Accident Analysis and Prevention 128 (2019) 1–7

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

A comparative analysis of black spot identification methods and road accident segmentation methods

Maen Ghadi*, Árpád Török

Budapest University of Technology and Economics, Budapest H-1111, Hungary

ARTICLE INFO

Keywords:
Segmentation method
Black spot identification
Spatial clustering
Consistency tests
Safety performance function

ABSTRACT

Indicating road safety-related aspects in the phase of planning and operating is always a challenging task for experts. The success of any method applied in identifying a high-risk location or black spot (BS) on the road should depend fundamentally on how data is organized into specific homogeneous segments. The appropriate combination of black spot identification (BSID) method and segmentation method contributes significantly to the reduction in false positive (a site involved in safety investigation while it is not needed) and false negative (not involving a site in safety investigation while it is needed) cases in identifying BS segments. The purpose of this research is to study and compare the effect of methodological diversity of road network segmentation on the performance of different BSID methods. To do this, four commonly applied BS methods (empirical Bayesian (EB), excess EB, accident frequency, and accident ratio) have been evaluated against four different segmentation methods (spatial clustering, constant length, constant traffic volume, and the standard Highway Safety Manual segmentation method). Two evaluations have been used to compare the performance of the methods. The approach first evaluates the segmentation methods based on the accuracy of the developed safety performance function (SPF). The second evaluation applies consistency tests to compare the joint performances of the BS methods and segmentation methods. In conclusion, BSID methods showed a significant change in their performance depending on the different segmentation method applied. In general, the EB method has surpassed the other BSID methods in case of all segmentation approaches.

1. Introduction and literature review:

The identification of hazardous road sections is one of the main parts of any successful road safety management process. The identification of hazardous road sections, high-risk accident locations or black spots (BS) receives great interest from road agencies and safety specialists. Black spot identification (BSID) can be defined as the process of searching for locations in transportation systems with a higher number of accidents than other similar locations mainly caused by local risk factors (Cheng and Washington, 2005; Elvik, 2007). Errors in BSID can result in many false positive and false negative cases. In other words, accidents can occur in both safe and unsafe locations, and the challenge is to avoid false positive and negative cases in identifying the most dangerous locations. Therefore, the correct identification of accident BS is essential for the greatest benefit of traffic safety projects (Montella, 2005).

There is a fairly extensive literature focusing on developing or applying BSID methods. Some of the studies rank BS locations by the accident frequency (AF) method (accidents per year or accident per km per year), some use the accident rate (AR) method (accidents per vehicle per kilometers or per entering vehicles), and some use a combination of the two (Borsos et al., 2016; Deacon et al., 1974; Yakar, 2015). These methods are usually not based on statistical models and can vary in the analyzed period. Some of the researchers have proposed using empirical Bayesian (EB) techniques (Cheng and Washington, 2008, 2005; Montella, 2010). The EB method combines the benefit of observed and predicted accident frequencies. The two values are weighted in a statistical model based on the reliability level of the predicted value, which is derived from a safety performance function (SPF) developed from historical accident data. SPF is a regression equation that estimates the average crash frequency for a given site. During the calibration of SPF, segmentation of accident dataset represents a crucial issue.

Data segmentation is a process of dividing and classifying a large and complex dataset into small and simple homogeneous groups or entities in which data within a group are very similar but dataset of the groups is dissimilar. Considering that traffic accident data is heterogeneous, in general, data segmentation is considered the first and most

* Corresponding author.
E-mail addresses: ghadi.maen@mail.bme.hu (M. Ghadi), arpad.torok@gjt.bme.hu (Á. Török).

https://doi.org/10.1016/j.aap.2019.03.002
Received 22 November 2018; Received in revised form 1 February 2019; Accepted 6 March 2019
Available online 04 April 2019
0001-4575/

important step of the BSID process. The more accurate the accident data is segmented, the more accurate the SPF will be (Cafiso et al., 2013), and this will consequently affect the performance of some BSID methods that rely on accident prediction models in their criterion. On the other hand, the segmentation process is important in the mechanism of locating BS segments and identifying their optimum lengths. Thomas (1996) argued that applying different lengths for segmenting road network can result in different definitions of hazardous locations which, in turn, affects the reliability of results. Koorey (2009) discussed the benefit of applying variable length segments and their effect on BS determination. It has been confirmed as well that selecting the appropriate segmentation method can have a significant effect on reducing the false positive and negative in the determination of BS segments. Most often, road segmentation methods are based on researchers' experiences, methodological decisions or objectives. According to the Highway Safety Manual (HSM) (American Association of State Highway and Transportation Officials, 2010), roadway segments should have consistent cross-sections and their endpoints can be marked by changes in Annual Average Daily Traffic (AADT) and other roadway features. In practice, it is not always easy to implement this type of segmentation as not all the variables are available (Fitzpatrick et al., 2006). Other researchers (Black and Thomas, 1998) used constant length segments. Constant length segmentation can result in a different BS set depending on the selected length and starting point of a segment (Thomas, 1996). Generally, segmenting roadways into homogeneous segments based on too many variables can result in very short average segment lengths, which can produce many zero-accident sections. In contrast, increasing the length of the segment would sacrifice the homogeneity. Recently, clustering techniques have begun to attract more and more attention in segmenting accident data (De Luca et al., 2012; Depaire et al., 2008; Ghadi et al., 2018). Clustering can be a useful technique to find hidden relationships and patterns for large datasets of accidents.

Compared with a large number of studies focused on the development of various BSID methods and various segmentation methods, considerably less research has been dedicated to comparing their performance. Cheng and Washington (2008) (Cheng and Washington, 2008) proposed three new consistency tests (site consistency test, method consistency test, and total rank differences test) to compare BSID methods. They applied these three tests in addition to two other tests to compare the performance of four commonly used BSID methods. They concluded that the EB showed the best consistency among the other BSID methods and should be the standard in the identification of BS. The same result was obtained by Montella (2010) who also proved the efficiency of the EB method compared to the other seven BSID methods. Cafiso et al. (2013) evaluated five different segmentation methods, which had segmented a roadway based on geometric and/or traffic-related attributes by comparing their performance in developing accident prediction models. The findings revealed that the method based on road design parameters (i.e. curvature characteristics) performed best since the resulting set of high-risk sections seemed to be well-correlated with the set of locations characterized by high accident density. On the other hand, very few studies have compared the performance of applying different segmentation methods in the BSID. Kwon et al. (2013) compared the performance of three dynamic segmentation methods (Sliding Moving Window, Peak Searching, and Continuous Risk Profile) in locating BS sites. The conclusion revealed that, although the input requirements for each of the three methods were identical, their performance varied markedly.

Based on the performed overall literature review, it can be said that no previous works have been done to study the combined impact of certain BSID and network segmentation methods on performance. This research aims to investigate and compare the joint performance between BSID methods and segmentation methods. Four commonly applied BS methods (EB, Excess EB (EEB), accident frequency (AF), and accident rate (AR)) have been evaluated against four different segmentation methods (spatial clustering, constant length, constant AADT, and the standard HSM segmentation method). Two tests have been used to compare the performance of the methods. The first evaluates the extent to which data segmented by different segmentation methods succeed in developing the more accurate SPF. The second applies the consistency tests to compare the joint performances between the BS methods and segmentation methods.

2. Description of segmentation methods

In the next section the four different segmentation methods mentioned are introduced. The model presented first is developed by the authors; introduction of the other models aims to be able to provide a comprehensive comparison.

2.1. The developed spatial clustering segmentation method

Cluster analysis is primarily intended to organize a large data set into a small number of homogeneous groups in which the degree of association between the objects of the same group is maximal. K-Means cluster is a popular clustering partitioning method. 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 technique reasonably commonly used to derive this error. The spatial location is a key classification variable in this article since it has a significantly higher probability that neighboring accidents have strongly related causes than in case of distantly located accidents (Flahaut et al., 2003). To reduce the demanded calculation capacity of the problem, instead of a bivariable representation of accidents' spatial location in the clustering algorithm, a linear referencing model seems to be appropriate. Linear referencing is used to locate objects along a curve from a reference point (RP). The linear referencing method has been applied to locate accidents along a single road where the road has been represented as a one-dimensional line, with a zero-distance starting point as a RP, as presented in Fig. 1. All accidents have been located along the line (road) by measuring their distances from the RP. For instance, in Fig. 1, the first accident (represented by a point counted from the left) is located 0.2 distance unit away from the RP. This approach results in a one-variable representation for K-Means clustering contrary to the two-variable geographical coordinates (the longitude and latitude), considering only road network in the clustering process (Ghadi et al., 2018). It also allows avoiding classifying accidents in the same cluster that are closely located but occurred on different roads. Accordingly, in the K-means clustering method, a cumulated distance parameter has been defined in one dimension space for each accident starting from the beginning of the road. So, each resulted cluster can be characterized by reasonably low heterogeneity regarding accident, traffic and infrastructure parameters. Fig. 1 presents an illustration of the accident clusters (SC1, SC2, SC3) generated by the proposed methodology.

The application of spatial clustering segmentation (Seg-1) can be very attractive since the road accident-based segmentation method makes it possible to relate the length of the identified segments to the clusters' lengths. Each cluster length depends on the number of accidents contained by the cluster and their spatial distribution. The length of a segment equals the distance between the first and last accidents

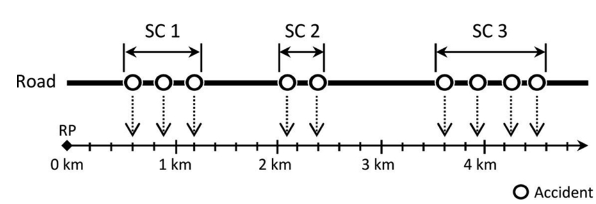

Fig. 1. A visual illustration of the spatial clustering.
Source: Ghadi et al. (2018).

within the cluster. In Fig. 1, for instance, the one-dimensional extent of SC1 spatial cluster can be represented by its length which is 0.6 distance unit (i.e. SC1 length = 1.2–0.6) and the number of accidents included by this cluster is 3. Short segments have been eliminated to avoid segments under 100 m, as recommended by the HSM (American Association of State Highway and Transportation Officials, 2010).

One of the crucial issues related to K-means clustering is that the number of clusters (k) has to be determined in advance, which can cause uncertainty without any prior knowledge. For determining the optimal estimated number of clusters (k_0), it is required to test different numbers of k and analyze the changes of the variances (or SSE) compared to cluster centers. A method of minimum variance partition has been used (Calinski and Harabasz, 1974). This method searches for a point, which splits the SSE function domain into two parts. Until this point, each added cluster results in a substantial reduction in the value of variance, and after the given point, any increase in k results in less-and-less reduction in the value of variance. This approach is analog to the second derivative of the variance as a function of the number of clusters.

The of accident segmentation process into smaller clusters is summarized as follows:

- Localizing accident objects based on GIS coordinates.
- Converting geographically referred roads into linearly referred one-dimensional objects.
- Localizing accident objects along linearly referred one-dimensional objects representing the investigated components of the road network, measured from the RP.
- Measuring and recording object distances providing input for the K-means algorithm.
- Identifying the optimal number of clusters K.
- Applying the K-means algorithm to define accident clusters and identifying their extension.

2.2. The other segmentation methods

The other three segmentation methods investigated in this article represent the most used methods; each of them differs in the criteria applied to determine homogeneous road segments. The methods are described as follows.

- Constant length segment (Seg-2): The road network is divided into equal length segments, while other variables do not affect the segmentation structure. The segment length has been chosen to be 750 m. It is an average length that is longer than the recommended minimum value of the HSM but short enough to provide homogeneity, so the characteristics of the road do not change too much within the length.
- Constant AADT (Seg-3): The resulted segments are homogenous in AADT while other variables do not affect the segmentation structure.
- HSM method (Seg-4): The generated segments are homogenous in AADT, roadside hazards, and presence of curves, as recommended by the HSM specifications for highway roads.

3. Description of BSID methods

Four commonly applied BSID methods have been tested and compared. Each method varies in the criteria and variables used to find BS sites, as described below.

- Empirical Bayesian method: The EB method combines the benefit of both observed and predicted accident frequency. The EB value can be used to estimate the expected average crash frequency (N_E) for both future and past periods (Eq. (1)), if only both observed (N_O) and predicted (N_P) number of accidents are available.

$$N_E = w \times N_P + (1 - w) \times N_O \tag{1}$$

  The weight factor (w) in Eq. (1) represents the degree of reliability in obtaining N_P, and it is inversely proportional with an over-dispersion parameter for the developed N_P model. Therefore, as the N_P value is more dispersed lower w will be obtained in Eq. (1) and vice versa. However, it can be noted that in Eq. (1), N_P is a very important variable. The predicted average crash frequency (N_P) can be estimated using SPF. The SPF is a regression equation that estimates the average accident frequency for a given site.
- Excess Empirical Bayesian method: EB can be used to calculate the Excess Empirical Bayesian (EEB), which is the difference between the estimated EB value (N_E) and the predicted accident frequency (N_P or SPF value). The EEB tries to rank BS sites according to their potential for safety improvements.
- Accident frequency (AF): AF is a straightforward method. This method ranks BS sites in descending order based on the observed number of accident per unit length during a given period (e.g. 1 year). However, this method only considers the number of accident in defining the risk level of a given infrastructure component without taking into account other important safety variables, which is an important drawback.
- Accident rate (AR): The AR improves some defects of AF. AR method considers traffic volume or entering vehicle volume as well. The result of this method is the descending order of the evaluated infrastructure components based on their estimated accident rank values. This method enjoyed broad applications by traffic agencies and road safety researchers. The output of this method still refers to historical data, so its applicability in prediction is frequently disputed.

4. Evaluation tests of performance

4.1. Segmentation methods evaluation criteria

The output of accident prediction models is significantly affected by the selected segmentation methodology. Safety performance function (SPF) is a commonly used model for predicting accident frequency. SPF represents a mathematical relationship between explanatory variables and accident frequency. A negative binomial regression model is usually used to fit accident data and define the SPF. The general form of the SPF is as follows (Eq. (2)) (Cafiso et al., 2013).

$$SPF = \exp\left[\alpha + \left(\sum_n \beta_n \times \ln(X_n)\right) + \ln(length)\right] \tag{2}$$

where; α is the intercept and β_n is the regression parameter of the corresponding explanatory variable X_n (i.e. AADT, speed, etc.).

To evaluate and compare goodness-of-fit between different developed SPFs, two different statistical methods are applied: the Quasi-likelihood under independence model criterion (QIC) (Cui, 2007; Pan, 2001) and the Pearson correlation coefficient (PCC) (Smyth, 2003). QIC is one of the most widespread well-established goodness-of-fit statistics for the generalized estimating equation (GEE), which is an extension of Akiake's information criterion (AIC) (Akaike, 1974). QIC can be used to select the most applicable estimation model structure in the generalized estimating equation (GEE) analyses. Like AIC, it balances the model fit with model complexity to select the most effective model. The models being compared do not need to be nested, which means that the parameters of the different evaluated models do not need to be the subsets of each other. Therefore, QIC can be used to compare and rank different models with different number of parameters. When QIC is applied to model evaluation, the model with the smallest QIC is preferred. However, QIC does not provide any test regarding a null hypothesis whether data fit the model results or not. The PCC works as a complementary test for evaluating the models. The PCC measures the linear correlation between the observed and predicted accident frequencies. The PCC has a value between +1 and −1. The higher the PCC absolute value is, the stronger the correlation between the observed and predicted values.

The selection of the explanatory variables for the SPF has been carried out by applying stepwise technique while testing the influence and significance of inserting or removing different variables for the SPF. The potential explanatory variables are described as follows:

- AADT: AADT is considered a major factor in predicting the number of accidents.
- Speed limit: speed is also an important factor in accident risk, especially for high-speed roads (Ghadi and Török, 2017).
- Percent of trucks: the percentage of trucks is measured compared to the total traffic in case of each specific segment.
- Horizontal deflection angle (HDA): HDA measures curvature change rate from a straight line horizontally.

4.2. BSID methods evaluation criteria

The performance of the four BSID methods (EB, EEB, AF, and AR) in identifying BS sites has been tested and compared based on different segmentation methods by applying the method of consistency test. The consistency tests include four quantitative evaluation tests (Cheng and Washington, 2008; Montella, 2010): the site consistency test, the method consistency test, the total rank differences test, and the total score test.

4.2.1. Site consistency test

The site consistency test (SCT) is applied to measure the ability of a BSID method to identify consistently high-risk sites over the successive observation periods. The test is based on the premise that a site which is identified as a high-risk site in the first period should also possess lower safety performance at a later period if the given infrastructure component has remained more or less unchanged compared to its baseline state. In a statistical perspective, the site consistency test assumes that the more efficient a BSID method is, the more the estimated future accident numbers (period 2) of the high-risk sites correlate to the current data (period 1).

4.2.2. Method consistency test

The method consistency test (MCT) is designed to evaluate BS methods by measuring the number of observed road segments that have been identified as high-risk sites in both periods (i.e. period 1 and period 2). The higher the rate (compared to the total number of segments considered) of sites identified as high-risk sites in both periods, the more consistent the BSID method performed. Thus, a good BSID method is the one that identifies the same high-risk sites over the two periods.

4.2.3. Total rank differences test

The total rank differences test (TRDT) is primarily based on the method consistency test. It takes into account the rank of the high-risk sites in both periods. The total rank differences test calculates the sum of rank total differences of the high-risk sites identified over the two periods. Therefore, contrary to the two previous tests, the higher the test value is, the less efficient the BSID method is.

4.2.4. Total score test

The total score test (TST) aims to represent the aggregated result of the previous three tests in a single value, assuming that all tests have the same weight. Accordingly, the test can be represented by the following formula:

$$TST = \frac{100}{3} \times \left[\left(\frac{SCT}{\max(SCT)} \right) + \left(\frac{MCT}{\max(MCT)} \right) + \left(1 - \frac{TRDT - \min(TRDT)}{\max(TRDT)} \right) \right] \tag{3}$$

The total score test facilitates making a comprehensive comparison of the BSID methods, considering all types of the mentioned consistency test approaches. The closer the total score test value of a BS method is to 100%, the more reliable the given analyzed BS method is.

During the consistency calculation method, the values of relative accident indicator have been defined proportionately to unit length and to a year-long period. The values of the consistency tests have been expressed proportionately to the number of segments. This can ensure an equivalent comparison between BSID methods applied to any segmentation method, as will be shown later in the “Results”.

5. Data description

In this study, traffic accident data of different Hungarian highway sections has been used. The length of the sample network is 1965 kilometers (in one direction) and the data timeframe of the analysis covers the period from 2013 to 2015. The investigated data includes the main parameters of the road cross-section, accident data, road design parameters, and traffic characteristics. In general, the cross-section of the investigated road category can be characterized by 2 lanes in each direction, divided by a median with barriers. The analysis focuses only on road segments; accordingly, intersections data is out of the scope of the investigation. The investigated dataset includes 2155 fatal and injury accidents related to the given period and the sample network.

6. Results

The proposed K-means clustering technique (Seg-1) and the other segmentation methods; constant length (Seg-2), constant AADT (Seg-3) and the standard HSM approach (Seg-4), have been applied for segmenting road accident network. The clustering method has resulted in a significantly lower number of segments (i.e. 304) with a higher average segment length (i.e. 6.76 km) in comparison with the other segmentation methods in case of the investigated roads regarding the analyzed period (2013–2015). Table 1 gives descriptive statistics of the applied segmentation methods.

After segmenting the road network with four different methods, it is required to compare the efficiency of each method and to evaluate their impact on BSID methods. In this research, two evaluation tests have been used.

6.1. Development of SPF results

The four segmentation approaches have been applied to develop a new process for the input of SPF. The stepwise approach has been used to check the significance of inserting or removing different explanatory variables for each SPF every time. Finally, AADT, speed, and HDA have been selected to be the input variables of the SPF's. Accident data from the first year (2013) of the study sample has been used to develop the models, while data of the year 2015 has been used for checking the performance of the models developed. The model calibration results for all segmentation methods are presented in Table 2.

Most of the estimated parameters (i.e. α, $\beta 1$, $\beta 2$ and $\beta 3$) of the SPFs presented in Table 2 are statistically significant to 95% confidence level and the rest to 90% confidence level. By evaluating the SPFs in terms of data fitting, it is evident from Table 2 that the Seg-1-based method shows greater potential for fitting accident data; this is obvious from its significantly lower QIC value ($QIC_1 = 324$). In contrast, Seg-2-based method is rather over-fitting the developed model, with the highest QIC value ($QIC_2 = 1721$).

Table 1
Descriptive statistics of segmentation results.

Parameters description	Seg-1	Seg-2	Seg-3	Seg-4
Segment length				
Min length (km)	0.22	0.75	0.20	0.42
Max length (km)	19.32	0.75	18.70	15.53
Average length (km)	5.12	0.75	5.39	2.37
Standard deviation	4.08	0.00	4.33	1.49
Total length (km)	1965	1965	1965	1965
Number of segments	304	2620	362	829
Accident frequency (fatality + injury)				
Min	0	0	0	0
Max	13	6	22	11
Mean	2.6	0.3	2.0	0.9
Standard deviation	2.6	0.6	2.1	1.1
Total	2155	2155	2155	2155
AADT				
Min	867	827	882	827
Max	22,163	56,140	55,677	56,176
Mean	6724	6473	7885	6651
Standard deviation	3499	4373	5680	4742
Horizontal deflection angle (HDA) (per km)				
Min	5	0	5	2
Max	449	1736	621	1207
Mean	89	106	102	124
Standard deviation	72	148	96	146
Speed				
Min	40	30	38	30
Max	90	100	97	97
Mean	76	74	72	74
Standard deviation	13	18	15	18

Table 2
Estimated values of SPF parameters and goodness-of-fit.

Parameters	Method of segmentation			
	Seg-1	Seg-2	Seg-3	Seg-4
α (Intercept) [*p*-value]	−4.721 [0.056]	−6.215 [> 0.001]	−3.579 [0.010]	−6.865 [> 0.001]
$\beta 1$ (AADT) [*p*-value]	0.628 [> 0.001]	0.653 [> 0.001]	0.670 [> 0.001]	0.675 [> 0.001]
$\beta 2$ (Speed) [*p*-value]	−0.625 [0.086]	−0.269 [0.076]	−0.871 [> 0.001]	−0.288 [0.036]
$\beta 3$ (DOA) [*p*-value]	0.184 [0.007]	0.148 [> 0.001]	0.058 [0.023]	0.216 [0.001]
Over-dispersion	1.387	1.717	1.108	1.096
QIC	324	1721	390	967
PCC	0.692	0.198	0.561	0.462

The Pearson correlation coefficient (PCC) method was also used to evaluate and compare the developed SPF's based on measuring their efficiency in predicting accident data for another year (2015) in the future. The values of the PCC are shown in Table 2. It is noted that the SPF developed by the Seg-2 method shows weakness in predicting accident frequency with the lowest PCC value (i.e. 0.20). Contrary to this, the proposed, clustering segmentation (Seg-1)-based model gives the best correlation between the predicted and observed data. This is described by its highest PCC value which is equal to 0.70.

6.2. Consistency tests results

Consistency tests have been fed by relative accident data regarding the unit long section of the segments and of one-year long period, as described previously, to achieve a comprehensive comparison between any BSID method and other segmentation methods. Consistency test results are described in the next section.

Table 3
Site consistency test results.

BS method	Segmentation method			
	Seg-1	Seg-2	Seg-3	Seg-4
5% risk				
EB	1.24	1.18	1.21	1.02
EEB	1.20	1.28	0.99	0.97
AF	1.41	1.01	1.23	0.96
AR	0.93	0.69	0.65	0.60
10% risk				
EB	1.03	0.94	1.10	0.92
EEB	102	0.95	0.89	0.78
AF	1.14	0.82	0.97	0.80
AR	0.88	0.64	0.64	0.60

6.2.1. Site consistency test

It is shown in Table 3, that from the point of view of the site consistency test, the best-performing models have been the EB and AF methods with reasonably close results. In the case of Seg-1, the AF method performs the best in identifying the top 5% and 10% of BS sites. Accordingly, the AF method has produced the highest equivalent accident frequencies for the same BS sites in case of the top 5% and 10% of BS sites (1.41, 1.14) between the compared baseline period (2013) and predicted period (2014–2015). In addition, the AF method has shown superiority with the Seg-3 method for the top 5% of BS sites. On the other hand, the EB method has also been able to excel in identifying the top 5% of BS sites, with a value of 1.02 for the standard HSM segmentation method (i.e. Seg-4), and also the top 10% of BS sites in case of Seg-3 and Seg-4, with 1.10 and 0.92.

In general, the highest value of the site consistency test for all segmentation methods and all BSID methods has been provided by the common application of the AF method and the Seg-1 method in both the top 5% and 10% of high-risk sites, with 1.41 and 1.14 equivalent accident frequencies. Contrary to this, the AR method performs worst in identifying the BS sites in all segmentation methods and especially for the Seg-4 method, resulting in the lowest accident frequencies between all BSID methods, with 0.6 accidents for both the top 5% and 10% BS sites.

6.2.2. Method consistency test

In the method consistency test, the EB method has outperformed the other BSID methods in most cases for the two study periods, based on the localized number of similar BS sites as shown in Table 4. In the case of the constant length segmentation method (Seg-2), the EB model has been the superior in both the 5% and the 10% risk cases, by identifying 35% and 39% of correlated BS sites between the two periods. Similarly, the highest test values in case of the Seg-4 method have also been provided by the EB method for the top 5% and 10% high-risk sites (34%

Table 4
Method consistency test results.

BS method	Segmentation method			
	Seg-1	Seg-2	Seg-3	Seg-4
5% risk				
EB	0.47	0.35	0.33	0.34
EEB	0.41	0.28	0.17	0.20
AF	0.53	0.23	0.33	0.20
AR	0.29	0.23	0.33	0.27
10% risk				
EB	0.45	0.39	0.44	0.43
EEB	0.45	0.33	0.28	0.27
AF	0.54	0.30	0.39	0.31
AR	0.37	0.24	0.36	0.34

Table 5
Total differences test results.

BS method	Segmentation method			
	Seg-1	Seg-2	Seg-3	Seg-4
5% risk				
EB	29	373	38	101
EEB	47	704	111	234
AF	41	406	53	190
AR	57	561	78	205
10% risk				
EB	40	465	37	142
EEB	56	986	113	265
AF	46	450	71	202
AR	56	576	83	222

and 43%). In the case of the Seg-1 method, the AF method has achieved the maximum values. This practically means 53% and 54% of sites characterized by similar accident frequency in case of the top 5% and 10% of BS sites, in case of the AF method.

The best results of the method consistency test have been produced by the Seg-1 method and have been achieved by the EB and AF methods, as presented in Table 4. The lower test results have been provided by the other BS methods strongly depending on the segmentation method applied. In the case of Seg-1 and Seg-2 methods, AR has performed the worst, whereas the Excess EB method has been the last BSID method for the rest of the segmentation methods (i.e. Seg-3 and Seg-4).

6.2.3. Total rank differences test

Table 5 illustrates that the EB method has significantly smaller total differences for all applied segmentation methods, which makes it the superior method in this test also followed directly by the AF methods. The lowest values of the EB method have resulted in the case of Seg-1 and Seg-3 methods for both the top 5% and 10% high-risk sites, respectively, with relatively close values, whereas, the poorest test values have been provided by the combined application of the Seg-2 method and the EEB model.

6.2.4. Total score test

The total score test combines the results of the previous three consistency tests (site consistency test, method consistency test, and total differences test).

Table 6 illustrates the values of the total score test that describe the total efficiency of each BSID method with respect to every segmentation method applied, for the two case studies (5% and 10% BS sites). The column titled "Total", presented in Table 6, represents the average performance for every BSID method considering all the investigated segmentation methods. The row titled "Total", presented in Table 6, represents the average performance for every segmentation method considering all the BS methods investigated.

Table 6
Total score test results.

BS method	Segmentation method	Total			
		Seg-1	Seg-2	Seg-3	Seg-4
5% risk					
EB	92.2	84.4	79.2	81.5	*84.3*
EEB	77.4	77.8	61.0	66.7	70.7
AF	93.1	72.8	79.0	66.9	77.9
AR	57.8	72.7	77.7	73.7	70.5
Total	*80.1*	76.9	74.2	72.2	
10% risk					
EB	96.8	81.9	78.3	84.7	*85.4*
EEB	82.0	77.0	65.0	71.5	73.9
AF	96.3	74.1	73.7	75.3	79.9
AR	72.6	69.5	71.3	77.1	72.6
Total	*86.9*	75.6	72.1	77.2	

In the two case studies (5% and 10% BS sites), the EB and AF methods have performed better than the other BSID methods. Generally, the best total score test values are included by Seg-1 column (see Table 6), except the AR model which has performed better with the constant AADT segmentation method (Seg-3). Contrary to this, most of the BSID processes which have been based on Seg-3 or Seg-4 methods, are performed reasonably worse. This is obvious, as the lowest total average performance of all BSID methods has been caused by Seg-4 (72.2) for the 5% high-risk sites and by Seg-3 (72.1) in case of the top 10% of the high-risk sites. Besides this, the AR method has provided the lowest average performance between the BSID methods considering all segmentation models.

7. Discussion of results

The previously presented results evaluate the performance of the introduced four commonly applied BS identification methods in case of the four analyzed road accident segmentation methods, using accident data from the Hungarian secondary main road network.

The data emerged from each applied segmentation method and has been fed into different accident prediction models (i.e. SPFs). Also, consistency tests have been used to evaluate and compare the performances of the four BSID methods in the localization of the high-risk accident sites. Based on the prepared analysis, the following conclusions can be drawn:

- The performance of SPFs differs, depending on the applied segmentation method. The more homogenous the accident groups' result, the less error characterizes the model. This result is supported by Ghadi and Török (2019) who discovered different goodness-of-fit of SPFs related to five different segmentation methods.
- The performance of some BSID methods, e.g. EB and EEB, which apply the SPF in its criteria, is hence significantly affected by the segmentation method applied in the model, as also approved by Kwon et al. (2013).
- The performance of each BSID method varies in case of the different segmentation methods.
- The performance of all BSID methods, except the AR, was the best in the consistency tests when the spatial segmentation (Seg-1) method has been applied to road accident data (see Table 6) (Ghadi and Török, 2018). This result is also supported by the SPF test (Table 2), where the datasets classified by the Seg-1 method has produced the best fitting SPF (Ghadi and Török, 2019). This is logical since the spatial segmentation classifies closely located accidents in groups that are expected to have the same causes, as argued by (Flahaut et al., 2003).
- The EB methods performed better than the other BSID methods in most tests. This is supported by a theoretical basis because the EB method takes advantage of the observed and predicted values in its statistics, which in turn increases the reliability of its results. In practice, the EB method proved its efficiency in this and other studies (Cheng and Washington, 2005; Elvik, 2008; Montella, 2010; Qu and Meng, 2014).
- The second best BSID method has been recorded by the AF, especially when it is applied with the spatial-based segmentation method (Seg-1). The performance of the AF method has been fairly close to the EB method and slightly outperformed it in some cases. Although, its performance has been poor in case of the Seg-4 method. Similar results have been obtained by Cheng and Washington (2008).
- The EEB and AR methods seem to be reasonably inconsistent in most of the cases. The EEB method is largely affected by the predicted value and consequently the validity of the developed SPF. As the predicted value of accident increases, the likelihood of a site being

selected as a BS increases. The main drawback of the AR method is that it incorrectly assumes a linear relationship between accident frequency and traffic volume, since accident intensity can be highly inferred from other factors, like speed (Ghadi and Török, 2017). This fact must be taken into consideration in case of using AR in road safety investigations. These results are also consistent with the results of other researches (Montella, 2010; Persaud et al., 1999).

8. Conclusion

The main purpose of this paper is to investigate and compare the influence of applying different highway segmentation methods on the performance of different black spot identification (BSID) methods. Beside this the article aims to examine the applicability of the newly developed segmentation method and its contribution to the performance of other BSID methods. Four different BSID methods have been evaluated against four different segmentation methods. The first and new segmentation method (Seg-1) applies clustering techniques to segment accident data based on their spatial distribution. The second and third methods (Seg-2 and Seg-3, respectively) apply a constant length and constant average annual daily traffic (AADT) segments, respectively, while the last method (Seg-4) is based on the specifications of the Highway Safety Manual (HSM), using curvature and AADT data. In the case of BS methods, the Empirical Bayesian (EB), Excess EB (EEB), accident frequency (AF), and accident rate (AR) methods have been applied. Two tests have been used to compare the performance of the methods in the case of a Hungarian highway. In the first test, the performance of safety performance functions (SPFs) developed from datasets generated by every method of segmentation has been statistically compared. In the second test, four consistency tests (site consistency test, method consistency test, total rank differences test, and total score test) have been used to compare the joint performances between the BSID methods and segmentation methods.

In the first test, which compares the performance of the segmentation methods, the best goodness-of-fit value has been obtained by the proposed spatial clustering segmentation model (Seg-1). This is rational since in this case numerous factors have been helped to optimize SPF calibration, like the consistency of the segmented datasets, the flexibility of segment length, quality, and other variables. This result has been supported by the second test, which has confirmed that all BSID methods, except AR method, have reached their best performance in the case of the spatial clustering segmentation (Seg-1) method. Contrary to this, the consistency evaluation tests have also shown that the EB method performed better than the other BSID methods in most of the cases. The test results highlight that the EB method is the most consistent method for identifying priority investigation sites. The AF has directly followed the EB, although it slightly outperformed in a few cases.

Beside this, another interesting result of the consistency tests is the performance variation of BSID methods in case of the different segmentation methods applied. For instance, the AF method has recorded its best performance, with 93.1% consistency, in case of the Seg-1 method, while it has achieved only 66.9% consistency, as a third performance, in the case of the Seg-4 methods, for the top 5% of the studied black spot sites (as presented in Table 6). In general, the performance of the EEB and AR methods has been the weakest in most segmentation cases. This is quite alarming, as many highway agencies and researchers use these methods.

In general, the results of this research try to highlight that data segmentation is an important and fundamental step in road accident data classification and configuration for subsequent operations. This has been demonstrated, in this research, by the different performance obtained by the BSID methods with different applied segmentation methods. However, there is a wide scope for using other segmentation and BSID methods and comparing them with more reliable approaches in future research.

References

Akaike, H., 1974. A new look at the statistical model identification. IEE Trans. Autom. Control 19 (6), 716–723. https://doi.org/10.1109/TAC.1974.1100705.

American Association of State Highway and Transportation Officials, 2010. Highway Safety Manual, 1st edition. .

Black, W.R., Thomas, I., 1998. Accidents on belgium's motorways: a network autocorrelation analysis. J. Transport Geogr. 6 (1), 23–31. https://doi.org/10.1016/S0966-6923(97)00037-9.

Borsos, A., Cafiso, S., D'Agostino, C., Miletics, D., 2016. Comparison of Italian and Hungarian black spot ranking. Transportation Research Procedia 14, 2148–2157. https://doi.org/10.1016/j.trpro.2016.05.230.

Cafiso, S., D'Agostino, C., Persaud, B., 2013. Investigating the influence of segmentation in estimating safety performance functions for roadway sections. TRB 92nd Annual Meeting, 15.

Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Commun. Stat. Theory Methods 3 (1), 1–27. https://doi.org/10.1080/03610927408827101.

Cheng, W., Washington, S., 2008. New criteria for evaluating methods of identifying hot spots. Transport. Res. Rec. J. Transport. Res. Board 2083 (1), 76–85. https://doi.org/10.3141/2083-09.

Cheng, W., Washington, S.P., 2005. Experimental evaluation of hotspot identification methods. Accid. Anal. Prev. 37 (5), 870–881. https://doi.org/10.1016/j.aap.2005.04.015.

Cui, J., 2007. QIC program and model selection in GEE analyses. Stata J. 7 (2), 209–220. https://doi.org/10.3760/cma.j.issn.1003-9406.2011.02.007.

Deacon, J.A., Zegeer, C.V., Deen, R.C., 1974. Identification of hazardous rural highway locations. Kentucky Transportation Center Research Report, vol. 29. Retrieved from https://www.uknowledge.uky.edu/ktcresearchreports/897.

De Luca, M., Mauro, R., Lamberti, R., Dell'Acqua, G., 2012. Road safety management using bayesian and cluster analysis. Procedia Soc. Behav. Sci. 54, 1260–1269. https://doi.org/10.1016/j.sbspro.2012.09.840.

Depaire, B., Wets, G., Vanhoof, K., 2008. Traffic accident segmentation by means of latent class clustering. Accid. Anal. Prev. 40 (4), 1257–1266. https://doi.org/10.1016/j.aap.2008.01.007.

Elvik, R., 2007. State-of-the-art approaches to road accident black spot management and safety analysis of road network. Ripcord-Iserest.

Elvik, R., 2008. Comparative analysis of techniques for identifying locations of hazardous roads. Transport. Res. Rec. J. Transport. Res. Board 2083, 72–75. https://doi.org/10.3141/2083-08.

Fitzpatrick, K., Schneider, W., Carvell, J., 2006. Using the rural two-lane highway draft prototype chapter. Transport. Res. Rec. J. Transport. Res. Board 1950, 44–54. https://doi.org/10.3141/1950-06.

Flahaut, B., Mouchart, M., San Martin, E., Thomas, I., 2003. The local spatial autocorrelation and the kernel method for identifying black zones. A comparative approach. Accid. Anal. Prev. 35 (6), 991–1004. https://doi.org/10.1016/S0001-4575(02)00107-0.

Ghadi, M., Török, Á., 2019. Comparison of different road segmentation methods. Promet Traffic Transport (in press).

Ghadi, M., Török, Á., 2017. Comparison different black spot identification methods. Transport. Res. Procedia 27, 1105–1112. https://doi.org/10.1016/j.trpro.2017.12.104.

Ghadi, M., Török, Á., 2018. A comparative analysis of spatial clustering segmentation method of road accidents. 8th International Scientific Conference CMDTUR. University of Zilina, Zilina, pp. 95–100.

Ghadi, M., Török, Á., Tánczos, K., 2018. Integration of probability and clustering based approaches in the field of black spot identification. Period. Polytech. Civil Eng. https://doi.org/10.3311/PPci.11753.

Koorey, G., 2009. Road data aggregation and sectioning considerations for crash analysis. Transport. Res. Rec. J. Transport. Res. Board 2103 (1), 61–68. https://doi.org/10.3141/2103-08.

Kwon, O.H., Park, M.J., Yeo, H., Chung, K., 2013. Evaluating the performance of network screening methods for detecting high collision concentration locations on highways. Accid. Anal. Prev. 51, 141–149. https://doi.org/10.1016/j.aap.2012.10.019.

Montella, A., 2005. Safety reviews of existing roads: quantitative safety assessment methodology. Transportation Research Record. J. Transport. Res. Board 1922, 62–72. https://doi.org/10.3141/1922-09.

Montella, A., 2010. A comparative analysis of hotspot identification methods. Accid. Anal. Prev. 42 (2), 571–581. https://doi.org/10.1016/j.aap.2009.09.025.

Pan, W., 2001. Akaike's information criterion in generalized estimating equations. Biometrics 57 (1), 120–125. https://doi.org/10.1111/j.0006-341X.2001.00120.x.

Persaud, B., Lyon, C., Nguyen, T., 1999. Empirical Bayes procedure for ranking sites for safety investigation by potential for safety improvement. Transport. Res. Rec. J. Transport. Res. Board 1665, 7–12. https://doi.org/10.3141/1665-02.

Qu, X., Meng, Q., 2014. A note on hotspot identification for urban expressways. Saf. Sci. 66, 87–91. https://doi.org/10.1016/j.ssci.2014.02.006.

Smyth, G.K., 2003. Pearson's goodness of fit statistic as a score test statistic. Science and Statistics: A Festschrift for Terry Speed 40 (March), 1–12. Retrieved from http://www.statsci.org/webguide/smyth/pubs/goodness.pdf.

Thomas, I., 1996. Spatial data aggregation: exploratory analysis of road accidents. Accid. Anal. Prev. 28 (2), 251–264. https://doi.org/10.1016/0001-4575(95)00067-4.

Yakar, F., 2015. Identification of accident–prone road sections by using relative frequency method. Promet Traffic Traffico 27 (6), 539–547. https://doi.org/10.7307/ptt.v27i6.1609.