Proceedings of 6th Transport Research Arena, April 18-21, 2016, Warsaw, Poland

A feasibility study for developing a transferable accident prediction model for Czech regions

Jiří Ambros [a,*], Jiří Sedoník [a]

[a] *CDV – Transport Research Centre (Centrum dopravního výzkumu, v.v.i.), Líšeňská 33a, 636 00 Brno, Czech Republic*

Abstract

Network safety ranking (or hotspot identification) is conducted in order to identify hazardous road locations and develop a priority list for application of effective safety treatments. In this process empirical Bayes approach, using accident prediction models, has been recommended. In order to be conducted efficiently, several questions need to be answered: for example which model function form should be adopted; which explanatory variables need to be collected and included in model; which time period is to be used. There is also a question whether each region requires its specific model or whether a transfer from elsewhere is feasible. These decision points motivated the presented study, which was conducted with data from national road network in two adjacent regions in the southeast of the Czech Republic (South Moravian region and Zlín region). For the purpose of network safety ranking, separate accident prediction models have been developed and applied in both regions. Several analysis steps were conducted in order to compare the model quality and performance, as well as both temporal transferability (between various periods) and spatial transferability (between the regions). The idea was to develop a transferable model, which would ease up on demands on data collection and modelling efforts and would thus increase effectiveness of the network safety ranking process. The presented study aims to assess the feasibility of developing such model.

Keywords: Road safety; network safety ranking; accident prediction model; transferability

1. Introduction

Hazardous road locations (also known as black spots or high risk sites) are the locations (road segments or intersections) which have insufficient level of safety and thus should be investigated and accordingly treated. The process of network safety ranking (also known as hotspot identification or network screening) is conducted in order to identify such locations. According to European Directive 2008/96/EC on road infrastructure safety management (European Commisson, 2008), network safety ranking is a method for identifying, analysing and classifying parts of the existing road network according to their potential for safety development and accident cost savings; it should result in a priority list of road sections where an improvement of the infrastructure is expected to be highly effective. In

* Corresponding author. Tel.: +420 541 641 362; fax: +420 541 641 712.
E-mail address: jiri.ambros@cdv.cz

order to fulfill these requirements the empirical Bayes approach using accident prediction models has been recommended (Elvik, 2008a).

In practice network safety ranking is in the responsibility of road agency; however statistical accident modelling is not an easy task and may be demanding in terms of data requirements and staff qualification (Eenink et al., 2008). In addition there are several important decisions to be taken during modelling process: for example which model function form to adopt; which explanatory variables to collect and include in model; which time period to use, etc. Last but not least there is a question whether each road agency needs to develop its own models.

The mentioned important decisions have motivated the presented paper. The feasibility study was conducted on national road network (1st class roads) in two adjacent regions in the southeast of the Czech Republic (South Moravian region and Zlín region). In 2014 road agencies in both regions commissioned CDV (Transport Research Centre) to identify hazardous locations on both networks, during which separate regional accident prediction models have been developed and applied according to empirical Bayes methodology. Several analysis steps were conducted in order to compare the models (and hazardous road locations identified with them). In addition both temporal transferability (between various periods) and spatial transferability (between the regions) have been studied. The idea was to develop a transferable model, which would ease up on demands on data collection and modelling efforts and would thus increase effectiveness of the network safety ranking process. The presented study aims to assess the feasibility of developing such model.

2. Research questions and study objectives

The empirical Bayes (EB) approach to road safety analysis combines local accident history with an expected accident frequency, estimated with accident prediction model (Hauer et al., 2002). This methodology increases the precision of estimation and corrects for the regression-to-mean bias, and thus has been recommended to be used in road safety studies, comparisons and evaluations. Accident prediction models (safety performance functions) are necessary tools in EB approach. However their use has been generally known and accepted, their development involves several decisions which have often not been sufficiently described. The main questions are as follows:

1. *Which model form should be adopted?* General form of accident prediction model is following:

$$P = \beta_0 \cdot AADT^{\beta_1} \cdot exp(\beta_2 x_2 + \beta_3 x_3 + \cdots) \quad (1)$$

2. i.e. expected (predicted) accident frequency P is estimated via multiplicative regression model, including explanatory variables of exposure (average annual daily traffic, $AADT$) and other risk factors x_i; β_i are regression coefficients to be estimated in modelling. Based on choice of explanatory variables, two types of models are distinguished: simple models involve exposure only; multivariate (or full) models use also further variables, usually geometric characteristics. Nevertheless it has often been found that additional predictors were not as beneficial as expected, and simple models may be sufficient (Peltola et al., 1994; Walmsley et al., 1998; Wood et al., 2013; Ambros et al., 2015; Ambros and Peltola, 2015; Ambros et al., 2016).
3. *Which function form or variables should be used?* Explanatory variables, used in modelling, may take on various function forms. Various approaches towards selection of the most suitable mathematical forms exist, including empirical integral functions and cumulative residual (CURE) graphs (Hauer and Bamfo, 1997), cumulative residuals box and whiskers plots (CRBW) (Kamińska, 2014) or regression splines (Gitelman et al., 2014). There is no universal guidance and various function forms are used: for example traffic volume is typically used in a power form, but some authors considered it jointly with an exponential form (Hauer et al., 2004; Reurings and Janssen, 2007; Elvik, 2008b; Kamińska 2011, 2014). Another example is segment length, usually applied in a power form, but sometimes also exponentially (Parajuli et al., 2006; Valentová et al., 2014; Ambros et al., 2015; Ambros and Peltola, 2015; Ambros et al., 2016) or as an offset, i.e. with regression coefficient = 1 (Jonsson, 2005; AASHTO, 2010; Cafiso et al., 2010; Jurewicz et al., 2014).
4. *Which time period should be used?* A period between 1 and 5 years is usually recommended for safety ranking, with 3-year period being the most frequent (Elvik, 2008a). Longer time periods may cause problems with instability of conditions which may not reflect current traffic situation anymore (Hauer and Persaud, 1984). Probably due to these issues no specific guidelines for time period choice are provided; usually some compromise between the need for quick detection and the need for accumulating a sufficient accident numbers to permit analysis is accepted (Elvik, 2010).

All the mentioned questions need to be answered during developing an efficient accident prediction model. Such model should be sufficiently reliable (describing safety performance of a modeled dataset) while also enough simple and parsimonious so that its future updating is not highly demanding.

In addition each accident prediction model is developed using data from specific time period and specific geographical area. It means that in time and across space data (and thus models) are different. This situation naturally leads to questions whether the models are transferable – could they be adapted (calibrated) to future time periods and different regions? Should this be the case, road agencies could rely on previously (and/or elsewhere) developed models and save their budgets. With this assumption, several studies related to transferability of models from US sources (*Interactive Highway Safety Design Model* and *Highway Safety Manual*) have been conducted. Some reported good transferability, mainly between US states (Sun et al., 2006; Xie et al., 2011; Bornheimer et al., 2012), however some were less successful (when applied abroad, for example in Canada or Italy) and recommended to develop their own specific models (Persaud et al., 2002; Sacchi et al., 2012; Young and Park, 2013).

Based on the previous text, the study objectives are following:

1. To decide on appropriate model form (simple or multivariate), function form of traffic volume and segment length, and time period for two Czech regions.
2. To decide on feasibility of transferring those two models in time and across space.

The two sets of tests will be applied: the first (related to the first objective) will assess model quality, the second (related to the second objective) will assess model performance.

3. Data and methods

The modelled network covers national road network (1st class roads) in two adjacent regions in the southeast of the Czech Republic: South Moravian and Zlín region (in the following text they will be referred to as SM and ZL). These roads contain rural and urban sections, both undivided and divided (single and dual carriageways). For the purpose of this study, only rural undivided sections (i.e. typical secondary roads), excluding intersections, were selected. Both samples are above 200 km of total length (258.7 and 222.4 km in SM and ZL, respectively) and represent more than 60% of both given networks.

3.1. Segmentation

Both samples were divided into homogeneous segments with respect to following variables: annual average daily traffic (AADT), speed limit reduction, road category, number of lanes, paved shoulder. A change of any of these variables marked the end of a segment and beginning of another one. The approach was inspired by other studies, such as Cafiso et al. (2010) and used in a previous study by Ambros et al. (2015).

The minimal length was set to 50 m. In order to obtain segment lengths which will be practical for follow-up safety inspections, segments longer than 500 m were divided into 250 m parts. After this segmentation total number of segments is above 1000 (1156 and 1168 in SM and ZL, respectively). Average segment length is approx. 200 m; most of them (above 80%) are up to 250 m long.

3.2. Variables

The segments were assigned specific values of response variable (accident frequency) and explanatory variables (exposure data, road and traffic characteristics, context and environment variables) which represent safety-related features. Recorded accident frequencies for 8-year period (2007 – 2014) were obtained from the Czech Traffic Police. They include all injury accidents, i.e. with slight, severe or fatal personal consequences. In total there were 1249 and 817 accidents in SM and ZL regions, respectively. Further explanatory variables were added:

- AADT (I) to represent the risk exposure. These data were acquired from Czech Road and Motorway Directorate, based on the results of national traffic census 2010.
- The segment length (L); its generation was described in the previous paragraph.
- Average curvature change rate (CCR), as a traditional alignment consistency indicator. It was computed as sum of angular changes divided by segment length.
- Other characteristics were also obtained from a Czech Road and Motorway Directorate database, reflecting the state in January 2013: density of intersections with minor roads (INT) and density of roadside facilities (FAC)

(computed as frequencies divided by segment lengths); road width categories (CAT); hard shoulders ($SHLD$); number of lanes ($LANE$); speed limit reductions (default speed limit is 90 km/h) ($SPEED$); tree alleys ($TREE$).

3.3. Model development

A negative binomial regression was applied to develop the models, using SPSS procedure GENLIN and a common form, introduced in equation (1) in the previous text. With the first study objectives in mind (deciding on appropriate model form, function form of AADT and segment length, and time period), several model variants were modelled. The process will be described in following paragraphs.

3.3.1. Time period

Depending on the number of explanatory variables, model type may be simple or multivariate. In order to study the suitable type within available time frame (8 years), investigation scheme with overlapping time periods was planned. 3-year period was taken as a minimum, and 6-year period as a maximum. For illustration see Fig. 1. Accident prediction models were developed for all 18 variants, in a backward elimination manner, with only the variables with at least 95% statistical significance being kept in the model.

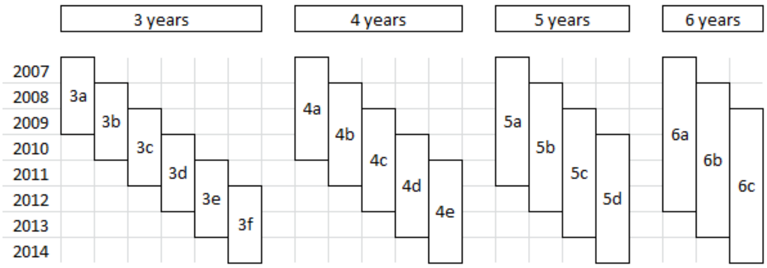

Fig. 1. Scheme of investigated variants of time periods.

During modelling several variants of 3-year models could not be built, due to non-invertible Hessian matrix. This error (caused by singularity or nonpositive definiteness of Hessian matrix) is 'frustrating but common occurrence in applied quantitative research' (Gill and King, 2004). Therefore in further analyses only 4-, 5- and 6-year models were used.

3.3.2. Function form

Regarding function form of AADT (I) and segment length (L), four variants were considered, see Table 1:

Table 1. Overview of model variants, based on power/exponential function forms of AADT and segment length.

	Segment length in a power form	Segment length in an exponential form
AADT in a power form	$P = \beta_0 \cdot I^{\beta_1} \cdot L^{\beta_2} \cdot \exp(\beta_3 x_3 + \beta_4 x_4 + \cdots)$	$P = \beta_0 \cdot I^{\beta_1} \cdot \exp(\beta_2 L + \beta_3 x_3 + \cdots)$
AADT in both power and exponential form	$P = \beta_0 \cdot I^{\beta_1} \cdot L^{\beta_2} \cdot \exp(\beta_3 I + \beta_4 x_4 + \cdots)$	$P = \beta_0 \cdot I^{\beta_1} \cdot \exp(\beta_2 I + \beta_3 L + \cdots)$

where β_i are regression coefficients to be estimated in modelling, and x_i are explanatory variables (curvature change rate (CCR) and others, described in paragraph 3.2).

3.3.3. Model form

In order to select the variables for simple model, the statistically significant variables for the 'least common denominator' (4-year variants) were summarized for both length function forms (see Fig. 2). In none of the variants, exponential form of AADT had statistically significant influence – it is therefore not further considered.

It is shown that variables AADT (I), length (L) and curvature change rate (CCR) are significant in all cases (across both regions and function forms), unlike the remaining variables which are often non-significant. Based on this finding, in the following text two model types are distinguished: *simple* (including the three mentioned variables I, L, CCR) and *multivariate* (including also other variables).

In order to assess quality of all models (according to described variants of time periods, segment length function forms and simple/multivariate models), various goodness-of-fit indicators may be used. For example Oh et al. (2003) used five different measures to assess the external validity (Pearson correlation coefficient between observed and predicted accident frequencies, mean prediction bias, mean absolute deviation, mean squared prediction error, mean squared error), while noting that they all should be considered jointly. For the sake of brevity, a single indicator was used here – proportion of systematic variation in the original accident dataset explained by the model ($\%SV$), also known as Elvik index (Fridstrøm et al., 1995; Kulmala, 1995).

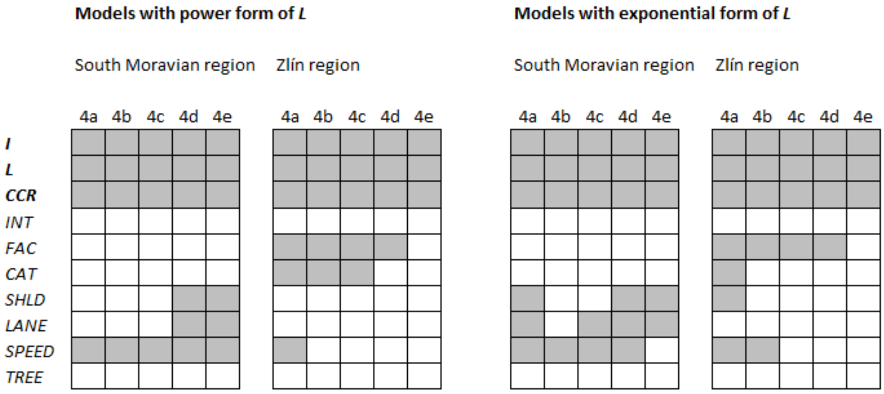

Fig. 2. Overview of achieved statistical significance of explanatory variables in 4-year models for both regions and function forms of segment length (grey cells = significant, white cells = non-significant).

3.4. Model performance

As stated in the introduction, the study focus is on network safety ranking, based on empirical Bayes method with accident prediction models. In principle the models were used to obtain predicted accident frequency for each segment. Empirical Bayes estimate of expected accident frequency was then calculated. Finally potential for safety improvement was obtained as a difference between predicted accident frequency and EB estimate.

$$EB_i = w_i \cdot P_i + (1 - w_i) \cdot R_i \tag{2}$$

$$w_i = \frac{k_i}{k_i + P_i} \tag{3}$$

$$k_i = \frac{k}{L_i} \tag{4}$$

$$PSI_i = EB_i - P_i \tag{5}$$

where EB_i are EB estimates, using weighted average (with weights w_i) of predicted and reported accident frequencies (P_i and R_i). Other letters denote length-dependent overdispersion parameter (k_i), segment length (L_i) and potential for safety improvement (PSI_i) (Persaud et al., 1999; Hauer et al., 2002).

Descending values of PSI were used for network safety ranking, resulting into the lists of segment numbers. Following previous studies, 1%, 3% and 5% upper tails (top parts of distribution) were further used in order to investigate the differences (1% = 12 segments, 3% = 35 segments, 5% = 58 segments).

The idea of model performance check is to assess the implications of choice of different modeling time periods on the results of network safety ranking (i.e. ranked lists of segment numbers). Assessment may be done in terms of 'consistency'. In literature various consistency tests have been used: for example percentage deviation and Spearman correlation coefficient (Miranda-Moreno et al., 2005), epidemiological criteria of sensitivity and specificity (Elvik, 2008c), or other consistency tests (Montella, 2010). In these studies consistency criteria have been used for comparison of different ranking methods. On the other hand in the present study the principle was adapted in order to compare the effect of using safety ranking based on models from different time periods. Therefore the original meaning of ‘identification with different methods’ is applied in the meaning of ‘identification in different time period’ (all with three different levels of upper tails). In total three criteria were selected from the mentioned studies. Their descriptions and definitions are summarized in Table 2. Note that in epidemiological diagnostic test, segments identified as hazardous in maximal (8-year) time period were considered as positives.

Table 2. Explanation of used consistency criteria.

Test	Test premise	Test criterion
Site consistency test	A segment identified as risky during time period i should also be identified as risky in time period $i+1$.	The method that identifies segments in time period i, with the highest accident frequency in a time period $i+1$, is the most consistent.
Method consistency test	A list of segments, which were identified as risky during time period i, should be similar to the list of segments, which will be identified as risky in time period $i+1$.	The method that identifies segments in time period i, whose list has the largest overlap with the list of segments identified in time period $i+1$, is the most consistent.
Epidemiological diagnostic test	The method should identify as many of the truly risky segments as possible (sensitivity), and as few of truly non-risky segments as possible (specificity).	The method with the highest sum of sensitivity and specificity is the most consistent.

3.5. Model transferability

Model transferability was tested in time (between time periods) and across space (between the two regions). The general process of transferring model from condition i to $i+1$ (where condition means time period or region) followed three steps:

- Using regression coefficients from condition i to predict accident frequencies in condition $i+1$.
- Calibration (re-scaling) of predicted values through multiplication by calibration factor $C = \sum R / \sum P$ (ratio of sum of recorded accidents to sum of predicted accidents) (Harwood et al., 2000; AASHTO, 2010).
- Comparison of original accident frequencies with calibrated accident predictions.

In literature various techniques have been used to assess the success of transferability, including goodness-of-fit indicators, visual plots and cumulative residuals graphs (Persaud et al., 2002; Oh et al., 2003; Young and Park, 2013). Here mean squared prediction errors (MSPE) and cumulative residuals (CURE) were used. MSPE is calculated as the sum of the squared differences between recorded accident frequencies and calibrated predictions, divided by sample size. Graphs of cumulative residuals (the differences between the actual and predicted values for each segment) are constructed for increasing values of AADT, as the most influential predictor. Ideal CURE graph should oscillate around zero and lie between the two standard deviation boundaries (Hauer and Bamfo, 1997).

4. Results

Results are reported in order of previously mentioned parts (Model quality, Model performance, Model transferability).

4.1. Model quality

The proportion of explained systematic variation ($\%SV$) was computed for all variants of 4-, 5- and 6-year models and their values were averaged. Typical values were between 60% and 70%, with visible increase between 4-year and 5-year models (see Fig. 3). Since $\%SV$ values do not significantly increase with 6-year models, 5-year models would seem as a suitable option.

Regarding comparison between model forms, simple models had $\%SV$ approx. only by 2-3% lower compared to multivariate models. It was therefore decided that simple models are sufficient and they were used in further analyses.

Above mentioned $\%SV$ values were computed for both variants of segment length function forms. Models with power function had higher values, with differences between 5% and 10% (see Fig. 3). Based on this finding, power function of segment length was chosen for further analyses.

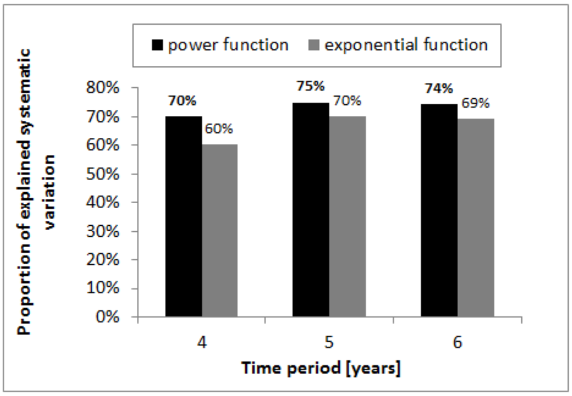

Fig. 3. Proportion of explained systematic variation according to time periods and two function forms of segment length.

4.2. Model performance

The consistency tests were applied on all 19 possible variation pairs of lists of segment numbers (10 pairs for 4-year models, 6 pairs for 5-year models, 3 pairs for 6-year models), for 3 upper tails (1%, 3%, 5%). The graphs in Fig. 4 show the results, averaged for specific combinations of upper tail and time period. It is evident that 6-year models achieve the best performance (the highest values) in all tests, with 5-year models being relatively close.

Together with Fig. 3, it seems that 5-year models provide acceptable results, which are close to 6-year models. For following analyses 5-year models were therefore used.

4.3. Model transferability

Temporal transferability of models was tested in following combinations of 5-year periods:

- transferring 5a to 5b (compared to 5b)
- transferring 5a or 5b to 5c (compared to 5c)
- transferring 5a, 5b or 5c to 5d (compared to 5d)

The results were virtually the same: the MSPE differences were below 1% and CURE graphs were very similar.

Spatial transferability was tested in following combinations:

- transferring model for SM region to ZL region and vice versa (labelled as SM $\rightarrow$ ZL and ZL $\rightarrow$ SM)
- transferring model, developed with combined data from both regions, to SM and ZL (ALL $\rightarrow$ SM, ALL $\rightarrow$ ZL)

Fig. 5 shows the results in terms of MSPE. It is seen that adoption of models across regions leads to higher MSPE (lower performance). On the other hand using combined model from both regions leads to comparable performance (in case of Zlín region) or even improves it (with lower MSPE) in South Moravian region.

Resulting cumulative residual graphs (CURE) are presented in Fig. 6. It is evident that, compared to original data (SM and ZL):

- performance of models from different regions (ZL $\rightarrow$ SM and SM $\rightarrow$ ZL) is worse (with ZL $\rightarrow$ SM graph being close to lower boundary and SM $\rightarrow$ ZL crossing upper boundary)
- performance of combined models (ALL $\rightarrow$ SM and ALL $\rightarrow$ ZL) is close to the original data (even lower for ALL $\rightarrow$ SM).

The findings from both MSPE and CURE are consistent: it is not beneficial to apply models across different regions; at the same time using combined models, which accumulate data from both regions, seems not to decrease the model performance.

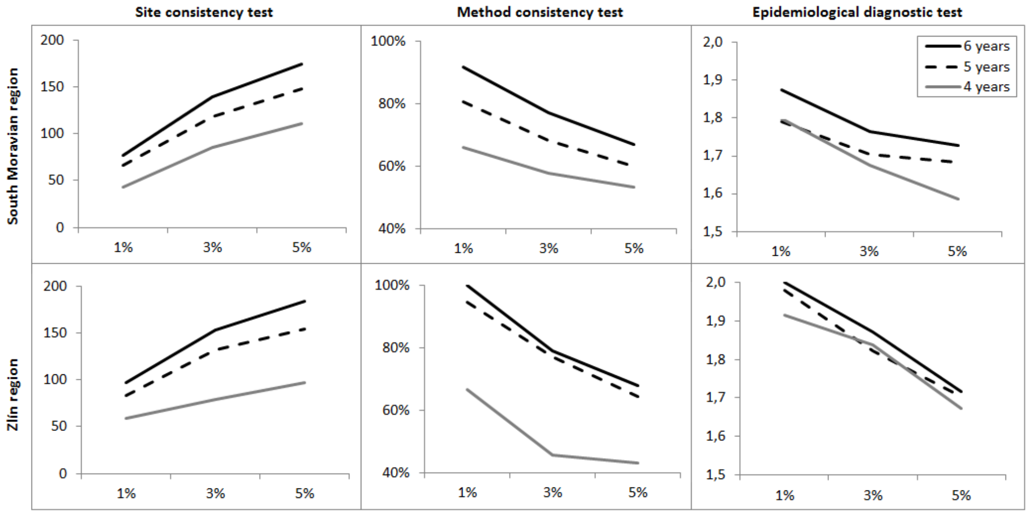

Fig. 4. Results of consistency tests, averaged for specific combinations of upper tails (1%, 3%, 5%) and time periods (4, 5, 6 years).

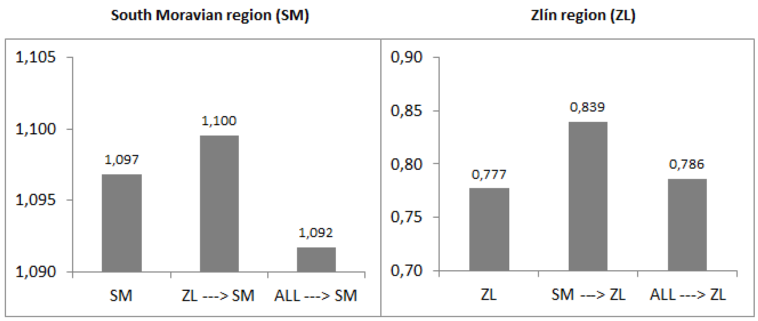

Fig. 5. Mean square prediction errors for original models (SM, ZL), transferred between regions (ZL $\rightarrow$ SM, SM $\rightarrow$ ZL) and with combined models (ALL $\rightarrow$ SM, ALL $\rightarrow$ ZL).

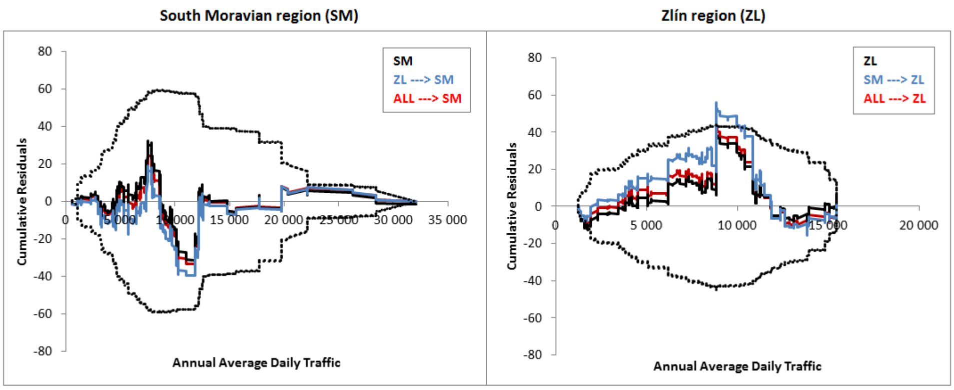

Fig. 6. Cumulative residual graphs for original models (SM, ZL), transferred between regions (ZL $\rightarrow$ SM, SM $\rightarrow$ ZL) and with combined models (ALL $\rightarrow$ SM, ALL $\rightarrow$ ZL).

5. Discussion and conclusions

The original idea was to develop a transferable model, which would be beneficial for road agencies in two Czech regions (SM and ZL), since it would ease up on demands on data collection and modelling efforts, needed for network safety ranking. The paper presents the analyses, which should answer whether development of such model is feasible. The two main tasks were:

1. To decide on appropriate model form (simple or multivariate), function form of AADT and segment length, and time period for two Czech regions.
2. To decide on feasibility of transferring those two models in time and across space.

A combination of power and exponential function forms of AADT was not successful. Since it should represent the fact that accident frequency may decrease after reaching certain threshold value of AADT, it proved significant mainly in data sets with high traffic volumes (maximum AADT: 68 500 in Hauer et al., 2004; 86 300 in Kamińska, 2011, 2014), which was not the case of data used in this study. For example in Kononov and Allery (2003), the 'turning point' was approximately at AADT = 25 000, which was exceeded only in 0.4% cases of South Moravian data.

Model quality was assessed in terms of proportion of explained systematic variation, with following findings:

- Simple models had almost the same explanatory power as multivariate models.
- Models with power function of segment length proved better, compared to variants with exponential function.
- In comparison of time periods, 5-year models seemed a suitable option.

In order to quantify model performance, three consistency tests were applied. 5-year models were found to perform second to best (6-year models), therefore 5 years were accepted as acceptable time period.

Transferability of 5-year models was tested using mean square prediction errors and cumulative residual graphs. Both methods proved that it is not beneficial to transfer models across different regions; at the same time using combined models, which accumulate data from both regions, seemed to be acceptable, in one case even better than original regional models.

The findings are relatively consistent with previous studies:

- Simple models have been used in several countries, e.g. Finland (Peltola et al., 1994), Denmark (Greibe, 2003) or the Netherlands (Reurings and Janssen, 2007). Recently it was indicated that for the purpose of safety ranking accident prediction models do not necessarily need to involve other than exposure variables (Srinivasan et al., 2013).
- Power function form of segment length is more often used in literature, compared to studies using the exponential form.
- 5-year may be a suitable time period. In a simulation study, Cheng and Washington (2005) concluded there is little gain in the safety ranking accuracy when using a period longer than 6 years. 5 years could then be an acceptable compromise.

Under these assumptions, 5-year simple models (using power form of AADT and length, and exponential form of curvature change rate) may be used in the studied regions. It was found that their mutual transferability is not ideal – so their models are not transferable between regions. However at the same time it was shown that model based on data from both regions combined may be comparable with original models or even better. This may be considered the most important finding of this study – it means that in time, as data are accumulated through developing models also for other Czech regions, combined model will become better alternative than using individual regional models. Such model will then be truly transferable and will fulfill the tasks of effective country-wide network safety ranking as was outlined in this study.

Acknowledgements

The study builds on the work conducted under the contracts for Road and Motorway Directorate of the Czech Republic (ŘSD). Appreciated is cooperation of Tomáš Vyhlídal and Karel Chudárek, who were contact persons at ŘSD in South Moravian region and Zlín region, respectively. The methods used were developed within research project IDEKO – Identification and treatment of high risk road spots and sections (VG20112015013), conducted for Czech Ministry of Interior.

References

AASHTO (American Association of State Highway and Transportation Officials), 2010. Highway Safety Manual. First Edition. AASHTO, Washington, USA.

Ambros J., Peltola H., 2015. Testing of 'application-based' goodness-of-fit measure of crash prediction model. 4th NTSA Annual Scientific Seminar, Espoo, Finland.

Ambros, J., Valentová, V., Janoška, Z., 2015. Investigation of difference between network screening results based on multivariate and simple crash prediction models. 94th Annual Meeting of the Transportation Research Board, Washington, USA, paper #15-1242.

Ambros, J., Valentová, V., Sedoník, J., 2016. Developing 'Updatable' Crash Prediction Model for Network Screening: A Case Study of Czech Two-Lane Rural Road Segments. 95th Annual Meeting of the Transportation Research Board, Washington, USA, paper #16-1817 (accepted).

Bornheimer, C., Schrock, S., Wang, M., Lubliner, H., 2012. Developing a regional safety performance function for rural two-lane highways. 91st Annual Meeting of the Transportation Research Board, Washington, USA, paper #12-4549.

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, 1072–1079.

Cheng, W., Washington, S., 2005. Experimental evaluation of hotspot identification methods. Accident Analysis and Prevention 37, 870–881.

Eenink, R., Reurings, M., Elvik, R., Cardoso, J., Wichert, S., Stefan, C., 2008. Accident Prediction Models and Road Safety Impact Assessment: recommendations for using these tools. SWOV Institute for Road Safety Research, Leidschendam, Netherlands.

Elvik, R., 2008a. A survey of operational definitions of hazardous road locations in some European countries. Accident Analysis and Prevention 40, 1830–1835.

Elvik, R., 2008b. The predictive validity of empirical Bayes estimates of road safety. Accident Analysis and Prevention 40, 1964–1969.

Elvik, R., 2008c. Comparative Analysis of Techniques for Identifying Locations of Hazardous Roads. Transportation Research Record 2083, 72–75.

Elvik, R., 2010. Assessment and applicability of road safety management evaluation tools: Current practice and state-of-the-art in Europe. Report 1113/2010. Institute of Transport Economics, Oslo, Norway.

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, L 319/59, Brussels, Belgium.

Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., Krogsgård Thomsen, L., 1995. Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. Accident Analysis and Prevention 27, 1–20.

Gill, J., King, G., 2004. What to Do When Your Hessian Is Not Invertible: Alternatives to Model Respecification in Nonlinear Estimation. Sociological Methods and Research 33, 54–87.

Gitelman, V., Doveh, E., Carmel, R., Pesahov, F., 2014. The relationship between road accidents and infrastructure characteristics of low-volume roads in Israel. 2nd International Conference on Traffic and Transport Engineering, Belgrade, Serbia.

Greibe, P., 2003. Accident prediction models for urban roads. Accident Analysis and Prevention 35, 273–285.

Harwood, D. W., Council, M., Hauer, E., Hughes, W. E., Vogt, A., 2000. Prediction of the Expected Safety Performance of Rural Two-Lane Highways. Report FHWA-RD-99-207. Federal Highway Administration, Washington, USA.

Hauer, E., Persaud, B. N., 1984. Problem of identifying hazardous locations using accident data. Transportation Research Record 975, 36–43.

Hauer, E., Bamfo, J., 1997. Two tools for finding what function links the dependent variable to the explanatory variables. ICTCT 97 Conference, Lund, Sweden.

Hauer, E., Harwood, D. W., Council, F. M., Griffith, M. S., 2002. Estimating Safety by the Empirical Bayes Method: A Tutorial. Transportation Research Record 1784, 126–131.

Hauer, E., Council, F. M., Mohammedshah, Y., 2004. Safety Models for Urban Four-Lane Undivided Road Segments. Transportation Research Record 1897, 96–105.

Jonsson, T., 2005. Predictive models for accidents on urban links: A focus on vulnerable road users. Bulletin 226. Lund University, Lund, Sweden.

Jurewicz, C., Steinmetz, L., Turner, B., 2014. Australian National Risk Assessment Model. Publication No. AP-R451-14. Austroads, Sydney, Australia.

Kamińska, J., 2011. Big cities influence on accident prediction modeling results for the whole country. Procedia – Social and Behavioral Sciences 20, 752–760.

Kamińska, J., 2014. Choice of Functional Form for Independent Variables in Accident Prediction Models. Logistics and Transport 21, 51–62.

Kononov, J., Allery, B., 2003. Level of Service of Safety: Conceptual Blueprint and Analytical Framework. Transportation Research Record 1840, 57–66.

Kulmala, R., 1995. Safety at rural three- and four-arm junctions: Development and application of accident prediction models. Publication No. 233. VTT Technical Research Centre of Finland, Espoo, Finland.

Miranda-Moreno, L. F., Fu, L., Saccomanno, F. F., Labbe, A., 2005. Alternative Risk Models for Ranking Locations for Safety Improvement. Transportation Research Record 1908, 1–8.

Montella, A., 2010. A comparative analysis of hotspot identification methods. Accident Analysis and Prevention 42, 571–581.

Oh, J., Lyon, C., Washington, S., Persaud, B., Bared, J., 2003. Validation of FHWA Crash Models for Rural Intersections: Lessons Learned. Transportation Research Record 1840, 41–49.

Parajuli, B., Persaud, B., Lyon, C., Munro, J., 2006. Safety Performance Assessment of Freeway Interchanges, Ramps, and Ramp Terminals. 2006 Annual Conference of the Transportation Association of Canada, Charlottetown, Canada.

Peltola, H., Kulmala, R., Kallberg, V.-P., 1994. Why use a complicated accident prediction model when a simple one is just as good? 22nd PTRC Summer Annual Meeting, Warwick, UK.

Persaud, B., Lyon, C., Nguyen, T., 1999. Empirical Bayes Procedure for Ranking Sites for Safety Investigation by Potential for Safety Improvement. Transportation Research Record 1665, 7–12.

Persaud, B., Lord, D., Palmisano, J., 2002. Calibration and Transferability of Accident Prediction Models for Urban Intersections. Transportation Research Record 1784, 57–64.

Reurings, M., Janssen, T., 2007. Accident prediction models for urban and rural carriageways (based on data from The Hague region). Report R-2006-14. SWOV Institute for Road Safety Research, Leidschendam, Netherlands.

Sacchi, E., Persaud, B., Bassani, M., 2012. Assessing International Transferability of Highway Safety Manual Crash Prediction Algorithm and Its Components. Transportation Research Record 2279, 90–98.

Srinivasan, R., Carter, D., Bauer, K., 2013. Safety Performance Function Decision Guide: SPF Calibration vs SPF Development. Report FHWA-SA-14-004. Federal Highway Administration, Washington, USA.

Sun, X., Li, Y., Magri, D., Shirazi, H. H., 2006. Application of "Highway Safety Manual" Draft Chapter: Louisiana Experience. Transportation Research Record 1950, 55–64.

Valentová, V., Ambros, J., Janoška, Z., 2014. A comparative analysis of identification of hazardous locations in regional rural road network. Advances in Transportation Studies 34, 57–66.

Walmsley, D. A., Summersgill, I., Binch, C., 1998. Accidents on modern rural single-carriageway trunk roads. TRL Report 336. Transport Research Laboratory, Crowthorne, UK.

Wood, A. G., Mountain, L. J., Connors, R. D., Maher, M. J., Ropkins, K., 2013. Updating outdated predictive accident models. Accident Analysis and Prevention 55, 54–66.

Xie, F., Gladhill, K., Dixon, K. K., Monsere, C. M., 2011. Calibrating the Highway Safety Manual Predictive Models for Oregon State Highways. Transportation Research Record 2241, 19–28.

Young, J., Park, P. Y., 2013. Benefits of small municipalities using jurisdiction-specific safety performance functions rather than the Highway Safety Manual's calibrated or uncalibrated safety performance functions. Canadian Journal of Civil Engineering 40, 517–527.