TRANSPORT
ISSN 1648-4142 / eISSN 1648-3480
2018 Volume 33(2): 478–488
doi:10.3846/16484142.2017.1289479

IDENTIFICATION OF HOTSPOTS ON ROADS USING CONTINUAL VARIANCE ANALYSIS

Dejan Anđelković[1], Boris Antić[2], Krsto Lipovac[3], Ilija Tanackov[4]

[1]*Faculty of Technical Sciences, University of Pristina, Kosovska Mitrovica, Serbia*
[2, 3]*Faculty of Transport and Traffic Engineering, University of Belgrade, Belgrade, Serbia*
[4]*Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia*

Submitted 23 January 2016; resubmitted 19 April 2016; accepted 15 May 2016;
published online 12 April 2017

Abstract. This paper presents a new statistical model for the identification of dangerous locations (subsections) on roads, also known as hotspots. The model is based on continual analysis of variance. The variance parameter has the potential for the synthesis of quantity and quality, especially regarding traffic accident frequencies and the consequences of traffic accidents within subsections and the significant comparison of the produced synthesis. The sensitivity of the suggested model can be adjusted with the level of disjunction and the length of subsections. A practical application of the new model is performed using a sample of 8442 traffic accidents, of which 6079 were Property Damage Only (PDO) accidents, 2041 resulted in injuries and 322 resulted in fatalities. The sample is for the period of 2001 to 2011 and is from an 'I class' two lane rural state road in the Serbia with total length of 284 kilometres. The results acquired using the continual analysis of variance were compared with previous results from four HotSpot Identification Methods (HSID) that are also based on the frequency of traffic accidents.

Keywords: traffic accidents; traffic safety; hotspots on the roads; continual analysis of variance; frequency of traffic accidents; HSID methods.

Introduction

Managing hotspots is one of the most important topics of road traffic safety, and the identification of hotspots (black spots) and their ranking represents a frequent subject of discussion among researchers. In the literature (Montella 2010; Anđelković *et al.* 2014), the identification of hotspots represents the first step in the process of road safety management.

Considering the nonexistence of a unique and generally accepted model (with universal benchmarks) for the identification and classification of hotspots, for the purpose of this research, a new statistical model is developed, which is based on the frequency of traffic accidents that are spatially and temporally distributed on certain road sections and subsections. Statistical-probabilistic methods are unavoidable in traffic safety research. The large number of applied models confirms it: Poisson models (Geedipally *et al.* 2014), negative binominal models (Miranda-Moreno *et al.* 2007; Ferreira, Couto 2013; Geedipally *et al.* 2014), generalized negative binominal models (Miaou, Song 2005; Russo *et al.* 2014), log-normal Poisson models (Miranda-Moreno *et al.* 2007), empirical Bayes models (Heydecker, Wu 2001; Miaou, Song 2005; Jiang *et al.* 2014), hierarchical Bayes models (Tunaru 2002), and many other models (Carey 2001). By using these models, prioritization is often based on the potential for reduction of accident risk and finding the locations of greatest risk.

Numerical markings can describe the number of accidents, the severity of consequences, the number of injured people and other parameters that characterize traffic accidents. The statistics of traffic accidents is the best data for the identification of dangerous locations. Based on the statistics of traffic accidents, it is possible, apart from identifying the locations were traffic accidents are frequent, to conduct a far more sensitive analysis of other significant factors connected with the occurrence of traffic accidents, and that analysis can include: determining the interaction between the vehicle characteristics, road characteristics and characteristics of individuals involved in traffic accidents (Savolainen *et al.* 2011; Sokolovskij, Prentkovskis 2013); determining the

Corresponding author: Dejan Anđelković
E-mail: *aaa.dejo@gmail.com*

Taylor & Francis
Taylor & Francis Group

road conditions, type of accident, and age and gender of the driver (Manner, Wünsch-Ziegler 2013; Russo *et al.* 2014); determining the effects of speed, congestion and horizontal curvature of the road (Wang *et al.* 2013); the behavior of young drivers, presence of passengers and involvement of vulnerable participants in traffic (Weiss *et al.* 2014; Alver *et al.* 2014); and many other factors.

It is important to emphasize that we are analysing random events, and it is well known that statistical methods that include random events provide certain regularities or patterns, especially when large numbers, i.e., large groups, are involved. Continual variance analysis was not used previously for the identification of hotspots, however, the analysis of variance was used within research regarding road traffic safety (Miaou, Lord 2003; Lord 2008; Jin *et al.* 2008; Antić *et al.* 2013; Qu *et al.* 2014).

The main goal of this paper is the introduction of a new model based on continual variance analysis for the identification of dangerous road subsections. The primary goal of the proposed method is the identification of hotspots, that is, the discovery of the most unsecure subsections with the highest concentrations of traffic accidents. A secondary goal is the identification of the safest subsections with the lowest concentrations of traffic accidents. Apart from that, this paper also contains a comparative analysis of the presented method and four HotSpot Identification Methods (HSID).

1. Methods

Most researchers agree that the empirical Bayes method is currently the most reliable method for the identification and ranking of hotspots (Heydecker, Wu 2001; Tunaru 2002; Miaou, Song 2005; Montella 2010; Aguero-Valverde 2013; Gregoriades, Mouskos 2013; Yu *et al.* 2014; Washington *et al.* 2014). Bayesian statistics includes three elements: (1) the previous distribution of accidents (historical data about accidents), (2) the distribution probability, and (3) additional (anticipated) distributions. It is most often based on the assumption that the occurrence of accidents follows a Poisson distribution and that the probability distribution of accident locations follows a gamma distribution (Cheng, Washington 2005). A Poisson distribution, in most cases, is not appropriate for accident data. The assumption made in simple Poisson models is that the mathematical expectations and variance are equal. Heterogeneity is mostly shown in the form of over dispersion, which implies that the variance is higher than the expectations. There are models that attempt to overcome that shortcoming, including hierarchical models (Hinde, Demétrio 1998; El-Basyouny, Sayed 2010), gamma models (Anastasopoulos, Mannering 2009; Connors *et al.* 2013; Zou *et al.* 2013), negative binominal models (Poch, Mannering 1996; Hinde, Demétrio 1998; Russo *et al.* 2014), log-normal models (Lord, Miranda-Moreno 2008; Connors *et al.* 2013) and other models (for example Miaou, Lord 2003).

Complex convolution of Poisson and gamma distributions is founded on the long-term mean value of accidents, which contains in its weighting factor the mandatory application of variance (Hauer 1997; Harwood *et al.* 2000; Vistisen 2002; Shen, Gan 2003). Variance is usually observed in three ways: as a fixed value; a varying function of the location characteristics; and as a value that varies randomly.

Generally, variance as a second central moment represents a strong statistical parameter introduced into the entire balance. It can be said that the greatest achievement of the use of variance is realized in the Pollaczek–Khinchine formula, which shows the influence of the variance on system function. Additionally, the variance indirectly introduces the Annual Average Daily Traffic (AADT) in the identification and ranking of hotspots, which is not the case here. The authors' idea is based on the fact that in poor countries and in developing countries, AADT data are often unavailable or inaccurate, so the authors want to create a model that can be used to identify hotspots in countries lacking precise AADT data. In that way, decision makers in such countries could use this simple model and the accompanying software to relatively quickly obtain classified and ranked hotspots without AADT data, and they would be able to allocate resources to the most dangerous locations and accordingly improve the level of road traffic safety. The inclusion of AADT data is one of the goals of a future expansion of the proposed model.

Often, the dominant characteristic in the identification of dangerous places is based on socio-economic factors (weighted factors) that represent a simple mathematical apparatus based on weights (weight factors) that statistically take into account only mathematical expectations.

The method of continual variance analysis represents a new model that can use weights and is based on the concept of sequentially searching for hotspots; that is, it is also a variance analysis model. The proposed model is conservative, and it is observed within the selected road (group).

1.1. Description of the Mathematical Model

The search for dangerous locations on a particular road, which is divided into smaller number of subsections, is based on observing two groups with mathematical expectations ε_1 and ε_2 with no significant difference (no difference between groups). When one variable is taken and then compared the variance from both groups, by using ANalysis Of the VAriance (ANOVA) of one group against the other group, certain values are obtained. The concept of a 'continual' variable synthetically contains factors that, in this case, are not differentiated and provide a complete platform. However, it can relate to a disjunctive subsection and create the possibility of a sensible calibration of the location if the subsections are overlapped. The simplest concept of a disjunctive subsection is taken to be an example of a model with continual variance analysis.

Let's consider a certain road, or more accurately, a section of a road, where during an observation period, Σ (sum) accidents happened. Of those, there were Σ_{PDO}

accidents with Property Damage Only (PDO), Σ_{INJ} accidents with injuries and Σ_{FAT} accidents with fatalities. The total number of accidents has to comply with the following condition:

$$\Sigma = \Sigma_{PDO} + \Sigma_{INJ} + \Sigma_{FAT}. \tag{1}$$

By dividing the (total) section into n disjunctive subsections, a sequence is created that distributes accidents according to the index of each subsection ($n_{PDO(k)}$ – number of traffic accidents with PDO in subsection k; $n_{INJ(k)}$ – number of traffic accidents with injuries in subsection k; $n_{FAT(k)}$ – number of traffic accidents with fatalities in subsection k).

$$\Sigma_{PDO} = \sum_{k=1}^{n} n_{PDO(k)};$$

$$\Sigma_{INJ} = \sum_{k=1}^{n} n_{INJ(k)};$$

$$\Sigma_{FAT} = \sum_{k=1}^{n} n_{FAT(k)}; \tag{2}$$

$$\Sigma = \Sigma_{PDO} + \Sigma_{INJ} + \Sigma_{FAT} = \sum_{k=1}^{n} n_{PDO(k)} + \sum_{k=1}^{n} n_{INJ(k)} + \sum_{k=1}^{n} n_{FAT(k)}. \tag{3}$$

The numbers of accidents in each subsection simultaneously provide a special sequence of accidents in a section that contains complementary values of subsections (Figs 1 and 2), and those are:

- when the number of accidents with material damage is excluded from the section, the complementary value at subsection k is $N_{PDO(k)} = \Sigma_{PDO} - n_{PDO(k)}$;
- when the number of accidents with injuries is excluded from the section, the complementary value at subsection k is $N_{INJ(k)} = \Sigma_{INJ} - n_{INJ(k)}$;
- when the number of accidents with fatalities is excluded from the section, the complementary value at subsection k is $N_{FAT(k)} = \Sigma_{FAT} - n_{FAT(k)}$.

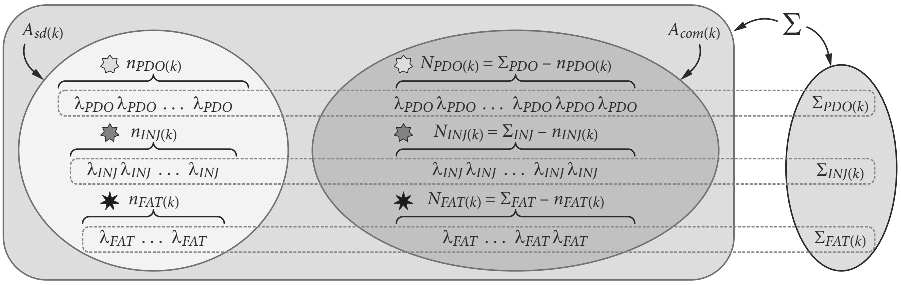
$A_{sd(k)}$
$A_{com(k)}$
Σ
$n_{PDO(k)}$
$\lambda_{PDO}\,\lambda_{PDO}\,\ldots\,\lambda_{PDO}$
$n_{INJ(k)}$
$\lambda_{INJ}\,\lambda_{INJ}\,\ldots\,\lambda_{INJ}$
$n_{FAT(k)}$
$\lambda_{FAT}\,\ldots\,\lambda_{FAT}$
$N_{PDO(k)} = \Sigma_{PDO} - n_{PDO(k)}$
$\lambda_{PDO}\,\lambda_{PDO}\,\ldots\,\lambda_{PDO}\,\lambda_{PDO}\,\lambda_{PDO}$
$N_{INJ(k)} = \Sigma_{INJ} - n_{INJ(k)}$
$\lambda_{INJ}\,\lambda_{INJ}\,\ldots\,\lambda_{INJ}\,\lambda_{INJ}$
$N_{FAT(k)} = \Sigma_{FAT} - n_{FAT(k)}$
$\lambda_{FAT}\,\ldots\,\lambda_{FAT}\,\lambda_{FAT}$
$\Sigma_{PDO(k)}$
$\Sigma_{INJ(k)}$
$\Sigma_{FAT(k)}$

Fig. 1. Display of observed groups with traffic accidents

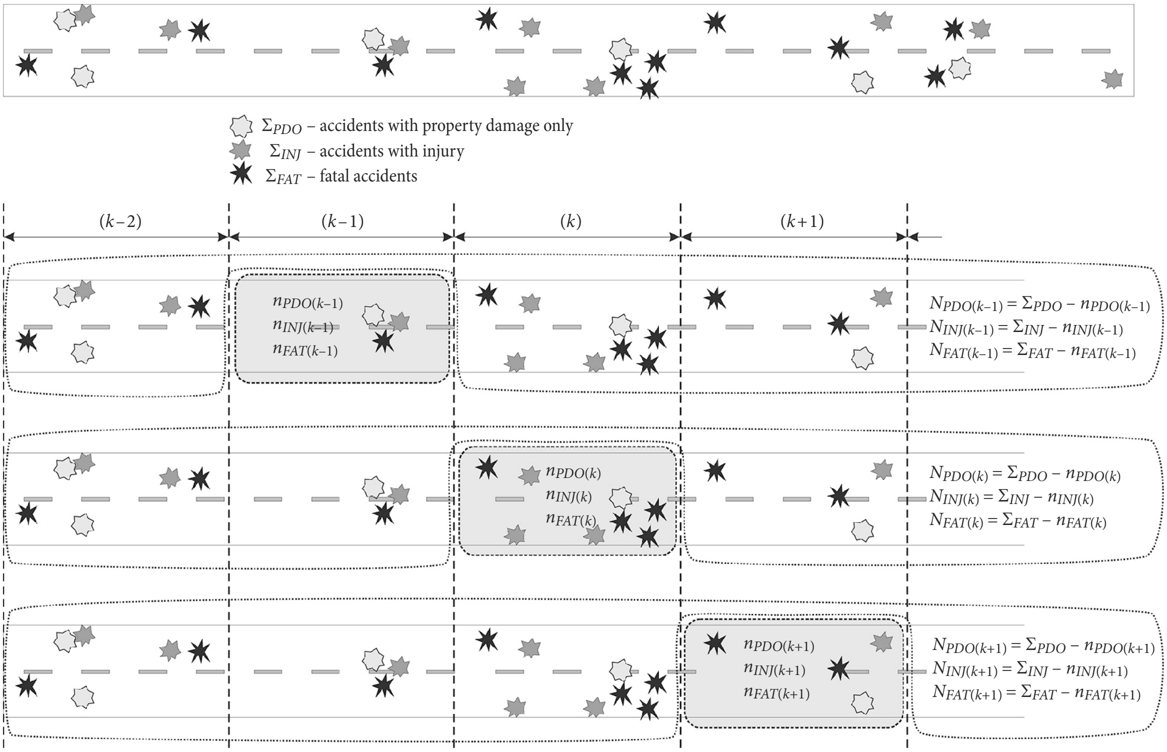
Σ_{PDO} – accidents with property damage only
Σ_{INJ} – accidents with injury
Σ_{FAT} – fatal accidents
$(k-2)$
$(k-1)$
(k)
$(k+1)$
$n_{PDO(k-1)}$
$n_{INJ(k-1)}$
$n_{FAT(k-1)}$
$N_{PDO(k-1)} = \Sigma_{PDO} - n_{PDO(k-1)}$
$N_{INJ(k-1)} = \Sigma_{INJ} - n_{INJ(k-1)}$
$N_{FAT(k-1)} = \Sigma_{FAT} - n_{FAT(k-1)}$
$n_{PDO(k)}$
$n_{INJ(k)}$
$n_{FAT(k)}$
$N_{PDO(k)} = \Sigma_{PDO} - n_{PDO(k)}$
$N_{INJ(k)} = \Sigma_{INJ} - n_{INJ(k)}$
$N_{FAT(k)} = \Sigma_{FAT} - n_{FAT(k)}$
$n_{PDO(k+1)}$
$n_{INJ(k+1)}$
$n_{FAT(k+1)}$
$N_{PDO(k+1)} = \Sigma_{PDO} - n_{PDO(k+1)}$
$N_{INJ(k+1)} = \Sigma_{INJ} - n_{INJ(k+1)}$
$N_{FAT(k+1)} = \Sigma_{FAT} - n_{FAT(k+1)}$

Fig. 2. Distribution of traffic accidents at disjunctive subsections

This procedure, from the basic group of accidents in a section $\left\{\Sigma_{PDO}, \Sigma_{INJ}, \Sigma_{FAT}\right\}$, separates two groups: (1) the group of accidents in a subsection k $A_{sd(k)} = \left\{n_{PDO(k)}, n_{INJ(k)}, n_{FAT(k)}\right\}$ and (2) the complementary group in a subsection k $A_{com(k)} = \left\{N_{PDO(k)}, N_{INJ(k)}, N_{FAT(k)}\right\}$.

If each accident with PDO is assigned a numerical marking λ_{PDO}, each accident with injuries (*INJ*) a numerical marking λ_{INJ} and each accident with fatalities (*FAT*) a numerical marking λ_{FAT}, the groups of accidents at a subsection $A_{sd(k)}$ and the complementary groups $A_{com(sd)}$ can be written in the following form:

$$A_{sd(k)} = \left\{ \underbrace{\lambda_{PDO},\ldots,\lambda_{PDO}}_{n_{PDO(k)}}, \underbrace{\lambda_{INJ},\ldots,\lambda_{INJ}}_{n_{INJ(k)}}, \underbrace{\lambda_{FAT},\ldots,\lambda_{FAT}}_{n_{FAT(k)}} \right\}; \tag{4}$$

$$A_{com(k)} = \left\{ \underbrace{\lambda_{PDO},\ldots,\lambda_{PDO}}_{N_{PDO(k)}}, \underbrace{\lambda_{INJ},\ldots,\lambda_{INJ}}_{N_{INJ(k)}}, \underbrace{\lambda_{FAT},\ldots,\lambda_{FAT}}_{N_{FAT(k)}} \right\}. \tag{5}$$

The mathematical expectation of the numerical markings of accidents and the variance of the numerical markings of accident in the total group on the entire observed section are equal to:

$$E_{\Sigma} = \frac{\lambda_{PDO}\Sigma_{PDO} + \lambda_{INJ}\Sigma_{INJ} + \lambda_{FAT}\Sigma_{FAT}}{\Sigma}; \tag{6}$$

$$V_{\Sigma} = \frac{a}{\Sigma - 1}, \tag{7}$$

where:

$$a = \sum_{i=1}^{\Sigma_{PDO}} \left(E_{\Sigma} - \lambda_{PDO}\right)^2 + \sum_{i=1}^{\Sigma_{INJ}} \left(E_{\Sigma} - \lambda_{INJ}\right)^2 + \sum_{i=1}^{\Sigma_{FAT}} \left(E_{\Sigma} - \lambda_{FAT}\right)^2.$$

The groups A_{sd} and the complementary group $A_{com(sd)}$ have the following parameters.

The mathematical expectation of the numerical markings of accident group $A_{sd(k)}$ at subsection k is:

$$\varepsilon_k = \frac{b}{n_{PDO(k)} + n_{INJ(k)} + n_{FAT(k)}}, \tag{8}$$

where:

$$b = \overbrace{\lambda_{PDO} + \lambda_{PDO} + \ldots + \lambda_{PDO}}^{n_{PDO(k)}} + \overbrace{\lambda_{INJ} + \lambda_{INJ} + \ldots + \lambda_{INJ}}^{n_{INJ(k)}} + \overbrace{\lambda_{FAT} + \ldots + \lambda_{FAT}}^{n_{FAT(k)}}.$$

The mathematical expectation of the numerical markings of accidents in the complementary group $A_{com(k)}$ is:

$$E_k = \frac{c}{N_{smš(k)} + N_{pov(k)} + N_{pog(k)}}, \tag{9}$$

where:

$$c = \overbrace{\lambda_1 + \lambda_1 + \lambda_1 + \ldots + \lambda_1 + \lambda_1 + \lambda_1}^{N_{smš(k)}} + \overbrace{\lambda_2 + \lambda_2 + \lambda_2 + \ldots + \lambda_2 + \lambda_2}^{N_{pov(k)}} + \overbrace{\lambda_3 + \lambda_3 + \ldots + \lambda_3 + \lambda_3}^{N_{pog(k)}}.$$

The variance of the numerical markings of accidents in group $A_{sd(k)}$ at subsection k is:

$$\overbrace{\left(\varepsilon_k - \lambda_{PDO}\right)^2 + \ldots + \left(\varepsilon_k - \lambda_{PDO}\right)^2}^{n_{PDO(k)}} = n_{PDO(k)}\left(\varepsilon_k - \lambda_{PDO}\right)^2 = \sum_{i=1}^{n_{PDO(k)}} \left(\varepsilon_k - \lambda_{PDO}\right)^2; \tag{10}$$

$$\overbrace{\left(\varepsilon_k - \lambda_{INJ}\right)^2 + \ldots + \left(\varepsilon_k - \lambda_{INJ}\right)^2}^{n_{INJ(k)}} = n_{INJ(k)}\left(\varepsilon_k - \lambda_{INJ}\right)^2 = \sum_{i=1}^{n_{INJ(k)}} \left(\varepsilon_k - \lambda_{INJ}\right)^2; \tag{11}$$

$$\overbrace{\left(\varepsilon_k - \lambda_{FAT}\right)^2 + \ldots + \left(\varepsilon_k - \lambda_{FAT}\right)^2}^{n_{FAT(k)}} = n_{FAT(k)}\left(\varepsilon_k - \lambda_{FAT}\right)^2 = \sum_{i=1}^{n_{FAT(k)}} \left(\varepsilon_k - \lambda_{FAT}\right)^2; \tag{12}$$

$$v\left(A_{com(k)}\right) = v_k = \frac{d}{n_{PDO(k)} + n_{INJ(k)} + n_{FAT(k)} - 1},$$

where:

$$d = \sum_{i=1}^{n_{PDO(k)}} \left(\varepsilon_k - \lambda_{PDO}\right)^2 + \sum_{i=1}^{n_{INJ(k)}} \left(\varepsilon_k - \lambda_{INJ}\right)^2 + \sum_{i=1}^{n_{FAT(k)}} \left(\varepsilon_k - \lambda_{FAT}\right)^2. \tag{13}$$

The variance of the numerical markings of accidents in the complementary group $A_{com(k)}$ is:

$$\overbrace{\left(E_k - \lambda_{PDO}\right)^2 + \ldots + \left(E_k - \lambda_{PDO}\right)^2}^{N_{PDO(k)}} = N_{PDO(k)}\left(E_k - \lambda_{PDO}\right)^2 = \sum_{i=1}^{N_{PDO(k)}} \left(E_k - \lambda_{PDO}\right)^2; \tag{14}$$

$$\overbrace{\left(E_k - \lambda_{INJ}\right)^2 + \ldots + \left(E_k - \lambda_{INJ}\right)^2}^{N_{INJ(k)}} = N_{INJ(k)}\left(E_k - \lambda_{INJ}\right)^2 = \sum_{i=1}^{N_{INJ(k)}} \left(E_k - \lambda_{INJ}\right)^2; \tag{15}$$

$$\overbrace{\left(E_k - \lambda_{FAT}\right)^2 + \ldots + \left(E_k - \lambda_{FAT}\right)^2}^{N_{FAT(k)}} =$$

$$N_{FAT(k)}\left(E_k-\lambda_{FAT}\right)^2=\sum_{i=1}^{N_{FAT(k)}}\left(E_k-\lambda_{FAT}\right)^2; \quad (16)$$

$$V\left(A_{sd(k)}\right)=V_k=\frac{e}{N_{PDO(k)}+N_{INJ(k)}+N_{FAT(k)}-1}, \quad (17)$$

where:

$$e=\sum_{i=1}^{N_{PDO(k)}}\left(E_k-\lambda_{PDO}\right)^2+$$
$$\sum_{i=1}^{N_{INJ(k)}}\left(E_k-\lambda_{INJ}\right)^2+\sum_{i=1}^{N_{FAT(k)}}\left(E_k-\lambda_{FAT}\right)^2.$$

This method for calculating the parameters of mathematical expectation and variance is conservative in comparison to the total sample-section or road we are testing.

The group of accidents at a subsection usually has a much smaller number of data points than the complementary group:

$$n_{PDO(k)}+n_{INJ(k)}+n_{FAT(k)}<<$$
$$N_{PDO(k)}+N_{INJ(k)}+N_{FAT(k)}. \quad (18)$$

We can expect that the number of data points on the complementary section will be stable and smaller than – but approximately equal to – the total number of accidents at the observed section:

$$\frac{N_{PDO(k)}+N_{INJ(k)}+N_{FAT(k)}}{\Sigma_{PDO}+\Sigma_{INJ}+\Sigma_{FAT}}\approx 1. \quad (19)$$

In case that there are no accidents on the subsection, the complementary group is equal to the basic group:

$$n_{PDO(k)}+n_{INJ(k)}+n_{FAT(k)}=0\Leftrightarrow$$
$$\frac{N_{PDO(k)}+N_{INJ(k)}+N_{FAT(k)}}{\Sigma_{PDO}+\Sigma_{INJ}+\Sigma_{FAT}}=1. \quad (20)$$

The complementary group has values of numerical markings and variance that are similar to the values of the basic group that have a small variation of mathematical expectation and variance:

$$E_\Sigma\approx E_k;\ V_\Sigma\approx V_k,$$
$$\forall k\notin\left(1,n\right). \quad (21)$$

The parameters of the numerical markings of a subsection $\left(\varepsilon_k,v_k\right)$ will have variation values of mathematical expectations and variance that will be directly influenced by the number of accidents and their associated numerical markings.

Along with the existence of the variance of the group of numerical markings of accidents at a subsection k and the complementary group, a null hypothesis can also be postulated regarding the equality of the mathematical expectations or variance of numerical markings of accidents with an importance threshold α. This hypothesis can be postulated in relation to the total group:

$$H\left(\varepsilon_k=E_\Sigma\right)_\alpha;\ H\left(v_k=V_\Sigma\right)_\alpha, \quad (22)$$

or, in relation to complementary group:

$$H\left(\varepsilon_k=E_k\right)_\alpha;\ H\left(v_k=V_k\right)_\alpha. \quad (23)$$

However, it should be noted that the group of numerical markings of subsection is a subgroup of a total group. Because of that, specificities (factors) that lead to accidents at a subsection exist even in the total group. If null hypotheses are placed in relation to the complementary group, then the specificities of the subsection are preserved and present even in the complementary group. This fact is key for the selection of the null hypothesis based on the parameters of the subsection and complementary group.

The research of the hypothesis regarding the equality of mathematical expectations is somewhat in accordance with the previously promoted weighted methods (PIARC 2004; Zein 2004; DoT 2006; Oh *et al.* 2010; Montella 2010; Vadlamani *et al.* 2011). However, it is known that the tests based on variances are more sensitive than tests based on mathematical expectations. Thus, the final choice is testing of the null hypothesis of the variance of a group of numerical markings from the subsection and complementary group:

$$H\left(v_k=V_k\right)_\alpha. \quad (24)$$

The problem arising from a subsection without accidents represents a potential advantage of the above mentioned approach. For the case in which there were no accidents on the subsection, the mathematical expectation and the variance of the numerical markings are equal to zero. The mathematical expectation and variance of the numerical markings of the complementary group are equal to those of the total group:

$$\varepsilon_k=0\wedge v_k=0\Leftrightarrow$$
$$E_k=E_\Sigma\wedge V_k=V_\Sigma. \quad (25)$$

To use this circumstance, it is necessary to attach the number '0' to the subsections that have no recorded accidents in databases of traffic accident data, which is explained in detail in the following paragraph. That completes the statistical group of accidents. For each subsection without traffic accidents, any variance test rejects the null hypotheses. The significant difference of the variance goes in favour of the desired safety.

The concept of testing the null hypothesis is based on disjunctive groups, that is, the subsection and complementary group, and it has a dual interpretation. If the null hypothesis is rejected, the subsection can be much more safe or unsafe. Significant safety is observed in cases with small numbers of accidents, while significant unsafe conditions are observed in cases with large numbers of accidents. When traffic accident weights are used, significant non-safety can be observed on certain subsections even in cases with smaller numbers of accidents but with more severe consequences.

The choice of statistical test for testing the null hypothesis regarding the equality of variances represents a key question. It is known that variance tests easily verify the null hypothesis if the number of data in the groups is small and that they have difficulties identifying the null hypothesis when the number of data in groups is large, which is the reason to select a large number of data. The number of data in groups of numerical markings of a subsection can be considerably smaller than the number of data in the complementary group. That brings us to the situation in which the null hypothesis is verified for two groups of which one has a small number of data and the other has a large number of data. There is also an issue when the complementary group has a small variation, in which case the null hypothesis is difficult to prove. Because of the above mentioned circumstances, a preliminary rating of the choice of statistical test could not be completed.

We mention one more time that we have adopted the application of the standard single-factor variance analysis (that is based on *F*-distribution), where, in this case, the continual factor is the length of the subsection, which must have a constant length; and the variable factor is the number of traffic accidents.

1.2. Source of Data

The model is tested with and based on official statistical data from the Ministry of Internal Affairs (Serbia) Traffic Police Department database. Data regarding the consequences, time of occurrence and location of traffic accidents (kilometre and meter of the road where accidents happened) were taken from the database. Modification of the existing categorization of traffic accident types needed for the research was made within the received database, with linear numerical accident markings assigned by bijection. This adjusted way of classifying accident types can be observed as an elementary weighting of traffic accidents. For research purposes, in an effort to complete a statistical group, a numerical marking of "0" is assigned for each subsection where, during one year, there were no traffic accidents. In this way, a statistical group is produced with adjusted traffic accident types joined to a certain kilometre section where the observed traffic accident occurred. The database completed in this manner for the application of the suggested model offers the possibility to choose different subsection lengths as well as different time periods of observation.

1.3. Selection of Road (Section) for Observation

In prior research (Lipovac *et al.* 2010) it was confirmed that in the Serbia on the 'I class' state road of Belgrade–Ribarice, which is 284 km in length, 27.5% of the kilometre-long sections were highly dangerous, so this road was chosen for analysis and application of the model. On the other hand, the selected road has a geo-strategic importance because it connects central Serbia with Montenegro and Southern Serbia–Kosovo. The selected road is considered to be a rural road (in a rural area). The speed limit on the selected road does not exceed 80 km/h.

1.4. Selection of the Subsection Length for Application of the Suggested Model

There is no clear indicator of the best length for determining dangerous segments or road sections, nor can the optimal length be defined. Lengths are chosen to limit heterogeneity within each road segment, but certain authors recommend constant lengths because interpretation of accident data can be complicated for different road lengths. Stern and Zehavi (1990) divide the road into 1-km-long sections, without any special reason for those lengths. Elvik (1988) suggests defining a dangerous road section so it is always the same length by moving a sliding window (slider) of a certain length along the road. Other researchers have used a road segmentation approach (Abdel-Aty, Radwan 2000; Cafiso *et al.* 2008). They often define road segments with fixed lengths or simply use the distance between two main intersections. Sadeghi *et al.* (2013) identified and segmented homogenous road sections based on accident factors.

Just as others have assumed in their work (AASHTO 2010; Harwood *et al.* 2010; Tegge *et al.* 2010) that highways in America have the same functional form, we assumed in this paper, just without the parameter estimate, that the entire length of a two-lane road has the same or approximately the same functional form; further, a constant subsection length is used, following the experience of others (Okamoto, Koshi 1989; Stern, Zehavi 1990; Abdel-Aty, Radwan 2000; Cafiso *et al.* 2008).

The proposed model can be applied with a constant subsection length. The selection of the subsection length is arbitrary. The model is flexible, and by using software, different subsections of interest can be studied. This paper presents the application of subsections with a constant length of 1 km. A constant subsection length less than 0.5 km could also be considered, as well as 0.5, 2, 3, 4, 5 km, or practically any length, using the software that was designed for this purpose. Within the suggested methodology there is an option to use the approach of the sliding moving window (Kwon *et al.* 2013).

1.5. Determining the Observational Time Period

The time period for the identification of dangerous locations in Serbia, within which every accident analysis is successfully completed is 5 years, and a detailed analysis and case studies happened within the last year. We are also aware that the most often used analysis period for the identification of dangerous places is from 3 to 5 years and that periods longer than 10 years should not be used; however, because of the increased sample size and estimates of the proposed model and verification of its capacity, a representative period of 11 years is selected.

It is well known, that variance tests can easily verify the null hypothesis if the number of data in groups is small and that the tests have a difficult time verifying the null hypothesis when the number of data in groups is large. We use this as justification for the use of such a long period of time as 11 years (2001–2011). This time period encompassed 6079 traffic accidents with PDO, 2041 traffic accidents with injuries and 322 traffic ac-

Table. Data and variations that were used in the number of Traffic Accidents (TA) for the period of 2001–2011

TA per one kilometre				Total number of TA for the section	
Type of TA	Min. number of TA per accident type	Max. number of TA per accident type	Mean	Total number of TA per accident type	[%]
PDO	0	222	21.40	6079	72.01
TA with injuries	0	91	7.19	2041	24.18
TA with fatalities	0	19	1.13	322	3.81
TA with casualties	0	110	8.32	2363	27.99
Σ	0	313	29.73	8442	100

cidents with fatalities, which in total is 8442 traffic accidents (Table). The use of this long time period is further justified by plans for a future comparison (research) of certain safety parameters of this section of a two-lane road to a highway with similar parameters whose construction began in 2012.

Because of the large number of data used within this study, to avoid slow and manual data processing, and to automate the data processing, all data are processed in mathematical software specially designed for this purpose, but also filtering of the anomalous observation is not taken into account like it was case in research performed by Russo *et al.* (2014), so it is limitation of presented method.

2. Results

By applying the mathematical expectation-mean value ε, variance v and F value in a continual analysis of variance-ANOVA, the primary and secondary goals of the proposed method are achieved. During testing of the proposed hypothesis $H(v_k = V_k)_\alpha$, the variance v, plays an important role in determining a large significance (significance threshold) between the observed subsection k and the complementary group K (Fig. 1). For the entire observed group of all 284 1-km-long subsections and the adopted significance (significance threshold) of $\alpha = 0.05$[1], for the continual variance analysis, two disjunction data groups are produced, that is, two subsection groups. The research produced 102 subsections that are significantly different from the complementary group for $\alpha < 0.05$ and 182 subsections (white circles[2]) that are not significantly different from the complementary group for $\alpha \geq 0.05$. Within the 102 subsections that are significantly different from the complementary group, there are 67 subsections that are significantly safe (green circles[2]) and 35 subsections that are significantly not safe (red circles[2]). All of these values were previously calculated by applying the described method in accordance with the adopted significance threshold of $\alpha = 0.05$. The mathematical expectation values and variance values shown in Fig. 3 represent the first filter, that is, the first stage in identification, which does not include all potentially significant subsections. The approximate limit values that note the significance of subsections are for values $\varepsilon < 0.9$ (significantly safe – green circles) and for $\varepsilon > 1.3$ (significantly not safe – red circles), while for group values ε {0.9, 1.3} we have insignificant subsections (white circles).

The F values (ANOVA) and the values of mathematical expectations shown in Fig. 4 represent the second filter, that is, the second stage of identification of potentially significant subsections. Significant subsections are indicated by $F > 4$, and for $\varepsilon < 0.9$, we have significantly safe subsections (green circles) and for $\varepsilon > 1.3$ we have significantly not safe subsections (red circles), for the significance threshold of $\alpha < 0.05$. Considering that there are subsections with the same mathematical expectation within the significant subsections, the F value of ANOVA will note those subsections with higher numbers of accidents in comparison to those with lower numbers of accidents, as illustrated by higher F values of the continual analysis of variance – ANOVA.

The F values of ANOVA and the variance values are shown in Fig. 5 and represent a third and final filter, that is, a final stage of identification of potentially significant subsections. In accordance with the primary goal of the paper (excluding non-significant subsections – white circles), the significantly unsafe subsections (red circles) are displayed, while the significantly safe subsections (green circles) are excluded from the figure.

Fig. 6 shows a unified three-dimensional zone image of all three observed values (ε, v and F) for all subsections of the observed group (road). The average value of the intensity of traffic accidents and the variance are independent variables, and the F value is a dependent variable. In accordance with the F values of ANOVA, a visual zone classification is created whereby all subsections are included with their respective colours, such that darker shades of each colour indicate significantly safer or more unsafe subsections. Considering the first dimension ε, framed by a yellow rectangle in Fig. 6 indicating the limiting values of ε as represented in Fig. 3, the group is divided into significantly safe subsections, significantly unsafe subsections and subsections that do not fall into either group. The same approach can be used to display the division of the group for the other two dimensions v and F within their respective limit values, as in Figs 4 and 5.

[1] Adopted limit (level) of significance in this paper is 0.05, but 0.01 or some other can also be adopted.

[2] The number of circles (subsections) of all colors in Figs 4–6 does not match completely with the listed exact calculated number of circles (subsections) because of complete overlap of a certain number of circles (subsection) with the same values.

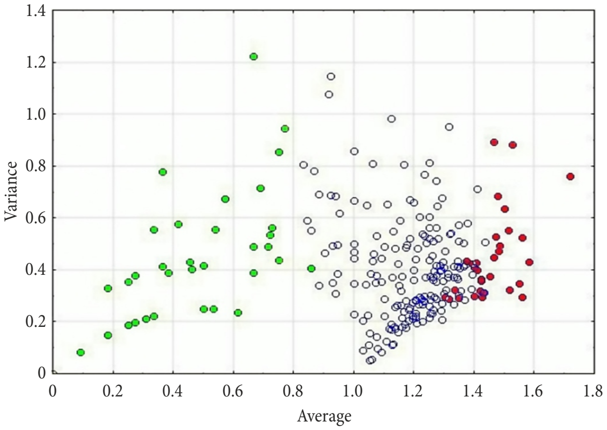

Fig. 3. Display of ε and v values of subsections that are significantly different

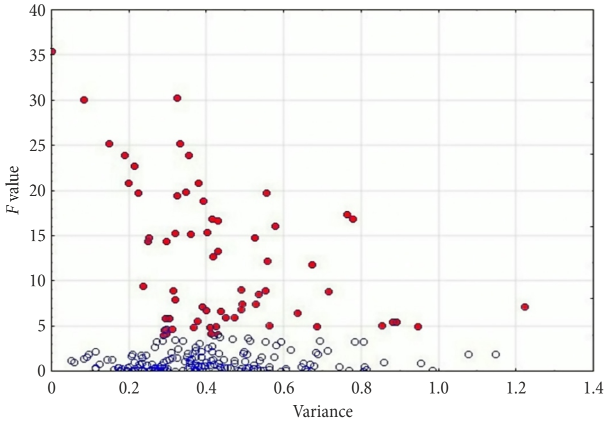

Fig. 5. Display of values F and v for subsections that are significantly different

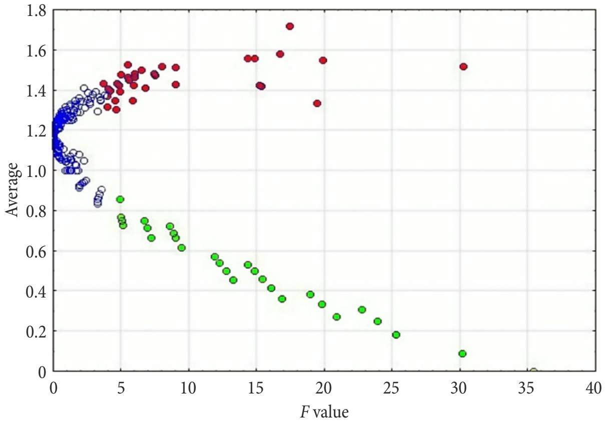

Fig. 4. Display of F and ε values of subsections that are significantly different

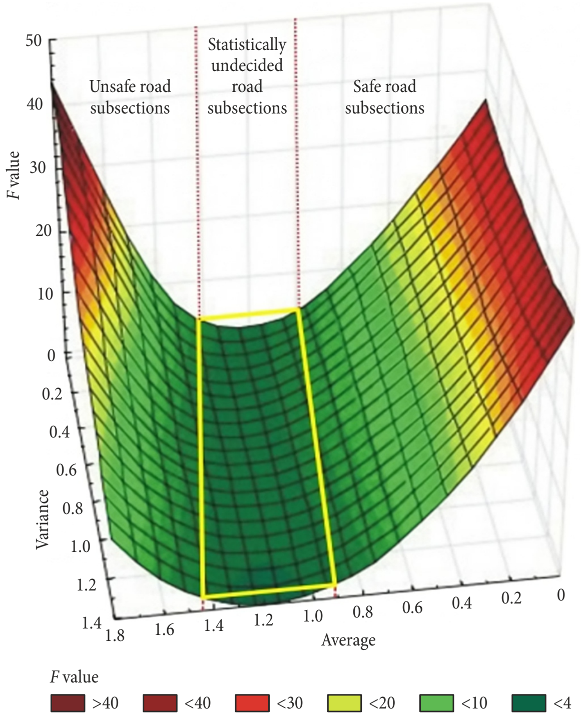

Fig. 6. Common three-dimensional display of ε, v and F values for all observed subsections

In this model, there is a calibration of weight accident factors. In this initial presentation of this model all factors are assumed to be linear (values 0, 1, 2, 3). They can be calibrated as an exponential, which would lead to an expansion of the zone of 'undetermined subsections' or as a logarithm, which would decrease the variance of the heterogenic group of accidents and establish a clearer border between the statistically 'undetermined' and unsafe subsections. A calibration was not done for this initial presentation of the model, and the previously described elementary, linear classification of traffic accidents was maintained. The reason is that this calibration enables a comparative analysis of the elementary conditions of the continual variance analysis with other HSID methods.

3. Discussion with Comparative Analysis

The suggested method SM is based on the frequency of traffic accidents, and because of that, a comparative analysis is completed using only methods that are also based on the frequency of traffic accidents. The same, previously mentioned accident data were used for the period from 2001 to 2011. The method is compared with the following HSID: ranking the frequency of accidents based on the total number of accidents *CF*; ranking the frequency of accidents based on the total number of accidents with casualties CF_1; ranking the frequency of accidents with PDO $EPDO_n$ with ordered numerical markings; and ranking the frequency of accidents with only equivalent PDO $EPDO_p$ with a weighted number of accidents (Montella 2010).

3.1. Results of the Comparison with HSID Methods

The identification of potentially dangerous places is completed in accordance with each of the previously mentioned HSID methods. Then, a comparison is made of the significantly safe subsections, whose number is limited to the number of subsections that, according to the suggested methodology, have a significance factor below (α < 0.05). In this case, 35 subsections are identified to be the least safe according to the suggested methodology, i.e., using the comparative methods described above. The results of the comparison of the suggested method with the other HSID methods show, to a certain extent, both good and bad overlaps in the number of subsections identified to be potentially least safe. The results of the suggested method mostly match (with 62.9% matching), those of the method for weighting traffic accidents $EPDO_p$ currently used in the Serbia; 28.6% match those for the method using the total number of traffic accidents CF, and there is a 54.3% match with the method using the total number of traffic accidents with casualties CF_1.

The suggested method, whose input accident data rely on an ordered numerical marking of traffic accidents, largely matches the $EPDO_p$ weighting method (with method CF_1 the next closest match), which indicates that the suggested method adequately represents the seriousness of traffic accidents and that it also includes the social costs of traffic accidents.

Conclusions

Within this study, a new methodology for the identification of dangerous locations on roads (disjunctive subsections) is presented, which uses the values of mathematical expectations ε, variance values v and F values in a continual analysis of variance – ANOVA. Due to the testing of the proposed hypothesis $H(v_k = V_k)_\alpha$, a comparison of the variance values V of each observed subsection k and the variance v of its complementary group K played an important role in determining the significance values, whose adopted limit in this paper was $\alpha = 0.05$.

The limiting value for significance was $\alpha < 0.05$, and we can increase or decrease the value if we wish to identify a larger or smaller number of potentially dangerous places. In other words, the problem of the wide zone of 'undetermined' subsections is solved by increasing the significance threshold, which shrinks the zone of 'undetermined' subsections.

The significantly safe and unsafe subsections are determined based on limiting values of the mathematical expectation ε and F value for a continual analysis of variance – ANOVA, whose limit values we can also reduce and increase according to predetermined needs.

For the purpose of evaluating the results, the output of the suggested method is compared with four HSID methods. The comparison showed that the number of significant unsafe subsections identified by the suggested method largely matches those of the weighting methods $EPDO_p$ (62.3%), CF_1 (54.3%), and $EPDO_n$ (34.3%), and the unweighted method CF (28.6%), which tells us that the suggested method (SM) has potential and can form the basis of a modern method for the identification of unsafe subsections that also accounts for the frequency and severity of accidents. At the same time, the suggested method provides a very good visualization, such as in Figs 3–6, of which subsections are significantly different (safest and most unsafe subsections) and which subsections are not significantly different.

Similar to the research of Bíl *et al.* (2013), the suggested method illustrated great simplicity regarding the input data necessary for analysis because it uses only the frequency and type of traffic accidents along with the kilometre markings of the road where the accident took place. In addition to the application of the method to rural areas (open roads), there is the possibility for testing, verification and expansion of the model for other roads with different characteristics (roads within city limits, highways, roads reserved for motor vehicles and other roads).

Further development of the presented method needs to be directed towards calibration of the observed numerical values, optimization of the subsection length (the length of subsection influences the variance value) and towards the application of continual analysis of variance on non-disjunction groups of subsections.

Finally, by presenting this method, we wish to attract the attention of other researchers, opening the door for more reliable interpretations and evaluations of the quality of results based on traffic accident frequency in an effort to identify dangerous locations (subsection) on roads.

References

AASHTO. 2010. *Highway Safety Manual*. 1st edition. American Association of State Highway and Transportation Officials (AASHTO), Washington, DC.

Abdel-Aty, M. A.; Radwan, A. E. 2000. Modeling traffic accident occurrence and involvement, *Accident Analysis & Prevention* 32(5): 633–642. https://doi.org/10.1016/S0001-4575(99)00094-9

Aguero-Valverde, J. 2013. Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: comparing the precision of crash frequency estimates, *Accident Analysis & Prevention* 50: 289–297. https://doi.org/10.1016/j.aap.2012.04.019

Alver, Y.; Demirel, M. C.; Mutlu, M. M. 2014. Interaction between socio-demographic characteristics: traffic rule violations and traffic crash history for young drivers, *Accident Analysis & Prevention* 72: 95–104. https://doi.org/10.1016/j.aap.2014.06.015

Anastasopoulos, P. C.; Mannering, F. L. 2009. A note on modeling vehicle accident frequencies with random-parameters count models, *Accident Analysis & Prevention* 41(1): 153–159. https://doi.org/10.1016/j.aap.2008.10.005

Anđelković, D.; Antić, B.; Pešić, D.; Subotić, M. 2014. Polazne osnove u identifikaciji opasnih mesta na putevima [Fundamentals for identification of dangerous places on the roads], *Put i saobraćaj* [*Journal of Road and Traffic Engineering*] (2): 45–52. (in Serbian).

Antić, B.; Pešić, D.; Vujanić, M.; Lipovac, K. 2013. The influence of speed bumps heights to the decrease of the vehicle speed – Belgrade experience, *Safety Science* 57: 303–312. https://doi.org/10.1016/j.ssci.2013.03.008

Bíl, M.; Andrášik, R.; Janoška, Z. 2013. Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation, *Accident Analysis & Prevention* 55: 265–273. https://doi.org/10.1016/j.aap.2013.03.003

Cafiso, S.; Di Graziano, A.; Di Silvestro, G.; La Cava, G. 2008. Safety performance indicators for local rural roads: comprehensive procedure from low-cost data survey to accident prediction model, in *TRB 87th Annual Meeting Compendium of Papers DVD*, 13–17 January 2008, Washington, DC, US, 1–19.

Carey, J. 2001. *Arizona Local Government Safety Project Analysis Model*. Final Report 504. Arizona Department of Transportation, Phoenix, AZ, US. 136 p. Available from Internet: https://apps.azdot.gov/ADOTLibrary/publications/project_reports/PDF/AZ504.pdf

Cheng, W.; Washington, S. P. 2005. Experimental evaluation of hotspot identification methods, *Accident Analysis & Prevention* 37(5): 870–881. https://doi.org/10.1016/j.aap.2005.04.015

Connors, R. D.; Maher, M.; Wood, A.; Mountain, L.; Ropkins, K. 2013. Methodology for fitting and updating predictive accident models with trend, *Accident Analysis & Prevention* 56: 82–94. https://doi.org/10.1016/j.aap.2013.03.009

DoT. 2006. *2005 Valuation of the Benefits of Prevention of Road Accidents and Casualties*. Highways Economic Note No. 1. Department for Transport (DoT), London, UK. 13 p.

El-Basyouny, K.; Sayed, T. 2010. Application of generalized link functions in developing accident prediction models, *Safety Science* 48(3): 410–416. https://doi.org/10.1016/j.ssci.2009.09.007

Elvik, R. 1988. Some difficulties in defining populations of "entities" for estimating the expected number of accidents, *Accident Analysis & Prevention* 20(4): 261–275. https://doi.org/10.1016/0001-4575(88)90054-1

Ferreira, S.; Couto, A. 2013. Traffic flow-accidents relationship for urban intersections on the basis of the translog function, *Safety Science* 60: 115–122. https://doi.org/10.1016/j.ssci.2013.07.007

Geedipally, S. R.; Lord, D.; Dhavala, S.S. 2014. A caution about using deviance information criterion while modeling traffic crashes, *Safety Science* 62: 495–498. https://doi.org/10.1016/j.ssci.2013.10.007

Gregoriades, A.; Mouskos, K. C. 2013. Black spots identification through a Bayesian networks quantification of accident risk index, *Transportation Research Part C: Emerging Technologies* 28: 28–43. https://doi.org/10.1016/j.trc.2012.12.008

Harwood, D. W.; Council, F. M.; Hauer, E.; Hughes, W. E.; Vogt, A. 2000. *Prediction of the Expected Safety Performance of Rural Two-Lane Highways*. Publication No. FHWA-RD-99-207. Federal Highway Administration (FHWA), US Department of Transportation, Washington, DC, US. 200 p. Available from Internet: https://www.fhwa.dot.gov/publications/research/safety/99207/99207.pdf

Harwood, D. W.; Torbic, D. J.; Richard, K. R.; Meyer, M. M. 2010. *SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites*. FHWA-HRT-10-063. Federal Highway Administration (FHWA). 305 p. Available from Internet: http://www.dot.ca.gov/newtech/researchreports/reports/2010/final_report_task_1601.pdf

Hauer, E. 1997. *Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Pergamon. 289 p.

Heydecker, B. G.; Wu, J. 2001. Identification of sites for road accident remedial work by Bayesian statistical methods: an example of uncertain inference, *Advances in Engineering Software* 32(10–11): 859–869. https://doi.org/10.1016/S0965-9978(01)00037-0

Hinde, J.; Demétrio, C. G. B. 1998. Overdispersion: models and estimation, *Computational Statistics & Data Analysis* 27(2): 151–170. https://doi.org/10.1016/S0167-9473(98)00007-3

Jiang, X.; Abdel-Aty, M.; Alamili, S. 2014. Application of Poisson random effect models for highway network screening, *Accident Analysis & Prevention* 63: 74–82. https://doi.org/10.1016/j.aap.2013.10.029

Jin, T. G.; Saito, M.; Eggett, D. L. 2008. Statistical comparisons of the crash characteristics on highways between construction time and non-construction time, *Accident Analysis & Prevention* 40(6): 2015–2023. https://doi.org/10.1016/j.aap.2008.08.024

Kwon, O. H.; Park, M.J.; Yeo, H.; Chung, K. 2013. Evaluating the performance of network screening methods for detecting high collision concentration locations on highways, *Accident Analysis & Prevention* 51: 141–149. https://doi.org/10.1016/j.aap.2012.10.019

Lipovac, K.; Jovanović, D.; Vuksanović, B. 2010. Uporedna analiza identifikacije opasnih mesta i rizičnih deonica na državnim putevima R Srbije, in *X međunarodni simpozijum 'Prevencija saobraćajnih nezgoda na putevima 2010'*, 21–22 Oktobar 2010, Novi Sad, Srbija. (in Serbian).

Lord, D. 2008. Methodology for estimating the variance and confidence intervals for the estimate of the product of baseline models and AMFs, *Accident Analysis & Prevention* 40(3): 1013–1017. https://doi.org/10.1016/j.aap.2007.11.008

Lord, D.; Miranda-Moreno, L. F. 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective, *Safety Science* 46(5): 751–770. https://doi.org/10.1016/j.ssci.2007.03.005

Manner, H.; Wünsch-Ziegler, L. 2013. Analyzing the severity of accidents on the German autobahn, *Accident Analysis & Prevention* 57: 40–48. https://doi.org/10.1016/j.aap.2013.03.022

Miaou, S.-P.; Lord, D. 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes methods, *Transportation Research Record: Journal of the Transportation Research Board* 1840: 31–40. https://doi.org/10.3141/1840-04

Miaou, S.-P.; Song, J. J. 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence, *Accident Analysis & Prevention* 37(4): 699–720. https://doi.org/10.1016/j.aap.2005.03.012

Miranda-Moreno, L. F.; Labbe, A.; Fu, L. 2007. Bayesian multiple testing procedures for hotspot identification, *Accident Analysis & Prevention* 39(6): 1192–1201. https://doi.org/10.1016/j.aap.2007.03.008

Montella, A. 2010. A comparative analysis of hotspot identification methods, *Accident Analysis & Prevention* 42(2): 571–581. https://doi.org/10.1016/j.aap.2009.09.025

Oh, J.; Washington, S.; Lee, D. 2010. Property damage crash equivalency factors to solve crash frequency-severity dilemma: case study on South Korean rural roads, *Transportation Research Record: Journal of the Transportation Research Board* 2148: 83–92. https://doi.org/10.3141/2148-10

Okamoto, H.; Koshi, M. 1989. A method to cope with the random errors of observed accident rates in regression analysis, *Accident Analysis & Prevention* 21(4): 317–332. https://doi.org/10.1016/0001-4575(89)90023-7

PIARC. 2004. *Road Safety Manual 2004*. Recommendations from the World Road Association (PIARC).

Poch, M.; Mannering, F. 1996. Negative binomial analysis of intersection-accident frequencies, *Journal of Transportation Engineering* 122(2): 105–113. https://doi.org/10.1061/(ASCE)0733-947X(1996)122:2(105)

Qu, X.; Yang, Y.; Liu, Z.; Jin, S.; Weng, J. 2014. Potential crash risks of expressway on-ramps and off-ramps: a case study in Beijing, China, *Safety Science* 70: 58–62. https://doi.org/10.1016/j.ssci.2014.04.016

Russo, F.; Biancardo S. A.; Dell'Acqua, G. 2014. Consistent approach to predictive modeling and countermeasure determination by crash type for low-volume roads, *The Baltic Journal of Road and Bridge Engineering* 9(2): 77–87. https://doi.org/10.3846/bjrbe.2014.10

Sadeghi, A.; Ayati, E.; Neghab, M. P. 2013. Identification and prioritization of hazardous road locations by segmentation and data envelopment analysis approach, *Promet – Traffic&Transportation* 25(2): 127–136.

Savolainen, P. T.; Mannering, F. L.; Lord, D.; Quddus, M. A. 2011. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives, *Accident Analysis & Prevention* 43(5): 1666–1676. https://doi.org/10.1016/j.aap.2011.03.025

Shen, J.; Gan, A. 2003. Development of crash reduction factors: methods, problems, and research needs, *Transportation Research Record: Journal of the Transportation Research Board* 1840: 50–56. https://doi.org/10.3141/1840-06

Sokolovskij, E.; Prentkovskis, O. 2013. Investigating traffic accidents: the interaction between a motor vehicle and a pedestrian, *Transport* 28(3): 302–312. https://doi.org/10.3846/16484142.2013.831771

Stern, E.; Zehavi, Y. 1990. Road safety and hot weather: a study in applied transport geography, *Transactions of the Institute of British Geographers* 15(1): 102–111. https://doi.org/10.2307/623096

Tegge, R. A.; Jo, J.-H.; Ouyang, Y. 2010. *Development and Application of Safety Performance Functions for Illinois*. FHWA-ICT-10-066. Illinois Department of Transportation, Springfield, IL, US. 181 p.

Tunaru, R. 2002. Hierarchical Bayesian models for multiple count data, *Austrian Journal of Statistics* 31(2–3): 221–229.

Vadlamani, S.; Chen, E.; Ahn, S.; Washington, S. 2011. Identifying large truck hot spots using crash counts and PDOEs, *Journal of Transportation Engineering* 137(1): 11–21. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000183

Vistisen, D. 2002. *Models and Methods for Hot Spot Safety Work*: PhD thesis. Technical University of Denmark. 168 p.

Wang, C.; Quddus, M. A.; Ison, S. G. 2013. The effect of traffic and road characteristics on road safety: a review and future research direction, *Safety Science* 57: 264–275. https://doi.org/10.1016/j.ssci.2013.02.012

Washington, S.; Haque, M.; Oh, J.; Lee, D. 2014. Applying quantile regression for modeling equivalent property damage only crashes to identify accident blackspots, *Accident Analysis & Prevention* 66: 136–146. https://doi.org/10.1016/j.aap.2014.01.007

Weiss, H. B.; Kaplan, S.; Prato, C. G. 2014. Analysis of factors associated with injury severity in crashes involving young New Zealand drivers, *Accident Analysis & Prevention* 65: 142–155. https://doi.org/10.1016/j.aap.2013.12.020

Yu, H.; Liu, P.; Chen, J.; Wang, H. 2014. Comparative Analysis of the spatial analysis methods for hotspot identification, *Accident Analysis & Prevention* 66: 80–88. https://doi.org/10.1016/j.aap.2014.01.017

Zein, S. 2004. *Canadian Guide to In-Service Road Safety Reviews*. Transportation Association of Canada, Ottawa, Ontario, Canada. 232 p.

Zou, Y.; Geedipally, S. R.; Lord, D. 2013. Evaluating the double Poisson generalized linear model, *Accident Analysis & Prevention* 59: 497–505. https://doi.org/10.1016/j.aap.2013.07.017