

Accident Prediction Models for Hungarian Two-Lane Rural First-Class Main Roads

A. Borsos

Széchenyi István University, Department of Transport Infrastructure
Egyetem tér 1., H-9026 Győr, Hungary
E-mail: borsosa@sze.hu

Abstract: This paper describes a modeling effort to define accident prediction models for first-class main roads outside built-up areas in Hungary using variables that are available and believed to exert an influence on safety performance. The first part of the paper describes data collection and the segmentation technique. Six years of accident data are gathered for segments only; intersections with a 200m radius are taken out in order to avoid having intersection related crashes in the data. Altogether 1357 homogenous sections are formulated based on AADT (Annual Average Daily Traffic), road width, posted speed, horizontal curve and shoulder width. Models are proposed using the Generalized Linear Modeling (GLM) approach assuming a negative binomial error structure. It is concluded that AADT, roadway width, horizontal curve and segment length significantly influence accident frequency. Estimated model parameters are explained by putting them into international context and it is concluded that the results are in accordance with previous research findings.

Keywords: road safety, safety performance, accident prediction, two-lane rural road

1. Introduction

According to the Hungarian regulations road infrastructure safety management applies not only to the trans-European road network as laid down in the Directive 2008/96/EC of the European Parliament and of the Council [1], but all national main roads from 1 January 2014 on as well as to all the roads with more than 10,000 PCU/day from 1 January 2015 on. In Hungary crash rates are being used for such analyses assuming a linear relationship between exposure and crash frequencies, which is easy to use for practitioners, however, do not sit on scientifically proven grounds. Safety Performance Functions (SPF) as accident prediction models provide professionals with data needed to conduct road safety impact assessment and network safety ranking and accounting for a more realistic relationship between traffic volume and accident occurrence.

To this end, this paper describes a modeling effort to define accident prediction models using variables that are available and believed to exert an influence on safety performance, including traffic volume. The first part of the paper gives a general

overview on design characteristics and accident statistics of two-lane rural first-class main roads. Then it describes data collection as well as how the network is split up into homogenous sections. Then models are proposed using the Generalized Linear Modeling approach (GLM) assuming a negative binomial error structure. Five variables are used to predict accident frequencies, AADT, posted speed, curve, roadway width and shoulder width. Traffic volume is used as an exposure to accidents and the length of the homogenous sections as an offset variable in all models.

2. Literature review

Safety performance functions have been widely used in the traffic safety field for analyzing how the safety performance of road facilities is related to various road characteristics. Roughly 25 years ago a few researchers (e.g. [2] and [3]) started to demonstrate problems associated with conventional regression techniques used for accident prediction. Miaou and Lum [3] for instance used four regression models and demonstrated that the conventional linear regression models are not appropriate to make probabilistic statements about vehicle accidents, and that if the vehicle accident data are found to be significantly overdispersed relative to its mean, then using the Poisson regression models may overstate or understate the likelihood of vehicle accidents on the road. To overcome these problems later on the negative binomial modeling method became a widely accepted modeling technique and was recommended by many researchers e.g. [4] and was also used in other road categories such as urban arterials in [5].

There have been several researchers dealing with the safety performance of two-lane rural roads [6], [7], [8], [9], [10], [11]. Cafiso et al. [6] fitted 19 models and considered 14 variables including geometric and operational ones (curvature change rate, paved width, tangent ratio, curve ratio, average operating speed), design consistency variables (e.g. variables considering speed differentials) and context-related variables (roadside hazard, driveway density). Three models were selected as recommended ones, where the first was a basic model with AADT only and the other two having the best fits used at least one variable pertaining to the four main groups of variables listed above. Abdel-Aty and Radwan [4] divided a 227 km long road into 566 segments with homogeneous characteristics in terms of traffic flow and geometry (degree of horizontal curvature, shoulder and median widths, rural/urban classification, lane width and number of lanes), all variables found to be strongly related to the accident occurrence. Mayora and Rubio [8] investigated which are the relevant variables for crash rate prediction and concluded that access density, average sight distance, average speed limit and the proportion of no-passing zones have the highest correlation. Access density is the variable that influences most the rate of head-on and lateral collisions, while in run-off the road and single vehicle crashes sight distance is decisive.

As far as the segmentation technique is concerned, there are a few methodologies and there is no preferred one. Cafiso et al. [12] did a comprehensive study on the application of five different methods and the influence of segmentation on the performance of safety performance functions in terms of goodness of fit and the variables that could be modeled. They concluded that the best results were obtained for the segmentation based on having two curves and two tangents in each segment and the segmentation with fixed

length. They also highlighted that a segmentation technique using constant values of all variables, therefore resulting in very short segments, led to the poorest model.

The Highway Safety Manual [13] recommends the use of homogeneous segments with respect to a number of parameters, such as the AADT, number of lanes, lane width, shoulder width, shoulder type, curvature, driveway density, roadside hazard rating, median width and clear zone width. According to the manual there is no prescribed minimum segment length, however a minimum of 0.10 miles (0.16 km) is suggested. Fitzpatrick et al. [14] pointed out that in practice, this type of segmentation is not always easy to achieve as not all the variables are available. Koorey [15] arrived at similar conclusions, namely that variable-length road segments seem intuitively more useful than fixed-length segments, because of the mixed attributes contained in the latter. He added that this advantage is less when shorter lengths are used, and fixed-length segments are computationally easier to create from constant-interval raw data.

3. General overview of first-class main roads

3.1. Design characteristics

The Hungarian road network can be divided into two major categories: the national (31,833 km) and the municipal road network (approximately 170,000 km). The national road network can be further divided into two major subcategories: main roads consisting of motorways, express roads, first-class and second-class roads; and minor (or secondary) roads. In this paper our safety performance analysis has been done for the rural sections of the first-class main roads, therefore the following criteria and figures apply to this road category.

The length of the first-class rural main roads in Hungary is 1659 km, which is 29% of the total rural main road network and 8.6% of the total rural road network. Due to the continuous expansion of the limited access road network, almost 80% of the first-class main roads have a parallel motorway or expressway, so the traffic volume in this road category is moderate (the AADT on these roads is roughly 7500 vehicles/day on average).

A few selected criteria according to design standards are listed below:

- design speed is 90 km/h (in some cases 80 km/h depending on the terrain and land development circumstances),
- minimum curve radius must be at least 250 m (80 km/h) or 340 m (90 km/h),
- maximum gradient is 6%,
- minimum cross-slope is 2.5% and maximum superelevation is 7%,
- roadway width is 7.5 (7.0) m, lane width is 3.50 (3.25) m, shoulder width is 2.50 (2.0) m.

(Figures in parentheses are permitted under special circumstances.)

The general posted speed limit is 90 km/h, 60-70 km/h adjacent to junctions and in dangerous curves. There are a few sections with higher speed limits of 100 and 110

km/h. These sections are exceptions where either the alignment of an already existing road was eligible for a higher posted speed limit or newly built sections where a more generous design was followed.

3.2. Accident statistics

In Hungary according to recent statistics approximately two-thirds of the accidents involving fatalities and personal injuries occur inside built-up areas, whereas 60% of fatal accidents occur outside built-up areas, 20% of which happen on first-class main roads. There has been a major decrease in the number of accidents and victims in the recent years, and in the road category investigated this decrease is above the average.

The number of fatalities was 1232 on the entire road network in 2007 and plummeted to 641 by 2012, which is more than a 47% decrease. In case of the rural first-class main roads these results are slightly better, in the same time period the number of fatalities dropped by almost 56%. As far as the accident types are concerned, run-off road accidents represent 27%, rear-end collisions 20%, head-on collisions 19% and side impact collisions 13% of the total number of accidents.

4. Data collection and segmentation

As a first step of the data survey, the following data were gathered from the National Road Databank [16] for the entire length (1659 km):

- roadway width (pavement width)
- radius of horizontal curve
- shoulder width
- number of lanes
- traffic volume
- posted speed limit

This database is organized in such a way that if there is a change in any of the parameters there is a new record (a new segment begins). As a result, these segment lengths can be really short, in some cases just a few meters; therefore the segmentation technique itself was an important issue to deal with.

In our analysis the following segmentation technique was used (Figure 1):

- As this study focused on rural sections only, all segments inside built-up areas were taken out (marked 1 in Figure 1).
- Crashes within a 200 meter radius of intersections were taken out in order to avoid having intersection related crashes in the data (marked 3 in Figure 1). Not only major intersections, but all minor intersections and connecting streets with a paved surface were filtered out (marked 2 in Figure 1). Intersections with connecting agricultural roads were not excluded from the sample.

- Segments with more than two lanes (four-lane sections, sections with climbing lanes) were taken out.
- Although using a fixed segment length might be the most practical application, it disregards the opportunity that the segment length itself can be a predictor of accidents. Homogenous sections were formulated according to AADT, road width, shoulder width and horizontal curves and posted speed limit.
- As it is indicated elsewhere in the literature [15], [12] too short segments can be biased due to the inaccurate identification of accident locations, therefore segments shorter than 200 m were eliminated from the sample.

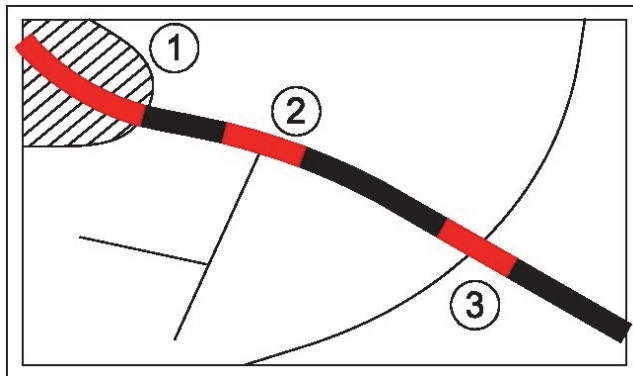


Figure 1. Segmentation technique

As a result of the segment delineation roughly 55% of the entire length (919 km of 1659 km) was used for further analysis. Although vertical alignment data were also collected, since the country has a relatively flat topography and because there were just a few sections with slopes this characteristic was omitted. Another variable that seemed to be an important predictor according to other researchers' work was access point density e.g. in [6]. In our case access points (driveways) were hardly present as our focus was on rural sections, where access points are very limited.

Six years of accident data (2007-2012) including fatal (fatality within 30 days as a result of the accident), serious (injury healing beyond 8 days) and light (recover within 8 days) injury accidents were gathered (altogether 2137 accidents). Property damage accidents were not taken into consideration.

5. Modeling

5.1. Regression technique and goodness of fit

Unlike Poisson models, negative binomial models assume that the conditional means are not equal to the conditional variances. Since the accident data were found to be significantly overdispersed relative to its mean, the Generalized Linear Modeling approach (GLM) with a negative binomial error structure was used. Modeling was done

in the R statistical software [17]. Several statistical measures were used to assess the goodness of fit of the models.

The Pearson χ^2 can be calculated by means of the following Equation (1):

$$\text{Pearson } \chi^2 = \sum_{i=1}^n \frac{[y_i - \hat{E}(y_i)]^2}{\text{Var}(y_i)} \quad (1)$$

The Pearson χ^2 of the model must be less than a critical value of χ^2 distribution value that is based on the model's degrees of freedom and a level of significance of α .

The Scaled Deviance (SD) is defined as the likelihood ratio test statistic measuring twice the difference between the log likelihoods of the studied model and the full or saturated model. The full model has as many parameters as there are observations so that the model fits the data perfectly. Therefore, the full model, which possesses the maximum log likelihood achievable under the given data, provides a baseline for assessing the goodness-of-fit of an intermediate model with p parameters [18].

The Akaike's Information Criterion (AIC) can be used for model selection and applicable to compare models. The AIC value is calculated according to Equation (2):

$$\text{AIC} = -2\log L + 2k \quad (2)$$

where $\log L$ is the maximum log-likelihood of the fitted model, k is the number of parameters in the model. The lower the value of AIC, the better is the model data fit. Given a dataset, several competing models may be ranked according to their AIC, with the one having the lowest AIC being the best.

Cumulative residual analysis was also used to evaluate the model form. The residual is equal to the difference between the observed and estimated values of the dependent variable. Cumulative residuals were plotted versus AADT for each homogeneous section. The closer the curve stays to the x-axis, the more appropriate the model form is, and a curve that stays within 2 standard deviations is considered to be satisfactory [19].

Simple models of accident frequencies were estimated using five predictor variables: AADT, posted speed, curve, roadway width, shoulder width. Traffic volume is used as an exposure to accidents and therefore present in all models. The prediction equation estimated is given in Equation (3):

$$E(Y) = e^{\alpha_0} \cdot l \cdot \text{AADT}^{\alpha_1} \cdot e^{\sum_{j=1}^m \beta_j x_j} \quad (3)$$

where $E(Y)$ is the expected accident count/year; l is the segment length (in kilometers), AADT is the Annual Average Daily Traffic (AADT) (vehicles/day); x_j is the any of m -additional variables; α_0 , α_1 , and β_j are the coefficients to be estimated. AADT is scaled to an annual value (multiplied by 365 and divided by 10^7) and included with an exponent, as it is known to have a nonlinear relationship with crash incidence [20]. Here the segment length is included without an exponent to normalize for the variability from one segment to another. However, later the segment length is also used as a predictor. Equation (3) is transformed logarithmically to (Equation 4):

$$\ln E(Y) = \alpha_0 + \ln(l) + \alpha_1 \ln \text{AADT} + \sum_{j=1}^m \beta_j x_j \quad (4)$$

5.2. Modeling results

As a result of the segmentation 1357 homogenous sections were formed. Table 1 and Table 2 show the variables used. AADT and length are used as continuous variables, whereas variables describing road geometry are clustered into categories for two reasons. Firstly, in many cases these characteristics change quite frequently: there are minor variations in the width of the cross-section elements within a short distance. From a road safety point of view, it can be doubted that for instance a few centimeter change in the width of the roadway would exert a significant influence on accident frequency. Secondly, had all these variations been taken into account it would have led to even shorter homogenous sections causing problems that already have been specified in the literature review section.

Table 1. Continuous variables and their descriptive statistics

Variable	Min	Max	Mean	Standard dev.
AADT	645	25438	6729	3165.53
Length (km)	0.200	6.245	0.677	0.619

Table 2. Categorical variables and their frequencies

Variable	Categories	Frequency
Horizontal curve	tangent	799
	< 750 m	84
	$750 \text{ m} \leq x < 1500 \text{ m}$	151
	$1500 \text{ m} \leq x < 3000 \text{ m}$	129
	$\geq 3000 \text{ m}$	194
Roadway width	< 7 m	64
	$7 \text{ m} \leq x \leq 7.5 \text{ m}$	923
	> 7.5m	370
Shoulder width	< 2 m	692
	$2 \text{ m} \leq x \leq 2.5 \text{ m}$	642
	> 2.5m	23
Posted speed limit	< 90 km/h	49
	90 km/h	1267
	> 90 km/h	41

Categories for horizontal curves were defined to have roughly even frequencies and as curves with small radii were underrepresented in the final sample, curve categories in the lower range were not further divided. For roadway width, shoulder width and posted speed limit category boundaries were defined based on the deviation from a reference level for design standards.

The modeling was conducted in two steps. First, each variable was entered into the model alone to see which ones significantly affect accident frequency. Then a full model was built using all the significant predictors. From the one-variable models it turned out that shoulder width and posted speed limit are not significant. The former one is believed to influence accident frequency so that increasing shoulder width has been found to reduce the number of accidents, mostly injury accidents [21]. Table 3 shows the model results for the base model (AADT only) and the two one-variable models with horizontal curve and roadway width. Not surprisingly, AADT is a highly significant predictor of accidents in all models. It has an exponent less than 1, which is in accordance with other research suggesting that the relationship between accident frequency and AADT is non-linear, i.e. shows a decreasing slope with increasing AADT. Model 1 has the highest AIC value, likely because other predictor variables are not used.

Roadway width is a significant predictor in both categories (Model 2), suggesting that narrower pavement width (compared to the design standards) increases the expected number of accidents, whereas wider roadways slightly decreases it. For horizontal curve the lower curve categories have significant coefficients and the expected signs show that accidents are more likely to occur in these curves, to a greater extent in sharp ones (with a radius of less than 750 m) than in curves between 750 and 1500 m. It also seems that higher curve categories (over 1500 m) are not significantly different from the tangent sections (Model 3), therefore it was decided to combine them and run a new model (Model 4).

These two significant variables (roadway width and horizontal curve) were also used in a full model (Model 5) as shown in Table 4.

It was also analyzed whether length as a predictor (in addition to the length as an offset variable) has an exponent significantly different from 1. To do so, the 799 tangent sections were used and length was added as a predictor to the base model. From Table 5 one can conclude that the tangent length itself is a significant predictor and it has a coefficient of 0.10006, which is statistically significant. This means that for each one-unit increase in length, the expected log count of accidents will increase by 0.10006. This parameter is positive, which is in line with our expectation, i.e. the longer a tangent segment is, accident frequency will slightly increase most likely due to speeding and overtaking maneuvers.

Table 3. Model parameters for the base model and one-variable models

No.	Model	Categories	Coefficient	Standard error	p value	<i>k</i>	AIC	SD (DF)
1.	AADT	constant	0.14816	0.08703	0.0887	0.217	3984.8	1458 (1355)
		lnAADT	0.76252	0.05929	<0.0001			
2.	AADT + Roadway width	constant	0.23719	0.09592	0.0134	0.211	3979.6	1456 (1353)
		lnAADT	0.81115	0.06156	<0.0001			
		< 7 m	0.28672	0.12839	0.0255			
		7 m ≤ x ≤ 7.5 m	reference					
		> 7.5m	-0.11543	0.06215	0.0633			
3.	AADT + Curve	constant	0.13870	0.08840	0.1167	0.213	3983.8	1454 (1351)
		lnAADT	0.76938	0.05914	<0.0001			
		tangent	reference					
		< 750 m	0.30441	0.12974	0.0190			
		750 m ≤ x < 1500 m	0.17915	0.09832	0.0684			
		1500 m ≤ x < 3000 m	-0.03309	0.09936	0.7391			
		≥ 3000 m	-0.04004	0.08465	0.6362			
4.	AADT + Curve	constant	0.12972	0.08694	0.1357	0.213	3980.1	1454 (1353)
		lnAADT	0.76898	0.05913	<0.0001			
		tangent	reference					
		< 750 m	0.31281	0.12885	0.0152			
		750 m ≤ x < 1500 m	0.18754	0.09714	0.0535			

Table 4. Model parameters for the full model

No.	Model	Categories	Coefficient	Standard error	p value	<i>k</i>	AIC	SD (DF)
5.	AADT + Roadway width + Curve	constant	0.21916	0.09572	0.0220	0.208	3975.4	1452 (1351)
		lnAADT	0.81636	0.06136	<0.0001			
		< 7 m	0.26784	0.12824	0.0367			
		7 m ≤ x ≤ 7.5 m	reference					
		> 7.5m	-0.11735	0.06192	0.0580			
		tangent	reference					
		< 750 m	0.30278	0.12896	0.0189			
		750 m ≤ x < 1500 m	0.18441	0.09699	0.0572			

Table 5. Model parameters for tangent sections

No.	Model	Categories	Coefficient	Standard error	p value	<i>k</i>	AIC	SD (DF)
6.	AADT + Length	constant	-0.02950	0.10417	0.777003	0.141	2509.8	1221 (1119)
		lnAADT	0.73713	0.06912	<0.0001			
		length	0.10006	0.03007	0.000875			

Cumulative residuals were plotted versus AADT. Figure 2 shows an example for the base model. Although there are some fluctuations in the cumulative residuals, the model can be considered satisfactory.

It was also tested, whether adding a variable is overall significant or not. To determine if curve and roadway width are overall statistically significant, we can compare a model with and without them. Table 6 shows that the two degree-of-freedom chi-square tests indicates that both curve and roadway width are statistically significant predictors of the number of accidents.

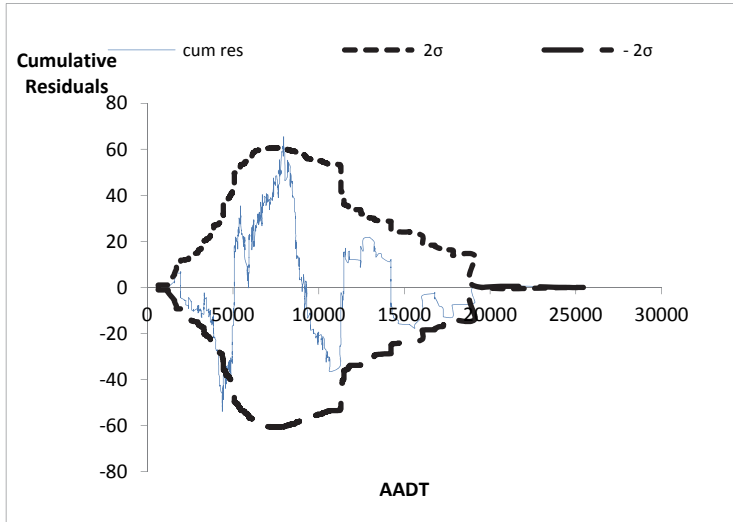


Figure 2. CURE plot of the base model.

Table 6. Likelihood ratio tests

Models	Resid. DF	2xlog-likelihood	DF	LR stat.	<i>p</i>
AADT	1355	-3978.837			
AADT + Curve	1351	-3969.845	4	8.992304	0.061
AADT	1355	-3978.837			
AADT + Roadway width	1353	-3969.637	2	9.20047	0.010

5.3. Numerical examples

To illustrate some of the models two numerical examples are given below. Consider the base model (Model 1) first. Here the coefficients given in Table 3 can be used according to the following Equation (5).

$$E(Y) = 1 \cdot \text{AADT}^{0.762} \cdot e^{0.148} \quad (5)$$

If on a road of this kind with a section length of 1 km the AADT is 10,000 vehicles per day, then the expected number of accidents per year will be 0.538 (Equation 6).

$$E(Y) = 1 \cdot \left(\frac{10,000 \cdot 365}{10^7}\right)^{0.762} \cdot e^{0.148} = 0.538 \text{ acc/year} \quad (6)$$

For the full model (Model 6) the coefficients given in Table 4 can be used. Consider a 1 km long tangent section with a roadway width of 7.6 m and an AADT of 10,000 vehicles per day. The expected number of accidents per year will be 0.486 (Equation 7).

$$E(Y) = 1 \cdot \left(\frac{10,000 \cdot 365}{10^7}\right)^{0.816} \cdot e^{0.219 - 0.117} = 0.486 \text{ acc/year} \quad (7)$$

6. Discussion and conclusions

This study focused on two-lane rural first-class main roads in Hungary with an objective to develop accident prediction models. The road network was segmented with a rigorous method and variables that were available and believed to potentially influence road safety were collected. Generalized Linear Modeling approach (GLM) assuming a negative binomial error structure was used. Expected injury accident frequency was predicted by exposure (AADT), length was used as an offset in all models. It was analyzed whether the variables roadway width, horizontal curve, shoulder width, posted speed limit significantly affect the accident frequency. It was also analyzed whether length as a predictor has an exponent significantly different from 1.

The base model only includes the AADT as an exposure and can be easily used by practitioners (auditors, road administration etc.) to conduct analyses such as network safety ranking. The coefficient of AADT (0.75~0.8) is in line with the international research results: Cafiso et al. [6] for instance found in their best performing models that for two-lane rural highways this value is 0.62 ~ 0.75. The full model with two more variables (horizontal curve and roadway width) is a more complex one, but can be still used in practice as the predictors used are relatively easy to collect. Also in the full model the variables showed reasonable coefficients and signs. Roadway widths narrower than what the design standards prescribe tend to increase ($e^{0.26784} = 1.31$), whereas wider roadways slightly decrease the number of accidents ($e^{-0.11735} = 0.89$). Curves below 1500 m increase accident frequency with a smaller value for radii between 750 m and 1500 m ($e^{0.18441} = 1.20$) and with a higher value in the lower range below 750 m ($e^{0.30278} = 1.35$). Our conclusion that accident frequency tends to rise in a non-linear fashion as the segment length increases, does also comply with international research results.

Here it should be noted that according to the international experience there are many more variables that can be used to calibrate sophisticated models, which have their own advantages and disadvantages. An accident prediction model with more variables can help us to better understand how these factors influence and describe road safety. However, it has been already pointed out by for instance Sawalha and Sayed [5] that many researchers find it tempting to include too many variables in a model. Such a model is not stable and will perform poorly when applied to a new sample. In this research our intention was to come up with simpler models that might help to improve road safety management. It is considered to be as a starting point to do similar and even more detailed studies for other road categories.

7. Acknowledgments

This research was supported by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP-4.2.4.A/ 2-11/1-2012-0001 ‘National Excellence Program’.

References

- [1] European Commission (2008): *Directive 2008/96/EC of the European Parliament and of the Council of 19 November 2008 on road infrastructure safety management*, Official Journal of the European Union, OJ L 319, p. 59, 29.11.2008
- [2] Hauer, E., Ng, J. C. N., and Lovell, J.: *Estimation of safety at signalized intersections*, Transportation Research Record, 1185, pp. 48-61, 1988
- [3] Miaou, S., Lum, H.: *Modeling vehicle accidents and highway geometric design relationship*, Accident Analysis and Prevention 25, pp. 689–709, 1993
- [4] Abdel-Aty, M. A., Radwan, A. E.: *Modeling traffic accident occurrence and involvement*, Accident Analysis and Prevention 32, pp. 633-642, 2000
- [5] Sawalha, Z., Sayed, T.: *Evaluating Safety of Urban Arterial Roadways*, Journal of Transportation Engineering, 127, pp. 151-158, 2001
- [6] Cafiso, S., Di Graziano, A., Di Silvestro, G., La Cava, G., Persaud, B.: *Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables*, Accident Analysis and Prevention, 42, pp. 1072-1079, 2010
- [7] Qin, X., Ivan, J. N., Ravishanker, N.: *Selecting Exposure Measures in Crash Rate Prediction for Two-lane Highway Segments*, Accident Analysis and Prevention, 938., pp. 1–9, 2003
- [8] Mayora, J. M. P., Rubio, R.L.: *Relevant Variables for Crash Rate Prediction in Spain's Two Lane Rural Roads*, In Proceedings of the 82nd Transportation Research Board Annual Meeting, Washington DC, USA, 2003
- [9] Pardillo, J.M., Llamas, R.: *Relevant variables for crash rate prediction in Spain's two lane rural roads*, In: Proceedings of the 82nd Transportation Research Board Annual Meeting, Washington DC, USA, 2003
- [10] Lamm, R., Beck, A., Zumkeller, K.: *Analysis and evaluation of interrelationships between traffic safety and highway geometric design on two-lane rural roads*, In Proceedings of the 2nd International Symposium on Highway Geometric Design, Mainz, Germany, pp. 557-570, 2000
- [11] Vogt, A., Bared, J.: *Accident models for two-lane rural segments and intersections*, Transportation Research Record, 1635, pp. 18–29, 1998
- [12] Cafiso, S., D'Agostino, C., Persaud, B.: *Investigating the influence of segmentation in estimating safety performance functions for roadway sections*, In Proceedings of the 92nd Transportation Research Board Annual Meeting, Washington DC, USA, 2013
- [13] American Association of State Highway and Transportation Officials (AASHTO): *Highway Safety Manual*, 1st Edition Volume 2., 2010
- [14] Fitzpatrick, K., Schneider, IV W., Carvell, J.: *Using the Rural Two-Lane Highway Draft Prototype Chapter*, In 85th Transportation Research Board Annual Meeting, Washington DC, USA, 2006
- [15] Koorey, G. F.: *Road Data Aggregation and Sectioning Considerations for Crash Analysis*, In 88th Transportation Research Board Annual Meeting, Washington DC, USA, 2009
- [16] Magyar Közút Nonprofit Zrt. (Hungarian Road Administration): *Országos Közúti Adatbank (National Road Databank)*

- [17] R Core Team, R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013
ISBN 3-900051-07-0, URL <http://www.R-project.org/>
- [18] McCullagh, P., Nelder, J. A.: *Generalized linear models*, Chapman and Hall, New York, 1989
- [19] Hauer, E., Bamfo, J.: *Two tools for finding what function links the dependent variable to the explanatory variables*, In Proceedings of 10th ICTCT workshop, Lund, Sweden, 1997
- [20] Ivan, J. N.: *New Approach for Including Traffic Volumes in Crash Rate Analysis and Forecasting*, Transportation Research Record, 1897, pp. 134-141, 2004
- [21] Elvik, R., Høye, A., Vaa, T., Sørensen, M.: *The handbook of road safety measures*, Second Edition, Emerald, 2009
ISBN: 978-1-84855-250-0