

Testing of 'application-based' goodness-of-fit measure of crash prediction model

Jiří Ambros



Harri Peltola



06/05/2015

4th NTSA Annual Scientific Seminar

VTT, Espoo, Finland

Crash prediction models (1/3)

- or 'safety performance functions'
- Crash annual frequency = function of traffic volume (AADT) and other risk factors (explanatory variables)
- Objective picture of safety performance of a set of intersections or sections
- Tools of rational road network safety management

Contact Information:

Jiří Ambros, CDV (Czech Republic), jiri.ambros@cdv.cz

Harri Peltola, VTT (Finland), Harri.Peltola@vtt.fi

Crash prediction models (2/3)

Which (and how many) variables to include?

$$\hat{y} = \left[\text{segment length for prediction} \times (X^{1.860} e^{-0.159X}) \times [2e^{-0.071 \times (\% \text{trucks})} + 0.037(\% \text{trucks})] \right] \quad \begin{matrix} \text{(Hauer et al., 2004)} \\ \text{(Vogt \& Bared, 1998)} \end{matrix}$$

$$E(m)_i = \left(\frac{ADTL}{1000} \right)^{1.242} \times LEN^{0.696} \times e^{(0.1955 \times LN - 0.1775 \times SHW + 0.2716 \times MT2 + 0.5669 \times TS - 0.1208 \times PTC - 0.0918 \times Y91)}$$

$$y = \alpha_{\text{year}} \times (0.803 \text{ if TWLTL present and } 1 \text{ otherwise}) \times [1 + 0.022 \cdot (\text{lane width} - 11 \text{ ft})] \quad \text{(Dell'Acqua \& Russo, 2010)}$$

$$Y_{ik} | \text{CHAR} = v_i \cdot \exp \left(\begin{matrix} \beta_{0,ik} + \beta_{1,ik} \cdot \ln(AADT_i) + \beta_{2,ik} \cdot \ln(HGV_i) + \beta_{3,ik} \cdot \text{BEND}_i + \dots \\ \beta_{4,ik} \cdot \text{SLP}_i^2 + \beta_{5,ik} \cdot \text{LAN}_i + \beta_{6,ik} \cdot \text{VEL}_i^2 + \beta_{7,ik} \cdot \text{EML}_i + \varepsilon_k \end{matrix} \right)$$

$$\begin{matrix} \times (1.26 \text{ if PC, } 1.75 \text{ if VS, } 1 \text{ otherwise}) \\ \times e^{-1.108(\text{segment length in model data})} \\ + 0.027 \times (\text{number of commercial driveways}) \end{matrix} \quad \text{(Deublein et al., 2013)}$$

Crash prediction models (2/3)

- Road section exposure: AADT and length
- Exposure only ... **simple** models
- Exposure + other variables ... **multivariate** models
- Area-wide availability? Explanatory power?
- **Are simple models good enough?**
- Quality measured by goodness-of-fit (GOF)

Contact Information:

Jiří Ambros, CDV (Czech Republic), jiri.ambros@cdv.cz

Harri Peltola, VTT (Finland), Harri.Peltola@vtt.fi

GOF measures

- ***Model quality***

- e.g. Akaike Information Criterion (AIC)
- no right/wrong, comparison only

- ***Model performance***

- e.g. 'degree of explanation'
- shows amount of systematic variation explained by model

Contact Information:

Jiří Ambros, CDV (Czech Republic), jiri.ambros@cdv.cz

Harri Peltola, VTT (Finland), Harri.Peltola@vtt.fi

Practical perspective

- Crash prediction models = tools
- Identification of hazardous road locations (network screening) is the most common practical application
- How to assess quality of a model for network screening?
- ***Do simple models help to identify the same locations as multivariate models?***



Contact Information:

Jiří Ambros, CDV (Czech Republic), jiri.ambros@cdv.cz

Harri Peltola, VTT (Finland), Harri.Peltola@vtt.fi

The study

- Secondary road sections (regional rural roads, paved, 2 lanes, undivided)
- Czech Republic (1 region) and Finland (whole)

	 CZ	 FI
Approx. total length [km]	1000	11 000
Approx. average AADT [veh/day]	2500	1400
Typical road width [m]	7.5	6.5
Typical speed limit [km/h]	90	80
Approx. crash density [km ⁻¹ ·year ⁻¹]	0.3	0.03



Contact Information:

Jiří Ambros, CDV (Czech Republic), jiri.ambros@cdv.cz

Harri Peltola, VTT (Finland), Harri.Peltola@vtt.fi

Model variables

- Dependent variable: 5-year frequency of injury crashes (CZ: 2010-2014, FI: 2007-2011)
- Explanatory variables:

MULTIVARIATE	SIMPLE	Czech Republic 		Finland 	
		AADT length curvature	AADT mileage (12 groups according to speed limit, width and population)	shoulder (yes/no) forest (yes/no) pavement quality	share of heavy goods vehicles region (north/south)



Contact Information:

Jiří Ambros, CDV (Czech Republic), jiri.ambros@cdv.cz

Harri Peltola, VTT (Finland), Harri.Peltola@vtt.fi

Model GOF



		
Model form	$a \cdot AADT^b \cdot \exp(c \cdot L + d \cdot C + \dots)$	$a \cdot M \cdot AADT^b \cdot \exp(c \cdot HGV + \dots)$
AIC	multivariate 5531 / simple 5554	multivariate 6052 / simple 6001
	increase by 23	increase by 51
Degree of explanation	multivariate 50% / simple 49%	multivariate 64% / simple 61%
	decrease by 1%	decrease by 3%

- AIC preference threshold ~ 10
- no guidance for degree of explanation ... both OK?
- but what is the effect on network screening?

Contact Information:

Jiří Ambros, CDV (Czech Republic), jiri.ambros@cdv.cz

Harri Peltola, VTT (Finland), Harri.Peltola@vtt.fi

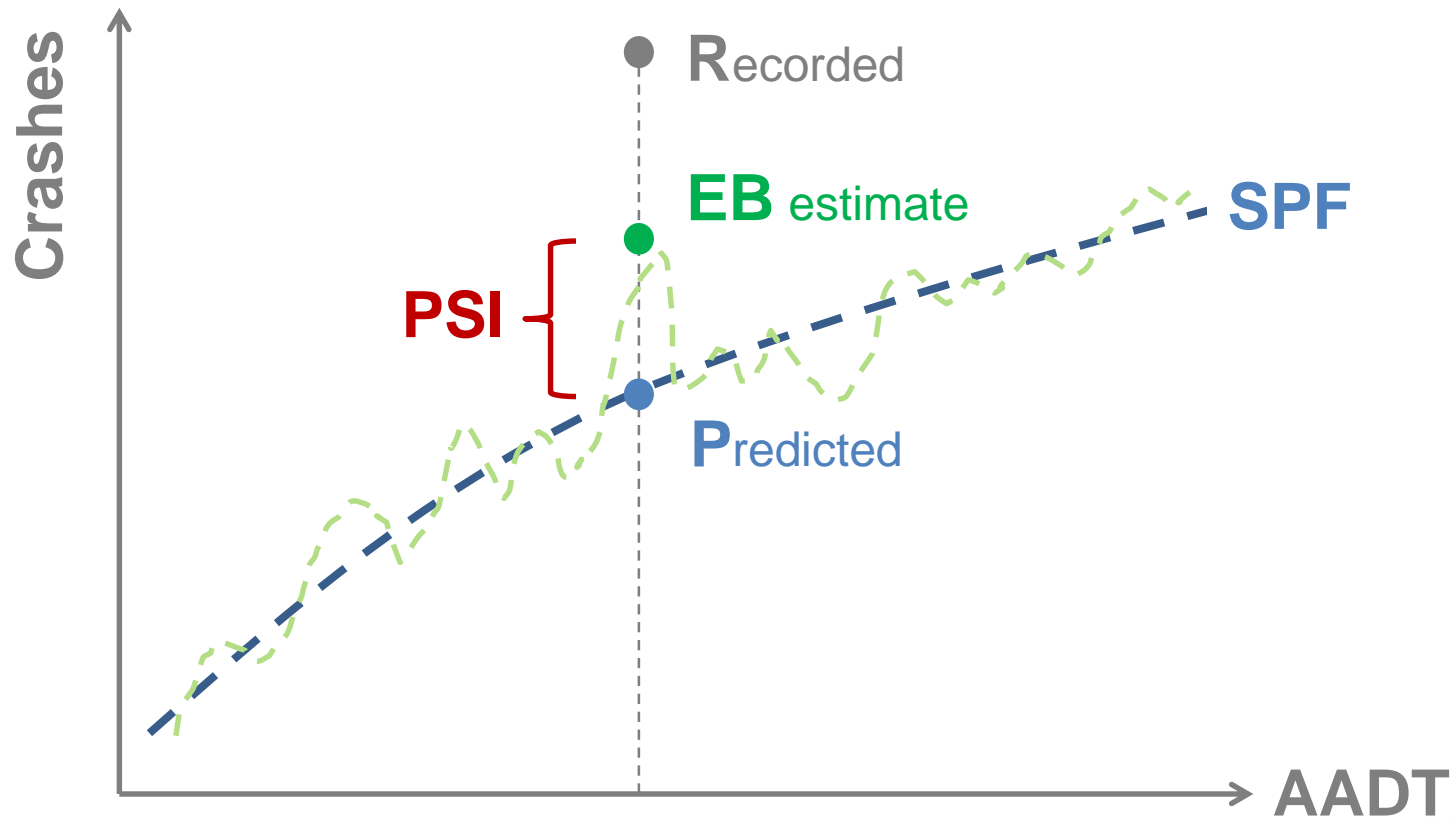
‘Application-based’ GOF

- How do results based on simple and multivariate models correlate?
- Sections ranked according to ‘potential for safety improvement’ (PSI) → 2 lists
- GOF = agreement in their % overlap
- 3 options tested: top 1%, top 2.5%, top 5%

Contact Information:

Jiří Ambros, CDV (Czech Republic), jiri.ambros@cdv.cz

Harri Peltola, VTT (Finland), Harri.Peltola@vtt.fi



$$EB = w \cdot P + (1 - w) \cdot R$$



$$PSI = EB - P$$

Contact Information:

Jiří Ambros, CDV (Czech Republic), jiri.ambros@cdv.cz

Harri Peltola, VTT (Finland), Harri.Peltola@vtt.fi

Results

	CZ 		FI 	
Top 1%	38 sections	93%	57 sections	70%
Top 2.5%	94 sections	92%	142 sections	70%
Top 5%	188 sections	95%	284 sections	78%

- Acceptable values?
- Cf. overlaps between time periods (EB, top 5%)



Cheng & Washington, 2008	47.3%
Montella, 2010	46.9%
Yu et al., 2014	46.2%

Contact Information:

Jiří Ambros, CDV (Czech Republic), jiri.ambros@cdv.cz

Harri Peltola, VTT (Finland), Harri.Peltola@vtt.fi

Discussion: potential differences

		
Homogeneity	one region	whole country
Segmentation	short sections	long sections
Level of safety	lower level more systematic variation less random variation (crash variance 0.8)	higher level less systematic variation more random variation (crash variance 0.4)

Contact Information:

Jiří Ambros, CDV (Czech Republic), jiri.ambros@cdv.cz

Harri Peltola, VTT (Finland), Harri.Peltola@vtt.fi

Conclusions

- Overall: similarity between results
- ***Simple models should be sufficient for network screening***
- No need to update and manage databases of special variables, e.g. pavement quality
 - ‘not new’ for VTT (Peltola et al., 1994)
 - ‘new’ for CDV (Ambros et al., 2015)

Contact Information:

Jiří Ambros, CDV (Czech Republic), jiri.ambros@cdv.cz

Harri Peltola, VTT (Finland), Harri.Peltola@vtt.fi

References



Ambros, J., Valentová, V., Janoška, Z. (2015). Investigation of difference between network screening results based on multivariate and simple crash prediction models. In *TRB 94th Annual Meeting*, Washington, D.C.



Peltola, H., Kulmala, R., Kallberg, V.-P. (1994). Why use a complicated accident prediction model when a simple one is just as good? In *22nd PTRC Seminar*, Warwick, UK.

Contact Information:

Jiří Ambros, CDV (Czech Republic), jiri.ambros@cdv.cz

Harri Peltola, VTT (Finland), Harri.Peltola@vtt.fi

Thank you for your attention!

Contact Information:

Jiří Ambros

CDV, Brno, Czech Republic
jiri.ambros@cdv.cz



Harri Peltola

VTT, Espoo, Finland
Harri.Peltola@vtt.fi

