Analytic Methods in Accident Research 15 (2017) 29–40

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Multivariate space-time modeling of crash frequencies by injury severity levels

Xiaoxiang Ma [a], Suren Chen [b,*], Feng Chen [a]

[a] Key Laboratory of Road & Traffic Engineering of the Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, China
[b] Department of Civil & Environmental Engineering, Colorado State University, Fort Collins, CO 80523, United States

ARTICLE INFO

Article history:
Received 15 September 2016
Received in revised form 10 June 2017
Accepted 10 June 2017
Available online 19 June 2017

Keywords:
Multivariate models
Spatial modeling
Space-time
Fine temporal scale

ABSTRACT

Road traffic crashes threaten thousands of drivers every day and significant efforts have been put forth to reduce the number and mitigate the impacts of traffic crashes. Although the last decade has witnessed substantial methodological improvements in crash prediction modelling, several methodological challenges still remain in terms of predicting crash frequencies of different injury severity levels. These challenges include spatial correlation and/or heterogeneity, temporal correlation and/or heterogeneity, and correlations between crash frequencies of different injury severity level. A framework of Bayesian multivariate space-time model is developed to address these challenges. A series of multivariate space-time models are proposed under the Full Bayesian framework with different assumptions on the spatial and temporal random effects. In addition to the ability to consider both temporal and spatial trends, the proposed framework is also capable of addressing complex correlations between crash types. It allows the underlying unobserved heterogeneity to be better captured and enables borrowing strength across spatial units and time points, as well as over crash types. The proposed methodology is illustrated using one-year daily traffic crash data from the mountainous interstate highway I70 in Colorado, which is categorized into no injury crash and injury crash. The results show that multivariate space-time model outperforms other alternatives, including multivariate random effects model and multivariate spatial models. The model comparison results highlight the importance to properly account for spatial effects, temporal effects and correlations between crash types.

1. Introduction

Road traffic crashes impose enormous emotional and economic burdens on human society due to the associated physical suffering, losses in life and financial burdens. To reduce the number and mitigate the impacts of traffic crashes, numerous studies have been conducted to improve the understanding and prediction of traffic crashes. As relatively rare events, highway traffic crashes are usually assessed by aggregating the crash counts over a certain time period (week, month, etc.) and within a specific geographical range (e.g. roadway segment, intersection). Since traffic crash data is always associated with certain spatial and temporal dimensions, both spatial and temporal correlations/heterogeneities are often present within the data. In addition, when crash frequencies of different injury severity levels are to be modeled, dependence among the counts

* Corresponding author.
E-mail addresses: shaun.ma.tj@gmail.com (X. Ma), suren.chen@colostate.edu (S. Chen), chenfeng.csu@gmail.com (F. Chen).

http://dx.doi.org/10.1016/j.amar.2017.06.001
2213-6657/

for a specific injury severity should also be accounted for. Despite the methodological advances during the past years, following methodological challenges still remain in terms of predicting crash frequencies of different injury severity levels: 1) spatial correlation and/or heterogeneity; 2) temporal correlation and/or heterogeneity; and 3) correlations between crash frequencies of different injury severity levels. In a recent review paper, Mannering et al. (2016) summarized the unobserved heterogeneity issues associated with accident data and pointed out that the spatial and temporal aspects of accident data are usually ignored. Although the last decade has witnessed substantial methodological improvements in crash prediction modelling, methods that can appropriately address all the above-mentioned challenges are not available. The model estimation results and the following inferences rely heavily on the underlying assumptions about crash prediction model and data structure. It therefore becomes critical to develop more advanced crash prediction models that can address these challenges appropriately.

Despite that crash data is in nature spatially and temporally distributed, most existing studies aggregated data over an extended time period (e.g. one or several years). These studies usually addressed the over-dispersion problem but assumed independence of observations from different roadway entities (Shankar et al., 1995; Ma et al., 2008; Malyshkina and Mannering, 2010). Neither spatial nor temporal correlations were considered in these literatures. Some studies attempted to use data with multiple temporal observations generated for each roadway entity, and treated the data as panel data. In that case, it results in a temporal correlation because unobserved heterogeneity associated with a specific roadway entity may be similar from time to time. Different panel data models were proposed to address temporal correlation, such as random effects models (Shankar et al., 1998; Chen et al., 2014), negative multinomial models (Shankar et al., 2003) and generalized estimating equations (Lord and Persaud, 2000). By the same token, these models were also applied to address spatial correlation (e.g. Wang and Abdel-Aty, 2006). However, a major drawback of these methods is the lack of correlation structure which can explicitly define the spatial and/or temporal correlations. As a result, these models do not easily lend themselves to explicitly analyzing temporal trend and spatial pattern of crash risks.

Recently, spatial modeling has gained wide recognition in evaluating traffic safety risks on various types of roadway entities. Spatial modeling deals with spatial correlation which may reflect unmeasured confounders. Different approaches were proposed for spatial modeling, including intrinsic conditional autoregressive (CAR) model (MacNab, 2004; Aguero-Valverde and Jovanis, 2008; Wang and Kockelman, 2013; Xie et al., 2014; Lee et al., 2015), spatial autoregressive model (Quddus, 2008), spatial error model (Quddus, 2008), and geographic weighted Poisson regression (Hadayeghi et al., 2010). Most of these studies adopted CAR model, which enables "spatial smoothing" by borrowing strength from neighboring sites. This was possibly because CAR model takes advantage of the flexibility of Bayesian hierarchical framework to incorporate spatial correlation and can be readily extended to address more complicated models. The 'Besag–York–Mollie' (BYM) model, an extension of CAR model, was proposed to address spatial correlation from neighboring effects as well as spatial heterogeneity (Besag et al., 1991). Many studies demonstrated that the BYM model is a proper tool for spatial modeling of traffic crashes, where study units span a large geographical area or comprise highways with varying functional classes (Aguero-Valverde and Jovanis, 2008; Quddus, 2008, 2010; Barua et al., 2014). Aguero-Valverde (2013) explored the multivariate conditional autoregressive (MCAR) model and the result indicated MCAR is superior over its univariate counterparts. Barua et al. (2014) developed three MCAR models and showed that the multivariate model with both spatial correlation and heterogeneity provides a better fit than other ones. Barua et al. (2016) later extended the MCAR model to a multivariate random parameter setting. Huang et al. (2017) proposed a MCAR model to account for spatial correlation and simultaneity among motor vehicle, bicycle and pedestrian crashes. Their results indicated the multivariate model outperformed the univariate one. Similar MCAR models were also applied for crash modeling at varying spatial units (segment, intersection, zones, etc.) in some other literature (Wang and Kockelman, 2013; Lee et al., 2015; Aguero-Valverde et al., 2016). Although abovementioned studies contributed to state-of-the-art methodologies for spatial modeling and offered valuable insights with regard to crash risk, none of them has accounted for temporal variation at the same time.

In contrast to the abundance of literature in spatial models, space-time studies in the traffic safety analysis are very limited. The first reported attempt to explore space-time model in traffic crash analysis was conducted by Miaou et al. (2003). In their study, Miaou et al. (2003) developed various county-level space-time crash risk models for Texas and they argued that temporal component was better modeled with fixed effects than with first-order autoregressive (AR1). Wang et al. (2013) developed a spatio-temporal model to study the impact of congestion on traffic safety. Their results showed that models with fixed time effects and first order random walk (RW1) time effects produced similar results for fatal and serious injury accidents. Aguero-Valverde and Jovanis (2006) adopted a space-time model proposed by Bernardinelli et al. (1995) to study county-level injury crashes in Pennsylvania. This model defines a spatio-temporal interaction that allows for different temporal trends for different locations. However, it restricts the temporal trends to be linear, which is clearly unrealistic for traffic safety studies especially those with fine temporal resolutions. Dong et al. (2016) also used the same space-time model for hotspot identification at the scale of traffic analysis zone (TAZ) to account for possible space-time interaction. There are, however, two major limitations of the space-time studies discussed above. First, they were all conducted over an extended time period (i.e. a year) with limited time points and thus suffer from loss of important time-varying information. Space-time study in fine temporal scales is yet to be conducted to bring new insights into crash analysis. Secondly, only separate univariate analyses were carried out even when crashes of different injury levels were studied. It has been well established in the literature that the correlation among injury severities is important and thus needs to be considered to avoid the potential statistical problem (Mannering and Bhat, 2014). This problem is likely to be carried over to space-time setting, creating a need for joint models of different injury severity levels with great complexity. Multivariate space-time analyses in traffic