Budapest University of Technology and Economics

Faculty of Transportation Engineering and Vehicle Engineering
Department of Automotive Technology

Methods of segmenting and analyzing of road accident data

A dissertation submitted by:

Ma’en Qaseem Ghadi

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Supervisor: Dr. Árpád Torok

Budapest, 2019-2020

1. Motivation

Indicating road safety-related aspects in the phase of infrastructure planning and operating is always a challenging task. The identification of hazardous road sections or black spots (BS) receives a great interest from road agencies and safety specialists. Black spot identification (BSID) can be defined as a process of searching for locations in transportation systems with a higher number of accidents than other similar locations that is mainly caused by local risk factors [1], [2]. Errors in BSID can result in many false positive and negative cases.

Considering that traffic accident data is heterogeneous, in general, data segmentation is considered the first and most important step of the BSID process. In other words, the success of any method applied in identifying BS on roads should depend fundamentally on how the data is organized into specific homogeneous segments. In road safety, road network segmentation is the process of organizing road infrastructure network data into homogenous entities. Data segmentation can also be defined as 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 data between groups are dissimilar.

Most often, road segmentation methods are based on researchers' experiences, methodological decisions or objectives. Thomas [3] argued that applying different lengths for segmenting road network can result in different definitions of the hazardous locations which, in turn, affects the reliability of the results. Koorey [4] discussed the benefit of applying variable segment length and their effect on BS determination. It has been confirmed as well that the more accurate the network segmentation is, the better performance the accident prediction model has, and this consequently affects the performance of the BSID methods that rely on accident prediction models in their criterion (i.e. EB method). On the other hand, the segmentation process plays also an important role in the mechanism of locating BS segments and identifying their proper lengths.

The main concept of the dissertation is based on the development of some novel statistical tools for clustering and segmenting large accident datasets, based on their spatial aggregations, into small and homogeneous groups. The idea is based on the assumption that spatially aggregated accidents are expected to be more homogeneous than distant accidents. Since, they are more likely to occur under similar environment conditions [5], [6].

The main purpose of the developed segmentation method has been to improve the accuracy of road network segmentation process. Thus, accident prediction and BSID reliability can also be improved. On the other hand, the dissertation has investigated the change in accident risk between different road segments, with different characteristics and road categories by developing a novel multilevel evaluation model.

2. Hypotheses

2.1 Hypothesis 1

In Thesis-1 two different BSID methods based on different segmentation models have been examined on two different road categories (i.e. high-speed motorway and low-speed local road).

The variation in performance of the applied methods for different applied road segment characteristics has provided the basis for the first sub-hypothesis that "The success of any method applied in identifying high-risk locations should depend fundamentally on how data being organized into specific homogeneous segments". And, the variation in the performance of the applied methods for different road categories and different input variables has provided the basis for the second sub-hypothesis.

2.2 Hypothesis 2

Thesis-2 has concentrated on developing a new road network accident data segmentation method aiming to group a large accident dataset into a number of small homogeneous groups so subsequent operations can easily be performed with more reliability.

The hypothesis behind the developed method is that accidents which occur within close distances, spatially and temporally, are more likely to have similar characteristics. So the developed method benefit from combining clustering and linear referencing techniques to identify homogeneous groups of spatially aggregated accidents.

2.3 Hypothesis 3

The third thesis investigates and compares the effect of methodological diversity of road network segmentation on the performance of different BSID methods.

The main hypothesis is based on the assumption that the performance of each BSID method varies in case of different segmentation methods.

2.4 Hypothesis 4

The combined uses of the developed segmentation method and the EB method have showed the best performance.

The hypothesis suggests that the practical application of both methods together could improve the process of BSID.

2.5 Hypothesis 5

Thesis-5 combines the benefit of identifying the spatial location of similar accidents within flexible length road segments with decision analysis tools.

According to, the hypothesis road segments locations with similar risk properties, depending on the conditions of their accident contents, can be exactly identified and discovered.

2.6 Hypothesis 6

When dealing with road safety issues, every single road with different characteristics needs to be investigated separately. Thesis 6 attempts to construct a general multilevel model to predict the accident frequency for each road segment located in any road category, at the micro and macro level.

This hypothesis assumes that the number of accidents is influenced by roadway categories and also the characteristics of their small segments. Every individual road segment has distinguished geometrical and traffic features. In contrast, every group of road segments within a single roadway can share similar characteristics that may differ from other road categories.

3. Research methodology

3.1 Research methodology related to hypothesis 1

During the research, the consistency of the SMW and SPA methods in identifying BS on two different road categories applying three different segmentation criteria has been examined. The proposed consistency test is used to measure the consistency of identifying road segments with high crash-risk in case of different applied segmentation characteristics (i.e. different segment length and sliding distance). The higher the proportion of shared BS sites is, the more consistent the applied BSID method is. the consistency is measured by the proportion of the shared sections related to the outputs of (i.e. n1, n2, .., n) the same BSID method with different segmentation models.

$$Consistency = \{n_1, n_2, .., n\}_{(BSID1)} \cap \{n_1, n_2, .., n\}_{(BSID2)} .. \cap \{n_1, n_2, .., n\}_{(BSIDn)} \quad (1)$$

Where *n* is the index of BS segments identified by the applied BSID method.

3.2 Research methodology related to hypothesis 2

As a new approach, K-means clustering has been used to find spatially aggregated accidents. Considering that the accidents are represented by the longitude and latitude variables on the road network, k-means algorithm could idenitfy inadequate spatial groups considering each coordinate as separately. Therfore a linear referencing model has been applied (Figure 3), instead of a bi-variable representation of accidents' spatial location in the clustering algorithm, which also contribute to reducing the demanded calculation capacity of the problem.

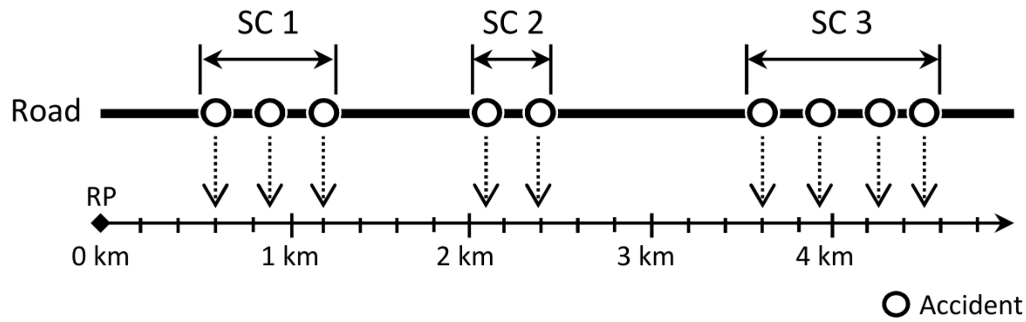

Figure 1: A visual illustration of the developed spatial clustering methodology

Safety performance function (Equation 2) has been used to model and evaluate the performance of the developed segmentation method compared with four other popular segmentation methods.

$$SPF = \exp[\alpha + (\sum_n \beta_n \times \ln(X_n)) + \ln(length)] \quad (2)$$

Where α is the intercept of the ordinate axis, and β_n is a regression coefficient of the corresponding explanatory variable X_n (i.e. β_1= AADT, β_2= speed, and β_3= HDA).

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) [7], [8] and the Pearson Correlation Coefficient (PCC) [9], [10].

3.3 Research methodology related to hypothesis 3

In the case of hypothesis 3, the methodology for comparing the joint performance between BSID methods and segmentation method is based on applying four consistency tests: Site consistency test (SCT), Method consistency test (MCT), Total rank differences test (TRDT), Total score test (TST). The total score test (TST) aims to represent the aggregated result of the three previous tests in a single value (Equation 3).

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

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 a balanced comparison among combined segmentation and BSID models.

3.4 Research methodology related to hypothesis 4

To reveal and rank BS, the methodology of standard Excess EB method with the suggested spatial segmentation model (as described above) has been applied..

3.5 Research methodology related to hypothesis 5

To analyze and classify road accident data considering both their spatial and environment attributes, a series of supervised and unsupervised data mining techniques have been applied.

To achieve the desired goals, accident data has been analyzed with the help of 3 major methods: spatial clustering, attribute-based classification (i.e. accident type, cause, etc.), and decision rule definition for the resulted clusters (Figure 4). In the first step, the proposed K-means clustering with the linear referencing technique has been applied to classify accidents based on their spatial dependence into several clusters. Then, databases of the resulted clusters have been used as input variables for the second classification. Finally, decision rules have been extracted from the resulted clusters.

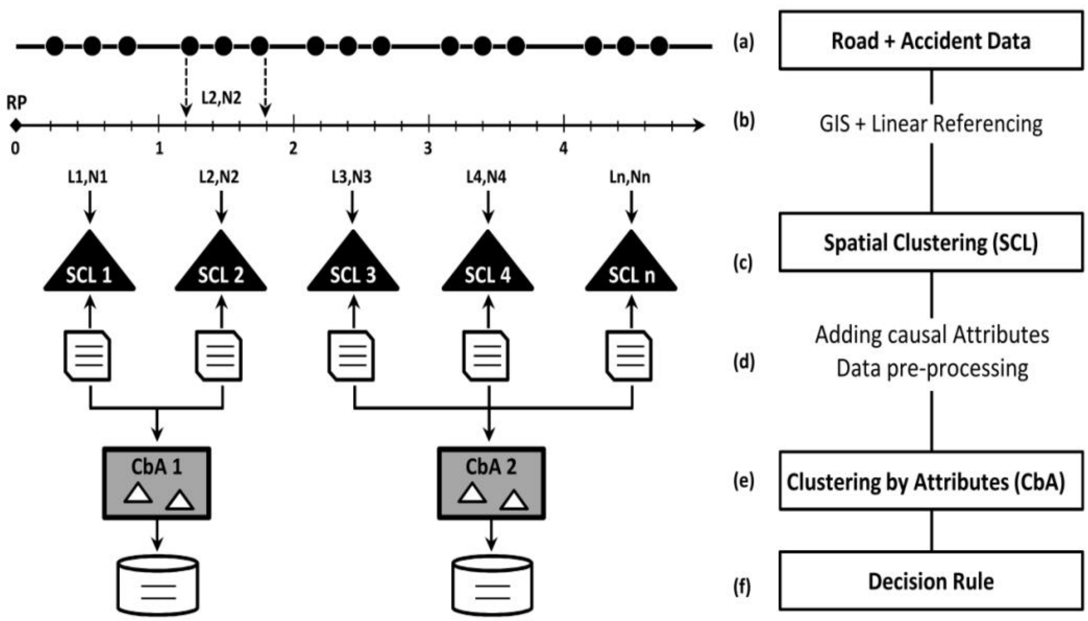

Figure 2: Proposed framework for the analyzing

3.6 Research methodology related to hypothesis 6

To analyze the accident frequency and reveal the correlation between explanatory variables at both micro- and macro-levels, the hidden effects of the hierarchical relationship must be investigated. Accident frequency is analyzed at two hierarchical information levels (Figure 5), including road level characteristics (i.e. category, level of service, average speed-limit) and small road segment level characteristics (i.e. AADT, speed limit, roadside hazard, curvature).

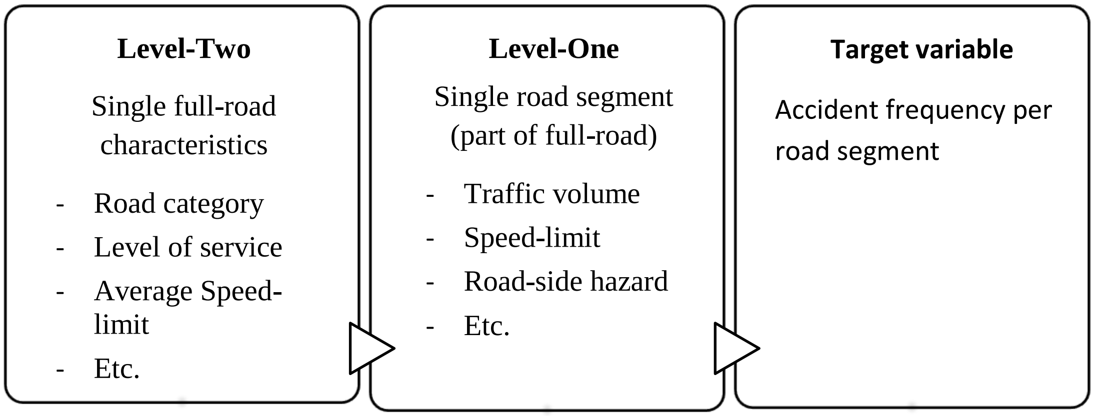

Figure 3: Multilevel hierarchical structure of road accident data

To do so, a two-level negative binomial multilevel approach has been used to examine the effect of the nested relationship between five different road categories and road segments, with distinguished characteristics, on accident frequency. The multilevel model treats individual road segments as parts of groups. Each group is corresponding to a specific road with distinguished characteristics (i.e. road category). Because individual segments of the same road are likely to share similar characteristics, they are more likely to respond in the same way compared with individuals of other roads, which in turn violating the assumptions of a single model. The general multilevel model is formed by combining level-1 and level-2 models in a single formula (Equation 4) [11].

$$\eta_{ij} = \gamma_{00} + \left(\sum_{i} \gamma_{i0} X_i \right) + \left(\sum_{j} \gamma_{0j} X_j \right) + u_{0j} \qquad (4)$$

Where: η_{ij} is equal to the natural logarithm of the expected accident count. γ_{00} is the fixed level-two intercept. X_i and X_j are fixed-effect predictors of accident frequency at level-one and level-two, respectively. γ_{i0} is a coefficient of X_i and γ_{0j} is a coefficient of X_j. and u_{0j} is the level-2 variance.

4. New scientific results

4.1 New scientific results of Thesis 1

- A new consistency test method is proposed in order to evaluate the consistency of the analyzed BSID methods in case of different segmentation characteristics.
- The use of different segment lengths (and sliding distance) seems to have a significant impact on the performance of the SMW method. This can be seen from the difference in the identified BS along the road in case of the different applied segmentation criteria, as

presented by the consistency test results. Generally, the SMW has shown a comparatively higher consistency in the case of the motorway compared with the urban roads.

- The application of the SPA method has also demonstrated a variation in performance with different segmentation criteria and different road categories. In the low-speed urban roads, accidents showed a higher spatial autocorrelation in the case of small segment lengths in comparison with motorway accidents. This can be explained by the nature of the accident distributions along the roadway.

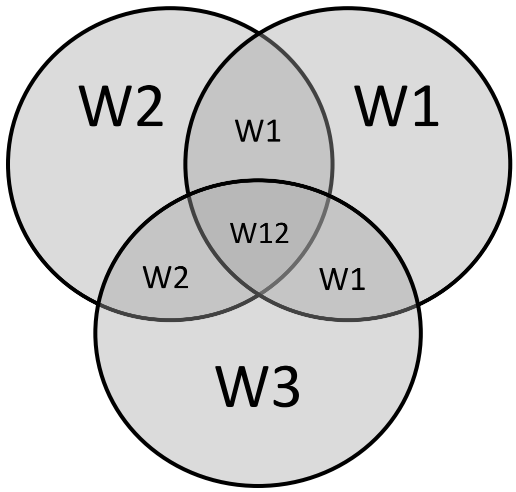

Figure 4: Illustration of the consistency measurement for three different segments criteria (W1, W2, W3)

Related publications Thesis 1: [12]

4.2 New scientific results of Thesis 2

My research work targeting the development of a novel accident dataset based network segmentation method based on the combination of K-mean clustering and linear referencing techniques has achieved scientifically new results in identifying homogeneous accidents clusters based on their spatial distribution (Figure 3). The developed method assumes that spatially and temporally coherent are more likely to occur within the same circumstances. In accordance with this, the developed method can divide road networks into flexible length road segments depending on the spatial distribution of accident clusters. Due to the favorable characteristics of the method, which allows us to adapt the segment lengths to the spatial distribution of the accidents, the identification and exclusion of empty road segments can efficiently support the reduction of computational demand of the BSID process.

A complex statistical technique has been used to evaluate the performance of the developed method compared to four other well-known segmentation models, including constant AADT segments, constant length segments, related curvature characteristics and a multivariable method suggested by the HSM (Table 2 and Figure 8). In the conclusion, the developed clustering based segmentation models gave the best goodness-of-fit statistics.

Tables 1: The resulted parameters and goodness-of-fit values of the five road segmentation methods.

Parameters	Methods of segmentation				
	Seg-1 (K-means clustering)	Seg-2 (HSM)	Seg-3 (Constant AADT)	Seg-4 (Constant length)	Seg-5 (Curvature)
α (Intercept) [p-value]	- 5.934 [0.062]	- 5.664 [0.013]	- 7.664 [<0.001]	- 14.269 [<0.001]	- 6.785 [<0.001]
β1 (AADT) [p-value]	0.784 [<0.001]	1.107 [<0.001]	0.961 [<0.001]	1.872 [<0.001]	1.009 [<0.001]
β2 (Speed) [p-value]	- 0.531 [0.080]	- 1.264 [0.018]	- 0.320 [0.10]	- 0.846 [0.019]	- 0.638 [0.003]
β3 (HDA) [p-value]	1.186 [0.007]	2.147 [0.034]	3.414 [0.063]	2.582 [0.057]	1.084 [0.067]
k	1.070	1.909	1.017	1.1	1.024
QIC	22	167	64	140	284
PCC [R-square]	0.794 [0.63]	0.300 [0.08]	0.760 [0.57]	0.474 [0.22]	0.684 [0.47]

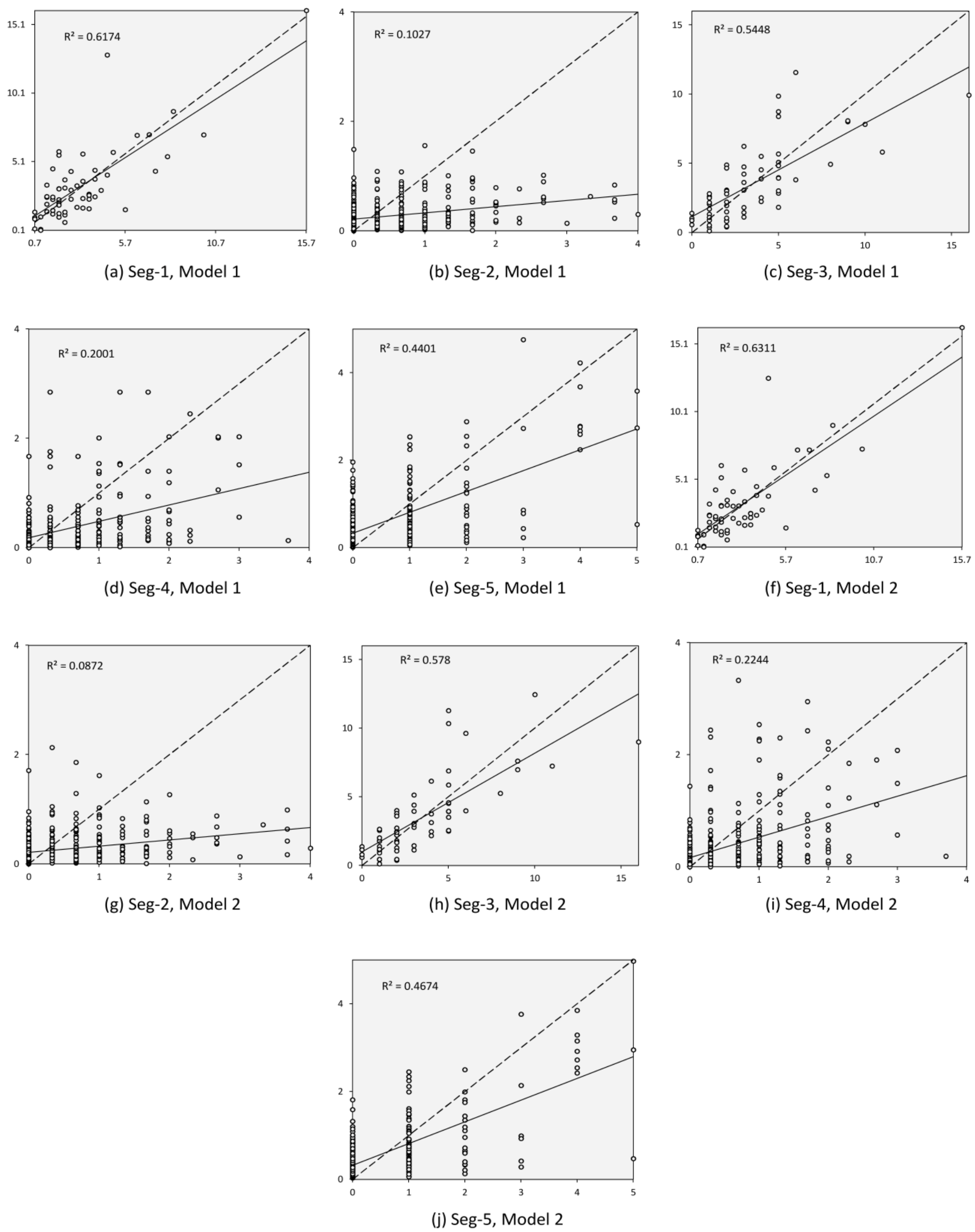

Figure 5: Graph the linear correlation between the observed accidents (x-axis) and the predicted accidents (y-axis) of the year 2016

Related publications Thesis 2: [13]–[15]

4.3 New scientific results of Thesis 3

As a new scientific result, I have revealed the effect of methodological diversity of road network segmentation methods on the performance of different black spot identification methods, and have performed a comprehensive comparison regarding their joint performance (Table 3).

Table 2: Total score test results for top 5% and 10% BS

BS method	**Segmentation method**				**Total**
	Seg-1 (K-means clustering)	Seg-2 (Constant length)	Seg-3 (Constant AADT)	Seg-4 (HSM)	
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	

The main achieved results are summarized below.

- The developed methodology makes it possible to evaluate
  - the joint performance of the investigated BSID and segmentation methods.
  - the performance of the different segmentation and BSID methods, independently in an integrated model environment.
- Demonstrated how the developed spatial clustering segmentation method can improve the performance of the different BSID methods.
- Practically, the applied methodology can significantly help to choose the best integrated BSID methodology.

Related publications Thesis 3: [16], [17]

4.4 New scientific results of Thesis 4

During the research a complex practical application has been implemented to examine the practical utility of the developed spatial clustering segmentation method. Based on the achieved outcomes it has become clear that the new approach can efficiently complement the EB model and the combined model framework can provide an outstandingly well performing BSID method. Beside this, the practical application has presented an addition key advantage of the developed integrated technique, which is the ability of grouping spatially coherent accidents in small segments, and so separating and excluding if required the empty and less risky road sections. After this it becomes possible to identify and rank all road segments which are affected by traffic accidents.

Related publications Thesis 4: [18]

4.5 New scientific results of Thesis 5

The new scientific achievements of the research provide an integrated methodology that can assist road safety analysis in identifying the patterns of accidents and BS locations along the road. The comprehensive methodology, presented in Figure 4, includes the application of the proposed segmentation method with other data mining techniques (i.e. decision tree).

Related publications Thesis 5: [19], [20]

4.6 New scientific results of Thesis 6

As a scientific novelty a multilevel negative binomial regression model has been built up for predicting accident frequency at the micro- and macro-level for different road categories.

$$\eta_{ij} = \gamma_{00} + \left(\sum_i \gamma_{i0} X_i\right) + \left(\sum_j \gamma_{0j} X_j\right) + u_{0j} + \log\left(L_{ij}\right) \qquad (5)$$

Where: β_{0j} is level-one intercepts. X_i and X_j are fixed-effect predictors of accident frequency at level-one (individual-level) and level-two (group-level), respectively. γ_{i0} is the coefficient of X_i and γ_{0j} is the coefficient of X_j. u_{0j} is level-two random effect variance.

The study has been based on the assumption that the number of accidents may vary according to the roadway category and the characteristics of their smaller segments. Every individual road segment has different geometrical and traffic features. In contrast, every group of road segments within a single roadway can share similar characteristics that may differ from other road categories. The resulted level-two variance components of the intercept verify that hierarchical structure resides within the data (Table 4), where correlation exists between accident frequencies of accident occurred on the same road, which supports multilevel modelling compared with single-level modelling (Table 5). Accordingly, the reached outcomes strongly support the assumption, that the road segment related accident frequency values are more likely to vary across different roadways.

Table 3: Variance values of the random effect components

Variance	Estimate	Std. Error	Z	Sig.
Intercept	0.084	0.033	2.569	0.010
Road category	0.071	0.028	2.555	0.011

Table 4: Comparing the goodness-of-fit of a single level and multilevel models

One-level Model	Akaike Corrected (AIC)	6682
	Bayesian (BIC)	6746
Multilevel Model	Akaike Corrected (AIC)	4377
	Bayesian (BIC)	4388

The developed multilevel model can contribute to the methodology of the road safety analysis with regard to the following aspects.

- Understanding the effect of risk change of accident existed on different road categories, also considering the effect of the environment characteristics of the different road segments.
- Predicting accident frequency with a higher accuracy taking also into account the nested relationships between different micro- and macro-level factors.

Table 6 summarizes the estimated fixed effect results of applying the multilevel model, considering that the length of the road segment as an offset.

Table 5: Resulted fixed effect coefficient values

Model Term	Odd ratio (Coefficient)	Significance
Intercept	1.47	0.00
Level-one road segment variables		
AADT	2.70	0.00
Speed-limit	0.69	0.00
Truck volumes	0.68	0.00
Road alignment= Consecutive-curves	1.49	0.00
Road alignment= One-curve	1.33	0.00
Road alignment= Straight	*	*
Resident= yes	1.90	0.00
Resident= No	*	*
Level-two full road variables		
Road Category= Local road	1.78	0.00
Road Category= Secondary main	1.25	0.06
Road Category= Arterial	1.93	0.00
Road Category= Expressway	0.20	0.00
Road Category= Motorway	*	*
Over-dispersion	0.53	

* Reference variable

Related publications Thesis 6: [21]

5. Application of the scientific results

5.1 Application of the scientific results related to Thesis 1

The main idea of Thesis-1 can be concluded as follows: “The success of any method applied in identifying BS locations of the road network should depend fundamentally on how the data is organized into segments”. In other words, the thesis has recognized and underpinned that the initial stage of BSID, represented by road network segmentation, should gain more attention in order to reduce the deficiencies of BSID. Furthermore, the thesis proposed a new consistency test to measure and compare the performance of different BSID method by measuring the consistency in identifying the same BS with each different segmentation characteristics. This achievement has underpinned the recognition that the following tasks should be completed during research:

a. Developing a comprehensive evaluation methodology applicable to investigate the interaction of segmentation and BSID methods also considering the effect of the integrated model environment.
b. Identifying a general concept combining both segmentation and BSID models would be necessary to improve the performance of the overall methodology.

This finding would be the basis for any related work, regarding the application and development of BSID methods.

5.2 Application of the scientific results related to Thesis 2

The developed road network segmentation method has a number of applications in the field of road traffic safety. In road safety, it can be used to improve the reliability of SPF which consequently affects the performance of the relevant BSID methods that rely on accident prediction models in their criterion. Thesis-2 provides also a methodology to assess and compare data segmentation methods. In the field of BS analysis, the developed segmentation method can contribute to the identification of proper segment lengths, which can help in fitting BS area to the characteristic of the accident distribution.

5.3 Application of the scientific results related to Thesis 3

The third thesis proposes a new methodology that can assist experts in selecting the proper BSID and segmentation method combination with the best joint performance. The thesis also call the attention of the experts that the application of some BSID (e.g. accident rate) and/or segmentation methods (e.g. constant length) can result in misleading conclusions.

5.4 Application of the scientific results related to Thesis 4

The fourth thesis introduces a complex practical methodology to investigate road safety more effectively. This methodology suggests the application of the developed segmentation method along with the EB method. Hence, it can be considered as a good integrated methodology in the area of BSID.

5.5 Application of the scientific results related to Thesis 5

The results of the data mining process can help road safety experts to make a comprehensive assessment of road accident patterns along a road by identifying road segments with similar risk features, so proper and comprehensive safety measures can be applied.

5.6 Application of the scientific results related to Thesis 6

The proposed multilevel analysis of Thesis-6 provides a general model that is able to identify accident frequency at any road segment, with different characteristics, and for any road category. Furthermore, the new model contributes to the improvement of accident prediction accuracy.

Beside this, the developed model can open up new horizons in road safety research, especially in connection with BSID.

6. Scope for future work

(1) The developed spatial segmentation method can be further improved in term of the applied clustering method and the criteria applied for identifying the number of clusters. On the other hand, the criteria applied in the comparison can include other statistical techniques with more predictor variables related to road segment characteristics.

(2) The other newly developed segmentation and BSID techniques of the future should also be investigated by the prepared methodology.

(3) It is reasonable to develop a software package related to the methodology proposed by Thesis-5 that can achieve the identified objective of accident pattern detection in a more simple way. On the other hand, a variety of variables can also be examined in the decision process.

(4) Related to Thesis-6, future work could be able to propose a more complex hierarchical model with an additional number of levels that could include more comprehensive variables to describe road accidents at the micro- or macro-level more accurately.

7. Summary list of Theses

Thesis (1) I have introduced new evaluation criteria of two different BSID methodologies applied with different segmentation characteristics and road categories. Based on the result of the investigation, it can be concluded that different road categories and different BSID methods can be characterized by significantly different consistency.

Thesis (2) I have developed a new road network segmentation method based on the spatial distribution of road accidents by identifying homogeneous accident clusters in terms of their spatial and temporal properties.

Thesis (3) I have revealed the effect of methodological diversity of road network segmentation on the performance of different black spot identification methods by adapting a methodology that is able to compare the performance of road segmentation methods, black spot identification methods, and the joint performance of both methods. I have demonstrated that the developed spatial clustering segmentation method can improve the performance of different black spot identification methods. The best result was yielded using the spatial clustering segmentation technique in combination with the Empirical Bayes method.

Thesis (4) I have revealed the practical utility of applying the developed spatial clustering segmentation method as a well applicable complementary approach to the EB method in providing the best-integrated methodology for BSID.

Thesis (5) I have developed a complex methodology to identify accident risk patterns along the road network from a spatial and causal perspective, including the application of supervised and unsupervised data mining techniques.

Thesis (6) I have developed a new multilevel negative binomial regression model to predict accident frequency at different road segments for different road categories. The multilevel analysis has modeled the relationship between the different groups of accidents on different roads at micro- and macro-level by identifying a hierarchical data structure that can utilize the advantages of the clustered dataset. In other words, the model can define the accident rate of the investigated road segments considering their environment features and road categories.

References

[1] W. Cheng and S. P. Washington, "Experimental evaluation of hotspot identification methods," *Accid. Anal. Prev.*, vol. 37, no. 5, pp. 870–881, 2005.

[2] R. Elvik, "State-of-the-art approaches to road accident black spot management and safety analysis of road network," 2007.

[3] I. Thomas, "SPATIAL DATA AGGREGATION: EXPLORATORY ANALYSIS OF ROAD ACCIDENTS," *Accid. Anal. Prev.*, vol. 28, no. 2, pp. 251–264, 1996.

[4] G. Koorey, "Road Data Aggregation and Sectioning Considerations for Crash Analysis," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2103, no. 1, pp. 61–68, 2009.

[5] B. Flahaut, M. Mouchart, E. San Martin, and I. Thomas, "The local spatial autocorrelation and the kernel method for identifying black zones. A comparative approach.," *Accid. Anal. Prev.*, vol. 35, no. 6, pp. 991–1004, 2003.

[6] B. Flahaut, "Impact of infrastructure and local environment on road unsafety: Logistic modeling with spatial autocorrelation," *Accid. Anal. Prev.*, vol. 36, no. 6, pp. 1055–1066, 2004.

[7] W. Pan, "Akaike's information criterion in generalized estimating equations," *Biometrics*, vol. 57, no. 1, pp. 120–125, 2001.

[8] J. Cui, "QIC program and model selection in GEE analyses," *Stata J.*, vol. 7, no. 2, pp. 209–220, 2007.

[9] G. K. Smyth, "Pearson's Goodness of Fit Statistic as a Score Test Statistic," *Sci. Stat. A Festschrift Terry Speed*, vol. 40, no. March, pp. 1–12, 2003.

[10] J. Lee Rodgers and W. Alan Nice Wander, "Thirteen ways to look at the correlation coefficient," *Am. Stat.*, vol. 42, no. 1, pp. 59–66, 1988.

[11] R. H. Heck, *Multilevel Modeling of Categorical Outcomes Using IBM SPSS*. 2014.

[12] M. Ghadi and Á. Török, "Comparison Different Black Spot Identification Methods," *Transp. Res. Procedia*, vol. 27, pp. 1105–1112, 2017.

[13] M. Q. Ghadi and Á. Török, "Comparison of Different Road Segmentation Methods," *PROMET - Traffic&Transportation*, vol. 31, no. 2, pp. 163–172, Apr. 2019.

[14] K. Tánczos, Á. Török, Z. Szabó, G. Pauer, and M. Ghadi, "Cluster analysis," in *Decision making methods in transportation*, First edit., Akadémiai Kiadó, 2018.

[15] M. Ghadi and Á. Török, "Development of Safety Performance Function for Highway by Cluster Analysis Segmentation Approach," in *Conference on Transport Sciences*, 2018, pp. 265–270.

[16] M. Ghadi and Á. Török, "A comparative analysis of spatial clustering segmentation method of road accidents," in *8th International Scientific Conference CMDTUR*, 2018, pp. 95–100.

[17] M. Ghadi and Á. Török, "A comparative analysis of black spot identification methods and road accident segmentation methods," *Accid. Anal. Prev.*, vol. 128, pp. 1–7, Jul. 2019.

[18] M. Ghadi, Á. Török, and K. Tánczos, "Integration of Probability and Clustering Based Approaches in the Field of Black Spot Identification," *Period. Polytech. Civ. Eng.*, Oct. 2018.

[19] M. Ghadi and Á. Török, "Analysis of Traffic Accident Black Spots: an Application of Spatial Clustering Segmentation Method," in *East-West Cohesion III. International Scientific Conference*, 2018.

[20] M. Ghadi and Á. Torok, "Evaluation of the impact of spatial and environmental accident factors on severity patterns of road segments," *Period. Polytech. Transp. Eng. Accept.*

[21] M. Ghadi, "Multilevel Analysis of Road Accident Frequency: The Impact of the Road Category," in *TRB 2020 - 99th Annual Meeting. Accepted*.