

## Research Article

# The Effects of Traffic Composition on Freeway Crash Frequency by Injury Severity: A Bayesian Multivariate Spatial Modeling Approach

Huiying Wen <sup>1</sup>, Jiaren Sun,<sup>1</sup> Qiang Zeng <sup>1</sup>, Xuan Zhang,<sup>1</sup> and Quan Yuan<sup>2</sup>

<sup>1</sup>School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, Guangdong 510641, China

<sup>2</sup>State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing 100084, China

Correspondence should be addressed to Qiang Zeng; zengqiang@scut.edu.cn

Received 27 March 2018; Accepted 5 July 2018; Published 7 August 2018

Academic Editor: Yair Wiseman

Copyright © 2018 Huiying Wen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study sets out to investigate the effects of traffic composition on freeway crash frequency by injury severity. A crash dataset collected from Kaiyang Freeway, China, is adopted for the empirical analysis, where vehicles are divided into five categories and crashes are classified into no injury and injury levels. In consideration of correlated spatial effects between adjacent segments, a Bayesian multivariate conditional autoregressive model is proposed to link no-injury and injury crash frequencies to the risk factors, including the percentages of different vehicle categories, daily vehicle kilometers traveled (DVKT), and roadway geometry. The model estimation results show that, compared to Category 5 vehicles (e.g., heavy truck), larger percentages of Categories 1 (e.g., passenger car) and 3 (e.g., medium truck) vehicles would lead to less no-injury crashes and more injury crashes. DVKT, horizontal curvature, and vertical grade are also found to be associated with no-injury and/or injury crash frequencies. The significant heterogeneous and spatial effects for no-injury and injury crashes justify the applicability of the proposed model. The findings are helpful to understand the relationship between traffic composition and freeway safety and to provide suggestions for designing strategies of vehicle safety improvement.

## 1. Introduction

As a primary component of roadway transportation system, the safety performance of freeways has attracted much attention in the field of traffic safety research in recent years [1–3]. To reveal crash occurrence mechanism, rank hotspots for safety improvement, and evaluate the safety benefits of countermeasures, crash prediction models (also called “safety performance functions”) are usually developed to link traffic volume, roadway attributes (e.g., length, speed limit, number of lanes, median and shoulder width, vertical grade, and horizontal curvature), and weather conditions (e.g., visibility, temperature, and precipitation) to the crash frequencies on freeway segments [2, 4, 5].

In addition to reducing crash occurrence, mitigating the degree of injury sustained by crash-involved road users is also a major emphasis of freeway agencies. Probably due to the high speed, freeway crashes tend to result in more severe

injury outcomes than those that occurred on other types of roadways (e.g., urban roadway). For example, the *Annual Statistical Report on Roadway Traffic Accidents* [6] shows that the fatality rate of freeway ranks first among all roadway types in China [6]. In 2015, nationwide, one third of roadway crashes resulting in over 10 people killed occurred on freeways. Thus, a more comprehensive evaluation on freeway safety performance necessitates modeling crash frequency by injury severity, which may reveal undetected deficiencies in total crash frequency prediction model and provide a more cost-effective approach for identifying crash hotspots [7, 8].

While there are several studies on analyzing crash frequency by severity, only a few of them focus on freeway safety. Ye et al. [9] propose a simultaneous equations model for analyzing frequencies of no-injury, possible injury, and injury and fatal crashes on 275 multilane freeway segments in the State of Washington. Ma et al. [1] use multivariate space-temporal approaches to modeling daily no-injury and

injury crash counts on the mountainous freeway, I70 in the State of Colorado. In these studies, the roadway geometry and environmental factors are used as the safety predictors. None of them has investigated the effects of traffic composition on crash frequency by injury severity. As indicated by Dinu and Veeraragavan [10], traffic composition can be defined as the involved vehicle types and the proportions of each vehicle type in the mixed traffic flow. Compared to urban traffic, freeway traffic generally consists of more motor vehicle types, varying from motorcycles, passenger cars, coaches, vans, to light/medium/heavy trucks and trailers. There may exist substantial differences in driving behavior, maneuver ability, and crash aggressivity and worthiness among these vehicle types, which would lead to considerable variability in crash risk and injury severity [11–13]. A number of previous studies have demonstrated that traffic composition has statistically significant effects on crash occurrence [3, 10, 14]. Therefore, it would be beneficial for modeling freeway crash frequency by injury severity to incorporate the proportions of each vehicle type in the mixed traffic into the explanatory variables.

From a methodological perspective, crash frequencies at different injury degrees can be modeled either separately or jointly [15]. Since they are often found significantly correlated, joint modeling approaches, including multivariate regression [16, 17], simultaneous equations [8, 9], joint-probability approach [18], two-stage multivariate model [19], and multinomial-generalized Poisson models [20–22], are widely used for analyzing crash frequency by severity. Among these approaches, multivariate regression is able to provide a fully general covariance structure for capturing the underlying correlation while accommodating some other characteristics of crash data, such as overdispersion and spatiotemporal correlation [1]. For freeway crash frequency modeling, spatial correlation is an important issue to be considered, because the safety performance of adjacent segments may be affected by common omitted factors [23]. The multivariate conditional autoregressive (CAR) model is one of the most advanced methods for multivariate spatial modeling under the Bayesian framework. It has been successfully applied to analyze crash frequency by severity or transportation mode at the macro level (e.g., canton and census tract) [24–26] and the micro level (e.g., roadway segment and intersection) [1, 27–29]. With the ability of accounting for heterogeneous and spatial effects and their correlations among crash severities, the multivariate CAR model is expected to improve the accuracy of identifying the contributing factors to freeway crash frequency by severity. Please refer to the review papers, Lord and Mannering [30] and Mannering and Bhat [31], for more descriptions and assessments on novel approaches for modeling crash frequency.

In this research, the key goal is to investigate the effects of traffic composition on freeway crash frequencies at different injury degrees. To achieve this goal, one-year crash-severity frequency data from Kaiyang Freeway in China are collected, and a multivariate CAR model is developed with the proportions of different vehicles and freeway-specific attributes as predictors for predicting crash frequency by severity simultaneously.

The remainder of the paper is as follows. The next section describes the collected data for the empirical analysis. In Section 3, the formulation of the multivariate CAR model is specified and the estimation process of the model is introduced. The results of parameter estimation are discussed in Section 4. Finally, Section 5 presents the concluding remarks and directions for future research.

## 2. Data Preparation and Preliminary Analysis

The crash, traffic, and roadway data on Kaiyang Freeway in Guangdong Province, China, in 2014 are collected for the current analysis. Kaiyang Freeway is about 125 km long, with a median barrier, four lanes, and 120 km/h speed limit along the whole freeway. The roadway geometry data, which cover the detailed information on vertical grade, horizontal curve, and the layout of bridges and ramps, are extracted from the Horizontal and Longitudinal Profile, provided by Guangdong Province Communication Planning and Design Institute Co., Ltd. The other roadway-specific attributes, including median width, lane width, roadway shoulder type and width, and pavement type, are constant along the freeway, thus not considered as safety predictors. According to two criteria—the homogeneity in horizontal and vertical alignments as well as the 150 m minimum length—the freeway is split into 154 segments. A more detailed description on the segmentation method can be found in Zeng et al. [3].

The crash data are obtained from the Highway Maintenance and Administration Management Platform of the Guangdong Transportation Group. The disaggregated crashes are mapped to these segments, according to their locations recorded as the mileages of the freeway in crash reports. In the original records, crash injury severities are categorized into four levels: no injury, slight injury, severe injury, and fatality. Due to the rareness of severe injury and fatality crashes which only account for 5.9% of the total observed crashes, two severity degrees are considered in this analysis: (1) no injury and (2) injury, which refers to slight injury, severe injury, and fatality. For each freeway segment, the aggregated crashes are further divided into two groups according to the crash severity. Therefore, the no-injury and injury crash counts on each segment in 2014 can be obtained.

The traffic data are acquired from the Guangdong Freeway Networked Toll System. The traffic volumes are originally recorded by vehicle category. In the Networked Toll System, vehicles are classified into five categories, based on four vehicle-size parameters: head height, axis number, wheel number, and wheelbase. The specific classification criteria are presented in Table 1. For freeway segment  $i$ , the annual average daily traffic (AADT) is calculated by using the weights 1, 1.5, 2, 3, and 3.5 for vehicle categories 1–5, respectively (the weights are officially recommended by the Transportation Department of Guangdong, according to the average sizes of the involved vehicle types in each of the categories. They are extensively employed in the field of transportation engineering, such as freeway toll and traffic volume prediction):

$$AADT_i = \frac{V_i^{(1)} + 1.5V_i^{(2)} + 2V_i^{(3)} + 3V_i^{(4)} + 3.5V_i^{(5)}}{365}, \quad (1)$$

TABLE 1: Vehicle categorization.

Category	Categorization criteria				Representative vehicle types
	Head height	No. of axes	No. of wheels	Wheelbase	
1	<1.3 m	2	2-4	<3.2 m	Passenger car, jeep, taxi-head truck
2	≥1.3 m	2	4	≥3.2 m	Minibus, minivan, light truck
3	≥1.3 m	2	6	≥3.2 m	Medium bus, large normal bus, medium truck
4	≥1.3 m	3	6-10	≥3.2 m	Large luxury bus, large truck, large trailer, 20-foot container vehicle
5	≥1.3 m	>3	>10	≥3.2 m	Heavy truck, heavy trailer, 40-foot container vehicle

TABLE 2: Descriptive statistics of variables in the model.

Variables	Description	Mean	SD	Min.	Max.
<i>Response variables</i>					
No-injury	No-injury crash count per segment in 2014	3.60	2.98	0	20
Injury	Injury crash count per segment in 2014	0.86	1.03	0	5
<i>Crash exposure variable</i>					
DVKT	Daily vehicle kilometers traveled ( $10^3$ km-pcu <sup>a</sup> )	44.3	16.9	9.98	129
<i>Risk factors</i>					
$Veh_i^{(1)}$	The percentages of vehicles of Category 1	38.3	0.70	37.6	40.3
$Veh_i^{(2)}$	The percentages of vehicles of Category 2	2.59	0.23	2.24	3.05
$Veh_i^{(3)}$	The percentages of vehicles of Category 3	21.8	0.31	21.0	22.4
$Veh_i^{(4)}$	The percentages of vehicles of Category 4	6.58	0.27	6.18	7.06
$Veh_i^{(5)}$	The percentages of vehicles of Category 5	30.7	0.87	28.9	31.6
Curvature	Segment horizontal curvature ( $0.1 \text{ km}^{-1}$ )	1.77	1.27	0	4.35
Gradient	Segment vertical gradient (%)	0.741	0.568	0	2.91
Bridge	A part of a bridge: Yes = 1, No = 0	0.500	0.502	0	1
Ramp	Presence of ramps Yes = 1, No = 0	0.210	0.407	0	1

<sup>a</sup>pcu: passenger car units

where  $V_i^{(1)}$ ,  $V_i^{(2)}$ ,  $V_i^{(3)}$ ,  $V_i^{(4)}$ , and  $V_i^{(5)}$  are the traffic volumes for vehicle categories 1-5, respectively, on freeway segment  $i$  in 2014. As in the previous studies [1, 2], the daily vehicle kilometers traveled (DVKT), measured as the product of AADT and segment length, are used as the crash exposure variable.

Moreover, the traffic composition, i.e., the percentages of the five vehicle categories,  $Veh_i^{(1)}$ ,  $Veh_i^{(2)}$ ,  $Veh_i^{(3)}$ ,  $Veh_i^{(4)}$ , and  $Veh_i^{(5)}$ , is calculated as

$$Veh_i^{(1)} = \frac{V_i^{(1)}}{V_i^{(1)} + 1.5V_i^{(2)} + 2V_i^{(3)} + 3V_i^{(4)} + 3.5V_i^{(5)}}, \quad (2)$$

$$Veh_i^{(2)} = \frac{1.5V_i^{(2)}}{V_i^{(1)} + 1.5V_i^{(2)} + 2V_i^{(3)} + 3V_i^{(4)} + 3.5V_i^{(5)}}, \quad (3)$$

$$Veh_i^{(3)} = \frac{2V_i^{(3)}}{V_i^{(1)} + 1.5V_i^{(2)} + 2V_i^{(3)} + 3V_i^{(4)} + 3.5V_i^{(5)}}, \quad (4)$$

$$Veh_i^{(4)} = \frac{3V_i^{(4)}}{V_i^{(1)} + 1.5V_i^{(2)} + 2V_i^{(3)} + 3V_i^{(4)} + 3.5V_i^{(5)}}, \quad (5)$$

$$Veh_i^{(5)} = \frac{3.5V_i^{(5)}}{V_i^{(1)} + 1.5V_i^{(2)} + 2V_i^{(3)} + 3V_i^{(4)} + 3.5V_i^{(5)}}. \quad (6)$$

The percentage of the 5th vehicle category,  $Veh_i^{(5)}$ , is set as the reference case.

Table 2 illustrates the definitions and descriptive statistics of the variables used in the model development. In SPSS, Pearson correlation tests for the risk factors are conducted. The results indicate that five variable pairs,  $Veh_i^{(2)}$  and  $Veh_i^{(3)}$ ,  $Veh_i^{(2)}$  and  $Veh_i^{(4)}$ ,  $Veh_i^{(3)}$  and  $Veh_i^{(4)}$ ,  $Veh_i^{(2)}$  and DVKT, and  $Veh_i^{(4)}$  and DVKT, are significantly correlated with the absolute values of their correlation coefficients over 0.6. To avoid the adverse effect of significant correlation,  $Veh_i^{(2)}$  and  $Veh_i^{(4)}$  are excluded from the analysis. The results of the multicollinearity diagnoses suggest that there is no significant collinearity in the other factors.

### 3. Methodology

**3.1. Model Specification.** In the present study, there are two injury severity degrees used for the empirical analysis. Therefore, a bivariate CAR model is developed, which can be extended to a multivariate version for simultaneously predicting  $K(K > 2)$  response variables [27]. In the bivariate CAR model, the observed no-injury ( $k = 1$ ) or injury ( $k = 2$ ) crash count  $Y_{k,i}$  on freeway segment,  $i$ , is assumed to follow a Poisson distribution:

$$Y_{k,i} | \lambda_{k,i} \sim \text{Poisson}(\lambda_{k,i}), \quad i = 1, 2, \dots, 154, \quad k = 1, 2. \quad (7)$$

A generalized linear relationship is assumed between the expected crash count  $\lambda_{k,i}$  and the observed risk factors  $\mathbf{X}_{k,i}$  (including traffic composition):

$$\ln \lambda_{k,i} = \ln e_{k,i} + \mathbf{X}'_{k,i} \boldsymbol{\beta}_k + \theta_{k,i} + \phi_{k,i}, \quad (8)$$

where  $\boldsymbol{\beta}_k$  are the regression coefficients corresponding to the risk factors for severity level  $k$ . To reveal their possible nonlinear relationship, the crash exposure  $e_{k,i}$  is formulated as a power of DVKT:

$$e_{k,i} = (DVKT_i)^{\alpha_k}, \quad (9)$$

where  $\alpha_k$  is an estimable parameter.

The residual terms  $\theta_{k,i}$  and  $\phi_{k,i}$  in the link function are used to capture the unstructured heterogeneous and structured spatial effects, respectively, for the expected crash frequency on freeway segment  $i$  at each severity level  $k$ . These two terms also account for the overdispersion characteristic commonly found in crash frequency data [1]. To accommodate the correlation between the heterogeneous effects of no-injury and injury crashes,  $\theta_{1,i}$  and  $\theta_{2,i}$  are assumed to be binormally distributed with zero means:

$$\begin{aligned} \boldsymbol{\theta}_i &\sim N_2(\mathbf{0}, \boldsymbol{\Sigma}), \\ \boldsymbol{\theta}_i &= \begin{pmatrix} \theta_{1,i} \\ \theta_{2,i} \end{pmatrix}, \\ \boldsymbol{\Sigma} &= \begin{pmatrix} \sigma_{1,1}^b & \sigma_{1,2}^b \\ \sigma_{2,1}^b & \sigma_{2,2}^b \end{pmatrix}. \end{aligned} \quad (10)$$

In the variance-covariance matrix  $\boldsymbol{\Sigma}$ , the diagonal elements ( $\sigma_{1,1}^b$  and  $\sigma_{2,2}^b$ ) represent the variances of heterogeneous effects for no-injury and injury crashes, respectively, while the off-diagonal elements ( $\sigma_{1,2}^b$  and  $\sigma_{2,1}^b$ ,  $\sigma_{1,2}^b = \sigma_{2,1}^b$ ) represent the covariance between the heterogeneous effects. To measure the correlation between them, the correlation coefficient is calculated as  $\rho = \sigma_{1,2}^b / \sqrt{\sigma_{1,1}^b \sigma_{2,2}^b}$ .

To capture the spatial effects correlated between crash severities, a bivariate two-dimensional CAR prior is advocated. Proximity structure is a critical element of the CAR prior. While various proximity structures have been investigated [32], the most prevalent structure, 0-1 first-order neighbor, is used to define the proximity matrix in this study. Specifically, if segments  $i$  and  $j$  are connected to each other directly,  $\omega_{i,j} = 1$ ; otherwise,  $\omega_{i,j} = 0$ . Given the 0-1 first-order neighbor structure, the bivariate CAR prior is expressed as follows:

$$\begin{aligned} \boldsymbol{\Phi}_i &\sim N_2\left(\bar{\boldsymbol{\Phi}}_i, \frac{\boldsymbol{\Omega}}{n_i}\right), \\ \boldsymbol{\Phi}_i &= \begin{pmatrix} \phi_{1,i} \\ \phi_{2,i} \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} \bar{\boldsymbol{\Phi}}_i &= \begin{pmatrix} \bar{\phi}_{1,i} \\ \bar{\phi}_{2,i} \end{pmatrix}, \\ \boldsymbol{\Omega} &= \begin{pmatrix} \sigma_{1,1}^{bs} & \sigma_{1,2}^{bs} \\ \sigma_{2,1}^{bs} & \sigma_{2,2}^{bs} \end{pmatrix}, \end{aligned} \quad (11)$$

where  $n_i = \sum_{i \neq j} \omega_{i,j}$  is the number of segments that are adjacent to segment  $i$  and  $\bar{\phi}_{k,i} = \sum_{i \neq j} \phi_{k,j} \omega_{i,j} / n_i$ .  $\boldsymbol{\Omega}$  is the variance-covariance matrix for spatial correlation, in which  $\sigma_{1,1}^{bs}$  and  $\sigma_{2,2}^{bs}$  reflect the spatial variances of no-injury and injury crash frequencies, respectively, and  $\sigma_{1,2}^{bs} (= \sigma_{2,1}^{bs})$  reflects the spatial covariance between them. The correlation coefficient  $\rho_s = \sigma_{1,2}^{bs} / \sqrt{\sigma_{1,1}^{bs} \sigma_{2,2}^{bs}}$  describes the correlation between the spatial effects.

The posterior proportion of extra-Poisson variation explained by the spatial correlation for each crash severity level is also of interest and is defined as follows [24]:

$$\eta_k = \frac{sd(\phi_k)}{sd(\theta_k) + sd(\phi_k)}. \quad (12)$$

**3.2. Model Estimation.** Due to the complexity of the bivariate CAR model, the parameters and hyperparameters are estimated by Bayesian methods using Markov chain Monte Carlo (MCMC) simulation available in freeware WinBUGS. The prior distributions of the (hyper)parameters, which reflect prior knowledge about the (hyper)parameters, should be specified, when conducting Bayesian estimation. In the absence of sufficient knowledge, noninformative (vague) prior distributions are usually used [33]. Specifically, a dif-fused normal distribution  $N(0, 10^4)$  is specified as the priors of the regression coefficients (i.e., the elements of  $\boldsymbol{\beta}_k$ ). A Wishart prior,  $W(\mathbf{P}, r)$ , is used for  $\boldsymbol{\Sigma}^{-1}$  and  $\boldsymbol{\Omega}^{-1}$ , where the scale matrix  $\mathbf{P}$  is an identity matrix and  $r = 2$  is the degrees of freedom [27]. A chain of 150,000 iterations of the MCMC simulation is set, and the first 100,000 iterations act as burn-ins. Visual inspection of the MCMC trace plots for the model parameters and monitoring of the ratios of the Monte Carlo errors relative to the respective standard deviations of the estimates are used to evaluate the MCMC convergence. The estimation results of the parameters and hyperparameters in the bivariate CAR models are presented in Tables 3 and 4, respectively.

## 4. Result Analysis

According to the results in Table 3, we can find that a larger proportion of Category 1 or 3 vehicles in the mixed freeway traffic would significantly decrease the probability of no-injury crashes, while significantly increasing the probability of injury crashes. The coefficients of  $Veh_i^{(1)}$  indicate that there are a 19.8% decrease in the expected no-injury crash frequency and a 22.1% increase in the expected injury frequency per 1% increase in Category 1 vehicles. Similarly, there are a 46.1% decrease in the expected no-injury crash frequency

TABLE 3: Parameter estimates in the bivariate CAR model<sup>a</sup>.

Variable	No injury				Injury			
	Mean	80%BCI <sup>b</sup>	90%BCI	95%BCI	Mean	80%BCI	90%BCI	95%BCI
Constant	95.24	<b>(90.26, 98.88)</b> <sup>c</sup>	<b>(89.71, 99.26)</b>	<b>(89.42, 99.54)</b>	-19.46	<b>(-24.12, -11.43)</b>	<b>(-24.89, -10.48)</b>	<b>(-25.42, -9.96)</b>
lnDVKT	0.95	<b>(0.70, 1.19)</b>	<b>(0.67, 1.21)</b>	<b>(0.65, 1.22)</b>	0.83	<b>(0.38, 1.17)</b>	<b>(0.29, 1.21)</b>	<b>(0.20, 1.23)</b>
$Veh_i^{(1)}$	-0.22	<b>(-0.31, -0.15)</b>	<b>(-0.32, -0.13)</b>	<b>(-0.33, -0.13)</b>	0.20	<b>(0.09, 0.29)</b>	<b>(0.06, 0.30)</b>	<b>(0.04, 0.31)</b>
$Veh_i^{(3)}$	-0.62	<b>(-0.74, -0.47)</b>	<b>(-0.77, -0.45)</b>	<b>(-0.79, -0.44)</b>	1.75	<b>(1.58, 1.95)</b>	<b>(1.55, 2.00)</b>	<b>(1.52, 2.03)</b>
Curvature	0.09	<b>(0.02, 0.17)</b>	(-0.004, 0.19)	(-0.02, 0.21)	0.10	(-0.04, 0.23)	(-0.08, 0.27)	(-0.11, 0.30)
Gradient	0.04	(-0.13, 0.20)	(-0.18, 0.25)	(-0.22, 0.29)	0.30	<b>(0.02, 0.57)</b>	(-0.05, 0.64)	(-0.11, 0.71)

<sup>a</sup>Bridge and Ramp are excluded, because none of their effects on crash frequencies with the two severity outcomes is statistically significant (less than 80% credibility level).

<sup>b</sup>BCI: Bayesian credible interval.

<sup>c</sup>Boldface indicates statistical significance at the corresponding credibility level.

TABLE 4: Hyperparameter estimates in the bivariate CAR model.

Hyperparameter	Mean	S.D.	80%BCI <sup>a</sup>	90%BCI	95%BCI
$\sigma_{1,1}^b$	0.12	0.04	<b>(0.08, 0.17)</b> <sup>b</sup>	<b>(0.07, 0.19)</b>	<b>(0.06, 0.22)</b>
$\sigma_{2,2}^b$	0.25	0.12	<b>(0.13, 0.41)</b>	<b>(0.11, 0.48)</b>	<b>(0.09, 0.54)</b>
$\sigma_{1,2}^b (= \sigma_{2,1}^b)$	0.01	0.05	(-0.05, 0.07)	(-0.06, 0.09)	(-0.07, 0.11)
$\rho$	0.06	0.25	(-0.28, 0.39)	(-0.36, 0.47)	(-0.43, 0.53)
$\sigma_{1,1}^{bs}$	0.17	0.08	<b>(0.09, 0.28)</b>	<b>(0.08, 0.32)</b>	<b>(0.07, 0.37)</b>
$\sigma_{2,2}^{bs}$	0.32	0.21	<b>(0.13, 0.59)</b>	<b>(0.11, 0.75)</b>	<b>(0.09, 0.87)</b>
$\sigma_{1,2}^{bs} (= \sigma_{2,1}^{bs})$	0.02	0.09	(-0.08, 0.13)	(-0.11, 0.18)	(-0.15, 0.23)
$\rho_s$	0.06	0.33	(-0.38, 0.49)	(-0.50, 0.58)	(-0.58, 0.65)
$\eta_1$	0.64	0.05	<b>(0.58, 0.70)</b>	<b>(0.56, 0.72)</b>	<b>(0.55, 0.73)</b>
$\eta_2$	0.68	0.06	<b>(0.61, 0.76)</b>	<b>(0.58, 0.78)</b>	<b>(0.57, 0.79)</b>

<sup>a</sup>BCI: Bayesian credible interval.

<sup>b</sup>Boldface indicates statistical significance at the corresponding credibility level.

and a 4.75 times increase in the expected injury frequency per 1% increase in Category 3 vehicles. These results may be attributed to the variations of the crash aggressivity and worthiness among these vehicle categories and the driving behavior and maneuver ability among their drivers [10]. For example, Huang et al. [11] have found that, compared to heavy trucks (Category 5), passenger cars (Category 1) and medium trucks (Category 3) possess much lower crash worthiness, which would increase the degrees of injury sustained by occupants in passenger cars or medium trucks. For freeway safety improvement, the findings suggest that more intensive education on freeway safety should be focused on drivers of small-size vehicles and more strict enforcement should be conducted on their common risky driving behaviors, such as speeding and no use of seat belt. Moreover, the findings also indicate the warrant of driving safety assistant system for the vehicles in Categories 1 and 3, to avoid injury (especially fatal) crashes.

Turning to the estimated safety effects of other explanatory variables, the natural logarithm of DVKT is found statistically significant at the 95% credibility level and positively associated with no-injury and injury crash frequencies. Specifically, a 1% increase in DVKT is expected to result in a 0.9% increase in no-injury crashes and a 0.8% increase in injury crashes. These results are reasonable and consistent to the findings in previous studies [1, 2], as more DVKT provide

more opportunities for crash occurrence at each severity degree.

The parameter estimates show that the effect of horizontal curvature is significant (at the 80% credibility level) on no-injury crashes only. The positive coefficient (mean=0.09) implies that a  $0.1 \text{ km}^{-1}$  increase in curvature is associated with a 9.4% increase in no-injury crashes. It is in line with the existing findings [34, 35]: freeway segments with great curvature generate limited sight distances for drivers and require strong centrifugal forces on vehicles negotiating the curve, which may lead to rear-end or sideswipe crashes. The outcomes of these two crash types are usually property damage only.

Vertical grade is found to have a significantly positive (at the 80% credibility level) effect on injury crash frequency. The coefficient (mean=0.30) indicates that a 0.01 increase in freeway grade may contribute to a 35% increase in injury crashes. The result conforms to engineering intuition and aligns with many previous studies which argue that a steep grade also has an adverse impact on stopping sight distances, thereby increasing crash risk [3, 34]. Moreover, some previous studies have demonstrated that vertical grade would increase crash injury severities [36, 37]. Due to the bad safety performance, steep grade is not recommended in freeway vertical design by almost all design manuals [4].

TABLE 5: The results of comparison between the bivariate CAR, univariate CAR, and (aspatial) bivariate Poisson log-normal models.

	Bivariate CAR	Univariate CAR	Bivariate Poisson log-normal
$\bar{D}$	918	970	953
$pD$	123	74	96
DIC	1041	1044	1049
$R^2$	0.60	0.52	0.55

With respect to the hyperparameters, the estimation results in Table 4 indicate that there are significant (at the 95% credibility level) heterogeneous ( $\sigma_{1,1}^b=0.12$  and  $\sigma_{2,2}^{bs}=0.25$ ) and spatial ( $\sigma_{1,1}^{bs}=0.17$  and  $\sigma_{2,2}^{bs}=0.32$ ) effects in both no-injury and injury crashes. The spatial correlations account for 64% and 68% of the extra-Poisson variations of no-injury and injury crash frequencies, respectively, as reflected by the estimates of  $\eta_1$  and  $\eta_2$ . These results demonstrate the strength of incorporating spatial correlation into freeway crash modeling. While the heterogeneous and spatial covariances and correlation coefficients between the two crash severities are positive, unfortunately, their magnitudes are very low and none of them is statistically significant (less than 80% credibility level). The results indicate that no-injury and injury crash frequencies are mutually independent. Further investigation on the finding is fully merited. A plausible cause is that their correlation may substantially be attributed to the safety effects of the observed factors. Nevertheless, to further verify the performance of the proposed model, we compare it with univariate CAR and (aspatial) bivariate Poisson lognormal models in terms of goodness of fit. The results in Table 5 show that the bivariate CAR model yields a lower deviance information criteria (DIC) value and a higher  $R^2$  value than the other two models, which suggests that the multivariate CAR model outperforms its univariate and aspatial counterparts. Therefore, the proposed multivariate CAR model is a good alternative for analyzing crash frequency by severity.

## 5. Conclusions and Future Research

This study investigates the effects of traffic composition on freeway crash frequency by injury severity based on a crash dataset collected from Kaiyang Freeway in Guangdong Province, China. In the empirical analysis, five vehicle categories and two severity levels (i.e., no injury and injury) are considered. To quantitatively estimate the safety effects of the percentages of different vehicle categories, a Bayesian multivariate CAR model is developed, with the effects of DVKT and freeway geometry controlled and the heterogeneous and spatial effects as well as the (aspatial and spatial) correlations between crash severities accommodated.

According to the results of Bayesian estimation conducted in WinBUGS, increasing the percentages of Categories 1 (e.g., passenger car) and 3 (e.g., medium bus, medium truck) vehicles would decrease the risk of no-injury crashes but increase the risk of injury crashes, compared to Category

5 (e.g., heavy truck) vehicles. Moreover, more (no-injury and injury) crashes are expected to occur on freeway segments with more vehicle kilometers traveled. Horizontal curvature is positively correlated with no-injury crashes, and vertical grade is positively correlated with injury crashes. The results are consistent to the findings of previous studies and conform to engineering experience. Significant heterogeneous and spatial effects are also found for both no-injury and injury crashes. These findings justify the applicability of the proposed model.

While the study is empirical in nature, it provides good overall understanding of the safety effects of freeway traffic composition and can bring inspiration to strategies of freeway vehicle safety improvement. However, there are some limitations in the current study, to which future research efforts can be devoted. First, each of the grouped vehicle categories contains several vehicle types. The crash risk of these vehicle types may also be different. More elaborated traffic data may provide a deeper insight on the relationship between traffic composition and freeway safety. Second, due to the limitation of data collection, weather conditions are not considered. Controlling for the weather factors may help further reveal the occurrence mechanism of freeway crashes. Last, while the proposed 0-1 first-order neighbor structure is the most prevalent method to define the proximity matrix in CAR models, more (such as distance-based) neighbor structures can be explored to find the most promising one for spatial modeling freeway crash frequency by severity.

## Data Availability

The copyrights of the used data belong to Guangdong Transportation Group and Guangdong Province Communication Planning and Design Institute Co., Ltd. Our rights of using these data are only limited to this research. Please contact the companies for gaining access to them.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was jointly supported by the National Natural Science Foundation of China (no. 51578247), the Natural Science Foundation of Guangdong Province (no. 2017A030310161), and the Fundamental Research Funds for the Central Universities (no. 2017BQ026).

## References

- [1] X. Ma, S. Chen, and F. Chen, "Multivariate space-time modeling of crash frequencies by injury severity levels," *Analytic Methods in Accident Research*, vol. 15, pp. 29–40, 2017.
- [2] R. Yu, M. Abdel-Aty, and M. Ahmed, "Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors," *Accident Analysis & Prevention*, vol. 50, pp. 371–376, 2013.
- [3] Q. Zeng, J. Sun, and H. Wen, "Bayesian hierarchical modeling monthly crash counts on freeway segments with temporal

- correlation,” *Journal of Advanced Transportation*, vol. 2017, Article ID 5391054, 8 pages, 2017.
- [4] M. Ahmed, H. Huang, M. Abdel-Aty, and B. Guevara, “Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway,” *Accident Analysis & Prevention*, vol. 43, no. 4, pp. 1581–1589, 2011.
  - [5] X. Ye, K. Wang, Y. Zou, D. Lord, and X. Ma, “A semi-nonparametric Poisson regression model for analyzing motor vehicle crash data,” *PLoS ONE*, vol. 13, no. 5, p. e0197338, 2018.
  - [6] Traffic Management Bureau of the Public Security Ministry of the People’s Republic of China, 2016. Annual Statistical Report on Roadway Traffic Accidents (2015).
  - [7] E. Hauer, B. K. Allery, J. Kononov, and M. S. Griffith, “How best to rank sites with promise,” *Transportation Research Record*, no. 1897, pp. 48–54, 2004.
  - [8] X. Ye, R. M. Pendyala, S. P. Washington, K. Konduri, and J. Oh, “A simultaneous equations model of crash frequency by collision type for rural intersections,” *Safety Science*, vol. 47, no. 3, pp. 443–452, 2009.
  - [9] X. Ye, R. M. Pendyala, V. Shankar, and K. C. Konduri, “A simultaneous equations model of crash frequency by severity level for freeway sections,” *Accident Analysis & Prevention*, vol. 57, pp. 140–149, 2013.
  - [10] R. R. Dinu and A. Veeraragavan, “Random parameter models for accident prediction on two-lane undivided highways in India,” *Journal of Safety Research*, vol. 42, no. 1, pp. 39–42, 2011.
  - [11] H. Huang, C. Siddiqui, and M. Abdel-Aty, “Indexing crash worthiness and crash aggressivity by vehicle type,” *Accident Analysis & Prevention*, vol. 43, no. 4, pp. 1364–1370, 2011.
  - [12] J. Tang, F. Liu, W. Zhang, R. Ke, and Y. Zou, “Lane-changes prediction based on adaptive fuzzy neural network,” *Expert Systems with Applications*, vol. 91, pp. 452–463, 2018.
  - [13] Q. Zeng, H. Wen, and H. Huang, “The interactive effect on injury severity of driver-vehicle units in two-vehicle crashes,” *Journal of Safety Research*, vol. 59, pp. 105–111, 2016.
  - [14] I. Grinberg and Y. Wiseman, “Scalable parallel simulator for vehicular collision detection,” *International Journal of Vehicle Systems Modelling and Testing*, vol. 8, no. 2, pp. 119–144, 2013.
  - [15] Q. Zeng, H. Huang, X. Pei, and S. C. Wong, “Modeling non-linear relationship between crash frequency by severity and contributing factors by neural networks,” *Analytic Methods in Accident Research*, vol. 10, pp. 12–25, 2016.
  - [16] J. Ma, K. M. Kockelman, and P. Damien, “A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods,” *Accident Analysis & Prevention*, vol. 40, no. 3, pp. 964–975, 2008.
  - [17] E. S. Park and D. Lord, “Multivariate poisson-lognormal models for jointly modeling crash frequency by severity,” *Transportation Research Record*, no. 2019, pp. 1–6, 2007.
  - [18] X. Pei, S. C. Wong, and N. N. Sze, “The roles of exposure and speed in road safety analysis,” *Accident Analysis & Prevention*, vol. 48, pp. 464–471, 2012.
  - [19] C. Wang, M. A. Quddus, and S. G. Ison, “Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model,” *Accident Analysis & Prevention*, vol. 43, no. 6, pp. 1979–1990, 2011.
  - [20] Y.-C. Chiou and C. Fu, “Modeling crash frequency and severity using multinomial-generalized Poisson model with error components,” *Accident Analysis & Prevention*, vol. 50, pp. 73–82, 2013.
  - [21] Y.-C. Chiou and C. Fu, “Modeling crash frequency and severity with spatiotemporal dependence,” *Analytic Methods in Accident Research*, vol. 5–6, pp. 43–58, 2015.
  - [22] Y.-C. Chiou, C. Fu, and H. Chih-Wei, “Incorporating spatial dependence in simultaneously modeling crash frequency and severity,” *Analytic Methods in Accident Research*, vol. 2, pp. 1–11, 2014.
  - [23] J. Aguero-Valverde and P. P. Jovanis, “Analysis of road crash frequency with spatial models,” *Transportation Research Record*, no. 2061, pp. 55–63, 2008.
  - [24] J. Aguero-Valverde, “Multivariate spatial models of excess crash frequency at area level: case of Costa Rica,” *Accident Analysis & Prevention*, vol. 59, pp. 365–373, 2013.
  - [25] M. T. Islam, K. El-Basyouny, S. E. Ibrahim, and T. Sayed, “Before-after safety evaluation using full Bayesian macroscopic multivariate and spatial models,” *Transportation Research Record*, vol. 2601, pp. 128–137, 2016.
  - [26] Y. Wang and K. M. Kockelman, “A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods,” *Accident Analysis & Prevention*, vol. 60, pp. 71–84, 2013.
  - [27] S. Barua, K. El-Basyouny, and M. T. Islam, “A Full Bayesian multivariate count data model of collision severity with spatial correlation,” *Analytic Methods in Accident Research*, vol. 3–4, pp. 28–43, 2014.
  - [28] S. Barua, K. El-Basyouny, and M. T. Islam, “Multivariate random parameters collision count data models with spatial heterogeneity,” *Analytic Methods in Accident Research*, vol. 9, pp. 1–15, 2016.
  - [29] H. Huang, H. Zhou, J. Wang, F. Chang, and M. Ma, “A multivariate spatial model of crash frequency by transportation modes for urban intersections,” *Analytic Methods in Accident Research*, vol. 14, pp. 10–21, 2017.
  - [30] D. Lord and F. Mannering, “The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives,” *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 291–305, 2010.
  - [31] F. L. Mannering and C. R. Bhat, “Analytic methods in accident research: methodological frontier and future directions,” *Analytic Methods in Accident Research*, vol. 1, pp. 1–22, 2014.
  - [32] J. Aguero-Valverde and P. P. Jovanis, “Spatial correlation in multilevel crash frequency models: effects of different neighboring structures,” *Transportation Research Record*, no. 2165, pp. 21–32, 2010.
  - [33] Q. Zeng and H. Huang, “Bayesian spatial joint modeling of traffic crashes on an urban road network,” *Accident Analysis & Prevention*, vol. 67, pp. 105–112, 2014.
  - [34] S. Labi, “Efficacies of roadway safety improvements across functional subclasses of rural two-lane highways,” *Journal of Safety Research*, vol. 42, no. 4, pp. 231–239, 2011.
  - [35] X.-S. Wang, Y. Song, H.-L. Huang, and D.-Y. Zhang, “Analysis of risk factors for suburban highways using hierarchical negative binomial model,” *China Journal of Highway and Transport*, vol. 27, no. 1, pp. 100–106, 2014.
  - [36] J. R. Klop and A. J. Khattak, “Factors influencing bicycle crash severity on two-lane, undivided roadways in North Carolina,” *Transportation Research Record*, no. 1674, pp. 78–85, 1999.
  - [37] M. A. Quddus, C. Wang, and S. G. Ison, “Road traffic congestion and crash severity: econometric analysis using ordered response models,” *Journal of Transportation Engineering*, vol. 136, no. 5, pp. 424–435, 2009.

