Accident Analysis and Prevention 119 (2018) 263–273

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

A Bayesian multivariate hierarchical spatial joint model for predicting crash counts by crash type at intersections and segments along corridors

Saif A. Alarifi*, Mohamed Abdel-Aty, Jaeyoung Lee

University of Central Florida, Department of Civil, Environmental, and Construction Engineering, Orlando, FL 32816, United States

ARTICLE INFO

Keywords:
Multivariate hierarchical spatial model
Corridor safety analysis
Joint model
Crash
types

ABSTRACT

The safety and operational improvements of corridors have been the focus of many studies since they carry most traffic on the road network. Estimating a crash prediction model for total crash counts identifies the crash risk factors that are associated with crash counts at a specific type of road entity. However, this may not reveal useful information to detect the road problems and implement effective countermeasures. Therefore, investigating the contributing factors for crash counts by different types is of great importance. This study aims to provide a good understanding of the contributing factors to crash counts by different types at intersections and roadway segments along corridors. Data from 255 signalized intersections and 220 roadway segments along 20 corridors have been used for this study. The investigated crash types include same direction, angle and turning, opposite direction, non-motorized, single vehicle, and other multi-vehicle crashes. Two models have been estimated, which are multivariate hierarchical Poisson-lognormal (HPLN) spatial joint model and univariate HPLN spatial joint model. The significant variables include exposure measures and some geometric design variables at intersection, roadway segment, and corridor levels. The results revealed that the multivariate HPLN spatial joint model outperforms the univariate HPLN spatial joint model. Also, the correlations among crash counts of most types exist at individual road entity and between adjacent entities. Additionally, the significant explanatory variables are different across crash types, and the magnitude of the parameter estimates for the same independent variable is different across crash types. The results emphasize the need for estimating crash counts by type in a multivariate form to better detect the problems and provide appropriate countermeasures.

1. Introduction

Corridors safety analysis is a major concern since they carry most traffic on the road network, and their safety and operational improvements have been the focus of many studies. Corridors mainly contain signalized intersections and roadway segments, and analyzing safety at corridors by having both components provides a good understanding of the crash risk factors at intersections and roadway segments along corridors. Crash risk factors along corridors include some geometric design features, traffic flow characteristics, and traffic control and signal information. Different types of countermeasures at the corridor level have been proposed (e.g. signal coordination, access management, and median treatments) to enhance the safety and operational efficiency at corridors.

Developing a crash prediction model for total crash counts identifies the crash risk factors that are associated with crash frequencies at specific locations. However, to implement effective countermeasures, it is required to investigate crash counts of different types. Also, different crash types are associated with traffic and geometric characteristics in different ways (Kim et al., 2006, 2007). Therefore, investigating the contributing factors to crash frequencies for different crash types is of great importance since it provides better explanatory power compared to a single total crash counts model. However, estimating a separate crash frequency model for each crash type may result in inefficient and biased parameters because different crash types may share unobserved or omitted variables (Ye et al., 2009; Aguero-Valverde et al., 2016). As a result, estimating a multivariate model, where crash counts by different types are modeled simultaneously, is necessary to handle the common unobserved factors and provide more accurate parameter estimates.

While analyzing intersections and roadway segments along corridors, there is a potential presence of spatial correlations among the road entities since these road entities have similarities in the roadway and driver characteristics, and accounting for this spatial correlation in the model is essential especially if the distance between the road entities is not large. Ignoring the spatial correlation may lead to biased model parameters (Guo et al., 2010; Lesage and Pace, 2009). In addition, it has

* Corresponding author.
E-mail address: Saif.a@knights.ucf.edu (S.A. Alarifi).

https://doi.org/10.1016/j.aap.2018.07.026
Received 8 January 2018; Received in revised form 13 July 2018; Accepted 21 July 2018
0001-4575/

been proven that the model performance has improved after considering the spatial effect among the adjacent road entities (Aguero-Valverde and Jovanis, 2010; Hadayeghi et al., 2010; Barua et al., 2014).

This study aims to provide a good understanding of the contributing factors to crash counts by different types at intersections and roadway segments along corridors by estimating a multivariate hierarchical spatial joint model. Also, it aims to evaluate the correlations among crash counts by types at the individual road entity and between adjacent road entities. Lastly, the estimation of univariate model, where the independence across crash counts of different types is assumed, is done for comparison purposes.

2. Literature review

A crash prediction model presents the relationship between the crash occurrence as dependent variable and the contributing factors as explanatory variables. Many advanced modeling approaches have been implemented over the past years to more preciously predict crash counts. Scholars have proposed advanced modeling approaches to overcome the data issues (for example, over-dispersed data, low sample size, low sample mean, and omitted variable bias) and the weaknesses that were found in the traditional modeling approaches (for example, interdependence outcomes, spatial correlation, and functional form). A review of the data issues and the methodological alternatives has been conducted in previous studies (Lord and Mannering, 2010; Mannering and Bhat, 2014).

Multivariate count models have been employed for predicting crash counts by severity levels (Park and lord, 2007; Ma et al., 2008; Aguero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009a; Barua et al., 2014, 2016), collision types (Ye et al., 2009; Mothafer et al., 2016; Aguero-Valverde et al., 2016; Cheng et al., 2017), or transportation mode (Lee et al., 2015; Huang et al., 2017). Many studies have incorporated the spatial correlation into multivariate models to predict crash counts (Wang and Kockelman, 2013; Aguero-Valverde, 2013; Barua et al., 2014; Aguero-Valverde et al., 2016; Huang et al., 2017; Bhat et al., 2017). Scholars have employed the multivariate model for predicting crash counts by different types at roadway segments (Aguero-Valverde et al., 2016; Mothafer et al., 2016) or intersections (Ye et al., 2009; Cheng et al., 2017).

Aguero-Valverde et al. (2016) analyzed a 7–year crash data for 832 rural two-lane roadway segments in Pennsylvania. They have estimated four different models, which are univariate model, univariate spatial model, multivariate model, and multivariate spatial model for crash counts by different types. They found that the multivariate Poisson-lognormal (MPLN) spatial model outperforms the other three models. With respect to explanatory variables, they only used the annual average daily traffic (AADT) as an independent variable. Ye et al. (2009) investigated a 2-year crash data for 165 rural intersections in the state of Georgia. They have estimated univariate and multivariate Poisson counts models for different crash types and found that the multivariate Poisson model provides a better fit of the data compared to the univariate model. Both studies revealed that the presence of shared unobserved factors across collision types is significant and should be considered in crash analysis.

Many corridor studies have explored the contributing factors to crashes at either intersections (Abdel-Aty and Wang, 2006; Guo et al., 2010; Xie et al., 2014) or roadway segments (El-Basyouny and Sayed, 2009b; Wang et al., 2014, 2015). Alarifi et al. (2017) addressed the whole corridor components. They developed a hierarchical joint model with random parameters to simultaneously identify the contributing factors to total crash counts at intersections and roadway segments. Another study that simultaneously modeled crash counts at intersections and roadway segments is by Zeng and Huang (2014). They estimated a spatial joint model to analyze crash counts at intersections and roadway segments at roadway network. Some corridor studies have accounted for spatial correlation to provide more accurate model parameters (Abdel-Aty and Wang, 2006; El-Basyouny and Sayed, 2009b; Guo et al., 2010; Xie et al., 2014). Also, many corridor studies have considered corridor variables in the model (Xie et al., 2014; Wang et al., 2014, 2015; Alarifi et al., 2017). The inclusion of corridor level variables to the hierarchical models incorporate the different levels and provide more reliable results than traditional models (Gelman and Hill, 2007; Huang and Abdel-Aty, 2010).

Wang et al. (2014) investigated 161 roadway segments along eight suburban arterials in China. They estimated a Bayesian hierarchical model for total crashes, and they also estimated a bivariate model for predicting minor and severe injury crash frequencies. It is the only corridor study, which estimated a bivariate model for severity levels, and there is no corridor study that dealt with multivariate modeling. Wang and Abdel-Aty (2006) explored 476 intersections along 41 corridors to model rear-end crash frequencies, and they employed generalized estimating equations with a negative binomial link function considering spatial correlation among the data.

Many studies have investigated the effect of different neighboring structures at micro and macro levels (Aguero-Valverde and Jovanis, 2010; Dong et al., 2015; Wang et al., 2016a, 2016b; Gill et al., 2017). Aguero-Valverde and Jovanis (2010) evaluated the effect of different spatial weight matrices in predicting crashes at roadway segments. The authors found that the roadway segments that belong to the same corridor have stronger correlation. Also, they found that the spatial correlation is important to be considered if the neighboring segments are within one mile. Dong et al. (2015) assessed the spatial effects in predicting crashes at traffic analysis zones using different weight matrices. Also, Gill et al. (2017) investigated many spatial-proximity matrices in predicting crashes at county-level data. Lastly, many studies have used the adjacency-based first-order neighboring structure when incorporating the spatial effect into the model (El-Basyouny and Sayed, 2009c; Wang and Kockelman, 2013; Barua et al., 2016).

In summary, while previous corridor studies have estimated different types of models to predict total, rear-end, minor, and severe injury crash counts, this study aims to a) estimate a multivariate hierarchical spatial joint model, which predicts crash counts by different types for the urban intersections and roadway segments and identifies the contributing factors for crash counts by different types simultaneously; and b) evaluate the correlations of crash frequencies by different crash types.

3. Methodology

This study proposes a Full Bayesian multivariate hierarchical spatial joint model to estimate crash counts by different type at intersections and roadway segments along urban corridors. The data structure in our study can be viewed as two-level hierarchy as shown in Fig. 1.

It can be seen that level one is the corridor level, and level two represents the intersection and roadway segment level.

Poisson-based models have been widely implemented in crash data analysis. The Poisson-lognormal (PLN) model incorporates heterogeneity random effect term to a typical Poisson model to account for

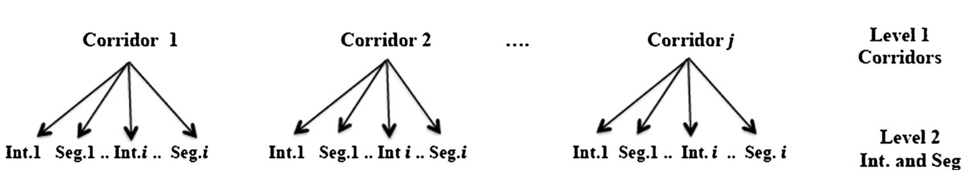

Fig. 1. The Hierarchical Structure of the Data.
a) Intersections.
b) Roadway Segments.