Accident Analysis and Prevention 92 (2016) 256–264

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Macroscopic hotspots identification: A Bayesian spatio-temporal interaction approach

Ni Dong[a,b], Helai Huang[b,*], Jaeyoung Lee[c], Mingyun Gao[d], Mohamed Abdel-Aty[c]

[a] School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, 610031, China
[b] Urban Transport Research Center, School of Traffic and Transportation Engineering, Central South University, Changsha, 410075, China
[c] Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL, 32816-2450, United States
[d] School of Science, Wuhan University of Technology, Wuhan, Hubei, 430063, China

ARTICLE INFO

Article history:
Received 30 August 2015
Received in revised form 5 March 2016
Accepted 3 April 2016
Available online 23 April 2016

Keywords:
Bayesian spatio-temporal interaction model
Hotspot identification
Ranking criteria

ABSTRACT

This study proposes a Bayesian spatio-temporal interaction approach for hotspot identification by applying the full Bayesian (FB) technique in the context of macroscopic safety analysis. Compared with the emerging Bayesian spatial and temporal approach, the Bayesian spatio-temporal interaction model contributes to a detailed understanding of differential trends through analyzing and mapping probabilities of area-specific crash trends as differing from the mean trend and highlights specific locations where crash occurrence is deteriorating or improving over time. With traffic analysis zones (TAZs) crash data collected in Florida, an empirical analysis was conducted to evaluate the following three approaches for hotspot identification: FB ranking using a Poisson-lognormal (PLN) model, FB ranking using a Bayesian spatial and temporal (B-ST) model and FB ranking using a Bayesian spatio-temporal interaction (B-ST-I) model. The results show that (a) the models accounting for space-time effects perform better in safety ranking than does the PLN model, and (b) the FB approach using the B-ST-I model significantly outperforms the B-ST approach in correctly identifying hotspots by explicitly accounting for the space-time variation in addition to the stable spatial/temporal patterns of crash occurrence. In practice, the B-ST-I approach plays key roles in addressing two issues: (a) how the identified hotspots have evolved over time and (b) the identification of areas that, whilst not yet hotspots, show a tendency to become hotspots. Finally, it can provide guidance to policy decision makers to efficiently improve zonal-level safety.

1. Introduction

Identifying a crash hotspot is a vital task for a safety improvement program, especially when highway agencies can afford to examine and improve only a limited number of road sites. Numerous studies have been conducted on investigating the suitability of various statistical methodologies for the development of effective hotspot identification and ranking criteria. A naïve statistical approach based on the estimation of safety by using historical traffic crash records, such as the crash frequency method (Deacon et al., 1975), the crash rate method, the rate quality control method (Stokes and Mutabazi, 1996), the crash severity methods and the safety index method (Tamburri et al., 1970), is often found to have serious limitations. This includes the fact that the approach is quite sensitive to random variations due to the random fluctuation and rarity of crash events, the absence of consideration of a phenomenon known as regression-to-the-mean (Hauer, 1997), the lack of examining crash dispersion (Elvik, 2007), the false assumption of a linear relation between crash count and traffic volume, etc.

Model-based approaches have offered several advantages to these estimation problems due to an increased precision by "borrowing strengths" across similar sites based on available auxiliary variables, especially in the case of small sample sizes (Miaou and Song, 2005; Cheng and Washington, 2005, 2008; Huang et al., 2009; Lee et al., 2015). Recently, the approaches based on empirical Bayesian (EB) adjusted safety performance measures have become popular. Such approaches make joint use of two clues to the safety performance of a road entity by accounting for both crash history and the predicted crash frequency of similar sites (Carlin and Louis, 1996). Thus, it is clear that the EB method is expected to be more reliable as compared to the traditional method as shown by several researchers (Hauer and Persaud, 1984; Hauer, 1996, 1997; Hauer

* Corresponding author.
E-mail addresses: dongni@home.swjtu.edu.cn (N. Dong), huanghelai@csu.edu.cn (H. Huang), jaeyoung@knight.ucf.edu (J. Lee), wh14.gao@126.com (M. Gao), mabdel@mail.ucf.edu (M. Abdel-Aty).

http://dx.doi.org/10.1016/j.aap.2016.04.001
0001-4575/

et al., 2002; Elvik, 2007). Nevertheless, the EB method may be criticized for implicitly requiring a large sample size of data to develop the reliability of safety performance functions and for ignoring the "uncertainty" of associations of covariates and safety (Lan and Persaud, 2011; El-Basyouny and Sayed, 2013).

Within the model-based approaches, the full Bayesian (FB) method has been explored in terms of whether it more reliably identifies hotspots as compared to EB (Huang et al., 2009; Jiang et al., 2014). Specially, the advantage of the FB approach is its explicit use of probability for quantifying the uncertainty of the associations of crash risk and various risk factors, which could be accommodated to estimate a posterior distribution representing the final safety estimate for a specific site on which various ranking criteria could be based (Carlin and Louis, 1996).

A further merit of FB is that it is able to allow probabilistic and functional structures of complex space-time heterogeneities of crash occurrence to be more realistically considered and tested by use of hierarchical model specification in statistical analysis. Several recent studies have indicated that crash rates may be better fitted by explicitly accounting for the spatial and temporal effects by using a hierarchical modeling technique (Miaou and Song, 2005; Huang and Abdel-Aty, 2010; Dong et al., 2014, 2015; Xu and Huang, 2015; Wang and Huang, 2016). As demonstrated by Levine et al. (1995), spatial variations in crashes exist among observations over space as crash data are typically collected with reference to a location dimension. In a similar vein, there can be a correlation over time because many of the observed effects associated with a specific site may remain the same over time or fluctuate periodically (Congdon, 2003; Lord and Mannering, 2010). Hence, the consideration of both space and time effects are of fundamental importance in the study of hotspot identification as evidenced both empirically and theoretically (Huang et al., 2009; Jiang et al., 2014).

Unfortunately the aforementioned FB hierarchical modeling approaches for hotspot identification deal with the spatial and temporal effects in crash data as distinct entities, thus ignoring the necessary interaction of space and time to crash occurrence. Most previous studies assumed the impact of temporal effects on crash occurrence was constant over space; that is, a time effect, which represents the variation of crash rates in time might be stationary from area to area. But this latent assumption has been denied in many recent studies (Law et al., 2014; Li et al., 2014) in which a spatio-temporal interaction effect arises when events located relatively close in geographic space occur at about the same time. In other words, it is possible that the variation of the rate in time has larger impacts in certain spatial units and has smaller impacts in others. Consequently, the possibility of accounting for this spatio-temporal interaction by allowing the time-trend to vary from area to area has considerable potential (Richardson et al., 2006).

Variations of the time-trend across areas could be referred to as space-time variation. In 1995, Bernardineilli et al. (1995) proposed a Bayesian hierarchical spatio-temporal interaction model to address this issue by adapting ideas from time series and spatial modeling to a joint analysis of space-time variation. This leads to a more precise estimation of the variability in parameters by accounting for random variability, especially in the case of small sample sizes. Particularly in the context of hotspot identification, Bayesian spatio-temporal interaction analysis contributes to a detailed understanding of differential trends through analyzing and mapping probabilities of site-specific crash trends differing from the mean trend. It could be adopted to highlight specific locations where crash occurrence is deteriorating or improving over time. Rather than simply identifying hotspots that occurred in the past at one period, the Bayesian spatio-temporal interaction approach provides insight into the processes influencing changing crash rates over time and enables engineering evaluation and safety improvement to predict where crashes might be increasing in the future. Thus, from a traffic planning and operational point of view, there is a need to explore the use of the Bayesian spatio-temporal interaction model in ranking and hotspot identification in the context of macroscopic safety analysis. This can provide guidance to policy decision makers to efficiently improve zonal-level safety.

This study aims at employing the Bayesian spatio-temporal interaction approach to analyze crash trends as area-specific and to identify crash hotspots in regional safety analysis. We propose a Bayesian spatio-temporal interaction model in which both an area-specific intercept and trends over time are modeled as random effects and interaction between them is allowed for. Comparison is conducted with the Bayesian spatial and temporal model with respect to safety ranking and hotspot screening to provide an empirical evaluation of alternative hotspot identification approaches.

2. Methodology

This section presents the alternative FB hierarchical approaches (i.e., the Poisson Lognormal model, the Bayesian spatial and temporal model, the Bayesian spatio-temporal interaction model), including model specification, comparison and assessment and a decision parameter in ranking and identifying hotspots.

2.1. Full Bayesian hierarchical modeling in hotspot identification

The essential characteristic of the FB approach in hotspot identification is the establishment of a probabilistic functional form of crash frequency and area-specific diverse risk factors (i.e., a safety performance function and a crash prediction model). The Poisson model has been applied for fitting the crash rate. However, the underlying assumption of the Poisson distribution of variance equal to mean is often violated in the crash count data. To account for this issue of over-dispersion, the over-dispersed Poisson model has generally been used as follows:

$$Y_{it}|\lambda_{it} \sim \text{Poisson}(\lambda_{it}) = \text{Poisson}(\mu_{it}\, e_{it})$$

$$\log(\mu_{it}) = \alpha + \boldsymbol{X}_{it}{}'\beta + \nu$$

where Y_{it} denotes the crash count at the i th areas ($i=1, \ldots N$) in the t th time period ($t=1, \ldots T$) as Poisson distributed with the mean crash frequency λ_{it}. Given the exposure component e_{it} and the observed risk factors X_{it} at a specific area, the mean crash rate μ_{it} varies by a stochastic component ν.

Conveniently, ν could be assumed independent of X_{it} and also independent of each other for different observations, denoted as ε_{it}. Standard Poisson-lognormal (PLN) models are obtained by specifying ε_{it} with a lognormal distribution.

Model 1. Poisson-lognormal model:

$$\log(\mu_{it}) = \alpha + \boldsymbol{X}_{it}{}'\boldsymbol{\beta} + \varepsilon_{it}$$

$$\varepsilon_{it} \sim \text{normal}(0,\sigma_{\varepsilon}^2)$$

Although the PLN model presented above is capable of capturing unstructured over-dispersion, it largely ignores any structured heterogeneities due to the spatial and temporal effects of crash data (Miaou and Song, 2005). To explicitly model the structured heterogeneities introduced in the data collection and clustering process, a hierarchical modeling technique has been found to be a better alternative in several traffic safety studies (Huang and Abdel-Aty, 2010; Xu et al., 2014; Zeng and Huang, 2014). For this purpose, a Bayesian spatial and temporal (B-ST) model could be proposed by replacing the cross-observation component ε_{it} with the spatial and temporal random effects component ϑ_{it} in the link function. Possible spatial and temporal dependence can be reflected in the