Use of Accident Prediction Models in Identifying Hazardous Road Locations

P. Šenk*, J. Ambros, P. Pokorný, R. Striegler
CDV – Transport Research Centre, Brno, Czech Republic
* Corresponding author: petr.senk@cdv.cz

ABSTRACT: The paper introduces the possibility of using accident prediction models for the identification of hazardous road locations. The application of this method is presented with an example of secondary rural roads in the South Moravian region which are classified into road segments homogeneous in terms of basic geometric and traffic characteristics. The prediction model is represented by a generalized linear model which, on the basis of the available data, determines the expected number of accidents for individual types of road segments. A critical road segment is defined as a segment where the reported number of accidents significantly exceeds the number of expected accidents on roads with similar geometric and traffic characteristics. This method can be used as an effective tool for road network safety management.

KEY WORDS: road safety inspection, hazardous road location, accident prediction model.

1 STUDY BACKGROUND AND OBJECTIVES

The first stage of road network safety management is to identify hazardous road locations. The research of literature shows that the majority of EU countries use the system of identification of hazardous road locations (so-called black spots, hot spots, high risk sites, etc.) and an analysis of accidents occurring there (Elvik, 2008). However, none of these systems succeeds in coming closer to the ideal theoretical approach which is described in e.g. Elvik's (2007) work, who claims that critical spots identification and treatment is to be performed as follows:

- Development of an accident prediction model which can be also used as the basis for the analysis of the road network when identifying critical spots;
- Production of an extensive list of road elements to which the analysis is applied, and their classification (road segments, junctions, curves, bridges, tunnels, etc.). This classification is important for several reasons, e.g. not to identify an excessive number of junctions as accident junctions simply due to the fact that usually more accidents occur at junctions than at road segments of a similar length (e.g. 100 m);
- Estimation of the expected number of accidents for each element;
- Application of an algorithm in order to identify road segments with a higher than common number of road accidents;
- Design of potentially effective measures leading to the improvement of safety of the elements.

This approach was used in the study which is introduced in the following text. The study is a part of IDEKO project ("Identification and solutions to critical spots and road segments on the road network which stimulate road users' illegal and improper behaviour, due to their arrangement") conducted by Centrum dopravního výzkumu, v.v.i. and funded by the Programme of Safety Research of the Czech Ministry of Interior. The study deals with the network of secondary rural roads in the South Moravian region since this category is, together with primary roads, the most hazardous road category in the Czech Republic.

The next part introduces the sources of data used in the study, methodological guidelines of designing the accident prediction model and the resulting model form. Subsequently, the use of the model for the identification of hazardous road locations is described. The final part contains the overview of the study results and a brief description of the work plan of the IDEKO project.

2   ACCIDENT PREDICTION MODEL

Accident prediction models have been extensively used in the domain of road infrastructure for the estimation of the expected number of accidents on road segments and junctions (Hauer et al., 1988; Mountain et al., 1996; Greibe, 2003; Daniels et al., 2009) as well as for the estimation of safety benefits (Kulmala, 1995; Carson & Mannering, 2001; Usman et al., 2010).

2.1 Data

Data on road traffic accidents

Data on road traffic accidents was obtained from the sources of the Czech Police. For the purpose of this study road traffic accidents which occurred on rural secondary roads in the South Moravian region in 2009 to 2011 were used. Furthermore, the accidents which occurred at junctions with tertiary and higher category roads were removed, so that the database only contained the accidents which occurred at non-junction road segments. The accidents at junctions with local and access roads (access to field and forest roads, access to car parks, petrol stations, etc.) were kept in the monitored road segments.

Manual checks of these accidents were performed and they found that some of the accidents which are reported at a junction with a local road actually occurred at a junction with a tertiary road or a road of higher category. These accidents were then removed. Furthermore, the accidents which were located further than 50 m off the nearest road were also removed – these accidents were probably incorrectly localized. The total number of accidents which were left in the database was 1408 (515 in 2009, 480 in 2010 and 413 in 2011).

Table 1 contains an overview of the values of attributes from road accident reports. Besides some exceptions, all registered accidents occurred at two-lane roads, most frequently on straight road segments (45 % of cases), on straight road segments up to 100 m off a horizontal curve (22 % of cases), and in horizontal curves (30 % of cases). Just 3 % of reports refer to accidents at junctions with local and access roads. The majority of accidents are the consequence of a collision, the crash of a vehicle with a solid obstacle, and a crash with another vehicle (84 % of cases). Collisions with animals represent 11 % of cases and collisions with pedestrians and other collisions represent 3 % or 2 % of cases respectively. Due to the high proportion of accidents with animals, the data on roads were complemented with an attribute of "road surroundings" which is used as a proxy variable for animal exposure. A closer look at the liability for collisions with animals reveals

that vehicle drivers are liable in the majority of cases (83 %), animals are liable in 11 %, and drivers of non-motor vehicles, pedestrians, and vehicle collisions in the remaining cases.

Table 1: Description statistics of road accident from 2009-2011.

Variable	*Frequency*
ACCIDENT LOCALITY	
urban areas	13
rural areas	1395
TYPE OF ACCIDENT	
crash with non-rail vehicle	562
crash with solid obstacle	294
crash with pedestrian	39
crash with animal	158
crash with train	2
crash with other objects	326
other type of accident	27
ACCIDENT CHARACTER	
accident with injury or fatal consequences	776
accident with property damage only	632
ACCIDENT LIABILITY	
motor vehicle driver	1173
non- motor vehicle driver	38
pedestrian	15
forest animals, domestic animals	158
road fault	3
vehicle technical breakdown	12
other liability	9
ROAD CLASSIFICATION	
two-lane	1390
other	18
ROAD ALIGNMENT	
straight segment	638
straight segment after curve	307
curve	418
junction	45

Data on infrastructure

Data on secondary roads in the South Moravian region were collected from the sources of the Road and Motorway Directorate (ŘSD) database. Regarding the objectives of the study, only rural road segments without junctions with primary, secondary and tertiary roads

were selected. Following the example of Cafiso et al. (2010), the selected road segments were further divided into segments, so that each segment would meet the following criteria:

- the length of segment of at least 50 m
- equal number of traffic lanes along the whole length of road segment
- same road category along the whole length of road segment
- existence/non-existence of paved verges along the whole length of road segment
- existence/non-existence of permanent speed limit reduction along the whole length of road segment
- existence/non-existence of continuous forest vegetation in the vicinity of road segment
- equal traffic volume along the whole length of road segment

After the road division into homogeneous segments, each of the segments was complemented with data on road segment length, curvature, proportion of heavy vehicles, and the number of junctions with local roads. Finally, each of the segments was assigned with information on the corresponding number of road traffic accidents. The resulting set includes 848 segments. Table 2 shows basic description statistics of the data file.

Table 2: Basic description statistics of road infrastructure data file.

Label	Description	Variable type	Source	Description statistics (mean value /standard deviation/ minimum/maximum or frequency)
KRAJNICE	Road with shoulder	Binary [1=yes]	ŘSD	1:170; 0: 678
KATPK	Road category	Categorical	ŘSD	S7,5: 642; S9,5: 142; S11,5: 64
KRIZENI	Number of junctions with local roads	Continuous	ŘSD	0,20 / 0,50 / 0 / 4
VYBAV	Road equipment	Continuous	ŘSD	0,47 / 0,94 / 0 / 13
CUMUL	Curvature	Continuous [gon/km]	ŘSD	116 / 174 / 0 / 1595
JPRUH	Number of traffic lanes	Categorical	ŘSD	1:1; 2:806; 3:39; 4:2
LES	Road vicinity (forest)	Binary [1=yes]	Cenia	1:199; 0: 649
RPDI	AADT	Continuous [veh./year]	ŘSD	3063 / 2931 / 91 / 18500
PODILTV	Proportion of heavy vehicles	Continuous	ŘSD	0,18 / 0,05 / 0,06 / 0,49
LEN	Segment length	Continuous [m]	ŘSD	1176 / 1120 / 51 / 6456
CRASH	Number of accidents	Continuous [acc./3 years]	PČR	1,66 / 2,9 / 0 / 20

Methodology part

When designing the accident model, we took into account the data specific features, specifically the Poisson division of the number of accidents per 1 km of road segment length (see Figure 1). The data of this type are modeled with the use of Poisson regression model or a negative binomial regression model in the case overdispersion is suspected. In this case, we selected a general version of negative binomial regression, which is reduced to the traditional Poisson regression in the case of non-significant overdispersion.
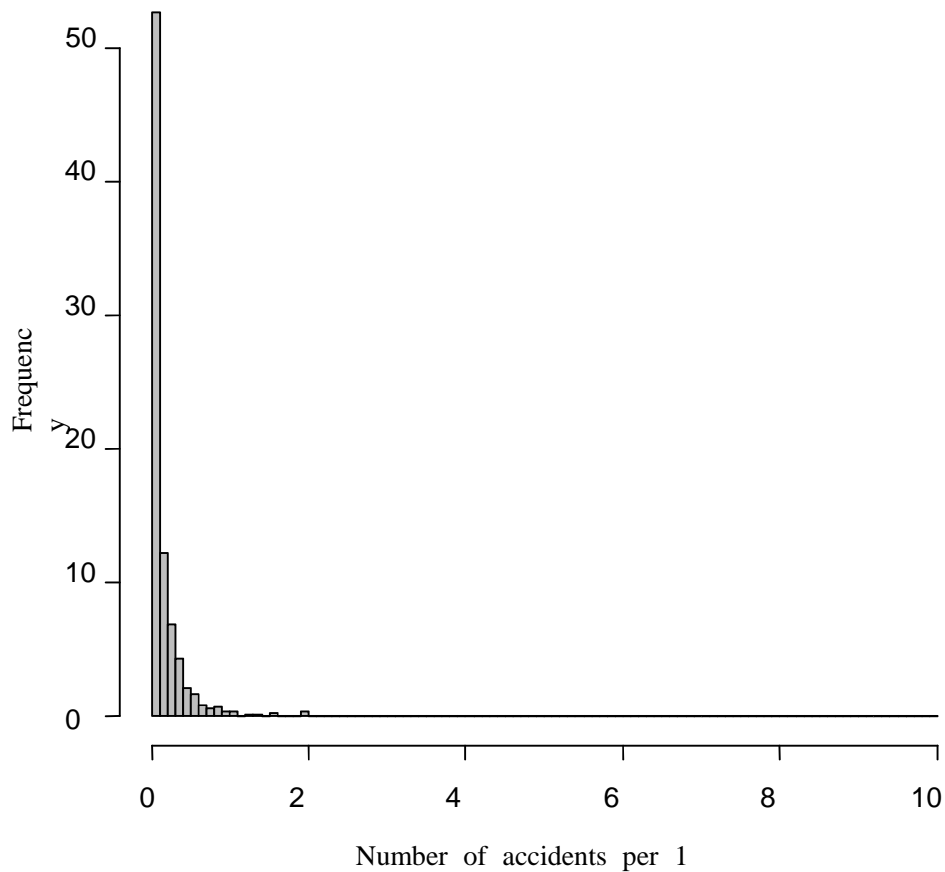
Figure 1: Histogram of the number of accidents per 1 km segment.

A detailed mathematical description of a negative regression model and its relationship to Poisson regression was published in a paper focused on the accident prediction model for roundabouts (Šenk & Ambros, 2011). It is to be noted that the negative binomial regression is a specific example of generalized linear regression, in which the model core is created by the following link function

$$\lambda = e^{(x\beta+\varepsilon)}$$

where e^{ε} is a random error with gamma distribution with the mean value of $E(e^{\varepsilon}) = 1$ and variance $Var(e^{\varepsilon}) = \alpha$. Integrating ε out of the above mentioned formula leads to negative binomial distribution of the described variable with the mean value of λ and variance $\lambda + \alpha\lambda^2$ Positive values of parameter α control for overdispersion of the response variable (number of accidents per road segment), while values close to zero reduce the model to the Poisson regression model. The estimation of parameters α and β is performed by the maximum-likelihood method.

$$E\,(CRASH) = e^{\top}((\beta_0 + \beta_1\,ln(RPDI) + \beta_2\,ln(LEN) + \sum_{i=3}^{n}(\beta_i \cdot x_i))}$$

where *RPDI* represents AADT of vehicles passing through the road segment, *LEN* represents the length of segment in metres, x_i geometric-traffic characteristics of the segment and β_i represents the corresponding regression coefficient. The ability of the model to represent empirical data was evaluated by the combination of the Akaike information criterion (AIC) and the likelihood ratio test.

Resulting regression model

To estimate the parameters of the regression model the statistical software R, specifically its function *glm.nb()* for negative binomial regression from the extension package *MASS*, was used. The final model version is the result of the following process:

- Creation of the initial/updated model version with basic variables *RPDI* and *LEN* (number of variables in model *j = 2*);
- Creation of a set of models N formed of *n* models, by the extension of the updated model version with one of *n* variables which is not included in the updated model version *(j = j + 1)*;
- The selection of a subset of models M out of N, so that for each model it is valid that coefficients of all variables are statistically different from zero at the significance level of 0.1. Furthermore, in comparison with the updated model version, the corresponding model version should explain the variability in a dependent variable better (assessed by likelihood ratio test). In case one of the conditions is not met, the modelling process is terminated and the updated model version is declared the final version.
- Model *m* out of a set M with the lowest AIC value is declared the updated model version.
- Return to point 2.

The final model version is shown in Table 3.

Table 3: Parameters of accident prediction model.

	β_i	SEM	z-score	Pr(>/z/)
γ	-13.64683914	0.632534442	-21.57485543	<0.001
ln(RPDI)	0.930655451	0.054874554	16.95969038	<0.001
ln(LEN)	0.949910152	0.052579181	18.0662791	<0.001
LES	0.419977676	0.095815101	4.383209652	<0.001
CUMUL	0.000417459	0.000233551	1.787444356	0.074
AIC	2357.3			
estimate α	2.08			
SEM α	0.28			
2 x log-likelihood	-2345.3			

A high positive value of dispersion parameter α and a low value of the standard error of the mean (SEM) confirm overdispersion of the response variable and the appropriateness to select the negative binomial model.

By substituting variables and corresponding values of parameters into the general function of the expected number of accidents, we receive the accident prediction model at road segments of secondary roads in the South Moravian region as follows

$$E(CRASH) = e^{-13,6468+0,9307\cdot ln(RPDI)+0,9499\cdot ln(LEN)+0,42\cdot LES+0,0004\cdot CUMUL}$$

3   MODEL APPLICATION

The previous chapter described the accident prediction model on the network of secondary rural roads in the South Moravian region. The following part introduces a potential application of the results of this model in the process of road network safety management, particularly from the viewpoint of road administrators. This process includes the identification of hazardous road locations (accident localities) and the determination of the priorities of their treatment.

3.1 Theoretical part

Pokorný and Striegler (2011) claim that there are currently a number of different definitions of an accident locality in the Czech Republic. There are more identification criteria for the determination of an accident locality. Within the literature research performed in the IDEKO project, a so-called *criterion of absolute difference* was selected in order to identify accident localities. This criterion focuses on localities with the highest potential for the reduction of the number of accidents. When using this criterion, it is necessary to determine what the absolute difference needs to be, so that a locality could be considered an accident locality. This depends on road safety policy, strategy, budget, and required accuracy level. Therefore, no single figure can be generally stated; however, two general rules can be used:

- The criterion for identification can be a determined figure which needs to exceed the potential (suitable for smaller territorial areas), or a certain percentage of the road network with the highest potential (suitable for larger territorial areas).
- Severity of accidents when identifying accident localities should not be taken into account.

Elvik (2007) defines hazardous road location as a location which has a higher expected number of accidents than other identical locations due to local risk factors under consideration that the local risk factors are particularly related to the road design.

3.2 Example of using the accident prediction model

Hazardous road locations are identified within the project on the road network of secondary roads in the South Moravian region (road segments without junctions only). These road segments are defined by their geometric and traffic characteristics which are represented by independent variables in the prediction model, described in part 2. The expected number of accidents, which is a result of prediction (and dependent variable as well), corresponds with the above mentioned definition. A suitable criterion to determine the severity (hazard rate) of road segments and priorities of their rehabilitation refers to the difference between the reported and expected number of accidents.

The risk being referred to by the number of accidents can be understood in two ways: individually or collectively. The individual risk represents a probability for drivers that a potential accident befalls a driver; it is expressed by accident rate. In contrast, the collective risk, expressed by the accident density, concerns all accidents. While, obviously, the individual risk is perceived by individual drivers, road administrators consider accident density a more suitable indicator (Ambros, 2012). In order to express the density, the mentioned difference between the reported (R) and expected number of accidents (E) was divided by the segment length:

$X = (R - E) / L$. Value X represents the above mentioned accident potential.

- Positive example: Two accidents occurred on a kilometre-long road segment, while three accidents were expected $\Rightarrow X = 2 - 3 = -1$.
- Negative example: Three accidents occurred on a kilometre-long road segment, while just two accidents were expected $\Rightarrow X = 1 - 2 = -1$.

Therefore, positive values of the accident potential indicate situations which should be targeted by subsequent treatment.

A demonstrative overview of the expected and real accident density of all segments on the network, which is the subject of this study, was created (see Table 4).

Table 4: Contingency table of the expected and real accident density.

expected and *real* accident density	0 – 0.5	0.5 – 1	1 – 1.5	1.5 – 2	2 – 2.5	2.5 – 3	3 – 3.5	3.5 – 4	4 – 4.5	4.5 – 5	5 – 5.5	5.5 – 6	6 – 6.5	6.5 – 7	7 – 7.5	7.5 – 8	8 – 8.5
0 – 0.5	100	120	83	56	33	14	12	7	5	7	0	3	3	2	3	1	1
0.5 – 1	8	21	28	13	1	1	2	1	0	0	0	0	0	0	0	0	0
1 – 1.5	2	25	16	12	3	3	2	0	1	0	0	0	0	0	0	0	0
1.5 – 2	7	16	15	8	6	2	2	2	0	0	0	0	0	0	0	0	0
2 – 2.5	3	9	13	8	3	1	0	2	1	0	0	0	0	0	0	0	0
2.5 – 3	0	6	6	7	4	2	2	0	2	0	0	0	0	0	0	0	0
3 – 3.5	0	2	8	8	3	2	2	0	2	1	0	0	0	1	0	0	1
3.5 – 4	0	1	1	3	3	3	0	0	1	0	1	0	0	0	0	0	0
4 – 4.5	0	1	2	1	5	2	2	0	0	0	1	0	0	0	0	0	0
4.5 – 5	0	1	0	1	3	0	0	0	1	0	0	0	0	1	0	0	0
5 – 5.5	0	0	0	3	1	2	0	0	2	0	0	0	0	1	0	0	0
5.5 – 6	0	0	0	2	1	0	1	0	1	0	1	1	1	0	0	0	0
6 – 6.5	0	1	1	2	0	0	0	0	0	1	0	0	0	0	0	0	0
6.5 – 7	0	0	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0
7 – 7.5	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0
7.5 – 8	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1
8 – 8.5	0	1	3	4	5	2	1	2	3	3	1	2	3	0	2	0	3

The rows are divided in intervals of the real accident density, and columns in intervals of the expected accident density. The number in each cell refers to the number of segments with the accident density in corresponding intervals.

The average real and expected accident density per 1 km of the analysed network is approximately 1.7. This is why the "below-the-average" road segments were highlighted, i.e. where the real, as well as the expected accident density, exceeded the value 2. This concerns 169 segments, i.e. approximately a fifth of the monitored network. It is obviously inconceivable to expect that such a large number of accidents can be treated. In the previous case, the criterion of "below-the-average" locality was applied. Another criterion can use a pre-determined proportion, e.g. upper 10 % of all values. According to this criterion, there are 83 hazardous segments occurring on the monitored network.

4　CONCLUSION

The article introduced the use of accident prediction models in order to identify hazardous road locations – both its theoretical background, and a specific example of its application on the network of secondary roads in the South Moravian region. Although the mentioned approach is recommended worldwide, it is the first known example of its application in the Czech Republic. The further project stage includes an accident analysis of selected spots and road safety inspection, in order to identify local risk factors. In this way the above mentioned theoretical definition of the critical segment will be met: it is a segment with a higher expected number of accidents than other similar segments, due to local risk factors. Subsequently, low-cost measures will be designed in order to remove these risk factors.

REFERENCES

Ambros, J., 2012. Jak měřit bezpečnost? – 1. část. *Silniční obzor*, 73 (3), pp. 78-82. (in Czech)

Cafiso, S., Di Graziano, A., Di Silvestro, G., La Cava, G., Persaud, B., 2010. Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables. *Accident Analysis and Prevention*, 42 (4), pp. 1072-1079.

Carson, J., Mannering, F., 2001. The effect of ice warning signs on accident frequencies and severities. *Accident Analysis and Prevention*, 33 (1), pp. 99-109.

Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2009. Explaining variation in safety performance of roundabouts. *Accident Analysis and Prevention*, 42 (2), pp. 393-402.

Elvik, R., 2007. *State-of-the-Art Approaches to Road Accident Black Spot Management and Safety Analysis of Road Networks,* report 883/2007 [online]. Oslo: Institute of Transport Economics. Retrieved from: www.toi.no/getfile.php/Publikasjoner/T%D8I%20rapporter/2007/883-2007/883-2007-nett.pdf

Elvik, R., 2008. A survey of operational definitions of hazardous road locations in some European countries. *Accident Analysis and Prevention*, 40 (6), pp. 1830-1835.

Greibe, P., 2003. Accident prediction models for urban roads. *Accident Analysis and Prevention*, 35 (3), pp. 273-285.

Hauer, E., Ng, J.C.N., Lovell, J., 1988. Estimation of safety at signalized intersections. *Transportation Research Record*, no. 1185, pp. 48-61.

Kulmala, R., 1995. *Safety at rural three- and four-arm junctions: Development and application of accident prediction models*. Publication no. 233. Espoo: VTT.

Mountain, L., Fawaz, B., Jarrett, D., 1996. Accident prediction models for roads with minor junctions. *Accident Analysis and Prevention*, 28 (6), pp. 695-707.

Pokorný, P., Striegler, R., 2011. Identifikace nehodových lokalit. *Dopravní inženýrství*, 6 (2), pp. 8-11. (in Czech)

Šenk, P., Ambros, J., 2011. Estimation of Accident Frequency at Newly-built Roundabouts in the Czech Republic. *Transactions on Transport Sciences*, 4 (4), pp. 199-206.

Usman, T., Fu, L., Miranda-Moreno, L.F., 2010. Quantifying safety benefit of winter road maintenance: Accident frequency modeling. *Accident Analysis and Prevention*, 42 (6), pp. 1878-1887.