IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 13, 2022, accepted April 2, 2022, date of publication April 5, 2022, date of current version April 13, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3165065

A Bayesian Bivariate Random Parameters and Spatial-Temporal Negative Binomial Lindley Model for Jointly Modeling Crash Frequency by Severity: Investigation for Chinese Freeway Tunnel Safety

MINGMAO CAI, FENG TANG, AND XINSHA FU
School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510640, China
Corresponding author: Xinsha Fu (fuxinsha_scut@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51978283 and Grant 51778242.

ABSTRACT As hazardous locations of a road, freeway tunnels have a higher risk of casualty than open roads. Therefore, it is necessary to seek a reliable crash prediction model and propose targeted improvement measures. However, existing studies on freeway tunnel crash models mainly suffer from the following problems: 1) They ignore the correlation between different injury severity levels of crashes; 2) They ignore the impact of excess zero observations; 3) They do not consider the influence of heterogeneity between samples and the spatio-temporal correlation. To solve the above problems, this paper has compiled a dataset with freeway tunnel design features, three years of traffic conditions, pavement conditions, and traffic crash data. Then, a bivariate random parameters negative binomial Lindley model (ST-BRPNB-L) is established for jointly modeling crash counts and injury severity levels, which consider excess zero observations by introducing Lindley parameters, characterize the heterogeneity, and spatial-temporal correlation between samples by introducing random parameters and spatio-temporal parameters. The Bayesian estimation results have shown that ST-BRPNB-L has the best goodness-of-fit among a series of comparison models, which verifies the superiority of the proposed model. On this basis, the influence of the risk factors on the frequency and severity of crashes was quantitatively analyzed based on the ST-BRPNB-L model's parameters estimation results, which provides a scientific basis for safety improvement measures of freeway tunnels.

INDEX TERMS Freeway tunnel, traffic safety, crash modeling techniques, crashes by severity, excess zero observations, spatio-temporal correlation, unobserved heterogeneity, bivariate random parameters Lindley model.

I. INTRODUCTION

Many factors may affect road traffic safety. Therefore, the authorities and researchers have long been searching for ways better to understand the influence of risk factors on crashes. The current research on road traffic safety risk factors mainly focuses on exploring the impact of road geometric design characteristics and traffic flow on the frequency of crashes on open roads. As a particular road structure in freeways, tunnels are limited by their rapidly changing lighting conditions, restricted cross-section width, and a narrowed vision field, resulting in a more complicated driving environment with a higher probability of road traffic crashes [1]. According to the 2019 statistical report of the Ministry of Transport of the People's Republic of China, freeway tunnels in China only account for 0.03% of the total road mileage, but the number of tunnel traffic crashes accounts for 0.24% of the entire road crashes. At the same time, the tunnels' relatively enclosed space would make it difficult for rescue operations to carry out and are likely to cause secondary accidents, thus aggravating the severity of tunnel accidents. Data shows that the fatality rate of freeway tunnel accidents in China

The associate editor coordinating the review of this manuscript and approving it for publication was Giambattista Gruosso.

is 0.55 persons per occurrence. Therefore, it is of great significance to the design and management of tunnels to seek suitable models of the frequency and severity of freeway tunnel crashes based on years of freeway tunnel operating data and analyze the influencing factors of tunnel crashes and injury severity levels.

A. EXISTING TUNNEL TRAFFIC SAFETY RESEARCH

Currently, there are mainly two types of research on tunnel safety. The first category focuses on the analysis of the spatio-temporal distribution characteristics of crashes by region using mathematical statistics, that is, dividing the tunnels into multiple regions according to the lighting conditions and driving environment, including the three-zone [3], [4], four-zone [5], five-zone [6], six-zone [7], and seven-zone analysis methods [8]. However, this research idea only probes into the characteristics of crash frequency distribution rather than further explores the causes of crashes and the acting mechanism of various risk factors on crashes.

On the other hand, the second category accurately correlates risk factors and crash variables by establishing safety performance functions(SPFs). Taking the over-dispersion of crash data into account, Caliendo, Guglielmo, and Guida [9] used a bivariate negative binomial (NB) model to study the factors affecting the frequency of freeway tunnel crashes in Italy. The results showed that, as the length of the tunnel, annual average daily traffic volume, the proportion of trucks, and the number of lanes increase, the crash frequency will increase. Caliendo, Guglielmo, and Guida [10] also used the random parameter negative binomial model to characterize the unobserved heterogeneity of tunnel crash data. In addition to data heterogeneity, other studies [11]–[13] further considered the interaction of heterogeneity using the relevant random parameter models, the results show that more precise estimations of crashes can be obtained when the random parameters are assumed to be correlated in statistical analysis. However, these models failed to consider the spatio-temporal correlation and the excess zero observations of crash panel data. Meanwhile, studies [14], [15] introduced Lindley parameters and random parameters, respectively, to fully observe the interference of excess zero observations and heterogeneity on the model performance, and then achieved considerable goodness of fit.

There are also some differences in the dependent variable. Most of research mentioned above used Univariate dependent variables, like crash frequency and crash rate. There are also some studies for specific types of traffic crashes, they used the univariate dependent variables, like study analyzed the truck crash frequency and crashes of serious injury severity level, study [16] explored the influence of contributing factors on tunnel truck crashes, the result shows that the gender, age of driver, mid-night to dawn and afternoon peak hours, weekdays, snowy or icy road conditions, the interior zone of a tunnel, the combination truck, overloaded trucks, and extra-long tunnels are associated with higher truck crash severity. Study [17] used a random forest algorithm with standard binomial regression to analyze the risk factor of serious injury severity, the adverse weather, fatigued and distracted drivers, collision type (i.e., head-on/angle/rear-end), tunnel exit, tunnel width, curve radius (radius less than 1800 m), and heavy vehicle positively influences the severity of crashes. Another type of dependent variable is multivariate dependent variables, which analyze the risk factors' contribution to crashes of different injury severity levels. Study [18] used a two-level binary logistic modeling approach to identify significant influential factors with tunnel crash safety, shows that speed limit, tunnel length, truck involvement, rear-end crash, rainy and foggy weather, and sequential crash were found to be positively associated with crash severity in freeway tunnels. Research [19] employed a random parameter logit model to examine the factors affecting the injury severity of the freeway tunnel group crashes, the result shows that the daytime, weekdays, entrance zone, downgrades, elder drivers, speeding, fatigue driving, and rollover collisions are positively associated, while winter, curves, and sideswipes are negatively associated with severe crashes.

After reviewing the above studies, it was found that there are still problems to be solved in the study of tunnel crash modeling, which is: 1) The safety performance functions only take heed of the relationship between crash frequency and risk factors. No model that considers the correlation between the severity of crashes and the number of crashes of different severity has been established; 2) Existing tunnel safety research rarely considers the characteristics of excess zero observations of crash data; 3) The mainstream crash models mainly uses random parameters to handle the heterogeneity between samples, but rarely discusses the spatio-temporal correlation.

B. EXISTING RESEARCH ON INFLUENCING FACTORS OF TUNNEL CRASHES

Many reasons can lead to traffic crashes, such as road design, traffic conditions, and pavement conditions. Among them, the traffic safety influencing factors related to road geometric design characters include roadside facilities and horizontal, vertical, and cross-sectional alignment indicators. In terms of the safety effect of horizontal alignment, the research results of the studies [20]–[25] consistently showed that the increase in the curvature of a road segment could reduce the frequency of crashes on the said road segment. The reason lies in the fact that the increase in the curvature can effectively put the drivers on guard, thereby reducing the risk of crashes. Research [12] believed that the increase in the curvature of a tunnel would lead to an increase in the frequency of crashes. In sharp turning segments of a tunnel, factors such as restricted cross-section, poor lighting, and poor visibility will aggravate driving risks. In terms of the safety effect of vertical alignment, studies [20]–[23], [26], [27] showed that slope length and gradient are positively correlated with the frequency of crashes. In particular, research [27] believed that the fewer gradient change points set in a road segment, the lower the frequency of accidents on the said road segment. Study [26] collected and analyzed a total of 567 collision accidents in a 50km-long tunnel in China, and it was discovered that the crash frequency of the descending road segment of the tunnel was higher than that of the ascending road segment. In terms of cross-section safety effects, the results of research [21], [23], [24] showed that the frequency of highway crashes decreases with the increase of the width of the median strip. Research [25], [28] indicated that road segments that have wider shoulder widths tend to have lower accident frequency. Research [26] showed that the frequency of highway accidents increases with the number of lanes. Whereas research [22] indicated that the farther the vehicle is from the tunnel wall, the more significantly the crash frequency decreases.

In terms of traffic conditions, research [21] believed that different vehicles are more likely to have crashes on low-traffic roads; as the traffic volume increases, the frequency of crashes on road segments decreases significantly, which is contrary to the results of studies [11], [24], [26]. Study [12] believed that the frequency of highway accidents is significantly related to the increase in the proportion of class 5 vehicles, which is consistent with the result of [25], [29]. Whereas research [30] believed that the frequency of crashes would initially increase with the increase in the proportion of heavy vehicles but decrease when the frequency of crashes when reaching a certain threshold. The reason is that when the traffic volume exceeds a certain value, the highway will be extremely congested, and the driving speed will be far below the speed limit. Research [25] indicated that setting a lower speed limit on a road segment can effectively reduce the frequency of crashes. In contrast, research [23] believed that a lower speed limit increases the probability of single-vehicle accidents.

Pavement performance also has a significant impact on crashes. Research [21]demonstrated that the higher the friction coefficient of the pavement, the lower the probability of slight injury crashes, but the possibility of severe injury related crashes increases. Research [32] indicated that increasing the friction coefficient of wet pavements can effectively reduce the crash frequency, which is of great reference significance for the design and maintenance of roads in rainy areas. Study [24] showed that the crash frequency would increase with the increase in the international smoothness index (IRI) value, while study [12] believed that the crash frequency would decrease with an increase in the IRI value. Study [22] held that the higher the Pavement surface condition index(PCI) value and Skidding resistance index(SRI) value, the lower the crash frequency of the road segment. Study [33] analyzed the relationship between the pavement surface conditions and the crash frequency, and according to the road damage index and IRI value of the road segments, concluded that flat roads are more likely to see fatal crashes because drivers tend to be more aggressive when driving on a flat road. Study [34] showed that segments with a combination of poor road conditions and steep slopes would double the probability of crashes. Study [35] analyzed the impact of pavement damage and traffic characteristics on the frequency of crashes at various damage stages over time and evaluated the safety effects of the roads in different damage stages.

In summary, traffic crashes are mainly due to road alignment design indicators, traffic conditions, and pavement performance. One-sided consideration of crash factors will lead to deviations in parameters estimation and may cause opposite inferences. Therefore, selecting all aspects of factors as comprehensively as possible and establishing a comprehensive data set is essential for identifying crash risk factors.

C. EXISTING RESEARCH ON EXCESS ZERO OBSERVATIONS MODELING TECHNOLOGY

Regarding crash modeling technology, the negative binomial model satisfies the requirements for non-negativity and randomness of crash frequency and those for the over-dispersion of most crash data. Thus, the prevailing crash modeling method adopts this model and its improved models as the basis for technical improvements. However, the freeway tunnel crashes are rare events, the vast majority of tunnel segments may never have had a crash during their operational period, and secondly, the study needs to count crashes at specific periodicities, the tunnel segments may not have had a crash during these periodicities. So, there are a large number of zero observations in the statistical crash data. The Poisson and negative binomial models are commonly used models that the Poisson distribution tends to under-estimate the number of zeros given the mean of the data, while the NB may over-estimate zeros, but under-estimate observations with a count [22]. From the above,concluding that excess zero observations result in a deviation between the actual distribution of crash frequency, neglecting the excess zero observations to construct a model will lead to bias or even to false inferences. The zero-inflated model [25], [36], [37] was first used to process excessive zero observations and divide road segments into absolute safety and non-absolute safety states using the Logit and Probit models. Then, this model adopted NB and other aggregate models to fit the zero and non-zero observations in the data of the non-absolute safety state to maximize the fitted data. Study [38] believed that the ZINB model arbitrarily divides road segments into absolute safety and non-absolute safety states can only fit data statistics instead of genuinely explaining the logic of the crashes. The Markov model proposed in study [39] allowed the road segments to shift between the absolute safety state and non-absolute safety state, but this model also has a shortcoming, which is its complex computing process [40]. Many studies have been conducted on processing excessive zero observations in recent years. The introduction of a new distribution combined with NB distribution can effectively process excessive zero observations and objectively explain the logic of crash occurrence. For example, the results of Study [41] indicated that the NB-L model has a better goodness-of-fit for data sets that contain a large number of zero values. On the other hand, study [42] used the NB-GE model, Poisson

TABLE 1. Overview of crash frequency modeling technology and the superiority of the proposed model.

Model	Previous research	Features	Limitations of the model
Poisson model	D. Lord,2005; P. Vangala, 2015	A good **logical explanation** of crashes	Can't deal with the **over-dispersion** character of the data.
Negative binomial model (Poisson-Gamma model)	C. Caliendo,2013	Can fit **over-dispersion** data and the process of **parameter estimation is simple**	-
Poisson-lognormal model	Y. Wang, 2013	Can deal with **over-discrete**	Requiring **high sample quality**.
Zero-inflated Poisson (ZIP) /zero-inflated negative binomial (ZINB) models	D. Lord, 2007	Effective in solving **excess zero observations** attributes	**Weak in explaining the logic** of crashes.
Random parameters count models	C. Caliendo,2015; C. Caliendo,2019; Q. Hou,2018	Capturing **heterogeneity** between samples	-
Spatio-temporal Bayesian model	Q. Zeng, 2017; F. Tang, 2021	Explaining **the spatio-temporal correlation** of crash frequency	-
Markov switching counting model/finite mixture/Latent -class models	N. V. Malyshkina, 2009; A. Behnood, 2016	Assuming states are transferable to account for **unobserved heterogeneity**	Model parameters are complex to estimate and **computationally costly**
Bivariate/multivariate models with random parameters	C. Dong, 2014; S. Chen, 2017; Q. Zeng, 2017	Can discuss the influencing factor's contributions to **different injury severity levels of crashes**.	-
Bivariate random parameters negative binomial model(the proposed model in this research)	-	Can discuss the influencing factor's contributions to **different injury severity levels of crashes**; capturing **heterogeneity** between samples; explaining the **logic** of crashes; Solving **excess zero observations** attributes; Explaining the **spatio-temporal correlation** of crash frequency	-

model, NB model, and NB-L model in data samples with over-dispersion and many zero observations. The results indicated that the performance of the NB-GE model is similar to the NB-L model and is significantly better than the NB model.

At the same time, those mentioned above excess zero observations modeling technology is only applicable in univariate analysis, without considering the severity of crashes and the correlation between different severity levels, and the applicability and portability of each model for freeway tunnels in China remain to be further verified.

D. EXISTING RESEARCH ON RANDOM PARAMETERS AND SPATIO-TEMPORAL MODELING

Freeway tunnel crashes are complex events that involve a variety of human responses to external stimuli, as well as complex interactions between the vehicle, roadway features/condition, traffic-related factors, and environmental conditions. With such a complex situation and limited collection methods, it is impossible to have access to all of the data that could potentially determine the likelihood of a freeway tunnel crash or its resulting injury severity. The absence of such important data can potentially present serious specification problems for traditional statistical analyses that can lead to biased and inconsistent parameter estimates, erroneous inferences, and erroneous accident predictions. In other words, the unobserved heterogeneity arises from misspecification of the model, randomness of covariates, and omission of independent variables. If a model ignores the unobserved heterogeneity and the impact of the observed variable on all road segments on the crash is fixed, that is, adopting the fixed-parameter model, it will result in biased estimation and may lead to completely different conclusions [43]. Therefore, study [26], [44] introduced a negative binomial model with random effects and assumed the intercept term to follow normal distribution so that the model has better goodness of fit compared to the NB model. However, only considering the intercept term is not enough to explain the unobserved heterogeneity [22]. Therefore, studies [12], [24] allowed the regression parameters of each variable to vary randomly to explain the heterogeneity caused by unobserved factors. At the same time, to further capture the influence of unobserved heterogeneity shared by continuous cycles and adjacent road segments, the relevant parameters of temporal correlation, spatial correlation, and spatio-temporal interaction were included in the link function of the model to express the spatial-temporal correlation. For example, study [15] incorporated temporal correlation into the model and found a significant temporal effect in the model. Research [45] considered the spatial correlation, found significant spatial effects and improved the goodness of fit of the model. Research [14] introduced spatio-temporal parameters and the spatio-temporal interaction effect, and the results indicated that the spatio-temporal parameters of the model are all significant, and the goodness of fit of the model considering the spatio-temporal effects is significantly higher than that of the control model. Although random parameters and spatio-temporal parameters have been used to explain the heterogeneity caused by various factors, it is still of great significance to introduce a bivariate model to analyze the mechanism of crashes of different injury severity levels.

E. RESEARCH PURPOSE AND METHODOLOGY

The current tunnel safety researches have the following problems: 1) The severity of crashes in freeway tunnels and its

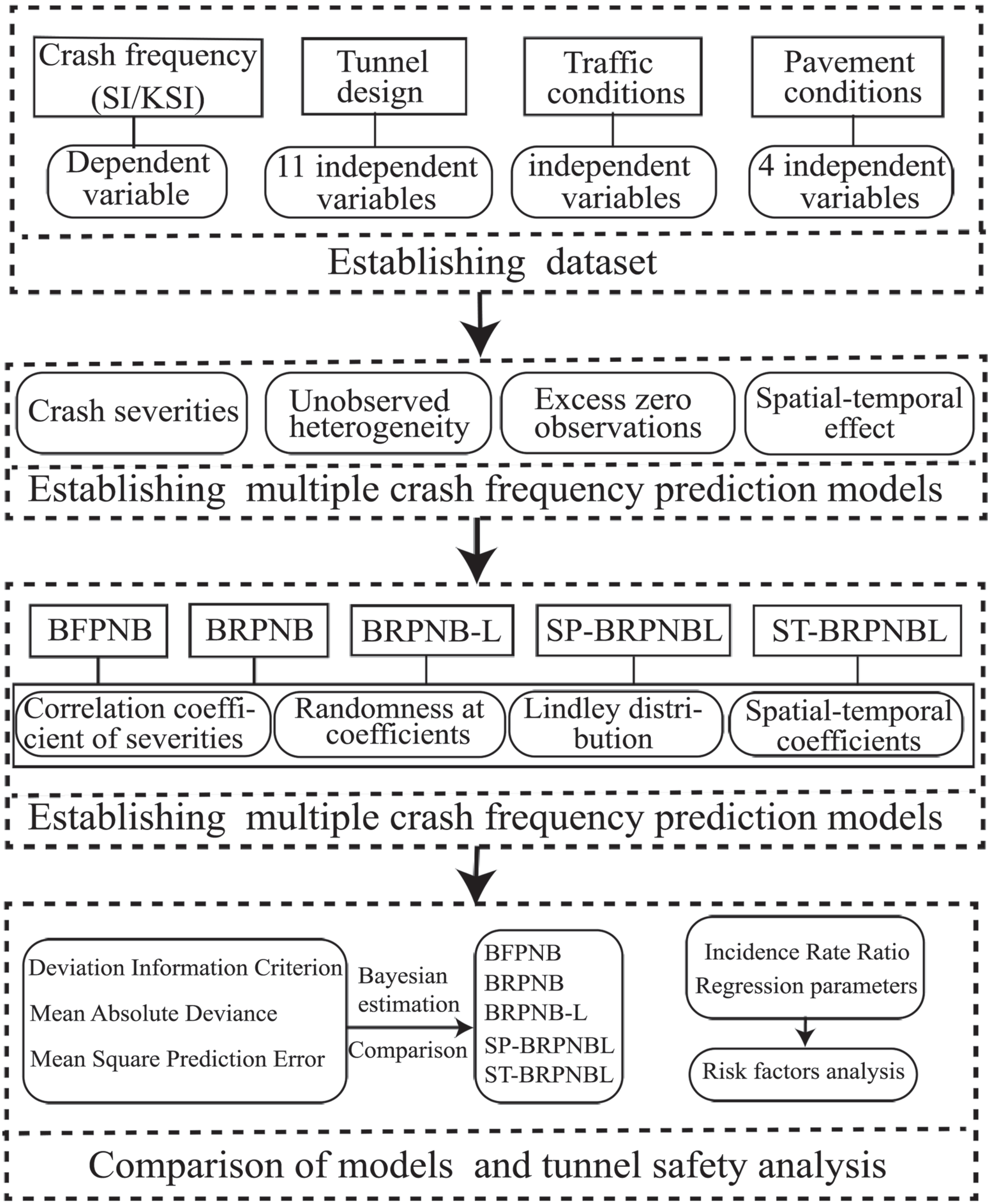

FIGURE 1. Technical roadmap in this paper. The top of the figure shows creating a dataset that includes different injury severity levels of crashes and three types of risk factors. The bottom of the figure presents the proposed ST-BRPNB-L model and the comparative SP-BRPNB-L, BRPNB-L, and BRPNB models. The middle of the figure illustrates the problem solved by the ST-BRPNB-L model and evaluation and analysis methods. The BRPNB, BRPNB-L, SP-BRPNB-L, and ST-BRPNB-L models represent the bivariate random parameters negative binomial model, bivariate random parameters negative binomial Lindley model, bivariate random parameters and spatial negative binomial Lindley model, and bivariate random parameters and spatial-temporal negative binomial Lindley model, respectively.

relationship with crash frequency, which have more apparent impacts on society, is not included in the scope of tunnel safety studies; 2) The modeling technology for excessive zero observations is rarely applied in tunnel studies; 3) Few tunnel safety studies adopt the random parameters method and the spatio-temporal parameters method to consider the unobserved heterogeneity and the spatial-temporal correlation. Because of the shortcomings of the above studies, this paper proposes to 1) establish a bivariate model to jointly consider the number of crashes of different severity in the tunnel; 2) introduce the Lindley distribution to consider the excess zero observations contained in data; 3) introduce random parameters and spatio-temporal correlation parameters to consider the unobserved heterogeneity and the spatial-temporal correlation.

II. DATA DESCRIPTION

The subjects included a total of 84 one-way tunnels of three major freeways in Guangdong Province, China, including (1) the Lianzhou-Huaiji section of the Erenhot-Guangzhou freeway; (2) the Huaiji-Sanshui section of the Erenhot-Guangzhou freeway; (3) Ma'an-Hekou section of the Guangzhou-Kunming freeway. In addition, the collected data includes information on crashes of freeway tunnel sections, tunnel design characteristics, traffic conditions, and pavement conditions, and was collected from 2015 to 2018. There was a total of 587 tunnel crashes and 20 influencing factors.

Another critical step in data processing is the division of tunnel segments. According to the study [12], areas of 100 m before the entrance and 100 m after the exit of a tunnel can be defined as its influence area, as shown in Figure 2. The prevailing classification methods are the fixed-length method and the homogeneity method. Relevant studies have shown that, despite its simplicity and feasibility, the fixed-length method has all indicators averaged, which is difficult to reflect the true impact of indicators on crashes freeway tunnels with complex driving conditions. In contrast, the various indicators of the homogeneity method are actual values, which are conducive to obtaining the existing relationship between traffic crashes and the causes. Therefore, this paper adopted the homogeneity method to divide the tunnel sections, that is, dividing the sections based on the consistency of speed limit, traffic volume, road width, number of lanes, and horizontal and vertical alignments for the same road segment. Finally, 217 homogeneous sections were obtained (the total samples collected in the three years is 217*3=651).

On the basis of dividing tunnel sections, the number of crashes of different severity and influencing factors were collected to establish a data set of tunnel crashes, including: 1) tunnel traffic crashes data provided by the Guangdong Provincial Freeway Administration; 2) daily traffic conditions data from freeways toll stations; 3) tunnel design characteristics indicators containing detailed tunnel design elements were provided by the Guangdong Provincial Communication Construction Group.; 4) pavement conditions indicators provided by the Guangdong Transportation Group Testing Center (GTGTC).

The statistical description of each indicator is shown in Table 2, and the following explanation is required: 1) This paper divides the severity of crash into two groups: kill and serious injury (KSI) and slight injury (SI). 2) The traffic volume data includes the proportion of the vehicles classified by 5 classes and the annual average daily traffic volume (AADT). Among them, the former was collected from the Network Tolling System of Guangdong Provincial Highway, which classifies vehicles according to the height of the vehicle, the number of axles, the number of wheels, and the wheelbase. The classification standards are shown in Table 3. Since the traffic volume between toll stations will not change, the AADT (annual average daily traffic) is the average number of vehicles that pass a roadway section each day in a particular year, calculating based on the conversion ratio of 1:1.5:2:3:3.5 for vehicles of classes 1 to 5 in accordance with the regulations of the Guangdong Provincial Department of Transport. For the specific calculation method, refer to the Traffic Administration Bureau of the Ministry of Public Security. 3) Referring to the Technical Standards for Highway Engineering of China, this paper defines a tunnel with a length of less than 500 m as a short tunnel, and a tunnel with a length of 500 m or more as a long tunnel. 4) Research [46], [47], found that the crash frequency for any tunnel segment with a longitudinal gradient greater than 2% is significantly higher than that of other tunnel segments. Therefore, this paper considers whether the gradient is greater than 2% as one of the influencing factors of the crash frequency. 5) The pavement conditions of the tunnels were tested annually by the GTGTC in accordance with the Highway Performance Assessment Standards. The test indicators include Pavement Surface Condition Index (PCI), Riding Quality Index (RQI), Skidding Resistance Index (SRI), and Subgrade Condition Index (SCI). Among them, PCI is used to determine the damage of pavement surface according to its cracks, potholes, ruts, and other distresses, and the data was stored in units of 20 m intervals. The higher the PCI value, the smaller the road damage. RQI was calculated according to the smoothness of the pavement surface to evaluate the impact of the pavement surface on driving comfort, thereby determining the driving quality of the road. The higher the RQI value, the greater the driving comfort. SRI was calculated according to the lateral force coefficient of the pavement surface, and is used to evaluate the anti-skid performance of the pavement surface. The higher the SRI value, the better the anti-skid performance of pavement surface. Whereas SCI was used to assess the degree of the roadbed damage, including road shoulder damage, slope collapse, and roadbed settlement. The SCI data was stored in unites of 50 m intervals. The higher the SCI value, the lower the degree of roadbed damage. Since the length of the divided road sections is greater than the sampling unit, the pavement conditions indicators of each road segment were replaced by the weighted average of the indicators for the sampling unit included.

TABLE 2. Descriptive statistics of the variables.

Variables	Continuous variables				Discrete variables	
	Mean	Std. Dev.	min	max	Count	Percentage
RESPONSE VARIABLE						
Slight crash counts	0.572	1.087	0	10	–	–
Killed and seriously injured crash count	0.171	0.654	0	4	–	–
EXPOSURE VARIABLE						
Segment length (km)	0.598	0.528	0.106	4.811	–	–
AADT (1000 veh/day)	8.166	4.437	3.926	17.809	–	–
TUNNEL DESIGN FEATURES						
Values of limited speed						
0, 80km/h	–	–	–	–	204	36.36%
1, 100km/h	–	–	–	–	357	63.64%
Length of tunnel						
0, Shorter than 500 m	–	–	–	–	163	29.06%
1, Longer than 500 m	–	–	–	–	398	70.95%
Tunnel entrance indicator						
0, Including the tunnel entrance	–	–	–	–	325	57.93%
1, Not including the tunnel entrance	–	–	–	–	236	42.07%
Tunnel exit indicator						
0, Including the tunnel exit	–	–	–	–	343	61.14%
1, Not including the tunnel exit	–	–	–	–	218	38.86%
Distance between two tunnels (km)	6.189	7.137	0.240	45.209	–	–
Curvature (1/km)	0.291	0.297	0	1.163	–	–
Steep upgrade indicator						
0, Less than 2% of the largest grade	–	–	–	–	443	78.97%
1, Greater than 2% of the largest grade	–	–	–	–	118	21.03%
Steep downgrade indicator						
0, Greater than -2% of the largest grade	–	–	–	–	475	84.85%
1, Less than -2% of the largest grade	–	–	–	–	85	15.15%
Number of one-way lanes						
0, Two lanes	–	–	–	–	1665	74.20%
1, Three lanes	–	–	–	–	579	25.80%
The type of pavement						
0, Cement pavement	–	–	–	–	266	47.46%
1, Asphalt pavement	–	–	–	–	295	52.54%
TRAFFIC CONDITIONS						
Proportion of class 1 vehicles (%)	74.07	4.189	63.041	83.598	–	–
Proportion of class 3 vehicles (%)	8.571	2.816	0.261	13.308	–	–
Proportion of class 4 vehicles (%)	2.368	1.398	0.544	5.289	–	–
Proportion of class 5 vehicles (%)	12.325	4.426	5.279	17.441	–	–
PAVEMENT CONDITIONS						
Pavement surface condition index (PCI)	76.462	7.543	60.431	100	–	–
Riding quality index (RQI)	89.576	5.194	65.919	99.383	–	–
Skiding resistance index (SRI)	86.100	6.483	62.355	98.240	–	–
Subgrade condition index (SCI)	99.544	2.271	83.184	100	–	–

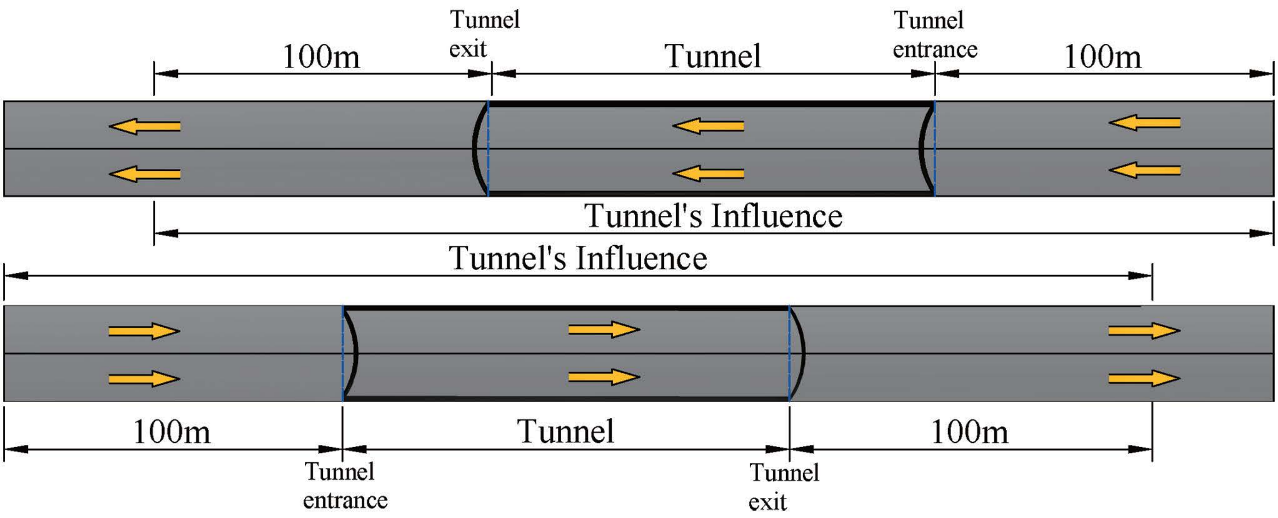

FIGURE 2. Schematic of a tunnel's influence.

TABLE 3. Vehicle classifification.

Class	Characteristics				Examples
	Height (m)	Number of axles	Number of wheels	Wheelbase (m)	
1	<1.3	2	2-4	<3.2	Passenger car, jeep, pick-up truck
2	$\geqslant$ 1.3	2	4	$\geqslant$ 3.2	Minibus, minivan, light truck
3	$\geqslant$ 1.3	2	6	$\geqslant$ 3.2	Medium bus, large ordinary bus, medium truck
4	$\geqslant$ 1.3	3	6-10	$\geqslant$ 3.2	Large luxury bus, large truck, large trailer, 20-feet container truck
5	$\geqslant$ 1.3	>3	>10	$\geqslant$ 3.2	Heavy truck, heavy trailer, 40-feet container truck

According to the literature [42], when the variance influence factor (VIF) of the candidate variable exceeds 5, there is multi-collinearity between the variable and other variables. Such variables should be excluded when selecting variables. The VIF is selected to test whether there is multi-collinearity between the variables (as shown in Table 4). The results in the table show that the VIF values of the variables selected in this paper are less than 5, so there is no significant multi-collinearity between the influencing factors.

TABLE 4. VIF values of the explanatory variables.

Variables	VIF	Variables	VIF
Log (Segment length)	1.64	Number of one-way lanes	2.87
Log (AADT)	1.32	The type of pavement	3.23
Values of limited speed	2.14	The proportion of class 1 vehicles	1.15
Length of tunnel	1.86	The proportion of class 3 vehicles	1.76
Tunnel entrance indicator	1.77	The proportion of class 4 vehicles	1.61
Tunnel exit indicator	1.56	The proportion of class 5 vehicles	1.28
Distance between two tunnels (km)	1.77	Pavement damage condition index (PCI)	1.46
Curvature (1/km)	2.11	Driving quality index (RQI)	1.74
Steep upgrade indicator	2.54	Skid resistance index (SRI)	1.98
Steep downgrade indicator	1.49	Subgrade condition index (SCI)	1.49

III. METHODOLOGY

A. BIVARIATE FIXED PARAMETERS NEGATIVE BINOMIAL MODEL

The number of freeway tunnel crashes is a non-negative integer, and the process of a crash is similar to that of the Bernoulli experiment, which can be regarded as a Poisson distribution when the number of Bernoulli experiments in the model is infinite, and the probability of the event is extremely small. Vehicles passing through the freeway tunnel segments can be approximated as a large number of repeated experiments, and occurring crashes on tunnel segments are small probability events compared to normal driving situations. Thus, many studies have considered the distribution type of crashes as Poisson distribution [36], [38]. But the Poisson distribution restricts the mean and variance to be equal. As shown in Table 2, the means of the SI and KSI crashes are 0.572 and 0.171, respectively. The variances are 1.182 and 0.428, respectively. Means and variances are not equal, and the means are less than variances, indicating samples have significant over-dispersion characteristics. The Poisson distribution can not fit crash frequency effectively. The negative binomial distribution model introduces discrete parameters φ_k based on the Poisson distribution, assuming that the discrete parameters φ_k obey the gamma distribution, and this form is considered more effective in dealing with the over-dispersion of the samples [26], [44]. Based on this, the negative binomial distribution was introduced to describe the distribution of freeway tunnel crash frequency. So This paper adopts the bivariate fixed parameters negative binomial model (BFPNB model) as the basic model to match the non-negative integer and over-dispersion concerning the number of crashes of different severity levels in freeway tunnel sections. Assuming that the crashes of different severity k ($k = 1$, SI crashes; $k = 2$, KSI crashes, respectively.)that occur on the road segment i within the period t obey the NB distribution with a mean value of $\lambda_{i,t,k}$ and a discrete parameter of φ_k. The probability of the BFPNB modelcan be expressed as follows:

$$p\left(Y_{i,t,k} = y_{i,t,k}\right) = \frac{\Gamma\left[\varphi_k + y_{i,t,k}\right]}{\Gamma\left(\varphi_k\right) y_{i,t,k}!}\left(\frac{\varphi_k}{\varphi_k + \lambda_{i,t,k}}\right)^{\varphi_k}\left(\frac{\lambda_{i,t,k}}{\varphi_k + \lambda_{i,t,k}}\right)^{y_{i,t,k}} \quad (1)$$

where, $\Gamma(\cdot)$ is the gamma distribution function, which is used to describe the non-linear relationship between the exposure variables and the number of crashes [48]–[50]; $\lambda_{i,t,k}$ refers to the mean value of the crashes of severity level k occurring on the tunnel segment i within period t:

$$\lambda_{i,t,k} = \alpha_0 F_{it}^{\alpha_1}\ \mathrm{EXP}\left(\beta_n^{i,t,k} X_n^{i,t,k} + \varepsilon_{i,t,k}\right) \quad (2)$$

where, F_{it} is the exposure variables of tunnel segment i at period t, which is the product of AADT and length of tunnel segment i at period t. $X_n^{i,t,k}$ is the n_{th} vector in the data on tunnel design characteristics, traffic condition characteristics, and pavement surface condition characteristics of tunnel segment i in relation to the severity k at time t, while $\beta_n^{i,t,k}$ is a vector of the estimable parameters corresponding to the risk factor n in relation to the crash with a severity level k on tunnel segment i at period t:

$$\beta_n^{i,t,k} = \left(\beta_n^{1,1,k}, \beta_n^{1,2,k}, \ldots, \beta_m^{217,3,k}\right), \quad k = 1, 2, n = 1, 2, \ldots, 20 \quad (3)$$

$EXP\left(\varepsilon_{i,t,k}\right)$ is the error term between the frequency of crash with a severity level k on tunnel segment i at period t. The parameters estimation result, which is used to reflect the non-structural heterogeneity caused by the individual effects of the samples. Assuming that it obeys a bivariate normal distribution with a mean value of 0.

$$\varepsilon_{i,t \sim N_2(0,\Sigma)}, \varepsilon_{i,t} = \begin{pmatrix} \varepsilon_{i,t,1} \\ \varepsilon_{i,t,2} \end{pmatrix}, \quad \Sigma = \begin{bmatrix} \delta_{1,1} & \delta_{1,2} \\ \delta_{2,1} & \delta_{2,2} \end{bmatrix} \quad (4)$$

Σ is the variance-covariance matrix, the main diagonal element $\delta_{k,k}$ represents the variance of the error term $\varepsilon_{i,t,k}$, and the remaining elements represent the covariance between $\varepsilon_{i,t,1}$ and $\varepsilon_{i,t,2}$. In order to measure the effect of non-structural heterogeneity, the standard deviation is calculated using formula $\sigma_k = \sqrt{\delta_{k,k}}$. The correlation coefficient between $\varepsilon_{i,t,1}$ and $\varepsilon_{i,t,2}$ was calculated using formula $\rho = \delta_{1,2}/\left(\sigma_1\sigma_2\right)$. This parameter indicates the correlation between the frequency of crashes of different injury severity levels. The larger the parameter, the stronger the correlation.

B. BIVARIATE RANDOM PARAMETERS NEGATIVE BINOMIAL MODEL

Freeway tunnel crashes are complex events that involve a variety of human responses to external stimuli and complex interactions between the vehicle, roadway features/condition, traffic-related factors, and environmental conditions. With such a complex situation, it is impossible to access all of the data that could potentially determine the likelihood of a freeway tunnel crash or its resulting injury severity. The absence of such important data can potentially present serious specification problems for traditional statistical analyses, leading to biased and inconsistent parameter estimates, erroneous inferences, and erroneous crash predictions.

This paper adopts the random parameters to deal with unobserved heterogeneity in relation to the number of crashes of different severity in freeway tunnel sections. Based on the BFPNB model, the bivariate random parameters negative binomial model(BRPNB model) was developed by introducing random parameters. The BRPNB model assumed that the crashes of different severity k that occur on the road segment i within the period t obey the NB distribution with a mean value of $\lambda_{i,t,k}$ and a discrete parameter of φ. The probability of crashes of different severity is consistent with eq.1, and the mean value is constant with eq.2. Where, $\beta_n^{i,t,k}$ is the random parameters corresponding to the risk factor n concerning the crash with a severity level k on road segment i at time t and is assumed to obey a normal distribution:

$$\beta_n^{i,t,k} \sim N\left(\overline{\beta_{n,k}}, \omega_{n,k}^2\right) \quad (5)$$

where, $\bar{\beta}_{n,k}$ and $\omega_{n,k}^2$ represent the mean value and variance of random parameters, respectively.

C. BIVARIATE RANDOM PARAMETERS NEGATIVE BINOMIAL-LINDLEY MODEL

The BRPNB model does not take the excess zero observations into account. Generally speaking, Freeway tunnel crashes are rare events. The vast majority of tunnel segments may never have had a crash during their operational period, and secondly, every study needs to count crashes at specific periodicities, the tunnel segments may not have had a crash during these periodicities. Figure 3 mainly shows the distribution of crash frequency of different injury severity levels. From the figure, it can be seen that there are a large number of zero observations in the data, 359 samples have never had a SI crash, and 593 samples have never encountered a KSI crash at the statistical periods. As presented in review of this article, most of the current models suffer from flaws in the goodness of fit or the logical interpretation of crashes when fitting data with excess zero observations, so proposing a model that can fit the excess zero observations and explain the crash's logic is essential. Based on this, we have explored the new distribution, the Lindley distribution. The Lindley distribution is characterized by a close to zero mean value and the observed value far from zero with a long tail. It's a mixture of exponential and gamma distribution, and the probability of the Lindley distribution can be defined as follows:

$$f\left(X = x_{i,t,k}; \theta_{i,t,k}\right) = \frac{\theta_{i,t,k}^2}{\theta_{i,t,k} + 1}\left(1 + x_{i,t,k}\right) e^{-\theta_{i,t,k} x_{i,t,k}} \quad (6)$$

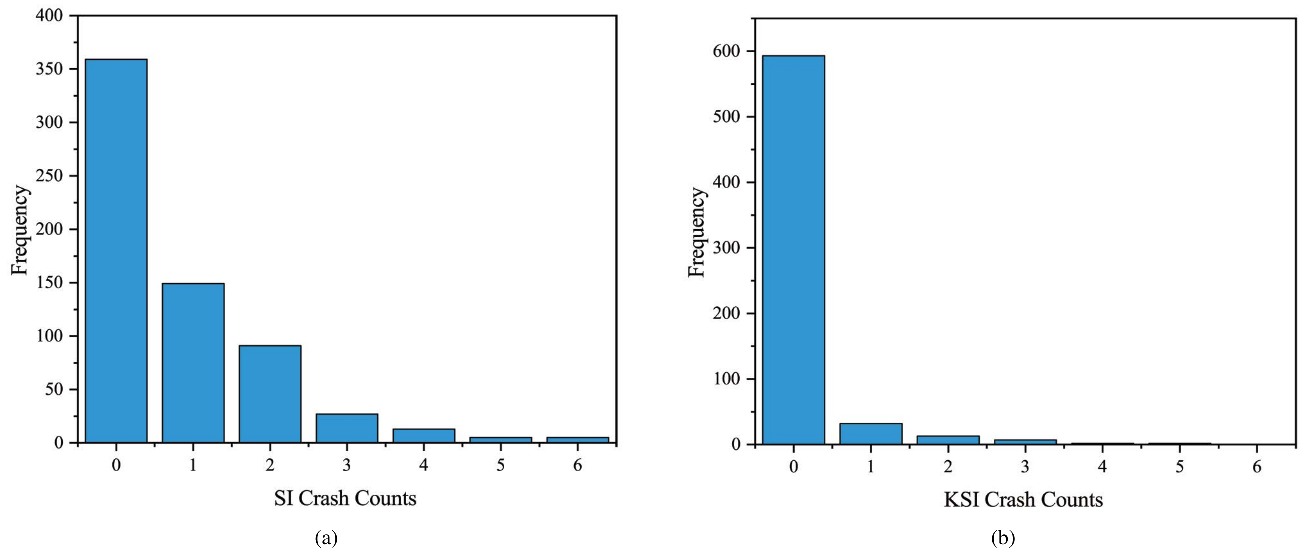

FIGURE 3. Distribution of the crash frequency. (a) respects the frequency distribution of SI crash counts, and (b) respects the frequency distribution of KSI crash counts. Most of the tunnel segments have never had a crash at both the SI and KSI severity levels. The zero observations account for a large proportion of all statistics.

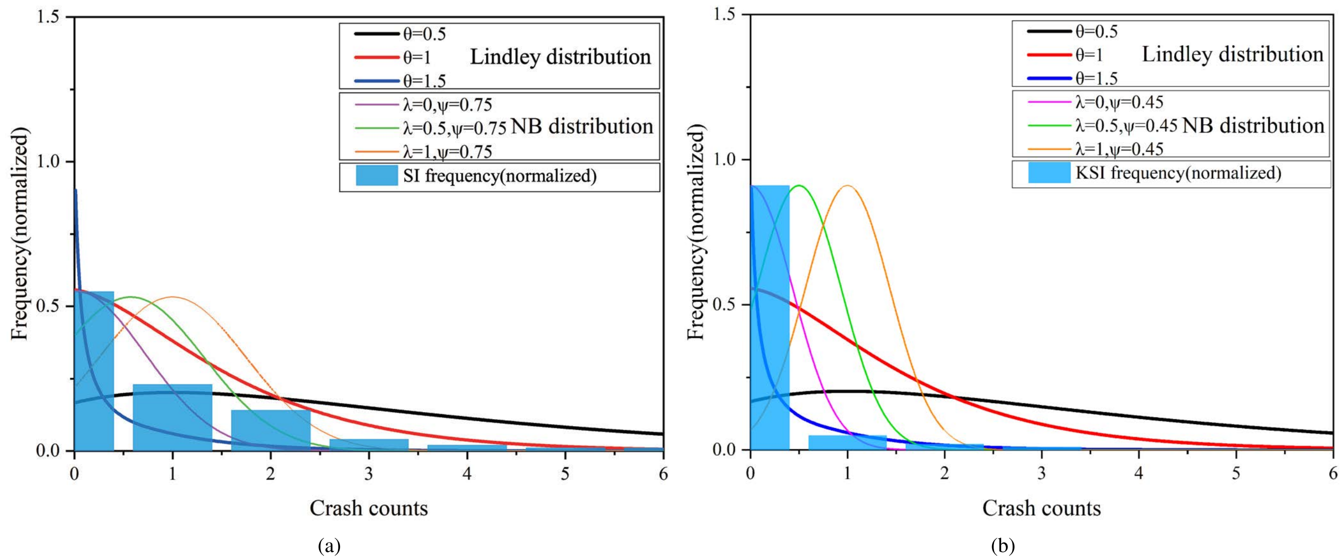

FIGURE 4. Comparison of Lindley distribution and NB distribution. Fitting the tunnel crash frequency distribution when different parameters are chosen for the Lindley distribution and NB distribution. (a) Lindley parameter θ is 0.5, 1, 1.5, respectively, and the parameters of NB distribution are $\lambda = 0, 0.5, 1$, respectively, the parameter φ is 0.75. (b) Lindley parameter θ is 0.5, 1, 1.5, respectively, and the parameters of NB distribution are $\lambda = 0, 0.5, 1$, respectively, the parameter φ is 0.45.

where $\theta_{i,t,k}$ is the Lindley parameter, and $x_{i,t,k}$ is the influencing factor of tunnel segment i in relation to the severity k at time t.

Figure 4 displays that both the NB distribution and Lindley distribution can fit the crash frequency. The figure changes the parameters of different distributions and initially find a distribution shape that is more closely to crash frequency, as shown in Figure 4, when the Lindley parameters θ are approximately 1 and 1.5 respectively, the Lindley distribution is similar to the distribution of SI crash frequency and KSI crash frequency, respectively. When the parameters of NB distribution (λ and φ) are 0 and 0.45, respectively, the NB distribution can fit the SI crash frequency well. When the λ and φ are 0 and 0.75, respectively, the NB distribution is close to the KSI crash frequency. Both the NB distribution and Lindley distribution can fit the zero observations of crash frequency, but it seems that the NB distribution tends to under-estimate observations with a count. From these preliminary comparisons, we speculate that Lindley distribution may have advantages in fitting crash frequency distribution. Therefore we suspect that the Lindley distribution may provide a new perspective on solving the excess zero observations. Inspired by this, introducing the Lindley parameter $\theta_{i,t,k}$ to construct the bivariate random parameters NB Lindley(BRPNB-L) model, the BRPNB-L model is a combination of the BRPNB model and Lindley distribution, and it can be expressed as a combination of the negative binomial distribution, Bernoulli distribution, and gamma distribution, combined with the random parameter method, and assuming that the frequency of crashes of different injury severity levels follows the negative binomial distribution which the mean value is $\lambda_{i,t,k}$ and the discrete parameters is φ_k, the probability density function is as follows.

$$p\left(Y_{i,t,k}=y_{i,t,k}\right) = \frac{\Gamma\left[\varphi_k+y_{i,t,k}\right]}{\Gamma\left(\varphi_k\right) y_{i,t,k}!}\left(\frac{\varphi_k}{\varphi_k+\theta_{i,t,k}}\right)^{\varphi_k}\left(\frac{\theta_{i,t,k}}{\varphi_k+\theta_{i,t,k}}\right)^{y_{i,t,k}} \tag{7}$$

where $\theta_{i,t,k}$ is the Lindley parameter, which is calculated as follows:

$$\theta_{i,t,k} = \lambda_{i,t,k}\psi_{i,t,k} \tag{8}$$
$$\psi_{i,t,k} \sim \text{Gamma}\left(1+\gamma_k, \chi_k\right) \tag{9}$$
$$\gamma_k \sim \text{Bernoulli}\left(1/\left(1+\chi_k\right)\right) \tag{10}$$

D. BIVARIATE RANDOM PARAMETERS SPATIAL NEGATIVE BINOMIAL-LINDLEY MODEL

The BRPNB-L model takes into account unobserved heterogeneity and wants to improve the fit goodness to the excess zero observations. However, as the divided freeway tunnel segments are interconnected, certain unobserved factors may have similar safety impacts on adjacent sections, i.e. there are spatial effects, and the BRPNB-L model ignores spatial effects between adjacent section units and also ignores the spatial effects in the same section with different correlation of spatial effects between crash frequencies with different injury severity [12]. To address this issue, a spatial correction term with a (Multivariate conditional autoregressive prior, MCAR) prior $\phi_{i,k}$ can be added to the link function:

$$\lambda_{i,t,k} = \alpha_0 F_{it}^{\alpha_1}\,\text{EXP}\left(\beta_{i,t,k}X_{i,t,k}+\varepsilon_{i,t,k}+\phi_{i,k}\right) \tag{11}$$

The adjacency structure is an essential component of the MCAR prior [50]. This study uses a 0-1 adjacency matrix. Specifically, the adjacency weights ω_{ij} between segments i and j is 1 when tunnel segment i and tunnel segment j share a common boundary, and otherwise $\omega_{ij}=0$. Based on the 0-1 adjacency matrix, the MCAR prior can be expressed as:

$$\begin{aligned}\Phi_i &\sim N_2\left(\Phi_i, \Omega_s/n_i\right),\\ \Phi_i &= \begin{pmatrix}\phi_{i,1}\\ \phi_{i,2}\end{pmatrix},\\ \Phi_l &= \begin{pmatrix}\phi_{l,1}\\ \phi_{l,2}\end{pmatrix},\\ \Omega_s &= \begin{bmatrix}\delta^s_{1,1} & \delta^s_{1,2}\\ \delta^s_{2,1} & \delta^s_{2,2}\end{bmatrix}\end{aligned} \tag{12}$$

where n_i is the number of sections adjacent to tunnel section i,$\overline{\phi_{l,k}} = \Sigma_{i\neq j}\phi_{j,k}\omega_{ij}/n_i$, Ω_s is the variance matrix of spatial correlations, where the main diagonal elements represent the variance of the spatial effects of slight injury, severe and fatal injury crashes, respectively. The remaining elements represent the covariance of the spatial effects of crashes between the two different severity levels.

E. BIVARIATE RANDOM PARAMETERS SPATIAL-TEMPORAL NEGATIVE BINOMIAL-LINDLEY MODEL

Due to the time-varying characteristics of data such as traffic conditions and road surface conditions, there are interactions between such time-varying variables and unobserved factors during the data collection period. Some of the unobserved factors do not change over time, leading to a certain correlation between the number of crashes on the same road section i at different periods, and ignoring this correlation will lead to bias in parameter estimation and thus to wrong inferences. At the same time, there is an interaction between spatial and temporal correlation, i.e., the spatial effect may change over time, and the temporal effect may vary with road section. To account for the spatio-temporal correlation, a spatial term, a temporal term, and a spatio-temporal interaction term are added to the link function of the SP-BRPNB-L model. The link function can be expressed as follows.

$$\lambda_{i,t,k} = \alpha_0 F_{it}^{\alpha_1}\,\text{EXP}\left(\beta_{i,t,k}X_{i,t,k}+\varepsilon_{i,t,k}+\phi_{i,k}+\tau_t\left(\alpha_k+\zeta_{i,k}\right)\right) \tag{13}$$

where τ_t represents the temporal scalar parameter over period t, α_k is the scalar parameter of the linear time trend of different severities for all tunnel segments, and $\zeta_{i,k}$ is the spatial component of the spatio-temporal interaction, obeying the MCAR prior.

$$\begin{aligned}\Phi_i &\sim N_2\left(\Phi_l, \Omega_s/n_i\right),\\ \Phi_i &= \begin{pmatrix}\phi_{i,1}\\ \phi_{i,2}\end{pmatrix},\\ \Phi_l &= \begin{pmatrix}\phi_{l,1}\\ \phi_{l,2}\end{pmatrix},\\ \Omega_s &= \begin{bmatrix}\delta^s_{1,1} & \delta^s_{1,2}\\ \delta^s_{2,1} & \delta^s_{2,2}\end{bmatrix}\end{aligned} \tag{14}$$

F. DESCRIPTION OF THE GOODNESS-OF-FIT INDICATOR

The criteria, Deviation Information Criterion(DIC), Mean Absolute Deviance (MAD), and Mean Square Prediction Error(MSPE), are often used to evaluate the goodness of fit of Bayesian models. The DIC is a hierarchical combination of the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion), which is a combined measure of model complexity and fit and reflects the estimation efficiency and accuracy of the model.

$$DIC = \bar{D} + PD \tag{15}$$

$\bar{D}$ is the posterior mean-variance, reflecting the fit of the model, and PD is the number of valid parameters, reflecting the complexity of the model. In general, the smaller the DIC of a model, the better its performance, and a DIC difference of 5 indicates significant differences between models [51].

The MAD and MSPE reflect the degree of deviation of the fitted crash frequency of the model from the actual crash frequency. To assess the goodness of fit of the model for crash frequencies at different severity levels, MAD and MSPE were calculated for different severity levels [49].

$$MAD_k = \frac{1}{3 \times 217} \sum_{i=1}^{i=217} \sum_{t=1}^{3} \left| Y_{i,t,k} - \lambda_{i,t,k} \right| \quad (16)$$

$$\mathrm{MSPE}_k = \frac{1}{3 \times 217} \sum_{i=1}^{i=217} \sum_{t=1}^{3} \left(Y_{i,t,k} - \lambda_{i,t,k} \right)^2 \quad (17)$$

G. EVALUATION INDICATORS FOR SAFETY EFFECTS

To understand the effect of observed factors on the crash frequency of different severity levels, the range of variation of crash frequency expected for a unit period increase in the independent variable was calculated [46], i.e., The Incidence Rate Ratio (IRR). Also, from the IRR value can be indirectly seen that significant risk factors are linearly or non-linearly correlated with crash frequency. When the IRR value is close to 1, the risk factors are linearly correlated with crash frequency; conversely, they are non-linearly correlated.

$$\begin{aligned}\mathrm{IRR_m} &= \frac{E\left(Y_{itk} \mid \tilde{X}_{\mathrm{i,t,k}}, L_{\mathrm{i}}, V_{i,t,k}, x_m + 1\right)}{E\left(Y_{itk} \mid \tilde{X}_{i,t,k}, L_{\mathrm{i}}, V_{i,t,k}, x_m\right)} \\ &= \exp\left(\beta_m\right) \end{aligned} \quad (18)$$

where $\tilde{X}_{i,t,k}$ denotes the influence of factors other than x_m for crashes with a period of t and severity k on segment i, β_m is the regression coefficient of the exposure variable x_m and the IRR value of the exposure variable represents the value of the change in crash frequency corresponding to a 1% shift in the exposure variable, with the equation:

$$\mathrm{IRR_e} = \frac{E\left(Y_{i,t,k} \mid X_{\mathrm{i,t,k}}, 1.01\, X_e\right)}{E\left(Y_{i,t,k} \mid X_{i,t,k}, X_e\right)} = 1.01^{\beta_e} \quad (19)$$

β_e is the regression coefficient of X_e.

IV. RESULTS AND DISCUSSION

A. PLATFORM AND PARAMETERS SETTING BASED ON THE BAYESIAN MODEL

With the advancement of computing methods, Bayesian inference has become more widespread. As there is no need for Bayesian inference to use the closed likelihood function, thereby reducing the complexity of the model parameters estimation, its application is becoming more popular [52]. All candidate models used in this paper were programmed, parameters estimated, and evaluated in OPENBUGS. This software was simulated using MCMC (Markov Chain Monte Carlo), which is currently widely represented by the algorithms of Gibbs sampling and Metropolis-Hastings [14].

The prior distribution of parameters is required for Bayesian estimation, and non-informative prior distributions are usually specified when the prior information is insufficient. In this study, the diffusion prior distribution N $\left(0, 10^4\right)$ was used as the prior average of random parameters $\beta_{k,m}(m = 1, 2, \ldots, 20; k = 1, 2)$ and scalar parameters $(k = 1, 2)$, and the diffusion gamma distribution $(0.001, 0.001)$ was used as the prior average of random parameters. The Wishart distribution (W(P,r)) was set as the prior of Σ^{-1}, Ω_s^{-1} and Ω_a^{-1}, where, P is the proportional matrix; r is the degree of freedom. A chain of 500,000 simulation iterations was constructed for each model, and the first 450,000 iterations were excluded as aging. The convergence of MCMC was evaluated according to the Gelman-Rubin statistics in openbugs. According to research [12], a parameter is defined as random if the Bayesian estimated standard deviation of the parameter is significantly different from zero, otherwise, it is estimated as a fixed parameter.

B. MODEL COMPARISON

The regression coefficients and goodness of fit indicators estimated by the candidate models in this paper are shown in Table 5. First, comparing the estimation results of the BRPNB model and BFPNB model, it can be seen that the BRPNB model has lower DIC, MAD, and MSPE values than the BFPNB model, indicating the goodness of fit of the BRPNB model is better, Showing that the proposed model introduced random parameters to consider the unobserved heterogeneity can primarily improve the performance of fitting crash data of different injury severity levels.

Secondly, according to the comparison results between the bivariate random parameters NB Lindley model and the bivariate random parameters NB model, regardless of the severity of the crash data set, the former has lower DIC, MAD, and MSPE values than the latter, indicating that the BRPNB-L model is more suitable for tunnel crash data sets with excess zero observations, thus obtaining better goodness of fit. At the same time, the Lindley parameter in the BRPNB-L model is significant (2.187) at 95% Bayesian credibility level, which also reflects the effect of Lindley distribution. This result is consistent with the research result of Tang Feng [20]. Both results support the idea that Lindley distribution can improve the model's ability to fit excess zero observations.

Thirdly, this study discusses the unobserved heterogeneity and spatio-temporal interaction by comparing the three models of BRPNB-L, SP-BRPNB-L, and ST-BRPNB-L. According to Table 5, the DIC value (3158), MAD value (SI: 0.31; KSI: 0.24), and MSPE value (SI: 0.21; KSI: 0.13) for SI and KSI crashes in the SP-BRPNB-L model are all lower than the DIC value (3611), MAD value (SI: 0.42; KSI: 0.30), and MSPE value (SI: 0.33; KSI: 0.20) in the BRPNB-L model, indicating the former is more suitable for the crash data set than the latter.

The spatial effects of SI and KSI crash frequency are significant at the 95% Bayesian credibility level, 4.19 and 2.69, respectively. Previous research has shown that [44], [53], [54], considering spatial effects can significantly improve the fit of the model. At the same time, the correlation between the spatial effects of different severity for crashes is relatively high ($\rho_s = 0.96$), indicating the necessity of using MCAR prior for bivariate spatial modeling. The strong correlation between spatial effects may be attributed to the spatial aggregation of factors such as terrain and socio-economics and are shared by crashes of different severity at the same time. After considering the spatial correlation, the non-structural heterogeneity between SI and KSI crashes was reduced by 18% and 9.22%, respectively. Literature [22], [55] found that part of the unobserved heterogeneity in the bivariate model is derived from the same unobserved heterogeneity in adjacent tunnel segments or the influence of spatial effects.

After incorporating temporal correlation and spatio-temporal interaction into discussion, the DIC value (=3014) and MAD values (SI=0.29, KSI=0.23) are lower than those of the SP-BRPNB-L model (DIC=3158, MAD-SI=0.31, MAD-KSI=0.24). The MSPE values of both models are equal in SI crashes, and the spatio-temporal model in KSI crashes appears to be smaller. It is evident that the spatio-temporal model significantly improves the goodness of fit compared to the spatial model. It can be seen from the results of the ST-BRPNB-L model that the scalar parameters of the linear time trend for the frequency of crashes with two severities are significant, indicating that there is a non-negligible temporal correlation in the data set. The scalar parameters indicate that the frequency of SI crashes, but that of SCI crashes will increase over time. The temporal correlation between SI and KSI crashes can be further demonstrated by the estimates of δa(1, 1) and δa(2, 1). P_a is estimated to be 0.99, indicating the strong spatial-temporal correlation among the frequency of crashes of different severity. This shows that the lack of factors has an impact on the safety of the tunnel. These factors are the same on the adjacent tunnel segments and are dependent on temporal changes. In addition, SI and KSI crashes may share the same spatial-temporal conditions.

In summary, the ST-BRPNB-L model has lower DIC, MAD, and MSPE values than BRPNB, BRPNB-L, and SP-BRPNB-L models when fitting SI and KSI crash data, indicating that the ST-BRPNB-L model is the optimal model.

C. ANALYSIS OF PARAMETERS ESTIMATION RESULTS

The Bayesian estimation results of the regression parameters of the candidate models are shown in Table 6. This section is based on the optimal model ST-BRPNB-L to analyze the parameters estimation results. Table 6 shows that, in the ST-BRPNB-L model, the length of the road segments, AADT, whether the road segment contains a tunnel entrance, a tunnel exit, is an ascending or a descending road segment, the proportions of class 3 vehicles and class 5 vehicles are positively correlated with the frequency of KSI crashes at the level of 95%, and distance between tunnels, curvature, PCI and SRI are negatively correlated with the frequency of KSI crashes at the level of 95%. While for the frequency of SI crashes, it is not significantly correlated with the proportion of type 3 vehicles and PCI, but is significantly positively correlated with the logarithm of tunnel segment length, the logarithm of AADT, whether the tunnel segment contains a tunnel entrance, a tunnel exit, is an ascending or a descending road segment, and the proportion of class 5 vehicles at the statistical level of 95%, significantly negatively correlated with the curvature at the level of 95%, and the distance between tunnels and SRI at a statistical level of 90%. The estimation results showed that, for SI crashes, the regression parameters for whether the road segment contains a tunnel entrance or is a descending road segment and for class 5 vehicles are all random parameters. For KSI crashes, the regression parameters for whether the road segment is a downhill section, the proportion of class 5 vehicles, and SRI are all random parameters. The standard deviations of these variables are significant at the 95% Bayesian credibility level, as shown in Table 7. A larger standard deviation of the regression parameters indicates more significant fluctuations of parameters estimation values for random variables among samples. This indicates that this variable and other variables have strong correlations and are significantly heterogeneous [12]. The order of standard deviation for each random parameter of SI crashes is as follows: tunnel entrance (0.742) > steep ascending road segment (0.448) > proportion of class 5 vehicle (0.245). For a KSI crash, the order of standard deviation for each random parameter is as follows: steep descending road segment (0.539) > proportion of class 5 vehicle (0.197) > SRI (0.109). The influence of the above random variables on the frequency of crashes of different severity will be explained in detail in the following section.

D. ANALYSIS ON THE SAFETY EFFECT OF SIGNIFICANT RISK FACTORS

Similarly, the parameters estimation results of the optimal model ST-BRPNB-L are explained in Figure 5, Table 6, Table 7, and Table 8. As shown in Table 8 and Figure 6, the IRR values are used to discuss the impact of significant risk factors on SI and KSI crashes.

1) EXPLANATION OF RANDOM PARAMETERS

The parameters estimation results show that the descending gradient indicator and proportion of the five classes of vehicles have a heterogeneous effect on the frequency of SI and KSI crashes. While the tunnel exit variable has a heterogeneous effect on the frequency of SI crashes, SRI only has a heterogeneous effect on the frequency of KSI crashes. The regression parameters for road segment length and AADT have a heterogeneous effect on the frequency of crashes of different severity only. The regression parameters of the logarithms of road segment length and AADT are significantly positive at the Bayesian credibility level of 95%. The regression parameters indicate that road segment length and AADT have a significant fixed effect on the frequency of crashes of different severity. Despite the same sign of the remaining essential factors, the coefficients in relation to both crashes of different injury severity levels are different.

TABLE 5. Estimation results and goodness of fit indexes of the candidate models in this paper.

Evaluation indexes	BFPNB model Estimates	BRPNB model Estimates	BRPNB-L model Estimates	SP-BRPNB-L model Estimates	ST-BRPNB-L model Estimates
Dispersion parameter φ	2.367	3.424	3.913	4.598	4.321
Lindley parameter θ	-	-	2.187	1.739	1.176
$\delta_{1,1}$	1.93	1.24	0.89	0.73	0.70
$\delta_{1,2}(=\delta_{2,1})$	0.54	0.36	0.21	0.17	0.11
$\delta_{2,2}$	2.18	1.98	2.17	1.97	1.23
ρ	0.26	0.23	0.15	0.14	0.12
$\delta_s 1,1$	-	-	-	17.52	37.21
$\delta_s 1,2(=\delta_s 2,1)$	-	-	-	10.81	24.49
$\delta_s 2,1$	-	-	-	7.21	16.89
ρ_s	-	-	-	0.96	0.98
α_1	-	-	-	-	8.99
α_2	-	-	-	-	-2.33
$\delta_a 1,1$	-	-	-	-	49.20
$\delta_a 1,2(=\delta_a 2,1)$	-	-	-	-	19.02
$\delta_a 2,1$	-	-	-	-	7.46
ρ_a	-	-	-	-	0.99
DIC	4037	3969	3611	3158	3014
MAD-SI	0.59	0.49	0.42	0.31	0.29
MAD-KSI	0.36	0.33	0.30	0.24	0.23
MSPE-SI	0.51	0.44	0.33	0.21	0.21
MSPE-KSI	0.35	0.24	0.20	0.13	0.11

TABLE 6. Estimation results of the regression coefficients in the bivariate random parameters spatio-temporal NB-L model.

Variables	Slight injury			Killed and serious injury		
	Estimates	95% BCI	90% BCI	Estimates	95% BCI	90% BCI
Constant	-28.9	**(-36.014,-22.368)**	**(-34.854,-22.496)**	-19.2	**(-23.780,-15.558)**	**(-22.886,-15.018)**
Log(Segment length)	1.128	**(0.924,1.339)**	**(0.930,1.331)**	1.174	**(0.893,1.380)**	**(0.957,1.359)**
Log(AADT)	1.987	**(1.601,2.424)**	**(1.657,2.399)**	2.679	**(2.113,3.290)**	**(2.212,3.181)**
Distance between two tunnels	-0.129	(-0.162, 0.101)	**(-0.157,-0.103)**	-0.198	**(-0.242,-0.156)**	**(-0.234,-0.167)**
Tunnel entrance indicator	0.304	**(0.235,0.361)**	**(0.244,0.360)**	0.412	**(0.321,0.486)**	**(0.329,0.475)**
Tunnel exit indicator	0.143	**(0.108,0.168)**	**(0.121,0.165)**	0.357	**(0.281,0.429)**	**(0.299,0.420)**
Curvature	-0.048	**(-0.059,-0.038)**	**(-0.056,-0.040)**	-0.029	**(-0.035,-0.023)**	**(-0.033,-0.023)**
Steep upgrade indicator	0.375	**(0.293,0.437)**	**(0.302,0.435)**	0.189	**(0.155,0.227)**	**(0.156,0.219)**
Steep downgrade indicator	0.228	**(0.171,0.268)**	**(0.190,0.263)**	0.429	**(0.331,0.521)**	**(0.358,0.520)**
Proportion of class 3 vehicle	-	-	-	0.059	**(0.046,0.071)**	**(0.049,0.079)**
Proportion of class 5 vehicle	0.072	**(0.054,0.087)**	**(0.059,0.086)**	0.084	**(0.064,0.102)**	**(0.065,0.098)**
PCI	-	-	-	-0.01	**(-0.007,-0.012)**	**(-0.008,-0.012)**
SRI	-0.033	(-0.031,0.037)	**(-0.026,-0.040)**	-0.021	**(-0.017,-0.025)**	**(-0.018,-0.025)**

Note: (1)BCI respects the bayesian credible interval.
(2)Boldface indicates statistical significance at the corresponding level.

Therefore, it is necessary to model the frequency of crashes with different injury severity levels to quantify different impacts of influencing factors on different injury severity levels.

In the SI crash analysis, the regression parameters of the steep downgrade indicator were estimated to obey a normal distribution with a mean value of 0.228 and a variance of 0.448, as shown in Figure 5(a), which indicates that, in 69.46% of the observed road segments, the frequency of SI crashes will increase as the descending gradient increases, while in the remaining 30.54% observed road segments, the frequency of SI crashes will decrease as the descending gradient increases. In the KSI crash analysis, the regression parameters of the steep downgrade indicator were estimated to obey a normal distribution with a mean value of 0.429 and a variance of 0.539, as shown in Figure 5(b). This indicates that, in 78.70% of the observations, the frequency of KSI crashes will increase as the descending gradient increases, and for the remaining 21.30% road segments, the frequency of KSI crashes will decrease as the descending gradient increases. These findings are consistent with studies [12], [22], and also in line with engineering experience, that is, on steep

TABLE 7. Estimated standard deviations of the random parameters in bivariate random parametersspatio-temporal NB-L model.

Slight injury				Killed and serious injury			
Random variable	Std. Dev.	90% BCI	95% BCI	Random variable	Std. Dev.	90% BCI	95% BCI
Tunnel entrance indicator	0.742	(0.594,0.885)	(0.561,0.925)	Steep downgrade indicator	0.539	(0.426,0.634)	(0.407,0.648)
Steep downgrade indicator	0.448	(0.379,0.517)	(0.368,0.551)	Proportion of class 5 vehicles	0.197	(0.162,0.239)	(0.150,0.242)
Proportion of class 5 vehicles	0.245	(0.192,0.285)	(0.190,0.294)	SRI	0.109	(0.091,0.131)	(0.090,0.132)

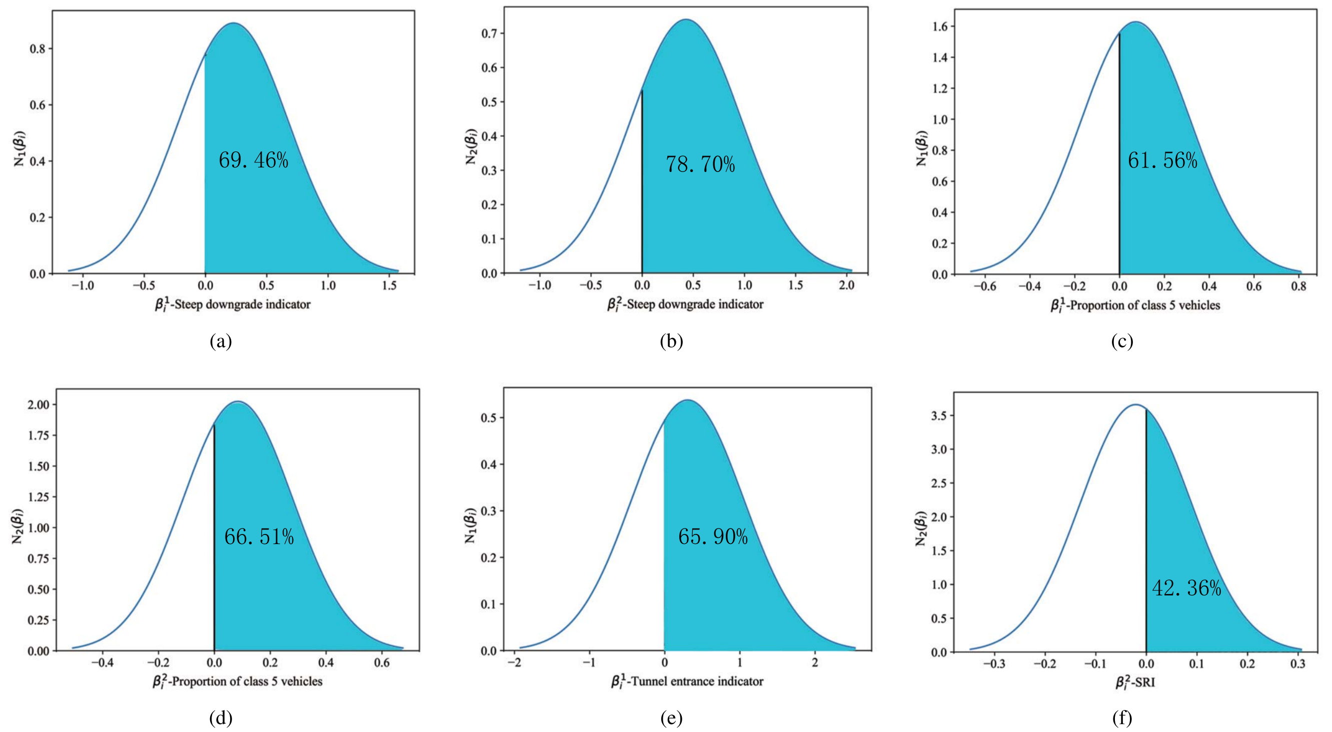

FIGURE 5. Distributions of the random parameters in the ST-BRPNB-L model. (a),(b) represent the random parameter distributions of steep downgrade indicator at SI and KSI severity level, respectively, (c),(d) represent the random parameter distributions of the proportion of class 5 vehicles at SI and KSI severity level, respectively,(e) respects the random parameter distributions of tunnel entrance indicator at SI severity level, and (f) respects the random parameter distributions of SRI at KSI severity level.

TABLE 8. IRR values of each significant variable in the bivariate random parameters spatio-temporal NB-L model.

Variables	Slight injury	Killed and serious injury
Log(Segment length)	1.011	1.012
Log(AADT)	1.020	1.027
Distance between two tunnels	0.879	0.820
Tunnel entrance indicator	1.355	1.510
Tunnel exit indicator	1.154	1.429
Curvature	0.953	0.971
Steep upgrade indicator	1.455	1.208
Steep downgrade indicator	1.256	1.536
Proportion of class 3 vehicle	–	1.061
Proportion of class 5 vehicle	1.075	1.088
PCI	–	0.990
SRI	0.968	0.979

descending road sections, with the increasing of vertical gradient, the probability of a crash which caused by brake failure due to frequent braking will increase. At the same time, the IRR value shows that the tunnel segment with a descending gradient greater than 2% is 25.6% and 53.6% higher than the remaining road segments with SI and KSI crashes.

In the SI crash analysis, the regression parameters for the proportion of class 5 vehicles were estimated to obey a normal distribution with a mean value of 0.072 and a variance of 0.245, as shown in Figure 5(c). This shows that, in 61.56% of the observations, the frequency of SI crashes will increase as the proportion of class 5 vehicles increases, and in the remaining 38.44% observations, the frequency of SI crashes will decrease as the proportion of class 5 vehicles increases. In the KSI crash analysis, the regression parameters for the proportion of class 5 vehicle were estimated to obey a normal distribution with a mean value of 0.084 and a variance of 0.197, as shown in Figure 5(d). The results showed that, in 66.51% of the observations, the frequency of SI crashes would increase as the proportion of class 5 vehicles increases, and in the remaining 33.49% observations, the frequency of KSI crashes will decrease as the proportion of class 5 vehicles increases. According to the research results of [14], the proportion of class 5 vehicles is related to the driver's

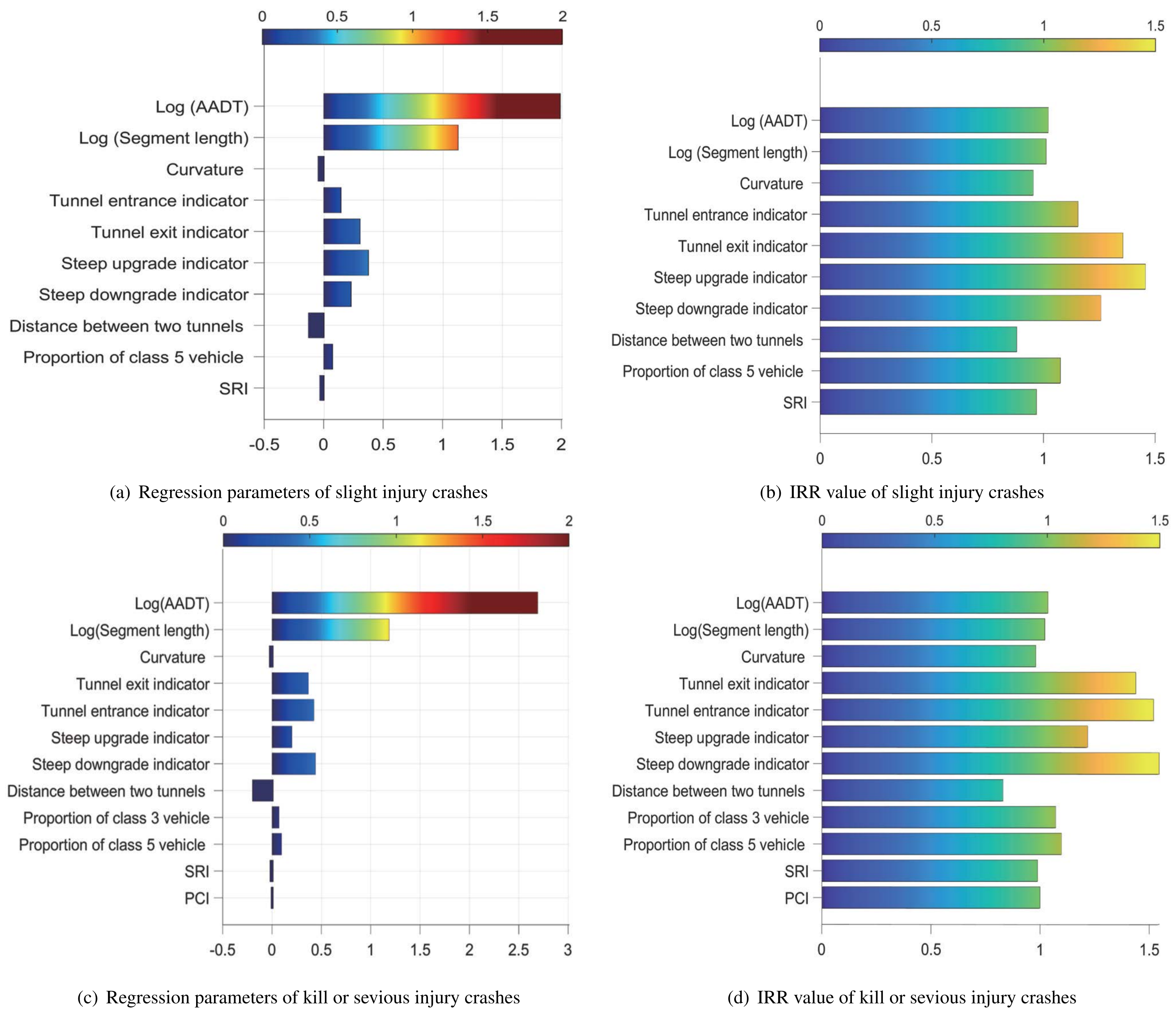

FIGURE 6. The estimated parameters and IRR values of significant variables based on the SP-MRPNB-L models. (a) represents the estimated parameters of significant variables at the SI crash severity level. (b) represents the IRR values of significant variables at the SI crash severity level. (c) represents the estimated parameters of significant variables at the KSI crash severity level. (d) represents the IRR values of significant variables at the KSI crash severity level.

age, pavement conditions, and the traffic environment, which jointly affect the frequency of crashes. Therefore, the proportion of class 5 vehicles as a random variable is mainly attributed to pavement conditions of each tunnel and drivers' driving habits. Some research [15], [56] believed that, as the proportion of class 5 vehicles increases, the frequency of highway crashes would see a significant decrease. As for the main reason, heavy-duty vehicle drivers are more skillful in driving and are more familiar with the established routes. In addition, the driving duration of heavy-duty vehicle drivers is strictly regulated in China. So, the proportion of class 5 vehicles is negatively correlated with the frequency of crashes. However, the above research is based on open roads on highways, and tunnel sections have more enclosed space. As the proportion of class 5 vehicles increases, 1) The driving speed in the tunnel is slower, which increases the possibility of lane change and overtaking by vehicles behind; 2) The width of the lateral field of view in the tunnel is narrower, and heavy-duty vehicle drivers will have a much stronger sense of oppression. The lateral position of the vehicle will be frequently adjusted to maintain its spacing from the tunnel wall, which will cause inference to adjacent lanes. 3) Heavy-duty vehicles may limit the field of view of small vehicles nearby, thereby adding the possibility of collisions between different vehicles. Research [12], [14] supported the ideas in this paper. Meanwhile, the IRR value for the proportion of class 5 vehicles (1.075 and 1.088 for SI and KSI accidents) indicated that the proportion of class 5 vehicles in the traffic volume is positively correlated with the incidence of traffic crashes of different severity. In particular, when the proportion of class 5 vehicles increases by 1%, the frequency of slight injury crashes will increase by 7.5% (=1.075-1), and the frequency of KSI crashes will increase by 8.8% (=1.088-1).

For SI crashes, the regression parameters of the tunnel entrance were estimated to obey a normal distribution with a mean value of 0.304 and a variance of 0.742, as shown in Figure 5(e). This showed that, in 65.90% of the observations, the segment with a tunnel entrance has a higher frequency of SI crashes, and in the remaining 34.10% observations, the segment with a tunnel entrance has a lower frequency of SI crashes. According to study [14], this random variable and the physiological conditions of the driver are correlated with each other. It is difficult for most drivers to adapt to the rapidly changing lighting conditions, and the violent psychological fluctuations and operating errors of the driver may lead to an increase in the frequency of crashes. In addition, the IRR values of the tunnel entrance index (1.355 and 1.510 for SI and KSI crashes) indicate that the frequency of SI and KSI crashes on the road segment with a tunnel entrance is 35.5% and 51% higher than that of other road segments, respectively.

The regression parameters of SRI were estimated to obey a normal distribution with a mean value of -0.021 and a variance of 0.109, as shown in Figure 5(f). This showed that, in 42.36% of the observations, the frequency of KSI crashes would increase as the SRI increases, and in the remaining 57.64% observations, the frequency of KSI crashes will decrease as the SRI increases. Also, the IRR value of SRI (0.968 and 0.979 for SI and KSI crashes) showed that SRI is negatively correlated with the incidence of traffic crashes of different severity. In particular, the frequency of SI crashes will decrease by 3.2% (=0.968-1), and that of KSI crashes will reduce by 2.1% (=0.979-1) when SRI increases by 1%.

2) EXPLANATION OF FIXED PARAMETERS

Firstly, the length of road segment and AADT are significantly positively correlated with the frequency of SI and KSI crashes at the 95% Bayesian credibility level. The regression parameters of road segment length that correspond to the frequency of SI and KSI crashes are 1.128 and 1.174, respectively, indicating that the frequency of SI and KSI crashes in the tunnel will become higher as the length of tunnel segments increases. The regression parameters of AADT that correspond to the frequency of SI and KSI crashes are 1.987 and 2.679, respectively, showing that the greater the traffic volume, the higher the frequency of crashes of different severity in the tunnel, which is consistent with engineering experience. Studies [14], [22], [24] also supported this viewpoint. When crashes occurred on each road segment and said segment is affected by other external factors, each tunnel segment will have an equal probability of crashes. Therefore, crash frequency will increase as the tunnel segment length increases. At the same time, the IRR values of road segment length that correspond to SI and KSI crashes are 1.011 and 1.012, respectively. This showed that, for every 1% increase in the Log(segment length), the frequency of SI crashes would increase by 1.1%, and that of KSI crashes by 1.2%. Literature [31] showed that the crash frequency on an open road would decrease as the AADT increases. The researchers believed that the speed of the traffic would become extremely low when there is congestion on the road, thereby avoiding traffic crashes. And literature [22], [24], [26], [31] described the principle of the impact of traffic volume on the crash rate, that is, the driver is disturbed more by surrounding vehicles when they drive on a road segment with heavy traffic volume. Also, the enclosed environment of the tunnel segment has aggravated the inference among traffic flow, with a higher probability of crashes. The IRR values of AADT corresponding to SI and KSI crashes are 1.020 and 1.027, respectively. This indicated that, for every 1% increase in Log(AADT), the frequency of SI crashes would increase by 2.0%, and that of SCI crashes by 2.7%.

Secondly, the impact of distance between adjacent tunnels on the frequency of crashes of different severity was discussed. The regression parameters of the distance between adjacent tunnels for SI and KSI crashes are −0.129 and −0.198, respectively, indicating the significant negative correlation of the distance between adjacent tunnels with the frequency of SI and KSI crashes, which is consistent with the results of [14]. Before entering the next tunnel, the driver has sufficient time to relive their mental state, reducing crash frequency. Meanwhile, study [12] showed that frequently changing lighting conditions within a short distance will make drivers exposed to greatly different lighting conditions (lighting conditions and natural lighting conditions in tunnels). As a result, the driver will become more nervous and feel it is difficult to adapt to the conditions, causing an increase in crash frequency. IRR values of this indicator that correspond to SI and KSI crashes are 0.879 and 0.820, respectively. This indicated that for an increase in every unit of the distance between adjacent tunnels, the frequency of SI crashes will decrease by 12.1% and that of KSI crashes by 18%.

A road segment that contains a tunnel entrance or exit is significantly positively correlated with the frequency of SI and KSI crashes at the 95% Bayesian credibility level. Among them, the regression parameters for whether the road segment contains tunnel entrances in SI and KSI crashes are 0.304 and 0.412, with IRR values of 1.355 and 1.510, respectively; the regression parameters for whether the road segment contains a tunnel exit in SI and KSI crashes are 0.143 and 0.357, with IRR values of 1.154 and 1.429 respectively. It shows that the frequency of SI and KSI crashes in the segment with tunnel entrance and exit is higher than that of other tunnel segments; and at the entrance and exit of the tunnel, the frequency of SI crashes is 35.5% and 15.4% higher than that of other tunnel segments; and the frequency of KSI crashes is 51.0% and 42.9% higher than that of other tunnel segments, which are consistent with the results of studies [12], [22]. Due to the rapidly changing lighting conditions at the entrance and exit of the tunnel, the driver's vision will respond fast. The physiological effect of the driver caused may lead to their improper operation and then result in traffic crashes. The frequency of SI and KSI crashes at the entrance of the tunnel are higher than that at the exit of the tunnel, showing high crash frequency at the tunnel entrance and higher severity of the crash, which is consistent with the results of [8]. This means that when the illumination in the drivers' field of vision changes dramatically in a short time, there will be temporary blindness, and they need a certain time interval to recognize the internal situation of the tunnel, which is commonly referred to as ''the black hole effect(strong light environment to weak light)'' and ''the white hole effect(weak light environment to strong light)''. Such phenomena will seriously affect traffic safety.

In the analysis for SI and KSI crashes, the regression parameters of curvature were estimated to be -0.048 and -0.029, respectively, indicating the negative correlation of curvature with the frequency of SI and KSI crashes. The results of this study are consistent with the conclusions of [24] and are opposite to the conclusions of [12], [25]. The reason lies in the fact that the selection of the design indicator values for freeways in China, especially freeway tunnel sections, is relatively conservative. The actual design geometry values are usually much higher than the value of design indicators, and the indicator selection is generally reasonable. Such crash is generally caused by the decrease in the driver's alertness due to the excessively high alignment index. Therefore, an appropriate increase in curvature of a turn improves tunnel traffic safety. The IRR values corresponding to SI and KSI crashes are 0.953 and 0.971, respectively. This indicates that, for every 1% increase in curvature, the frequency of SI crashes in the tunnel segment decreases by 4.7%, and that of KSI decreases by 2.9%.

The proportion of class 3 vehicles is significantly positively correlated with the frequency of KSI crashes at the 95% Bayesian credibility level, with a regression parameter of 0.059 and IRR value of 1.061. This indicates that, for every 1% increase in the proportion of class 3 vehicles, the frequency of KSI crashes increases by 6.1%.

PCI has a significant effect on KSI crashes and is negatively correlated with the frequency of KSI crashes, with a regression parameter of -0.010. The results of this study are consistent with those of [12], [14]. That is, a better pavement condition indicates a lower frequency of crashes on the tunnel segment. The IRR value is 0.990, indicating that for every 1% increase in PCI indicator, the frequency of KSI crashes decreases by 1%.

V. CONCLUSION

A bivariate spatio-temporal analysis method was adopted in this study. The frequency of crashes in the tunnel section was classified by severity levels in combination with the mixed distribution method that considers the excess zero observations and the random parameters method. In addition, a bivariate random parameters negative binominal Lindley model that considers the spatio-temporal effects was established, aiming to explore the influence of tunnel design characteristics, traffic conditions, pavement conditions, and other factors by crashes of different injury severity levels on freeway tunnels in China. In order to demonstrate the superiority of the ST-BRPNB-L model, this model was compared with the BFPNB(widely used by researchers), BRPNB, BRPNB-L, and SP-BRPNB-L models. The results showed that the ST-BRPNB-L model reduces the estimation bias caused by excessive zero observations through the introduction of the Lindley distribution; introducing the independent spatial effect (specified by MCAR prior), temporal effect (specified by linear temporal trend), and spatio-temporal effect (formed by the product of the temporal trend and the prior of spatial term and MCAR) reduce estimation bias caused by spatio-temporal correlation. At the same time, through the structured heterogeneity, it can be found that the correlation coefficients between different injury severity levels are small. This indicates that, after the spatio-temporal parameters are considered, the unobserved heterogeneity of the model will be small, and the spatio-temporal effect and its interaction will be an essential part of structured heterogeneity. Thus, the ST-BRPNB-L model has the best goodness of fit among all models. According to the ST-BRPNB-L estimation results, the length of the road segment, AADT, whether the road segment has an entrance, an exit, is an ascending road segment or a descending road segment, proportion of class 3 vehicles, and proportion of class 5 vehicles are significantly positively correlated with the frequency of KSI crashes at the 95% level. The distance between tunnels, curvature, PCI, and SRI values are negatively correlated with the frequency of KSI crashes at the level of 95%. The frequency of SI crashes are not significantly correlated with the proportion of class 3 vehicles and PCI, significantly positively correlated with the road segment length, AADT, a tunnel segment with an entrance or an exit, is an ascending road segment or a descending road segment, and proportion of class 5 vehicles at the statistical level of 95%, significantly negatively correlated with curvature at the 95% level, and significantly negatively correlated with the distance between tunnels and SRI at the statistical level of 90%. It can be known from the IRR values that the distance between two tunnels, tunnel entrance/exit, and whether the ascending/descending road segment have a more substantial influence on the frequency of crashes of different injury severity levels. The results of this study allow a better understanding of the influence of tunnel design characteristics, traffic characteristics, and pavement conditions on the frequency of crashes of different severity in freeway tunnels and provide suggestions for safety measures of tunnel sections.

REFERENCES

[1] S. Bassan, ''Overview of traffic safety aspects and design in road tunnels,'' *IATSS Res.*, vol. 40, no. 1, pp. 35–46, Jul. 2016.

[2] (2019). Ministry of Transport of the People's Republic of China. *Statistical Bulletin of the Development of the Transportation Industry*. Analysis Bulletin. Government of the People's Republic of China. [Online]. Available: http://www.gov.cn/xinwen/ 2020-05/12/content_5510817.htm

[3] J. S. Yeung and Y. D. Wong, "Road traffic accidents in Singapore expressway tunnels," *Tunnelling Underground Space Technol.*, vol. 38, pp. 534–541, Sep. 2013.

[4] H. Sun, Q. Wang, P. Zhang, Y. Zhong, and X. Yue, "Spatialtemporal characteristics of tunnel traffic accidents in China from 2001 to present," *Adv. Civil Eng.*, vol. 2019, pp. 1–12, Aug. 2019.

[5] F. H. Amundsen and G. Ranes, "Studies on traffic accidents in Norwegian road tunnels," *Tunnelling Underground Space Technol.*, vol. 15, no. 1, pp. 3–11, Jan. 2000.

[6] J. Wang, A. Pervez, Z. Wang, C. Han, L. Hu, and H. Huang, "Crash analysis of Chinese freeway tunnel groups using a five-zone analytic approach," *Tunnelling Underground Space Technol.*, vol. 82, pp. 358–365, Dec. 2018.

[7] A. Pervez, H. Huang, C. Han, J. Wang, and Y. Li, "Revisiting freeway single tunnel crash characteristics analysis: A six-zone analytic approach," *Accident Anal. Prevention*, vol. 142, Jul. 2020, Art. no. 105542.

[8] A. Pervez, H. Huang, L. Jaeyoung, C. Y. Han, J. Wang, and X. Zhang, "Crash analysis of expressway long tunnels using a seven-zone analytic approach," *J. Transp. Saf. Secur.*, vol. 13, no. 1, pp. 108–122, 2021.

[9] C. Caliendo, M. L. De Guglielmo, and M. Guida, "A crash-prediction model for road tunnels," *Accident Anal. Prevention*, vol. 55, pp. 15–107, Jun. 2013.

[10] C. Caliendo, M. L. De Guglielmo, and M. Guida, "Comparison and analysis of road tunnel traffic accident frequencies and rates using random-parameter models," *J. Transp. Safety Secur.*, vol. 8, no. 2, pp. 177–195, 2016.

[11] C. Caliendo, M. L. De Guglielmo, and I. Russo, "Analysis of crash frequency in motorway tunnels based on a correlated random-parameters approach," *Tunnelling Underground Space Technol.*, vol. 85, pp. 243–251, Mar. 2019.

[12] Q. Hou, A. P. Tarko, and X. Meng, "Analyzing crash frequency in freeway tunnels: A correlated random parameters approach," *Accident Anal. Prevention*, vol. 111, pp. 94–100, Feb. 2018.

[13] C. Caliendo, M. Guida, F. Postiglione, and I. Russo, "A Bayesian bivariate hierarchical model with correlated parameters for the analysis of road crashes in Italian tunnels," *Stat. Methods Appl.*, vol. 31, no. 1, pp. 109–131, Mar. 2022.

[14] F. Tang, X. Fu, M. Cai, Y. Lu, and S. Zhong, "Investigation of the factors influencing the crash frequency in expressway tunnels: Considering excess zero observations and unobserved heterogeneity," *IEEE Access*, vol. 9, pp. 58549–58565, 2021.

[15] Q. Zeng, H. Wen, H. Huang, X. Pei, and S. C. Wong, "Incorporating temporal correlation into a multivariate random parameters tobit model for modeling crash rate by injury severity," *Transportmetrica A, Transp. Sci.*, vol. 14, no. 3, pp. 177–191, Mar. 2018.

[16] S. Jung and X. Qin, "A data-driven approach to strengthening policies to prevent freeway tunnel strikes by motor vehicles," *Accident Anal. Prevention*, vol. 157, Jul. 2021, Art. no. 106171.

[17] S. Chen, S. Zhang, Y. Xing, and J. Lu, "Identifying the factors contributing to the severity of truck-involved crashes in Shanghai river-crossing tunnel," *Int. J. Environ. Res. Public Health*, vol. 17, no. 9, p. 3155, May 2020.

[18] Z. Zhou, F. Meng, C. Song, N. N. Sze, Z. Guo, and N. Ouyang, "Investigating the uniqueness of crash injury severity in freeway tunnels: A comparative study in guizhou, China," *J. Saf. Res.*, vol. 77, pp. 105–113, Jun. 2021.

[19] A. Pervez, H. Huang, J. Lee, C. Han, Y. Li, and X. Zhai, "Factors affecting injury severity of crashes in freeway tunnel groups: A random parameter approach," *J. Transp. Eng., A, Syst.*, vol. 148, no. 4, Apr. 2022, Art. no. 04022006.

[20] R. Fu, Y. Guo, W. Yuan, H. Feng, and Y. Ma, "The correlation between gradients of descending roads and accident rates," *Saf. Sci.*, vol. 49, no. 3, pp. 416–423, Mar. 2011.

[21] P. Anastasopoulos, V. N. Shankar, J. E. Haddock, and F. L. Mannering, "A multivariate tobit analysis of highway accident-injury-severity rates," *Accident Anal. Prevention*, vol. 45, pp. 9–110, Mar. 2012.

[22] F. Tang, X. Fu, M. Cai, Y. Lu, S. Zhong, and C. Lu, "Applying a correlated random parameters negative binomial Lindley model to examine crash frequency along highway tunnels in China," *IEEE Access*, vol. 8, pp. 213473–213488, 2020.

[23] R. Yu and M. Abdel-Aty, "Multi-level Bayesian analyses for single- and multi-vehicle freeway crashes," *Accident Anal. Prevention*, vol. 58, pp. 97–105, Sep. 2013.

[24] P. Anastasopoulos and F. L. Mannering, "A note on modeling vehicle accident frequencies with rando-m-parameters count models," *Accident Anal. Prevention*, vol. 41, no. 1, pp. 9–153, Jan. 2009.

[25] M. Hosseinpour, A. S. Yahaya, and A. F. Sadullah, "Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: Case studies from Malaysian federal roads," *Accident Anal. Prevention*, vol. 62, pp. 22–209, Jan. 2014.

[26] Z. Ma, H. Zhang, S. I.-J. Chien, J. Wang, and C. Dong, "Predicting expressway crash frequency using a random effect negative binomial model: A case study in China," *Accident Anal. Prevention*, vol. 98, pp. 214–222, Jan. 2017.

[27] A. Montella and L. L. Imbriani, "Safety performance functions incorporating design consistency variables," *Accident Anal. Prevention*, vol. 74, pp. 44–133, Jan. 2015.

[28] B. Wali, A. J. Khattak, J. Waters, D. Chimba, and X. B. Li, "Development of safety performance functions: Incorporating unobserved heterogeneity and functional form analysis," *Accident Anal. Prevention*, vol. 2672, no. 30, pp. 9–20, Dec. 2018.

[29] L. Garach, J. de Ona, G. Lopez, and L. Baena, "Development of safety performance functions for Spanish two-lane rural highways on flat terrain," *Accident Anal. Prevention*, vol. 95, pp. 65–250, Oct. 2016.

[30] Y. Lao, G. Zhang, Y. Wang, and J. Milton, "Generalized nonlinear models for rear-end crash risk analysis," *Accident Anal. Prevention*, vol. 62, pp. 9–16, Jan. 2014.

[31] N. V. Malyshkina and F. L. Mannering, "Zero-state Markov switching count-data models: An empirical assessment," *Accident Anal. Prevention*, vol. 42, no. 1, pp. 30–122, Jan. 2010.

[32] J. M. P. Mayora and R. J. Pina, "An assessment of the skid resistance effect on traffic safety under wet-pavement conditions," *Accident Anal. Prevention*, vol. 41, no. 4, pp. 6–881, Jul. 2009.

[33] P. Buddhavarapu, A. Banerjee, and J. A. Prozzi, "Influence of pavement condition on horizontal curve safety," *Accident Anal. Prevention*, vol. 52, pp. 9–18, Mar. 2013.

[34] M. M. Ahmed, M. Abdel-Aty, and R. Yu, "Assessment of interaction of crash occurrence, mountainous freeway geometry, real-time weather, and traffic data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2280, no. 1, pp. 51–59, Jan. 2012.

[35] J. Park, M. Abdel-Aty, and J.-H. Wang, "Time series trends of the safety effects of pavement resurfacing," *Accident Anal. Prevention*, vol. 101, pp. 78–86, Apr. 2017.

[36] D. Lord, S. Washington, and J. N. Ivan, "Further notes on the application of zero-inflated models in highway safety," *Accident Anal. Prevention*, vol. 39, no. 1, pp. 7–53, Jan. 2007.

[37] J. Lee and F. Mannering, "Impact of roadside features on the frequency and severity of run-off-roadway accidents: An empirical analysis," *Accident Anal. Prevention*, vol. 34, no. 2, pp. 149–161, Mar. 2002.

[38] D. Lord, S. P. Washington, and J. N. Ivan, "Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory," *Accident Anal. Prevention*, vol. 37, no. 1, pp. 35–46, Jan. 2005.

[39] N. V. Malyshkina, F. L. Mannering, and A. P. Tarko, "Markov switching negative binomial models: An application to vehicle accident frequencies," *Accident Anal. Prevention*, vol. 41, no. 2, pp. 26–217, Mar. 2009.

[40] A. Behnood and F. L. Mannering, "An empirical assessment of the effects of economic recessions on pedestrian-injury crashes using mixed and latent-class models," *Analytic Methods Accident Res.*, vol. 12, pp. 1–17, Dec. 2016.

[41] S. R. Geedipally, D. Lord, and S. S. Dhavala, "The negative binomial-Lindley generalized linear model: Characteristics and application using crash data," *Accident Anal. Prevention*, vol. 45, pp. 65–258, Mar. 2012.

[42] P. Vangala, D. Lord, and S. R. Geedipally, "Exploring the application of the negative binomial–generalized exponential model for analyzing traffic crash data with excess zeros," *Analytic Methods Accident Res.*, vol. 7, pp. 29–36, Jul. 2015.

[43] F. L. Mannering, V. Shankar, and C. R. Bhat, "Unobserved heterogeneity and the statistical analysis of highway accident data," *Analytic Methods Accident Res.*, vol. 11, pp. 1–16, Sep. 2016.

[44] T. C. Coburn, "Statistical and econometric methods for transportation data analysis," *Technometrics*, vol. 46, no. 4, pp. 492–493, 2004.

[45] Q. Zeng, H. Wen, S. C. Wong, H. Huang, Q. Guo, and X. Pei, "Spatial joint analysis for zonal daytime and nighttime crash frequencies using a Bayesian bivariate conditional autoregressive model," *J. Transp. Saf. Secur.*, vol. 12, no. 4, pp. 566–585, Apr. 2020.
[46] Q. Hou, X. Huo, and J. Leng, "A correlated random parameters tobit model to analyze the safety effects and temporal instability of factors affecting crash rates," *Accident Anal. Prevention*, vol. 134, Jan. 2020, Art. no. 105326.
[47] 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 Anal. Prevention*, vol. 43, no. 4, pp. 9–1581, Jul. 2011.
[48] S. Mitra and S. Washington, "On the nature of over-dispersion in motor vehicle crash prediction models," *Accident Anal. Prevention*, vol. 39, no. 3, pp. 68–459, May 2007.
[49] S. Washington and M. M. Haque, "On the commonly accepted assumptions regarding observed motor vehicle crash counts at transport system locations," in *Proc. 92nd Annual Meeting Transp. Res. Board (TRB)*, Washington DC, USA, 2013, pp. 1–19.
[50] R. Rusli, M. M. Haque, A. P. Afghari, and M. King, "Applying a random parameters negative binomial Lindley model to examine multi-vehicle crashes along rural mountainous highways in Malaysia," *Accident Anal. Prevention*, vol. 119, pp. 80–90, Oct. 2018.
[51] S. Barua, K. El-Basyouny, and M. T. Islam, "A full Bayesian multivariate count data model of collision severity with spatial correlation," *Analytic Methods Accident Res.*, vols. 3–4, pp. 28–43, Oct. 2014.
[52] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives," *Transp. Res. A, Policy Pract.*, vol. 44, no. 5, pp. 291–305, 2010.
[53] H. Wen, X. Zhang, Q. Zeng, and N. N. Sze, "Bayesian spatial-temporal model for the main and interaction effects of roadway and weather characteristics on freeway crash incidence," *Accident Anal. Prevention*, vol. 132, Nov. 2019, Art. no. 105249.
[54] Y. Wang and K. M. Kockelman, "A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods," *Accident Anal. Prevention*, vol. 60, pp. 71–84, Nov. 2013.
[55] S. Barua, K. El-Basyouny, and M. T. Islam, "Multivariate random parameters collision count data models with spatial heterogeneity," *Anal. Methods Accident Res.*, vol. 9, pp. 1–15, Mar. 2016.
[56] Q. Zeng, Q. Guo, S. C. Wong, H. Wen, H. Huang, and X. Pei, "Jointly modeling area-level crash rates by severity: A Bayesian multivariate random-parameters spatio-temporal tobit regression," *Transportmetrica A, Transp. Sci.*, vol. 15, no. 2, pp. 1867–1884, Nov. 2019.

MINGMAO CAI was born in Fuzhou, China, in 1997. He received the B.S. degree in traffic engineering from the Qingdao University of Technology, Qingdao, China, in 2019. He is currently pursuing the M.S. degree with the School of Civil and Transportation Engineering, South China University of Technology, Guangzhou, China. His research interests include traffic safety analysis and intelligent transportation systems (ITSs).

FENG TANG was born in Huaihua, China, in 1992. He received the B.S. degree in civil engineering from the Changsha University of Science and Technology, Changsha, China, in 2015. He is currently pursuing the Ph.D. degree with the School of Civil and Transportation Engineering, South China University of Technology, Guangzhou, China. His research interests include traffic safety analysis and intelligent transportation systems (ITSs).

XINSHA FU was born in Changsha, China, in 1955. He received the B.S. degree in highway and bridge engineering from the Changsha University of Science and Technology, Changsha, in 1981, and the M.S. degree in highway engineering and the Ph.D. degree from Chang'an University, Xi'an, China, in 1987 and 2008, respectively. From 1981 to 2000, he was a Professor with the Changsha University of Science and Technology. Since 2000, he has been a Professor with the South China University of Technology, Guangzhou, China. His research interests include theoretical research on highway alignment design, research on highway alignment and computer-aided design, road safety audit, and intelligent transportation systems (ITSs). He served for the China Highway Society Youth Expert Committee and the China's Institutions of Higher Learning Civil Engineering Professional Guidance Committee, the Director of the Institute of Computer Application, an Academic Committee Member of Road of China Highway Society, China Faces the Transport in the 21st Century Version of the Institutions of Higher Learning Materials (Highway) Editorial Committee, the Computer-Aided Engineering Staff, and the Central South Highway Project Staff.