a constrained spatial

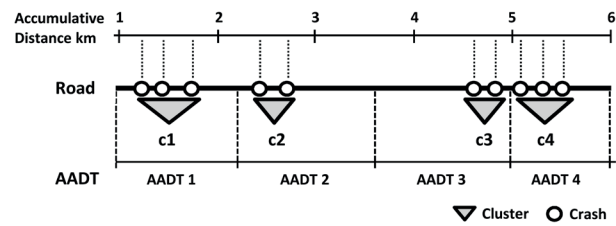


Fig. 1 Illustration of K-mean clustering (c1, c2, c3 and c4 are clusters)

interval, with other words a section of the road - can be characterized by reasonably low heterogeneity regarding AADT. Fig. 1 represents a visual explanation on how accident clusters generated by the above introduced methodology can also be applied to define specific segmentation of the road based on homogeneity criterion regarding accident and infrastructure parameters (the road is clustered according to its accident distribution and AADT). In the same figure (Fig. 1) cluster generation is strongly depending on AADT (different cluster is identified for different AADT as in case of cluster c3 and c4) and accident free sections are excluded from the segmentation process (empty intervals between clusters are not involved in section generation process).

The application of K-mean algorithm can be very attractive; since the road accident based segmentation method makes it possible to relate the length of the identified segments strongly to the clusters' lengths. Therefore, empty sections between clusters with no accident history will be excluded from the evaluation process. Each cluster length depends on the number of accidents contained by the cluster and their spatial distribution. This approach can help in minimizing false negatives and false positives during the estimation process related to BS location identification. Another reasonable advantage of K-mean clustering in road accident classification is its easy implementation. Simplicity of application even in case of large data sets makes the advantageous characteristic of the model even more articular (e.g. compared to hierarchical methods, density-based methods). On the other hand in case of K-mean clustering method the number of clusters (k) has to be determined in advance, which can cause uncertainty when no prior knowledge is available. Accordingly the question now is as follows: how the number of clusters (k) of road accidents can be determined?

2.1 Determining the number of clusters (k_0)

According to the basic approach, to determine the number of clusters (k_0) using K-mean clustering method it is required to test different numbers of k and analyze the

changes of the variances (or SSE) compared to cluster centers. Clustering method that minimizes SSE is often called minimum variance partition. This method is used to measure the compactness degree of clustering. SSE represents the sum of squared differences between each cluster element and its cluster's center. SSE can be written in the following algebraic formula (equation 2 and 3),

$$SSE = \frac{1}{2} \sum_{k=1}^K N_k S_k, \quad (2)$$

where,

$$S_k = \frac{1}{N_k^2} \sum_K |x_i - x_j|^2, \quad (3)$$

where $|x_i - x_j|$ is the absolute distance between each object (accident) j and its cluster's center i for each cluster.

In case of minimizing SSE the number of clusters (k_0) becomes reasonably close or equal to the number of the investigated objects (accidents), since the value of SSE becomes 0 if each individual object has an own cluster with a cluster centre coinciding with location of the given accident. Accordingly each data point will end up with its own little cluster, and the distance from each data point to its cluster centre is going to be zero, which is the best lower variance (or SSE) value. To avoid over fragmentation an additional methodological step needs to be introduced in the process. The process of segmentation can be characterized by the changes of SSE value depending on the changes of cluster numbers. In this case, obviously, the velocity of SSE is investigated projected to the unit change of the number of clusters (if the analyzed functions were continuous, SSE function differentiated by the number of clusters would be investigated). In light of the above mentioned aspects, when the function of SSE depending on number of clusters is analyzed the environment of the point where the slope of the curve's tangent changes an order of magnitude (e.g. when its absolute value is less than 1) is critical. At this point the number of clusters starts to increase faster and faster so this point seems to be appropriate to be applied as a constraint of the number of clusters.

3 Empirical Bayesian

After applying K-mean clustering to divide the road into segments based on the similarity of the analyzed accidents, as described in section 2, it is required now to evaluate the level of risk of the resulted segments in order to identify BS segments on the road. EB method is considered as a

state-of-the-art technique for doing that. This approach combines both the observed and predicted crash frequencies, for a specific road network, in one statistical model, applying the equation (4) below.

$$N_E = wN_p + (1 - w)N_0, \quad (4)$$

where N_E is the expected number of crashes; N_p is the predicted number of crashes; and N_0 is the observed number of crashes. N_E is usually measured for three years to avoid any dispersion in the data, i.e. due to any special events or improvements in the road. The weight adjustment factor w is a function of N_p over-dispersion parameter (See equation 5) [20]. Therefore w represents the degree of reliability in obtaining N_p for the studied three years.

$$w = \frac{1}{1 + q \times \left\{ \sum_{all-Study-years} N_p \right\}}, \quad (5)$$

q = over-dispersion parameter from the associated N_p model.

The process of identifying BS using EB method is started by dividing the road into segments, that has already been done (in clustering section), and followed by calculating the predicted number of crashes N_p for each segment characterized by the number and type of accidents, AADT and length of the given section, by using the formula in equation (6), which allows us to calculate N_E (equation 4),

$$N_p = \exp(a + b \cdot \ln(AADT) + \ln(L)), \quad (6)$$

where a and b are the regression parameters, their value depends on the category of road and type of crash [20]. Finally, the excess risk derived by EB method is calculated as the difference of N_E and N_p in which a positive value can refer to a BS. This technique also allows us to arrange the BS segments according to the degree of dangerousness.

4 Data

In our case study, the motorway road M3 in Hungary is examined. This road has a total length of 281 km and it connects the capital Budapest with Nyíregyháza city located in the northeastern part of the country. Each accident occurred on this road from the year 2013–2015 is recorded regarding its coordinates, type and severity (excluding ramp and intersection accidents). Traffic volume (AADT) for different sections of the road is also included in the dataset for the same period. The accident data is included fatal serious and slight injury accidents together in accordance with HSM's recommendations (AASHTO 2010).

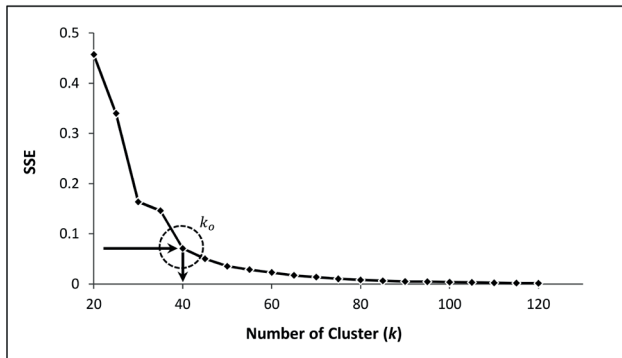


Fig. 2 Defining the value of k_0 based on the “Number of clusters-SSE” function

5 Results and discussion

A cluster analysis has been applied initially in order to divide accidents into homogeneous groups based on their spatial distribution assuming homogenous traffic and geometric conditions. To do so, the first step is to identify the applied number of clusters (k_0). In order to determine k_0 an initial k_0 value is required. In our study, the maximum value of section length has been defined of 10 km, in this case the number of clusters is about 30 ($281/10 \approx 30$). Accordingly, the initial cluster number is 30. In the next step “number of cluster-SSE” function is defined based on the calculation of SSE values derived from the unit change of the cluster number. Then, optimal value of k_0 is derived (See Fig. 2).

It can be noted from figure (2) that as the number of clusters (k) is increasing the variance is decreasing until it achieves zero. So, the best k point is there, where the difference quotient of function values and domain values changes an order of magnitude (becomes small enough - e.g. when its absolute value is less than 1). Up to this point each added cluster results in a substantial reduction in the value of variance but after that point, increasing k will result less-and-less reduction in the value of variance. As mentioned above, this consideration is similar to take the second derivative of the variance as a function of the number of clusters (indicating how much the difference changes as another cluster is added). In accordance with the result of the analysis, $k = 40$ seems to be the optimal cluster number.

The resulted number of accidents in each cluster and its corresponding segment length is presented in Table I. It is important to mention that some clusters are omitted since they just represented a single accident (their cluster lengths equal to zero since their accidents are far enough from any other accident). The next step is to evaluate the resulted segments, using EB approach (equations 4, 5 and 6) to

calculate the expected number of crashes N_E for a given section. Then in case of a positive excess risk derived by EB method (the difference value between N_E and N_p).

Table 1 Cluster information and the final BS identification process

Cluster ID	Observed number of rashes	Length (km)	Excess EB	Result
1	8	2,78	-29,10	
2	20	9,27	-14,28	
3	26	6,92	-17,16	
4	20	5,58	-24,84	
5	54	10,44	-42,42	
6	46	6,94	-7,20	
7	24	4,36	3,71	BS
9	6	0,06	0,87	BS
10	30	4,36	-6,30	
11	34	6,05	-29,28	
13	34	8,67	-16,78	
14	10	11,41	-7,12	
15	32	4,31	0,81	BS
16	24	5,29	-9,09	
17	24	3,09	-2,01	
18	18	4,39	-18,75	
19	14	4,41	-10,97	
20	4	3,65	-19,80	
21	26	6,05	-14,02	
22	14	3,17	-7,02	
23	44	6,76	-16,06	
24	28	4,52	-12,12	
25	4	0,40	0,04	BS
28	14	2,91	-12,15	
29	20	3,88	-7,64	
31	22	8,53	-10,05	
32	22	2,69	-2,61	
33	8	4,15	-25,09	
36	24	5,73	-17,98	
38	18	5,58	0,77	BS
39	40	6,58	7,70	BS
40	24	5,15	-3,89	

Note: cluster 8,12, 26, 27, 30, 34, 35, 37 are excluded since it represent a single point (crash) cluster

6 Conclusions

Previous researches have demonstrated that K-mean algorithm can be a useful tool to cluster large data sets in many scientific fields. This paper illustrates the benefits of applying this type of clustering technique in segmenting roads according to the number and distribution of accidents occurred on the road considering traffic situation

and parameters of road geometry in each section homogeneous. According to our results, clustering can be a well applicable complementary approach of EB method to provide an effective and simple BS identification technique. Traditional road segmentation methods, like screening methods can result in different estimated values regardless of which evaluation method is used [9]. Since K-mean clustering can overcome these deficiencies it seems to be a good initial methodological step before implementing risk calculation of the sections. As a key advantage of K-mean clustering it identifies all road sections which are affected by traffic accidents, by grouping homogeneous accidents separately in as small segments as possible and thus excluding empty and less risky road sections between these clusters. This consequently results in re-sectioning the whole road network according to the spatial distribution of accidents. Therefore, all potential segments can be examined and classified according to its risk level. To evaluate the robustness of the model; a case study has been implemented. A road of 281 km length with about 220 fatal, serious and slight injury accidents has been examined. Clustering method is started with an initial value of predetermined cluster number (k). Then the optimized number of clusters has been identified (based on the rate of change in variance or SSE) (See Fig. 1). Each resulted cluster has had its own specifically defined length depending on the number of accidents and their spatial distribution, and also on AADT values of the sections. The final result of the investigated case study has presented 40 spatially segmented accident clusters with different lengths, among them 6 clusters have been classified as BS with a different level of risk. The resulted BS (Table II) can be further investigated according to its level of risk, since sections with higher positive excess risk values derived by EB method are more risky. However, Huzjan et al (2017) [27] found that harmonization of regulated speed limits on motorways can highly affect crash potential, accordingly, speed and another accident related factors should be considered in future works to find more homogeneous cluster of accidents.

References

- [1] Cheng, W., Washington, S. P. "Experimental Evaluation of Black Spot Identification Methods", *Accident Analysis & Prevention*, 37(5), pp. 870–881, 2005.
<https://doi.org/10.1016/j.aap.2005.04.015>
- [2] Montella, A. "A comparative analysis of hotspot identification methods", *Accident Analysis & Prevention*, 42(2), pp. 571–581, 2010.
<https://doi.org/10.1016/j.aap.2009.09.025>
- [3] Qu, X., Meng, Q. "A note on hotspot identification for urban expressways", *Safety Science*, 66, pp. 87–91, 2014.
<https://doi.org/10.1016/j.ssci.2014.02.006>
- [4] Sipos, T. "Spatial statistical analysis of the traffic accidents", *Periodica Polytechnica Transportation Engineering*, Vol. 45(2), pp. 101–105, 2017.
<https://doi.org/10.3311/PPtr.9895>
- [5] Medury, A., Grembek, O. "Dynamic Programming-Based Black Spot Identification Approach for Pedestrian Crashes", *Accident Analysis & Prevention*, 93, pp. 198–206, 2016.
<https://doi.org/10.1016/j.aap.2016.04.037>
- [6] Bédard, M., Guyatt, G. H., Stones, M. J., Hirdes, J. P. "The independent contribution of driver, crash, and vehicle characteristics to driver fatalities", *Accident Analysis & Prevention*, 34(6), pp. 717–727, 2002.
[https://doi.org/10.1016/S0001-4575\(01\)00072-0](https://doi.org/10.1016/S0001-4575(01)00072-0)
- [7] Greibe, P. "Accident prediction models for urban roads", *Accident Analysis & Prevention*, 35(2), pp. 237–285, 2002.
[https://doi.org/10.1016/S0001-4575\(02\)00005-2](https://doi.org/10.1016/S0001-4575(02)00005-2)
- [8] Black, W. R., Thomas, I. "Accidents on Belgium's motorways: a network autocorrelation analysis", *Journal of Transport Geography*, 6(1), pp. 23–31, 1998.
[https://doi.org/10.1016/S0966-6923\(97\)00037-9](https://doi.org/10.1016/S0966-6923(97)00037-9)
- [9] Kwon, O. H., Park, M. J., Yeo, H. "Chung K. Evaluating the performance of network screening methods for detecting high collision concentration locations on highways", *Accident Analysis & Prevention*, 51, pp. 141–149, 2013.
<https://doi.org/10.1016/j.aap.2012.10.019>
- [10] Harwood, D. W., Torbic, D., Bauer, J., Persaud, K. M., Lyon, B. N., Hauer, C. A. E. "Safety analyst: Software tools for safety management of specific highway sites", *White Paper for Module 1, Network Screening*, Federal Highway Administration, Washington, D.C., United States, 2010.
- [11] Nguyen, H., Taneerananon, P., Koren, C., Iamtrakul, P. "The significance of developing a new black spot safety management approach for the local road traffic nature of Ho Chi Minh City", *Acta Technica Jaurinensis*, 6(2), pp. 91–106, 2013.
- [12] Elvik, R. "The predictive validity of empirical Bayes estimates of road safety", *Accident Analysis & Prevention*, 40(6), pp. 1964–1969, 2008.
<https://doi.org/10.1016/j.aap.2008.07.007>
- [13] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. "From data mining to knowledge discovery: an overview", *AI Magazine*, 17(3), pp. 37–54, 1996.
- [14] Sun, B., Cheng, W., Bai, G., Goswami, P. "Correcting and Complementing Freeway Traffic Accident Data Using Mahalanobis Distance Based Outlier Detection", *Technical Gazette*, 24(5), pp. 1597–1607, 2017.
<https://doi.org/10.17559/TV-20150616163905>
- [15] Depaire, B., Wets, G., Vanhoof, K. "Traffic accident segmentation by means of latent class clustering", *Accident Analysis & Prevention*, 40(4), pp. 1257–1266, 2008.
<https://doi.org/10.1016/j.aap.2008.01.007>
- [16] Truong, L. T., Somenahalli, S. V. C. "Using GIS to Identify Pedestrian-Vehicle Crash Black Spots and Unsafe Stops", *Journal of Public Transportation*, 14(1), pp. 99–114, 2011.

- [17] Marek, L., Paszto, V., Tucek, P., Dvorsky, L. "Spatial clustering of disease events using Bayesian methods", In: 14th International Workshop on Current Trends in Databases, Information Retrieval, Algebraic Specification and Object Oriented Programming, DATESO 14, Olomouc, Czech Republic, 2014, pp. 25–34.
- [18] McCallum, K. J., Ionita-Laza, I. "Empirical Bayes scan statistics for detecting clusters of disease risk variants in genetic studies, *Biometrics*", 71(4), pp. 1111–1120, 2015.
<https://doi.org/10.1111/biom.12331>
- [19] Luca, D. L., Muro, R., Lamberti, R., Acqua, G. D. "Road safety management using Bayesian and cluster analysis", *Procedia - Social and Behavioral Sciences*, 54, pp. 1260–1269, 2012.
<https://doi.org/10.1016/j.sbspro.2012.09.840>
- [20] AASHTO, Highway safety manual, 1st Edition. American Association of State Highway and Transportation Officials, Washington, D.C, 2010.
- [21] Sadeghi, A., Ayati, E., Neghab, M. P. "Identification and prioritization of hazardous road locations by segmentation and data envelopment analysis approach", *Promet - Traffic & Traffico*, 25(2), pp. 127–136, 2013.
<https://doi.org/10.7307/Ptt.V25i2.1295>
- [22] Levine, N. E. D., Kim, K. E., Nitz, L. H. "Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns", *Accident Analysis & Prevention*, 27(5), pp. 663–74, 1995.
[https://doi.org/10.1016/0001-4575\(95\)00017-T](https://doi.org/10.1016/0001-4575(95)00017-T)
- [23] Flahaut, B., Mouchart, M., Martin, E. S., Thomas, I. "The local spatial autocorrelation and the kernel method for identifying black zones", *Accident Analysis & Prevention*, 35(6), pp. 991–1004, 2003.
[https://doi.org/10.1016/S0001-4575\(02\)00107-0](https://doi.org/10.1016/S0001-4575(02)00107-0)
- [24] Anderson, T. K. "Kernel density estimation and K-means clustering to profile road accident hotspots", *Accident Analysis & Prevention*, 41(3), pp. 359–364, 2009.
<https://doi.org/10.1016/j.aap.2008.12.014>
- [25] Yu, H., Liu, P., Chen, J., Wang, H. "Comparative analysis of the spatial analysis methods for hotspot identification", *Accident Analysis & Prevention*, 66, pp. 80–88, 2014.
<https://doi.org/10.1016/j.aap.2014.01.017>
- [26] Everitt, B. S., Landau, S., Leese, M., Stahl, D. "Cluster analysis", 5th ed., John Wiley & Sons, London, UK, 2011.
- [27] Huzjan, B., Mandžuka, S., Kos, G. "Real-time traffic safety management model on motorways", *Technical Gazette*, 24(5), pp. 1457–1469, 2017.
<https://doi.org/10.17559/TV-20150331093202>