PROACTIVE IDENTIFICATION OF RISK ROAD LOCATIONS USING VEHICLE FLEET DATA: EXPLORATORY STUDY

JIŘÍ AMBROS, ONDŘEJ GOGOLÍN, JAN KUBEČEK, RICHARD ANDRÁŠIK, MICHAL BÍL

{jiri.ambros, ondrej.gogolin, jan.kubecek, richard.andrasik, michal.bil}@cdv.cz

CDV – Transport Research Centre, Líšeňská 33a, 636 00 Brno, Czech Republic

ABSTRACT

Identification of risk locations within road network is the primary task of its safety management. In this regards traffic accidents have been used as a traditional indicator; however it has also been known that accidents provide retrospective view only, their occurrence is random and their collection is time consuming. On the other hand, various alternative sources which can be used in the process of identification of risk locations have appeared recently, including vehicle fleet data (floating car data). Such data may be used for proactive safety evaluation and identification of risk locations, supposing that a close relationship between speed and accidents exists.

In the presented study, position data (series of GPS points representing driving of more than 1000 vehicles) are being collected continually on Czech national rural road network. Registered positions enable calculation of speed and derivation of other indicators. The road network data were analyzed in order to identify tangents and curves. The analysis focused on investigation of relation between accident frequency and speed derived from fleet data, using consistency measures (transitions between two successive road segments and degree of related speed or curvature changes).

The potential results may contribute to proactive rural safety monitoring, evaluation and ranking. Nevertheless the entire procedure involves several important decisions which influence quality of collected data and its following post-processing and interpretation (how to obtain free-flow speed, how to process GPS locations and trajectories, how to conduct road alignment segmentation, how to link consistency measures with objective safety, etc.). The study explores these issues and proves practical feasibility of the procedure. The first results from the study are presented and discussed, as well as their applicability and implications for further research.

Keywords: road safety, risk location, proactive identification, vehicle fleet, exploratory study

1. INTRODUCTION

Identification of risk locations is the primary task of road network safety management (AASHTO, 2010). In this regards traffic accidents have been used as a traditional indicator; however it has also been known that accidents provide retrospective view only, their occurrence is random and their collection is time consuming. On the other hand, various alternative sources which can be used in the process of identification of risk locations have appeared recently, including vehicle fleet data (floating car data) (Bekhor et al., 2013; Riguelle, 2015; Mousavi et al., 2015; Agerholm et al., 2015).

Collection and processing of floating car data (FCD) enables determination of operating speed. According to the TRB synthesis (TRB, 2011), the assessment of operating speeds enables to assess the expected speed changes of individual vehicles over successive road segments (tangents and curves). Inconsistency of operating speeds is regarded one of the symptoms that violates driver's expectation. Such violations may then develop into accidents, mainly in curves. In order to quantify the related unsafety, differences between speeds or curvature change rates (between tangent and curve) have often been used (Lamm et al., 2002; Hassan, 2004; Cafiso et al., 2010). Should it be possible to derive these indicators from vehicle fleet data, the results may be used for proactive safety evaluation and identification of risk locations.

In the presented study, position data (series of GPS points representing driving of more than 1000 vehicles) are being collected continually on national rural road network. Registered positions (4 times per second) enable calculation of speed and derivation of other indicators. The road network data were analyzed in order to identify straight road sections (tangents) and curves. The analysis focused on investigation of relation between accident frequency and speed derived from fleet data, using consistency measures (transitions between two successive road segments and degree of related speed or curvature changes).

The potential results may contribute to proactive rural safety monitoring, evaluation and ranking. Nevertheless the entire procedure involves several important decisions which influence quality of collected data and its following post-processing and interpretation (how to obtain free-flow speed, how to process GPS locations and trajectories, how to conduct road alignment segmentation, how to link consistency measures with objective safety, etc.). The study explores these issues and proves practical feasibility of the procedure. The first results from the study are presented and discussed, as well as their applicability and implications for further research.

2. DATA AND METHOD

Czech private company Princip a.s. is collecting FCD for the purposes of vehicle fleet management. The vehicle fleets are monitored through Vetronics electronic units with GPS receiver (for details see http://www.princip.cz/en/products/vetronics/).

Regarding the presented safety-related study, the focus was set on the most critical settings of Czech road network. This applies, according to the previous analyses (Ambros et al.,

2015), to rural parts of national roads. Therefore we purchased FCD collected on this subset from Princip a.s. and used them in the presented analysis.

The dataset consisted of GPS data points from 1172 company vehicles, collected between October 2014 to May 2015 (i.e. in 8 months). GPS position was registered 4 times per second. In the following text, the individual processing steps are described.

2.1 Selection of road sections

In order to obtain representative information, data from more drives in the selected sections are needed. Since no universal guidance exists, experience was sought in the previous studies. For example TRB synthesis (TRB, 2011: Table A-2) reports the parameters of operating speed studies, including the number of observations; a large number of them used the criterion of "at least 100 per site". We applied the same criterion in this study (in both driving directions), leading to approx. 50% reduction of number of sections.

2.2 Segmentation

Given the study focus, it was necessary to divide the current dataset (including both sections and intersections) into tangents (straight sections) and curves. In this regards various authors have used different approaches for obtaining alignment parameters from GPS data (Biagioni and Eriksson, 2012; Garach et al., 2014; Li et al., 2015). Nevertheless each method has its disadvantages such as limited accuracy or dependency on manual processing. Often a combination of manual and automatic identification is used; a universal automated method of road alignment inference does not exist. We applied our own segmentation methodology (Andrášik and Bíl, in review), consisting of the following steps:
1. pre-processing with Douglas-Peucker algorithm for data generalization
2. calculation of explanatory variables (e.g. an angle between three consecutive points, radius of a circumscribed circle)
3. discriminant analysis
4. post-processing with least squares method for radii computation (with empirical threshold of radius above 2100 m indicating a tangent)

This way we created tangents and curves. In the next step, we wanted to make sure that we use only the segments with sufficient length, in terms of spatial characteristics and location of accidents. "Highway Safety Manual" (AASHTO, 2010) recommends using minimum segment length 0.1 mi – we rounded this value to the minimum length 200 m and discarded all the shorter tangents and curves. For all these segments, curvature change rate (CCR, i.e. the sum of angular changes divided by the segment length), was also calculated for the further analyses.

2.3 Determination of speed

Speed was calculated from the GPS location and time interval between the points, given by the recording frequency. This way we assigned speed to each drive and data point.

For the analyses of relationships between speed, road geometry and traffic safety, it is necessary to use the "free-flow speed" (FFS), i.e. the speed which is not constrained by congestion, traffic devices or adverse weather conditions. Traditional approach to estimating

FFS relies on field studies, where speeds of isolated ("uninfluenced") vehicles are measured manually, for example by laser speed guns (TRB, 2011). However, with area-wide collected floating car data, another approach is necessary. Typically data from off-peak hours are believed to represent FFS (Wang et al., 2006; Bekhor et al., 2013; Riguelle, 2015). Since company vehicles often travel during peak hours, this approach could lead to enormous data loss; therefore we aimed to apply another method, which consists of the following steps:

1. For each point, we divided speed values into two groups (influenced/uninfluenced speed) based on cluster analysis (k-means method, with k = 2).
2. We calculated V_{85} as the 85th percentile of uninfluenced speed.
3. We used weighted average of V_{85} per segment, with weight given by number of drives through each segment (tangent or curve).

For each segment, we calculated two indicators of consistency:

- speed consistency, i.e. the relative speed difference between tangent and curve,
- alignment consistency, i.e. the relative CCR difference between tangent and curve,

where the values are subtracted as $consistency_{curve} - consistency_{segment}$. The process of obtaining differences (of both speed and CCR) is visualized in Fig. 1. In direction A, the difference $2A - 1A$ is calculated; in direction B, the difference is $2B - 3B$. The stars indicate the accident potential related to the mentioned differences.
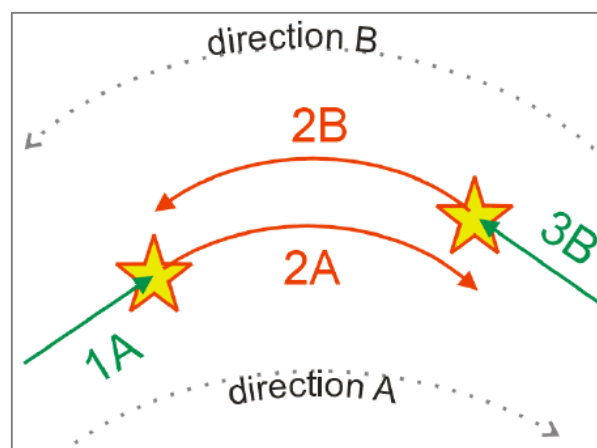

Fig. 1 Visualization of calculated differences between curve (in red) and tangents (in green).

In order to test the contribution of the whole process, the following analysis step was aimed at validation of the relationship between consistency and safety.

2.4 Validation

The original idea is that curve accident frequency is related to a difference of speeds (or curvature change rate) between tangent and curve, i.e. speed consistency (or alignment consistency) (Lamm et al., 1999). The validation therefore aimed at comparison between the safety level, based on speed consistency and/or alignment consistency, and objective safety level of each curve. For objective safety, we used frequency of accidents, which were recorded by the Czech Traffic Police in the 6-year period (2009 – 2014). Consistently with other studies (Dietze and Weller, 2011; Turner et al., 2012; Ambros et al., 2016), we selected only single-vehicle accidents, since they are more clearly linked to road geometry and speed.

In the next step we applied accident prediction model (safety performance function) model in the following form:

$$ACC = \exp(\beta_0) \cdot AADT^{\beta_1} \cdot L^{\beta_2} \cdot \exp(\beta_3 \cdot CCR)$$

where:
- ACC ... expected (predicted) number of single-vehicle accidents
- $AADT$ ... annual average daily traffic from the national traffic census 2010 [veh/day]
- L ... curve length [km]
- CCR ... curvature change rate [gon/km]
- β_i ... regression parameters to be estimated

The model was used according to the empirical Bayes methodology (Hauer et al., 2002), which combines both predicted (expected) and recorded accident frequency. Weighted average is used in this combination, with weight representing the model quality. The results are referred to as empirical Bayes estimates of expected accident frequency (in short "EB estimates") (Hauer et al., 2002).

When a surrogate safety measure is to be used for risk evaluation, it should comply with two general objectives (Cafiso et al., 2011): it must produce a safety evaluation correlated with accident history, and it must consistently rank safety problems. Another validation approach is comparison of safety levels within the empirically set categories, as used for example when comparing crash rates and star ratings according to road assessment programmes (Harwood et al., 2010; Lawson, 2011; McInerney and Fletcher, 2013).

3. RESULTS

According to the previous paragraph, there are three possible validation test approaches: correlation, ranking consistency, and category comparison. While all of them were tested, we were successful only with the last one.

Categories were defined according to the empirical thresholds, which have been often used in the literature, based on the works of Lamm et al. (1999, 2002, 2007):
- speed differences 10 and 20 km/h
- CCR differences 180 and 360 gon/km

In these categories, average of EB estimates was calculated – the results are presented in Fig. 2.
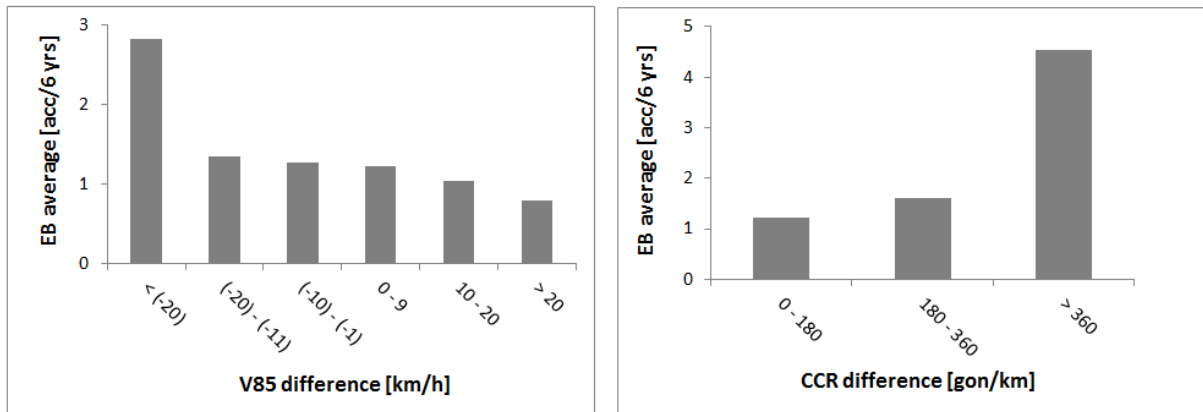
Fig. 2 Comparison of speed and alignment consistency level categories with average values of empirical Bayes estimates of expected accident frequency

The graphs show that objective safety (in terms of EB estimates):
- is increasing (there is less accidents) with increasing speed consistency (when there are smaller speed differences)
- is decreasing (there are more accidents) with decreasing alignment consistency (when there are larger CCR differences)

Both trends are consistent with general knowledge: low consistency may lead to risky situations when entering the curve, where drivers react with more pronounced braking.

4. DISCUSSION AND CONCLUSIONS

Both speed consistency and alignment consistency were able to classify the curves according to objective safety. This finding shows the potential of using the data obtained from floating vehicles as an alternative method of risk location identification. Such approach is proactive, which means that it may be applied even before occurrence of accidents.

The exploratory study, testing the applicability of the presented approach, proved successful – the proposed method is feasible and results are valid. Nevertheless there are several limitations – mainly due to the fact that most of the analysis points (for example choosing necessary sample size, estimating free-flow speed, or applying method of segmentation) are lacking guidance in literature. At the same time these decisions have significant influence on quality and precision of the final results. It may be the reason why the more detailed validation approaches (such as comparison of location rankings) were not successful.

To this end, improvement of the procedures is in progress. The revision will for example include testing segmentation options (aggregating several short segments into longer ones) or calculating consistency measures from a number of profiles within a curve (instead of averaging them per curve), as recommended by TRB (2011).

ACKNOWLEDGEMENTS

REFERENCES

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

Agerholm, N., Jensen, R., Andersen, C.S. (2015). Identification of Hazardous Road Locations on the basis of jerks. In *28th ICTCT Workshop*, Ashdod, Israel.

Ambros, J., Valentová, V., Oríšek, P. (2015). Safety Ranking of Rural Curves Based on Design Consistency Measures. In *5th International Symposium on Highway Geometric Design*, Vancouver, B.C., Canada.

Ambros, J., Valentová, V., Sedoník, J. (2016). Linking Rural Road Environment, Speed and Safety Factors with a 'Two-Stage' Model: A Feasibility Study. In *TRB 95th Annual Meeting*, Washington, D.C.

Andrášik, R., Bíl, M. (in review). Efficient Road Geometry Identification from Digital Vector Data. *Journal of Geographical Systems*.

Bekhor, S., Lotan, T., Gitelman, V., Morik, S. (2013). Free-Flow Travel Speed Analysis and Monitoring at the National Level Using Global Positioning System Measurements. *Journal of Transportation Engineering*, Vol. 139, pp. 1235–1243.

Biagioni, J., Eriksson, J. (2012). Inferring Road Maps from Global Positioning System Traces Survey and Comparative Evaluation. *Transportation Research Record*, No. 2291, pp. 61–71.

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*, Vol. 42, pp. 1072–1079.

Cafiso, S., La Cava, G., Montella, A. (2011). Safety Inspections as Supporting Tool for Safety Management of Low-Volume Roads. *Transportation Research Record*, No. 2203, pp. 116–125.

Dietze, M., Weller, G. (2011). *Applying speed prediction model to define road sections and develop accident prediction models: A German case study and a Portuguese exploratory study*. RISMET project deliverable 6.2.

Garach, L., de Oña, J., Pasadas, M. (2014). Mathematical formulation and preliminary testing of a spline approximation algorithm for the extraction of road alignments. *Automation in Construction*, Vol. 47, pp. 1–9

Harwood, D.W., Bauer, K.M., Gilmore, D.K., Souleyrette, R., Hans, Z.N. (2010). Validation of U.S. Road Assessment Program Star Rating Protocol: Application to Safety Management of U.S. Roads. *Transportation Research Record*, No. 2147, pp. 33–41.

Hassan, Y. (2004). Highway Design Consistency: Refining the State of Knowledge and Practice. *Transportation Research Record*, No. 1881, pp. 63–71.

Hauer, E., Harwood, D.W., Council, F.M., Griffith, M.S. (2002). Estimating Safety by the Empirical Bayes Method – A Tutorial. *Transportation Research Record*, No. 1784, pp. 126–131.

Lamm, R., Psarianos, B., Mailänder, T. (1999). *Highway Design and Traffic Safety Engineering Handbook*. McGraw-Hill, New York.

Lamm, R., Psarianos, B., Cafiso, S. (2002). Safety Evaluation Process for Two-Lane Rural Roads: A 10-Year Review. *Transportation Research Record*, No. 1796, pp. 51–59.

Lamm, R., Beck, A., Ruscher, T., Mailänder, T., Cafiso, S., La Cava, G. (2007). *How To Make Two-Lane Rural Roads Safer: Scientific Background and Guide for Practical Application*. WIT Press, Southampton, UK.

Lawson, S. (2011). *Crash rate – Star Rating comparisons: Review of available evidence*. Working Paper 504.2. EuroRAP, Basingstoke, UK.

Li, Z., Chitturi, M.V., Bill, A.R., Zheng, D., Noyce, D.A. (2015). Automated Extraction of Horizontal Curve Information for Low-Volume Roads. *Transportation Research Record*, No. 2472, pp. 172–184.

McInerney, R., Fletcher, M. (2013). *Relationship between Star Ratings and crash cost per kilometre travelled: the Bruce Highway, Australia*. Working Paper 504.20. iRAP, Basingstoke, UK.

Mousavi, S.-M., Parr, S.A., Pande, A., Wolshon, B. (2015). Identifying High-Risk Roadways Through Jerk-Cluster Analysis. In *2015 Road Safety & Simulation International Conference*, Orlando, FL, pp. 1139–1147.

Riguelle, F. (2015). Can Floating Car Data be used to monitor the speed behaviour? Results of a pilot study in Belgium. In *International Symposium on Road Safety Behaviour Measurements and Indicators*, Brussels, Belgium.

TRB (Transportation Research Board) (2011). *Modeling Operating Speed: Synthesis Report*. TR Circular E-C151. TRB, Washington, D.C.

Turner, S., Singh, R., Nates, G. (2012). *The next generation of rural road crash prediction models: final report*. Research report 509. NZ Transport Agency, Wellington, New Zealand.

Wang, J., Dixon, K., Li, H., Hunter, M. (2006). Operating-Speed Model for Low-Speed Urban Tangent Streets Based on In-Vehicle Global Positioning System Data. *Transportation Research Record*, No. 1961, pp. 24–33.