Advances in Transportation Studies: an international Journal Section B 44 (2018)

How to simplify road network safety screening?

J. Ambros J. Sedoník Z. Křivánková

CDV – Transport Research Centre
Líšeňská 33a, 636 00 Brno, Czech Republic
email: jiri.ambros@cdv.cz

subm. 9th August 2017 *approv. after rev. 11st December 2017*

Abstract

In order to conduct network screening (hotspot identification), safety performance functions (also known as accident prediction models) are required. However, their development is demanding, since they require knowledge of traffic volume data for all evaluated segments and intersections. In addition, the whole screening process as well as its output is separated into segments and intersections, which may not be the most practical step from the perspective of a road agency. In this regard, using a number of intersections per segment length is a potential simplification which allows omitting separate modelling for intersections; however, its performance has been rarely tested. The aim of the paper is to follow the original UK application of the concept, using it in different geographical conditions and also adding an assessment of the method consistency, which is important for the quality of network screening. The method was found feasible: predictions from a simplified model were closely correlated with predictions based on a combination of segment and intersection models, and consistency in terms of overlapping between two rankings of the final segment lists was also sufficient. The simplified approach may thus increase the efficiency of network screening and enable wider practical application for Czech regional agencies.

Keywords – road safety, network screening, segments, intersections, regions

1. Introduction

The first step of the road network safety management process is network screening (or hotspot identification), defined as a process by which the road network is screened to identify potentially hazardous sites [1].

According to the recommended practices [2-4], network screening employs safety performance functions (SPFs), also known as accident prediction models, and empirical Bayes (EB) approach. In the end a list is produced which enables ranking the locations based on their potential for safety improvement [5].

It has become a standard to develop SPFs for separate entities, typically intersections and road segments [5-8]; however, this custom leads to two issues:

1. Network screening, based on such SPFs, is also separate for intersections and segments, and yields two lists in the end. During the allocation of the budget to safety improvements, choosing between investments into countermeasures on, say, 5 segments or 5 intersections may be difficult.
2. Input data include traffic volumes (AADT), which are usually available for all major roads (from periodical traffic censuses), but rarely available for all minor roads. In order

to develop intersection SPFs, the modeller then has to conduct additional traffic surveys to complete AADT data on all intersection legs, which increases the time and budget demands.

For both analysts and funding road agencies it would thus be beneficial to use such model development approach that would allow considering intersections and segments jointly, without having to collect AADT data for all missing intersection legs. In fact, such approach was explored as part of study by Mountain et al. [9] – using data for approx. 3800 km of UK roads with more than 5000 intersections with minor roads (i.e. local distributors or access roads), they developed and compared two approaches: (1) modelling of segment and intersection accidents separately while summing the predictions, and (2) modelling of total accidents on segments including a predictor of the number of minor intersections per a kilometre, i.e. intersection density. The conclusion was that "there is nothing to choose between these approaches in terms of the quality of the estimates obtained".

In addition, this approach is consistent with current Safe System frameworks, which recommend moving from traditional black spot management, focusing on the most critical but isolated parts of road network, to more proactive and systemic network safety management, with longer road segments including intersections [10]. It is of interest that, that with one exception in Norway [11], this modelling design has not become practically used.

The paper presents a study that focused on proving feasibility of the original UK approach on Czech road network. In order to extend the application in a different country with different safety performance, vehicle fleet, weather, and under various conditions, case study was conducted for the network of national roads. Given the practical focus on road agencies' network screening, the method consistency was tested additionally. The following section presents the data and methods followed by results, discussion and conclusions.

2. Data and methods

Data from national roads in Zlín region (Czech Republic) was used for developing SPFs. Roughly half of the intersections were not covered by the National Traffic Census, i.e. lacking AADT data – this was the original motive to explore simplified approaches to network screening.

The study approach followed the original UK study [9] in developing 3 SPFs:

1. for total accidents on segments, including intersection density (combined SPF)
2. for segment-only accidents, i.e. excluding intersection accidents (SPF 2a)
3. for intersections (SPF 2b)

The samples are visualized in Figure 1. The study objective was to compare predictions 1 (based on the combined SPF) and 2 (i.e. a sum of results from SPFs 2a and 2b). Explanatory values were:

- for combined SPF: segment AADT, segment length, intersection density
- for segment-only SPF: segment AADT, segment length
- for intersection SPF: AADT on major and minor roads, number of legs, presence of any turn lane

Consistently with the previous research (e.g. [12-18], including authors' work [19-21]), the following model forms were adopted:

$$N_{combined} = e^{b_0} \cdot \left(AADT_{segment}\right)^{b_1} \cdot (Length)^{b_2} \cdot e^{(b_3 \cdot Intersection\_density)} \tag{1}$$

$$N_{segment} = e^{b_0} \cdot \left(AADT_{segment}\right)^{b_1} \cdot (Length)^{b_2} \tag{2}$$

$$N_{intersection} = e^{b_0} \cdot \underbrace{\left(AADT_{major}\right)^{b_1}}_{F1} \cdot \underbrace{\left(\frac{AADT_{minor}}{AADT_{major}+AADT_{minor}}\right)^{b_2}}_{FR} \cdot e^{(b_3 \cdot \#Legs + b_4 \cdot Turn\_lanes)} \quad (3)$$

where $N_{combined}$, $N_{segment}$, $N_{intersection}$ are accident frequencies for combinations of explanatory variables segment AADT ($AADT_{segment}$), segment length ($Length$), intersection density, major and minor road AADT at intersections ($AADT_{major}$ and $AADT_{minor}$), number of intersection legs ($\#Legs$) and presence of turn lanes ($Turn\_lanes$); e is natural logarithm base, and b_i are regression parameters to be estimated in modelling. For the sake of brevity, AADT terms in Equation (3) will be referred to as $F1$ and FR, as originally used by Vieira Gomes et al. [17].

Injury accidents, reported by Traffic Police in a 6-year period (2009 – 2014), were used for modelling. In order to determine intersection-related accidents, GPS location assigned by Traffic Police was used to define the influence area as the radius of a circle around the intersection centre. A radius threshold of 250 ft (approx. 76 m) was often used (e.g. [22-24]). However, Avelar et al. [25] studied the relationship between the accident locations and the probability of the accidents being associated with intersections and found that a threshold of 300 ft (approx. 92 m) minimizes the risk of underestimating the accident frequency. This threshold, rounded up to 100 m, was also used in the presented study: all accidents within 100-metre area around the intersection centre were considered intersection accidents. The study only used typical (unsignalized) intersections – the signalized, interchanges, roundabouts and other types were discarded. The descriptive characteristics of accident data are reported in Table 1.

A part of traffic volume data was retrieved from the National Traffic Census; the missing part was additionally surveyed using stationary radars.

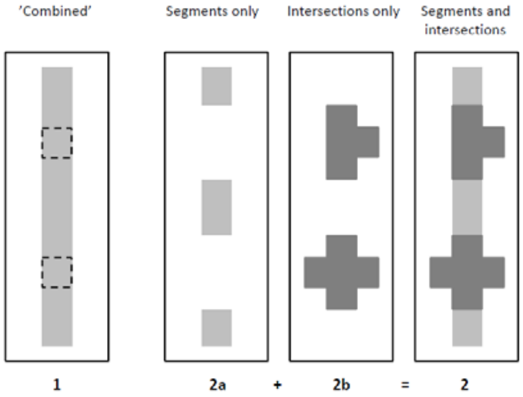

Fig. 1 - Visualization of principle of combined SPF, where intersections are considered only in terms of their frequency (1); segment-only SPF (2a); and intersection SPF (2b)

Tab. 1 - Descriptive characteristics of accident data

	Min.	Max.	Mean	SD
$N_{combined}$	2	25	9.71	7.07
$N_{segment}$	1	21	8.24	6.08
$N_{intersection}$	0	7	1.80	2.07

Collected short-term counts were factored up according to national guidelines [26] in order to obtain the AADT. Other variables were gathered from Road and Motorway Directorate databases or on-line map sources. Table 2 lists descriptive characteristics of all explanatory variables.

In order to compare the samples with the original UK study [9], two comparable characteristics were selected. The UK study used road categories A-single and A-dual (urban and rural); only the rural were used for further comparisons:

- Mean intersection density (1.6 in Zlín sample) was approx. 1.2 for A-roads.
- Ratio of intersection accidents to all accidents (15% in Zlín sample) was approx. 24% for A-roads.

Both characteristics are roughly comparable, which allows using them for the following comparisons. Inter-correlations were checked and generally found low, therefore all explanatory variables were used. SPFs were built using a generalized linear modeling procedure in IBM SPSS, applying a negative binomial error structure with the logarithmic link function; explanatory variables with a power form in Equations. (1)–(3) thus took the form of natural logarithms. Parameters of resulting SPFs are reported in Table 3.

Tab. 2 - Descriptive characteristics of explanatory variables

			Min.	Max.	Mean
Continuous variables	Segment AADT [veh/day]		1,398	15,041	7,325
	Segment length [km]		0.23	9.81	3.11
	Intersection density [km^{-1}]		0.24	4.41	1.61
	Major road AADT [veh/day]		1,398	15,041	8,026
	Minor road AADT [veh/day]		27	4,131	887
				Freq.	%
Categorical variables	Number of legs	= 3		49	87.5
		= 4		7	12.5
	Turn lanes	= Yes		23	41.1
		= No		33	58.9

Tab. 3 - Parameters of safety performance functions (SPFs)

	Parameters	B	SE	Sig.
SPF 1 (combined)	(Intercept)	-3.047	1.076	0.005
	ln_AADT	0.448	0.114	0.000
	ln_Length	0.991	0.157	0.000
	Intersection density	0.207	0.104	0.046
	(Overdispersion)	0.025		
SPF 2a (segments)	(Intercept)	-2.512	1.227	0.041
	ln_AADT	0.417	0.134	0.002
	ln_Length	0.887	0.169	0.000
	(Overdispersion)	0.062		
SPF 2b (intersections)	(Intercept)	-8.557	2.475	0.001
	Legs = 3	-0.656	0.319	0.039
	Legs = 4	0		
	ln_F1	1.176	0.274	0.000
	ln_FR	0.344	0.144	0.017
	(Overdispersion)	0.266		

Note: B – regression parameters, SE – standard errors, Sig. – level of statistical significance. Parameter of categorical variable Legs is to be compared to a reference category (Legs = 4).

All variables had systematic influence at the level of statistical significance ≤ 5% (0.05). The signs of regression coefficients confirm expectations: AADTs and lengths are positive, i.e. their increase is associated with increasing accident frequency; the same holds for intersection density. The number of intersection legs has a negative coefficient suggesting that 3-leg intersections have lower accident occurrence compared to 4-leg ones. The effect of turn lanes was found insignificant.

Regression coefficients may also be compared to the values in the original UK study. However, note that road categories may not be fully compatible, since UK A-roads were described as roads of national or regional importance, i.e. a potential mixture of national roads and regional roads [9]. For the purpose of a comparison, A-roads values (averages of values for A-single and A-dual roads), were considered alternative to national. Regression coefficients of explanatory variables AADT, length, intersection density were compared, i.e. b_1, b_2, b_3 from Equations. (1) and (2). The values are reported in Table 4 and visualized in Figure 2: the magnitudes are relatively comparable.

Using the developed SPFs, mean accident frequency predictions (m) were obtained (m_1 and m_{2a} for each segment, m_{2b} for each intersection). However, as mentioned before, the objective was to assess the performance of the proposed simplified approach in terms of network screening. For this purpose, mean predictions were further adjusted according to empirical Bayes (EB) methodology (for more information, see [27]), which combines predicted and reported accident frequencies:

$$EB_i = w_i \cdot P_i + (1 - w_i) \cdot R_i \tag{4}$$

$$w_i = \frac{k_i}{k_i + P_i} \tag{5}$$

where EB_i are EB estimates computed using the weighted average (with weights w_i) of predicted and reported accident frequencies (P_i and R_i). The obtained EB estimates for two segment SPFs (1 and 2a) were compared in terms of consistency – i.e. an overlap between the segment list rankings based on SPFs 1 and 2 (for more information on consistency tests, see i.e. [28-31]).

Tab. 4 - Regression coefficients of described safety performance functions (SPFs)

	Combined SPFs			*Segment-only SPFs*	
	b_1	b_2	b_3	b_1	b_2
Zlín (national roads)	0.448	0.991	0.207	0.417	0.887
UK (A-roads)	0.614	0.986	0.104	0.644	0.957

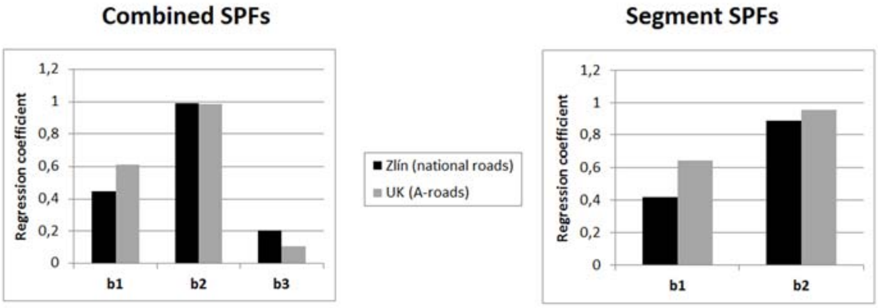

Fig. 2 - Graphical comparison of regression coefficients of described safety performance functions (SPFs)

3. Results and discussion

EB Estimates EB_{2a} and EB_{2b} were summed together and the results (EB_2) were compared to EB estimates EB_1. Their relationship was found very close and positive, with statistically significant Pearson correlation coefficient and coefficient of determination R^2 = 0.955.

Consistently with the original UK study, the developed simplified SPF seems to perform sufficiently in terms of both goodness-of-fit and correlation with traditional SPF. In addition, the method consistency was assessed for top segments with the highest EB estimates. The selected 10 segments in list 1 (segments ranked according to combined SPF 1) overlapped with 9 segments in list 2a (segments ranked according to segment-only SPF 2a), which equals to 90% consistency.

While the results seem encouraging, there is a limiting small size of the used sample. However, the 3800-km UK sample from the original study is incomparable with conditions in the Czech regions, which include on average only several hundreds of kilometres. Nevertheless, these reductions reflect real conditions of small road networks and the study is thus illustrative. While the sample extension may be beneficial, it would need to rely on data from other conditions (road categories, intersection categories, regions, etc.), which could reduce the sample homogeneity. Future studies may verify the approach in different conditions, for example on regional or urban roads.

4. Conclusions

For state-of-the-art network screening, safety performance functions (SPFs) are needed. However, their practical development is demanding, since it requires having AADT data for all evaluated units, i.e. road segments and intersections. In addition, the output, as well as the whole screening process, is separated into segments and intersections, which may not be the most practical from the road agency perspective. Using the number of intersections per segment length (i.e. intersection density) is a potential simplification, which would allow omitting data collection for intersection models – however, its performance has rarely been tested.

The aim of the paper was to follow the example of the original UK study [9] and verify the applicability of this approach in Czech conditions. The method was found feasible – estimates from SPFs with intersection density were very close to estimates based on a combination of segment and intersection SPFs; goodness-of-fit of simplified SPFs even improved, and consistency, in terms of an overlap between two rankings of the final segment lists, was also satisfactory.

Based on these results, the approach based on intersection density seems promising. It allows performing the network screening without having to conduct additional traffic surveys to complement the missing AADT data on intersections. This simplification will increase efficiency of network screening and allow a wider practical application for Czech regional agencies. These will in turn provide material for study extensions in the future with enlarged sample sizes.

Acknowledgments

The authors appreciate data collection and post-processing work of their colleagues Vladimír Dufek, Ondřej Gogolín, Martin Kyselý, Stanislav Řehák, Richard Turek and Lucie Vyskočilová. The study was conducted with the financial support of Czech Ministry of Education, Youth and Sports under the National Sustainability Programme I project of Transport R&D Centre (CZ.1.05/2.1.00/03.0064), using the research infrastructure from the Operation Programme Research and Development for Innovations.

References

1. Hancock, K. L., Zhang, W., Sardar, H., Wang, Y. (2016). A two-step geo-grid screening process to identify locations for safety improvements. *Advances in Transportation Studies*, 2, pp. 89-102.
2. Dissanayake, S., Ratnayake, I. (2006). Statistical modelling of crash frequency on rural freeways and two-lane highways using negative binomial distribution. *Advances in Transportation Studies*, 9, pp. 81-96.
3. Elvik, R. (2008). A survey of operational definitions of hazardous road locations in some European countries. *Accident Analysis & Prevention*, 40, pp. 1830-1835.
4. Hauer, E., Kononov, J., Allery, B., Griffith, M. S. (2002). Screening the Road Network for Sites with Promise. *Transportation Research Record*, 1784, pp. 27-32.
5. *Highway Safety Manual. First Edition*. AASHTO, Washington, D.C., 2010.
6. *Crash Estimation Compendium. First edition*. NZ Transport Agency, Wellington, 2016.
7. Srinivasan, R., Bauer, K. (2013). *Safety Performance Function Development Guide: Developing Jurisdiction-Specific SPFs*. Report FHWA-SA-14-005. FHWA, Washington, D.C.
8. Vogt, A., Bared, J. (1998). Accident Models for Two-Lane Rural Segments and Intersections. *Transportation Research Record*, 1635, pp. 18-29.
9. Mountain, L., Fawaz, B., Jarrett, D. (1996). Accident prediction models for roads with minor junctions. *Accident Analysis & Prevention*, 28, pp. 695-707.
10. Sørensen, M., Elvik, R. (2007). *Black Spot Management and Safety Analysis of Road Networks – Best Practice Guidelines and Implementation Steps*. Report 919/2007. Institute of Transport Economics, Oslo.
11. Høye, A. (2016). *Development of crash prediction models for national and county roads in Norway (2010-2015).* Report 1522/2016. Institute of Transport Economics, Oslo.
12. Barbosa, H., Cunto, F., Bezerra, B., Nodari, C., Jacques, M. A. (2014). Safety performance models for urban intersections in Brazil. *Accident Analysis & Prevention*, 70, pp. 258-266.
13. Garach, L., de Oña, J., López, G., Baena, L. (2016). Development of safety performance functions for Spanish two-lane rural highways on flat terrain. *Accident Analysis & Prevention*, 95, pp. 250-265.
14. Hadi, M. A., Aruldhas, J., Chow, L.-F., Wattleworth, J. A. (1995). Estimating safety effects of cross-section design for various highway types using negative binomial regression. *Transportation Research Record*, 1500, pp. 169-177.
15. Reurings, M., Janssen, T. (2007). *Accident prediction models for urban and rural carriageways*. Report R-2006-14. SWOV Institute for Road Safety Research, Leidschendam.
16. Sawalha, Z., Sayed, T. (2006). Traffic accident modeling: some statistical issues. *Canadian Journal of Civil Engineering*, 33, pp. 1115-1124.
17. Vieira Gomes, S., Geedipally, S. R., Lord, D. (2012). Estimating the safety performance of urban intersections in Lisbon, Portugal. *Safety Science*, 50, pp. 1732-1739.
18. Walmsley, D. A., Summersgill, I., Binch, C. (1998). *Accidents on modern rural single-carriageway trunk roads*. TRL Report 336. Transport Research Laboratory, Crowthorne.
19. Ambros, J., Havránek, P., Valentová, V., Křivánková, Z., Striegler, R. (2016). Identification of Hazardous Locations in Regional Road Network – Comparison of Reactive and Proactive Approaches. *Transportation Research Procedia*, 14, pp. 4209-4217.
20. Ambros, J., Sedoník, J. (2016). A Feasibility Study for Developing a Transferable Accident Prediction Model for Czech Regions. *Transportation Research Procedia*, 14, pp. 2054-2063.
21. 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, pp. 57-66.
22. Kim, D., Washington, S., Oh, J. (2006). Modeling Crash Types: New Insights into the Effects of Covariates on Crashes at Rural Intersections. *Journal of Transporation Engineering ASCE*, 132, pp. 282-292.
23. Mitra, S., Washington, S. (2012). On the significance of omitted variables in intersection crash modeling. *Accident Analysis & Prevention*, 49, pp. 439-448.
24. 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, pp. 41-49.

25. Avelar, R. E., Dixon, K. K., Escobar, P. (2015). Evaluation of Signalized-Intersection Crash Screening Methods based on Distance from Intersection. *Transportation Research Record*, 2514, pp. 177-186.
26. Bartoš, L., Martolos, J. (2012). *Stanovení intenzit dopravy na pozemních komunikacích (II. vydání).* Technical guidelines 189. EDIP s.r.o., Plzeň.
27. 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, pp. 126-131.
28. Ambros, J., Valentová, V., Sedoník, J. (2016). Developing Updatable Crash Prediction Model for Network Screening: Case Study of Czech Two-Lane Rural Road Segments. *Transportation Research Record*, 2583, pp. 1-7.
29. Manepalli, U. R. R., Bham, G. H., 2016. An evaluation of performance measures for hotspot identification. *Journal of Transportation Safety & Security*, 8, pp. 327-345.
30. Montella, A. (2010). A comparative analysis of hotspot identification methods. *Accident Analysis & Prevention*, 42, pp. 571-581.
31. Yu, H., Liu, P., Chen, J., Wang, H. (2014). Comparative analysis of the spatial analysis methods for hotspot identification. *Accident Analysis & Prevention*, 66, pp. 80-88.