

## Research Article

# Predicting Crash Frequency for Urban Expressway considering Collision Types Using Real-Time Traffic Data

Hui Zhang <sup>1,2</sup>, Siyao Li,<sup>1,2</sup> Chaozhong Wu <sup>1,2</sup>, Qi Zhang <sup>1,2</sup> and Yafen Wang<sup>3</sup>

<sup>1</sup>Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan 430063, China

<sup>2</sup>Engineering Research Center of Transportation Safety, Wuhan 430063, China

<sup>3</sup>Road Traffic Safety Risk Prevention and Control Center, Wuhan Traffic Management Bureau, Wuhan Municipal Public Security Bureau, Wuhan, China

Correspondence should be addressed to Chaozhong Wu; [wucz@whut.edu.cn](mailto:wucz@whut.edu.cn)

Received 27 December 2019; Revised 28 January 2020; Accepted 4 February 2020; Published 20 March 2020

Guest Editor: Feng Chen

Copyright © 2020 Hui Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Current studies on traffic crash prediction mainly focus on the crash frequency and crash severity of freeways or arterials. However, collision type for urban expressway crash is rarely considered. Meanwhile, with the rapid development of urban expressway systems in China in recent years, traffic safety problems have attracted more attention. In addition, the traffic characteristics are considered to be a potentially important predictor of traffic accidents; however, their impact on crashes has been controversial. Therefore, a crash frequency predicting model for urban expressway considering collision types is proposed in this study. The loop detector traffic data and historical crash data were aggregated based on the similarities of the traffic conditions 5 minutes before crash occurrence, among which crashes were divided by collision type (rear-end collision and side-impact collision). The impact of traffic characteristics along with weather variables as well as their interactions on crash frequency was modelled by using negative binomial regression model. The results indicated that the influence of traffic and weather factors on two collision types shared similar trend, but different level. For rear-end collisions, crash frequency increased with lower average speed and high traffic volume under low speed limit. And when the speed limit is high, higher average speed coupled with larger volume increases the probability of crash. Higher average speed and traffic volume increase the probability of side-impact collisions, without being affected by the speed limit. The findings of the present study could help to determine efficient safety countermeasures aimed at improving the safety performance of urban expressway.

## 1. Introduction

With the rapid development of Chinese cities, residents' demand for travel and the increasing number of motor vehicles put forward higher requirements for the operation efficiency of cities. Urban expressway is a key part of the city roadway networks, which is of great significance to improve the travel efficiency. Compared with main arterial roads, urban expressway is characterized by large traffic volume and higher speed [1], resulting in frequent traffic crashes in recent years. In 2017, 6652 road crashes occurred causing 1673 deaths and 6862 injuries along the urban expressway [2]. Analysing the influence factors of urban expressway traffic crashes and establishing a crash

prediction model play important roles for improving traffic safety.

A large number of studies on crash prediction models were mainly carried out from the aspects of crash frequency estimation and the crash severity prediction at the macro level (e.g., yearly, monthly) [3–6]. The Highway Safety Manual (HSM) prediction model represents the most widely used approach for road safety assessment, developing the safety prediction procedures for rural highways, urban and suburban arterials in the 2010 version of the HSM [7]. Taking factors such as traffic flow, road geometry, and so on into consideration, the HSM provides a prediction method for estimating the expected average frequency of single- and multiple-vehicle fatal-and-injury crashes.

Establishing the traffic crash model based on the historical crash data to analyse the relationship between crash frequency and the relevant risk factors such as roadway geometry variables has been a focus for a long time. Based on crash data of freeway, Ma et al. analysed the influence of road length, traffic environment, and other risk factors on crash rate [8]. Xie et al. analysed the relationship among crash probability and intersection characteristics and traffic volume based on the data of signalized intersections in Shanghai. The results showed that the number of lanes and average speed at intersections would have a significant impact on the crash probability [9]. There are also some scholars who have studied the prediction methods based on dividing crashes into single- and multivehicle crashes [10], but there are few studies conducted on the collision type (rear-end collision, side-impact collision, etc.), particularly in China. There is strong empirical evidence that the prediction model of crash frequency based on collision type can help better understand the influence of the crash occurrence contributing factors on specific collision type, especially in real-time risk assessment [11, 12].

Due to the impact of real-time driving environment data such as traffic flow on traffic accidents, and with the technology progress of traffic data detection and storage, real-time crash risk assessment has become a research hotspot in the field of traffic safety. Chen et al. adopted a zero-inflated, negative binomial regression model to estimate hourly crash frequency using real-time environmental and traffic data [13]. In order to explore the complex interactions among characteristics, the mixed logit model was adopted for his further research [14], which showed that environment and traffic were critical to the likelihood of collisions. Choudhary et al. explored the relationships between traffic characteristics and the occurrence of crashes based on traffic data collected through inductive loop detectors. The results indicated that crash probability is related to higher speeds, greater volume, and high between-lanes speed variation [15]. Shi et al. explored the impact of real-time traffic flow on urban expressway crash probability [16]. The crash risk analysis models were developed for total crashes and time-related crashes to reveal the significant factors that affect crash risk [17]. In addition, typical scenarios leading to an accident were found from the modelling results. In the following research, Yu et al. investigated the impacts of data aggregation schemes on the relationships between operating speed and traffic safety; additionally, a U-shaped relationship between operating speed and crash occurrence was identified [18].

Traffic characteristics are widely adopted as important indicators of crash frequency, but the research results on its safety effect are not consistent. The inconsistencies among the results may be relevant to the different data sources, low quality of data, and different analytical methods. Speed is considered to be one of the most important factors leading to crashes. Speed management interventions such as Variable Speed Limits (VSL) or fundamental speed limit settings are introduced to improve road capacity and safety. The application of speed limit settings relies on the in-depth understanding of the quantitative relationships between

operating speeds and traffic safety to determine strategies to reduce the crash risk. Previous studies have shown positive associations of average speed with crashes [19], while single-vehicle crashes and fatal-and-injury crashes involving multiple vehicles increase with the increase of the average speed [20]. However, others have shown a negative or an insignificant relationship between average speed and crash risk [21, 22]. In addition, some scholars have suggested that the safety effect of average speed would vary with road types [23]. As for the impact of speed limit on crash risk, some studies have found that speed limit is positively correlated with crash frequency and severity [24], and high speed limit road sections are often associated with high crash risk. However, Gou et al. mentioned that the crash frequency may be reduced due to better road facilities with high speed limit sections [25]. What is more, the safety effect of speed seemed to be related to other traffic variables such as traffic flow. For example, Kononov et al. have shown that higher crash risk is associated with high-speed driving in high-density traffic flow [26].

In summary, most studies on traffic crash prediction mainly focus on the crash frequency and crash severity on freeway or intersections, while the study considering the collision type of urban expressway crash is rare. Moreover, the relationship between traffic characteristics and crash probability needs to be further discussed. Weather conditions have been found to be associated with accident risk, especially rainy weather [27]. This study explores the impact of traffic characteristics and weather variable on crash frequency. Data from Wuhan urban traffic management systems were utilized here and crashes are divided by collision type (rear-end collision, side-impact collision; collision type statistics reveal that rear-end collision and sideswipe collisions are the most common type of collision [28], and other collision types are excluded from the analysis because of the limited number). Data aggregated following a condition-based approach to reflect the conditions prior to crash occurrence are more accurate [18]. Finally, the relationships between relevant factors and traffic safety were conducted using negative binomial regression model.

The remaining of the paper is organized as follows. The data are collected and the negative binomial regression model is proposed for crash risk assessment. Sections 3 presented and discussed the results and the verification of the model. The paper ended with conclusions and limitations and looked forward for further study.

## 2. Data Collection and Preparation

To predict the crash frequency of urban expressway, historical crash, real-time traffic, and weather data have been utilized. Two urban expressways crossing over the river in Wuhan were chosen to be the data collection area, totalling 3,986 meters per direction. Inductive loop detectors and video monitoring were installed in the study area to provide real-time traffic flow data. Considering the design characteristics of urban expressways, the two-way lanes were considered to be independent of each other. Three datasets were used to build the database: (1) historical crash data

from January 1, 2018 to October 31, 2018. It was obtained from the traffic accident archives recorded by Wuhan Municipal Public Security Bureau in detail, including information concerning crash occurrence time, location, weather, and collision types. During the study, there were 536 crashes in total, in which 321 were rear-end collisions and 135 were side-impact collisions. Considering the high frequency of rear-end and side-impact collision, this study only discussed the impact of the crash occurrence contributing factors on these two kinds of collisions and finally obtained a total of 466 collision samples. Moreover, weather conditions were also extracted from this dataset, which were divided into two categories in this study to indicate whether it was rainy or not. When a crash occurs, its weather data is usually recorded accurately by professional traffic police based on the weather information collected by the nearest weather station; (2) roadway posted speed limit. The speed limit on the road, as the upper limit of real-time speed, has critical impact on crash risk. The speed limits of the urban expressway sections studied are 50 km/h and 70 km/h, respectively, and remained consistent during the study period; (3) traffic data detected by Loop Detectors (LDs) located along the study area. Lane-based average speed and total volume at 5-minute interval were provided by LDs data, which could provide analysis with high-quality traffic flow data. For each crash, Mile Maker (MM) or Chinese characters associated with the crash records were used to describe its location, and the traffic data can be exactly matched with crash data according to the occurrence location and time recorded.

In the raw data of the LDs, there are abnormal values caused by equipment problems and other random errors, which can be identified by setting appropriate value interval of parameters, so as to screen out the unrealistic value [29]. The invalid values of parameters include (1) speed < 0 km/h or speed > 100 km/h; (2) speed > 0 km/h and volume = 0; and (3) volume > 0 and speed = 0 km/h. Previous studies have shown that data aggregation has a significant impact on the results of the crash modelling. Abdel-Aty et al. compared the crash risk prediction effect of data aggregated at intervals of three minutes and five minutes and found that the latter provided more information for analysis and had better research significance [30]. The random noise can be effectively reduced by aggregating the data into five-minute intervals. Considering that the weather conditions generally do not change in a short period of time, the hourly collected weather conditions were matched with crashes to provide the weather information for model building [15]. In addition, traffic characteristics (average speed, traffic volume, etc.) 5 minutes just prior to the crash time had great contribution to their occurrence [29]. Therefore, the traffic data five minutes prior to the reported crash time and the weather conditions at the reported crash time were identified and defined as the precrash condition to construct the crash prediction model and finally obtain a total of 466 precrash conditions. For example, the crash occurred at 9:30; then traffic data from 9:20 to 9:25 (a 5-minute window) collected by the LDs closest to the crash occurrence location and the weather information extracted from the traffic

accident archives were extracted to form the precrash condition as shown in Figure 1.

### 3. Methods

*3.1. Variable Setting.* Based on the impact of data aggregation on crash modelling, a scenario-based data aggregation approach of significance has been proposed to identify the traffic conditions just prior to the crashes which might lead to crashes. For the scenario-based analysis, crashes were aggregated based on the similarity of the traffic conditions just before the crashes as employed by Yu et al. [18]. Precrash conditions, coupled with speed limit, were used as the control variable to define the potential crash scenarios, which were classified into equal frequency categories [15]. Firstly, the average speed was divided into 25 equal groups with a 4-percentile step; then each speed quantile was divided into 3 categories with a 33-percentile step for traffic volume. Further, the sequence was followed by splitting the speed limit into two equal quantiles for each quantile of traffic volume. Similarly, the weather (rain/no rain) conditions were divided into two equal-frequency groups for each volume category. Finally, a total of 300 scenarios (i.e.,  $25 \times 3 \times 2 \times 2$ ) were created, representing all the possible traffic conditions just prior to the crashes.

The 300 condition-based scenarios were matched with crashes that occurred under these traffic conditions and the crash frequency of each scenario was expressed by collision types (rear-end collisions and side-impact collisions). Therefore, an analysis dataset was formed by aggregating crashes into the same scenario. The traffic characteristics of each scenario were represented by the median values of average speed and traffic volume. Since it makes sense to study the relationship between traffic variables and crashes further, the potential interaction among traffic variables also needs to be taken into account. In this study, multiple interaction terms were utilized to study their impact on crash frequency in addition to the individual traffic variables. Table 1 presents the summary statistics of the scenario-based dataset.

### 4. Modelling

Statistical counting models have been widely utilized for crash frequency prediction. The Poisson regression model is a basic model to analyse the impact of potential factors on crash frequency and requires the mean and variance of crash frequency to be equal [31]. However, the dispersion and low probability of traffic accidents do not satisfy the assumption of Poisson's distribution. This phenomenon of variance bigger than the mean value is usually called overdispersion. The negative binomial regression model is an extension of the Poisson regression model to handle this problem by introducing an error term that obeys Gamma distribution on the basis of Poisson's regression model. This study explored the relationship between independent variables and rear-end collision and side-impact collision by establishing two negative binomial regression models. The expected crashes can be expressed as follows:

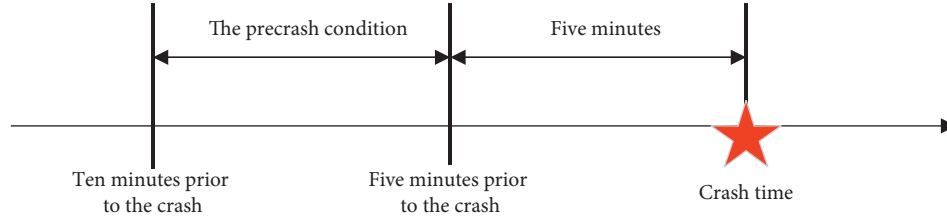


FIGURE 1: The precrash condition.

TABLE 1: Summary statistics of the scenario-based dataset.

Variable	Description	Mean	SD	Min	Max
Average speed (km/h)	Median speed for the crash occurrence scenario	37.68	8.04	20.98	80.77
Volume (vehicles in 5 min interval per lane, vehicles)	Median volume per lane for the crash occurrence scenario	126.97	44.19	12	209.875
Speed limit (km/h)	50 km/h or 70 km/h	60	60	50	70
Weather	=0 if rain; =1 otherwise	0.5	0.5	0	1
Average speed * volume (km/h * vehicles)	—	4644.11	1565.61	547.68	8143.15
Average speed * speed limit (km/h * km/h)	—	2178.36	641.01	1049	5653.9
Volume * speed limit (vehicles * km/h)	—	7148.54	2283.65	840	10756.67
Crash frequency	Rear-end collisions	1.07	2.09	0	10
	Side-impact collisions	0.45	1.41	0	6

$$\lambda_{ik} = \exp(\beta_{\mu k} X_{\mu k} + \varepsilon_{ik}), \quad (1)$$

where  $\lambda_{ik}$  refers to the expected crash frequency for  $k$  collision type of the scenario  $i$ .  $X_{\mu k}$  is a set of explanatory variables for  $k$  collision type, such as traffic volume, speed limit, etc.  $\beta_{\mu k}$  is the corresponding regression parameters to be estimated for  $k$  collision type.  $\varepsilon_i$  is the error term, which follows the Gamma distribution, with a mean value of 1 and a variance of  $\alpha^2$ . So, the variance of the crash frequency distribution is

$$\text{VAR}[y_{ik}] = E[y_{ik}][1 + \alpha E[y_{ik}]] = \lambda_{ik}(1 + \alpha \lambda_{ik}), \quad (2)$$

where  $y_{ik}$  represents the observed crash frequency for  $k$  crash type of the scenario  $i$ ,  $\text{VAR}[y_{ik}]$  represents its variance, and  $E[y_{ik}]$  is the expected crash frequency. Then, the probability density function of the negative binomial regression model is formulated as follows:

$$P(y_{ik}) = \frac{\Gamma(y_{ik} + (1/\alpha))}{\Gamma(y_{ik} + 1)\Gamma(1/\alpha)} \left( \frac{\alpha \lambda_{ik}}{1 + \alpha \lambda_{ik}} \right)^{y_{ik}} \left( \frac{1}{1 + \alpha \lambda_{ik}} \right)^{1/\alpha}. \quad (3)$$

Among them,  $\Gamma(\cdot)$  is a Gamma function. When  $\alpha = 0$ , the negative binomial regression model is the same with Poisson's regression model; when  $\alpha > 1$ , the data deviation is large; when  $\alpha < 1$ , the data deviation is small. Since the large deviation of traffic crash data, the negative binomial model is widely used for traffic crash prediction.

## 5. Performance Evaluation of Prediction

Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to assess the overall goodness-of-fit. Note that lower values of these measures signify better statistical fit. To further evaluate the models, two forecasting accuracy measures were adopted: Mean Absolute Deviation (MAD) and Mean Squared Error (MSE) [32].

MAD describes the magnitude of average bias in model prediction:

$$\text{MAD} = \frac{1}{n} \sum_{i=1}^n |y_{ik} - \lambda_{ik}|, \quad (4)$$

where  $y_{ik}$  is the observed crash frequency for  $k$  crash type of the scenario  $i$  and  $\lambda_{ik}$  is the predicted one.

On the other hand, MSE refers to the mean value of the square misprediction of the estimated models. MSE is computed as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_{ik} - \lambda_{ik})^2. \quad (5)$$

MSE and MAD can be used to describe the accuracy of the model fitting. In general, the lower the value indicated, the better the prediction of observed data. However, the range of their values is not limited, and the validity of the model is usually tested by artificially defining a reasonable range. Therefore,  $R^2$  is further introduced to describe the accuracy of the model, and its value ranges from 0 to 1. The larger  $R^2$ , the better the model fitting effect, and when  $R^2$  is greater than 0.4, the model is considered to have a good fit.

$$R^2 = \left( \frac{\sum_{i=1}^n \lambda_{ik} - \overline{y_{ik}}}{\sum_{i=1}^n (y_{ik} - \overline{y_{ik}})^2} \right)^2, \quad (6)$$

where  $\overline{y_{ik}}$  is the observed average crash frequency for  $k$  crash type of the scenario  $i$ .

## 6. Results and Discussion

Given the 0.05 significance level, the models were fitted by the Maximum Likelihood Estimation (MLE) method with the help of STATA 15.0 software. All the traffic variables along with their multiplicative interaction combinations and



rain were taken as explanatory variables in both multivariate models. The best fitted variable combination for crash frequency prediction included all traffic and weather variables plus the interaction between average speed and speed limit. The performance of prediction was evaluated and Table 2 presents the values of these measures.

The result shows that the negative binomial model has good prediction performance for the rear-end collision and the side-impact collision. The significant variables in the two models include average speed, traffic volume, weather, speed limit, and the interaction between average speed and speed limit. In Table 2, a positive (negative) sign for a variable in the crash count component indicates that an increase in the variable is likely to result in more (less) vehicle crashes. Some variables have similar influences on the two types of crashes. For instance, the parameter for traffic volume reveals a positive association with crash proportion in both models, indicating that the frequency of rear-end collisions and side-impact collisions will increase with the increase of traffic volume [7, 15, 18]. The estimated result of weather variable implies that the presence of rain has a negative effect on crash rates, which is consistent with previous studies [33–35]. The high crash risk in rainy weather may have a certain relationship with poor road conditions and reduced visibility, which leads to longer stopping distance and longer reaction time. However, the parameter estimate value of the same significance variable is different in the two models, indicating that the same variable has different effects on the two types of crash frequency. For example, in the rear-end collision prediction model, the weather coefficient is  $-1.4399$ , while being  $-0.9787$  in the side-impact collision prediction model, indicating that compared with the side-impact collision, rainfall has a greater impact on the rear-end collisions. The probability of rear-end collisions in a rainy environment is relatively high. It can be explained that when driving in rainy days, the stopping distance becomes longer due to the wetness of pavement, which are more likely to lead to rear-end collision.

Further analysis of the impact of road average speed and speed limit on the crash frequency was conducted. Due to the existence in interaction terms, although the estimation results of these two parameters were negative, we cannot draw the conclusion that they are negatively related to the crash frequency. In order to observe the contribution of the interaction between average speed and speed limit to the crash frequency, the crash rates are plotted in Figures 2(a) and 2(b) for the rear-end collision and the side-impact collision, respectively. The speed limit design of urban expressway generally ranges from 60 km/h to 100 km/h in China. In addition, considering that there was a research object with speed limit of 50 km/h in this study, the range of speed limit range was determined to be 50 km/h–100 km/h, and the corresponding average speed range is 0 km/h–100 km/h. The results show that the influence of their interaction on rear-end collision and side-impact collision is different. For the rear-end collision, when the speed limit is low, the crash frequency will decrease with the increase of speed limit coupled with a low average speed. However, when the speed limit

TABLE 2: Estimation results of negative binomial model.

Variables	Rear-end collisions		Side-impact collisions	
	Mean	P value	Mean	P value
Intercept	11.4970	0.0000	13.1129	0.0015
Average speed	-0.1561	0.0080	-0.1236	0.0195
Volume	0.0027	0.0121	0.0044	0.0431
Rain	-1.4399	0.0000	-0.9787	0.0006
Speed limit	-0.1917	0.0000	-0.2565	0.0007
Average speed * speed limit	0.0028	0.0067	0.0026	0.0134
AIC	490.61		392.37	
BIC	510.58		412.66	
MAD	1.42		1.23	
MSE	2.63		3.45	
R <sup>2</sup>	0.75		0.63	

continues to increase, the crash frequency is positively related to the average speed. But for side-impact collision, no matter what the speed limit is, the average speed limit performs positive, indicating that the crash frequency increases with the rise of average speed.

The relationship between average speed, traffic volume, and crashes is further analysed, and the surface relationship diagrams of average speed and traffic volume with rear-end collision and side-impact collision under different speed limit conditions were plotted, respectively, as shown in Figure 3. For rear-end collision, the curves show that high crash risk associates with low average speed when the speed limit is low. When the speed limit is low, vehicles tend to drive at a lower speed, which to some extent promotes the occurrence of traffic congestion. Many studies, including those of Golob et al. and Christoforou et al., have shown that traffic congestion is one of the most significant precursors of rear-end collisions [11, 13, 36]. In the case of traffic congestion, drivers must adjust the speed in short time and short distance, making it more likely to lead to a rear-end collision, which is consistent with the results of this paper. When the speed limit is high, crash seems to be triggered at higher average speed and volumes. The vehicle tends to drive at a high speed under high speed limit. The vehicle inertia is too large to brake within a safe distance, leading to higher probability of a rear-end collision crash. In addition, it is clear that rear-end collision potential increase when the traffic volume is higher. This may be due to more interaction between vehicles at higher flow conditions leading to an increased tendency for rear-end collisions. Similar findings are also found in other studies [37].

The results in Figure 4 show that for the side-impact collisions, when traffic volume increase, the frequency of side-impact collision increases, which may be related to more frequent lane change behaviour of vehicles in high-flow conditions. This finding is similar to the work of Christoforou et al. who reported that side-impact collisions are more likely to be positively correlated with traffic volume [11]. Further, side-impact collisions are more probable to occur for high levels of traffic average speed [38]. One possible explanation could be that side-impact collisions are usually associated with lane change operations, which are more prone to crash at high speeds. The relationship

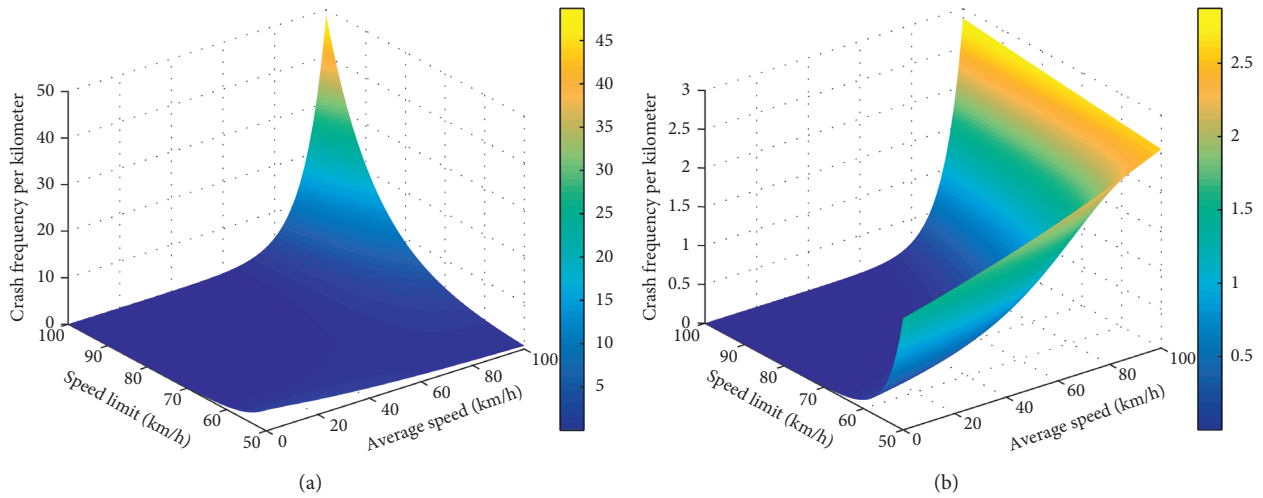


FIGURE 2: The impact of the interaction between average speed and speed limit on the crash frequency. (a) For the rear-end collision. (b) For the side-impact collision.

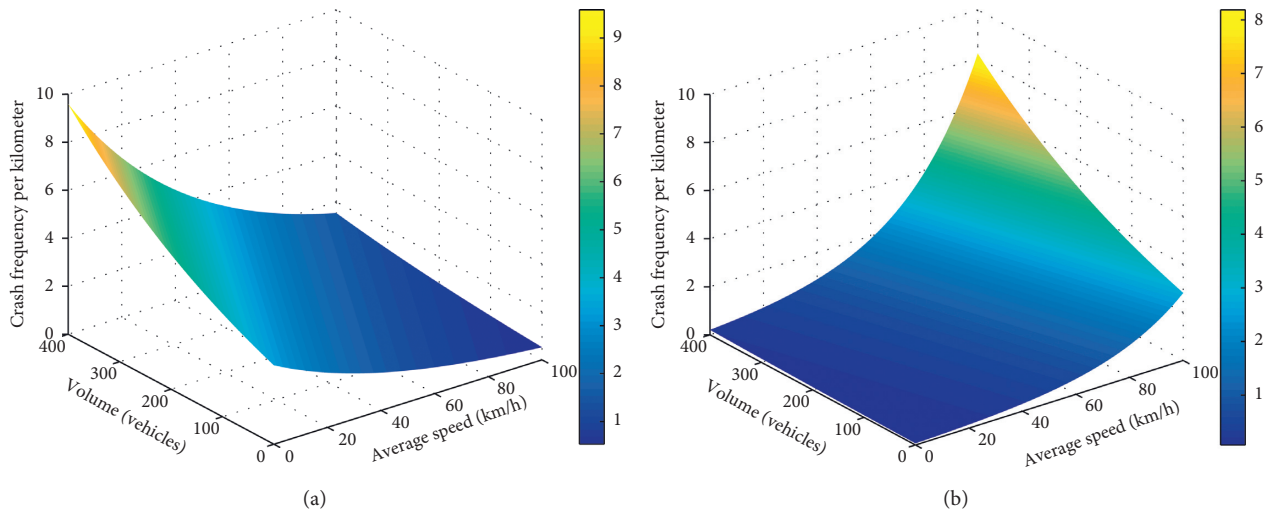


FIGURE 3: Relationships between speed, volume, and rear-end collision probability. (a) Speed limit = 50 km/h. (b) Speed limit = 70 km/h.

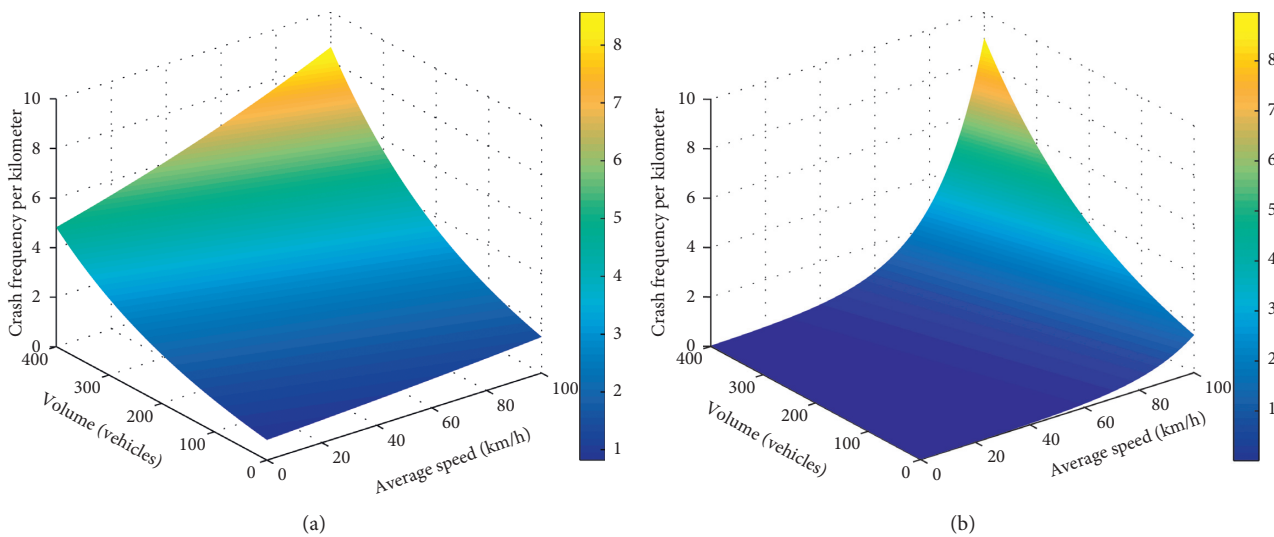


FIGURE 4: Relationships between speed, volume, and side-impact collision probability. (a) Speed limit = 50 km/h. (b) Speed limit = 70 km/h.

between these variables and the frequency of different crash types can be applied on safety improvement and management.

## 7. Conclusions

Traffic crash prediction is a hot topic in the field of traffic safety and keeps developing. In view of the shortage of urban expressway crash prediction and the controversial issue of the influence of speed on safety, the traffic and weather variables were used to predict the crash risk. This study developed crash risk analysis models for rear-end collision and side-impact collision with loop detector traffic data and historical crash data. For the purpose of exploring traffic conditions that leads to accidents, the data of Wuhan urban expressway were aggregated based on the similarity of traffic conditions in 5 minutes prior to crash occurrence. The impact of traffic variables along with their interactions and the weather condition on crash frequency was conducted by the negative binomial regression model. The results show that average speed, traffic volume, speed limit, and weather have significant impact on crash frequency and the influence of factors on two collision types is similar but with different degree. Specifically, the existence of rain increases the crash risk, especially for rear-end collisions. Higher traffic volume increases the crash risk. The impact of average speed on rear-end collision and side-impact collision on urban expressway is different. Specifically, under the lower speed limit, the probability of rear-end collision increases with the decrease of road average speed. At the higher speed limit, the probability of rear-end collision increases with the rise of road average speed. Regardless of the speed limit, with the increase of road average speed and traffic flow, the frequency of side-impact collision will increase.

In general, the specific combinations of traffic characteristics increase the probability of crash occurrences. These results are helpful to understand the crash risks under different traffic conditions and provide a basis for formulating traffic management countermeasures effectively. The results can be applied to real-time traffic management, where drivers can be warned via variable message sign once it is defined to be vulnerable to existing traffic conditions. In addition, the crash prediction model can also be applied to the evaluation and planning of road improvement projects when it is used to monitor the road safety level in real time. However, there are some limitations in this study. For example, due to the limited number of single-vehicle crashes, this study did not conduct the prediction model for it. Moreover, in order to better understand the impact of traffic characteristics on traffic crashes, further study of collision types is needed. At last, due to the existence of traffic heterogeneity, further crash risk analysis should consider the heterogeneous influence of various factors on traffic safety.

## Data Availability

The crash data and traffic data are available upon request for research purpose.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Authors' Contributions

Chaozhong Wu and Hui Zhang conceived and designed the study. Siyao Li, Hui Zhang, and Yafen Wang collected the data. Siyao Li, Hui Zhang, and Qi Zhang analyzed and interpreted the results. Siyao Li and Hui Zhang prepared the manuscript draft. All authors reviewed the results and approved the final version of the manuscript.

## Acknowledgments

This research was supported by the National Natural Science Foundation of China (61603282, u1624262) and the National Key Research and Development Program of China (2017YFC0804802). All authors also want to thank the staff from Wuhan Municipal Public Security Bureau for their kindly assistance in the data collection.

## References

- [1] C. Xu, X. Wang, H. Yang, K. Xie, and X. Chen, "Exploring the impacts of speed variances on safety performance of urban elevated expressways using GPS data," *Accident Analysis & Prevention*, vol. 123, pp. 29–38, 2019.
- [2] Transportation Bureau, Ministry of Public Security, *The Road Traffic Accidents Statistics Report in China*, Transportation Bureau, Ministry of Public Security, Beijing, China, 2017.
- [3] E. Sacchi and T. Sayed, "Bayesian estimation of conflict-based safety performance functions," *Journal of Transportation Safety & Security*, vol. 8, no. 3, pp. 266–279, 2016.
- [4] T. Ahmed and S. Tarek, "Models to evaluate the severity of pedestrian-vehicle conflicts in five cities," *Transportmetrica A: Transport Science*, vol. 15, no. 2, pp. 354–375, 2019.
- [5] H. L. Huang, P. P. Xu, M. Ma, and M. Abdel-aty, "Recent theoretical researches on transportation safety planning," *China Journal of Highway and Transport*, vol. 27, no. 09, pp. 90–118, 2014.
- [6] Q. Zeng and H. Huang, "A stable and optimized neural network model for crash injury severity prediction," *Accident Analysis & Prevention*, vol. 73, pp. 351–358, 2014.
- [7] AASHTO, *Highway Safety Manual*, American Association of State Highway and Transportation Officials, Washington, DC, USA, 2010.
- [8] C. Ma, S. R. Zhang, Z. L. Ma, and J. Wang, "Nonlinear negative binomial regression model of expressway traffic accident frequency prediction," *China Journal of Highway and Transport*, vol. 31, no. 11, pp. 176–185, 2018.
- [9] K. Xie, X. Wang, H. Huang, and X. Chen, "Corridor-level signalized intersection safety analysis in Shanghai, China using Bayesian hierarchical models," *Accident Analysis & Prevention*, vol. 50, pp. 25–33, 2013.
- [10] B. W. Dong, X. X. Ma, F. Chen, and S. R. Chen, "Investigating the differences of single-vehicle and multivehicle accident probability using mixed logit model," *Journal of Advanced Transportation*, vol. 2018, Article ID 2702360, 9 pages, 2018.
- [11] Z. Christoforou, S. Cohen, and M. G. Karlaftis, "Identifying crash type propensity using real-time traffic data on freeways," *Journal of Safety Research*, vol. 42, no. 1, pp. 43–50, 2011.

- [12] T. Bhowmik, S. Yasmin, and N. Eluru, "Do we need multivariate modeling approaches to model crash frequency by crash types? A panel mixed approach to modeling crash frequency by crash types," *Analytic Methods in Accident Research*, vol. 24, p. 100107, 2019.
- [13] F. Chen, S. Chen, and X. Ma, "Crash frequency modeling using real-time environmental and traffic data and unbalanced panel data models," *International Journal of Environmental Research and Public Health*, vol. 13, no. 6, p. 609, 2016.
- [14] F. Chen, S. Chen, and X. Ma, "Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data," *Journal of Safety Research*, vol. 65, pp. 153–159, 2018.
- [15] P. Choudhary, M. Imprialou, N. R. Velaga, and A. Choudhary, "Impacts of speed variations on freeway crashes by severity and vehicle type," *Accident Analysis & Prevention*, vol. 121, pp. 213–222, 2018.
- [16] Q. Shi, M. Abdel-aty, and R. Yu, "Multi-level Bayesian safety analysis with unprocessed automatic vehicle identification data for an urban expressway," *Accident Analysis & Prevention*, vol. 88, pp. 68–76, 2016.
- [17] R. Yu, X. Wang, K. Yang, and M. Abdel-aty, "Crash risk analysis for shanghai urban expressways: a Bayesian semi-parametric modeling approach," *Accident Analysis & Prevention*, vol. 95, pp. 495–502, 2016.
- [18] R. Yu, M. Quddus, X. Wang, and K. Yang, "Impact of data aggregation approaches on the relationships between operating speed and traffic safety," *Accident Analysis & Prevention*, vol. 120, pp. 304–310, 2018.
- [19] L. Aarts and I. van Schagen, "Driving speed and the risk of road crashes: a review," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 215–224, 2006.
- [20] M.-I. M. Imprialou, M. Quddus, and D. E. Pitfield, "Predicting the safety impact of a speed limit increase using condition-based multivariate Poisson lognormal regression," *Transportation Planning and Technology*, vol. 39, no. 1, pp. 3–23, 2016.
- [21] M. Quddus, "Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and GIS," *Journal of Transportation Safety & Security*, vol. 5, no. 1, pp. 27–45, 2013.
- [22] X. Wang, R. Yu, and C. Zhong, "A field investigation of red-light-running in Shanghai, China," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 37, pp. 144–153, 2016.
- [23] M. H. Cameron and R. Elvik, "Nilsson's Power Model connecting speed and road trauma: applicability by road type and alternative models for urban roads," *Accident Analysis & Prevention*, vol. 42, no. 6, pp. 1908–1915, 2010.
- [24] A. Vadeby and Å. Forsman, "Traffic safety effects of new speed limits in Sweden," *Accident Analysis & Prevention*, vol. 114, pp. 34–39, 2018.
- [25] Y. Guo, Z. Li, P. Liu, and Y. Wu, "Modeling correlation and heterogeneity in crash rates by collision types using full Bayesian random parameters multivariate Tobit model," *Accident Analysis & Prevention*, vol. 128, pp. 164–174, 2019.
- [26] J. Kononov, C. Durso, D. Reeves, and B. K. Allery, "Relationship between traffic density, speed, and safety and its implications for setting variable speed limits on freeways," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2280, no. 1, pp. 1–9, 2012.
- [27] F. Chen and S. Chen, "Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1677–1688, 2011.
- [28] C. Thanapong, J. Sajjakaj, C. Vuttichai, K. Ampol, and R. Vatanavongs, "Analysis of rear-end crash on Thai highway: decision tree approach," *Journal of Advanced Transportation*, vol. 2019, Article ID 2568978, 13 pages, 2019.
- [29] Y. Kui, X. S. Wang, and R. J. Yu, "A Bayesian dynamic updating approach for urban expressway real-time crash risk evaluation," *Transportation Research Part C: Emerging Technologies*, vol. 96, pp. 192–207, 2018.
- [30] M. Abdel-aty, A. Pande, A. Das, and W. J. Knibbe, "Assessing safety on Dutch freeways with data from infrastructure-based intelligent transportation systems," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2083, no. 1, pp. 153–161, 2008.
- [31] Q. Hou, A. P. Tarko, and X. Meng, "Investigating factors of crash frequency with random effects and random parameters models: new insights from Chinese freeway study," *Accident Analysis & Prevention*, vol. 120, pp. 1–12, 2018.
- [32] T. Bhowmik, S. Yasmin, and N. Eluru, "A joint econometric approach for modeling crash counts by collision type," *Analytic Methods in Accident Research*, vol. 19, pp. 16–32, 2018.
- [33] C. Xu, W. Wang, and P. Liu, "Identifying crash-prone traffic conditions under different weather on freeways," *Journal of Safety Research*, vol. 46, pp. 135–144, 2013.
- [34] F. Xing, H. Huang, Z. Zhan et al., "Hourly associations between weather factors and traffic crashes: non-linear and lag effects," *Analytic Methods in Accident Research*, vol. 24, p. 100109, 2019.
- [35] S. Jung, X. Qin, and D. A. Noyce, "Rainfall effect on single-vehicle crash severities using polychotomous response models," *Accident Analysis & Prevention*, vol. 42, no. 1, pp. 213–224, 2010.
- [36] T. F. Golob, W. W. Recker, and V. M. Alvarez, "Freeway safety as a function of traffic flow," *Accident Analysis & Prevention*, vol. 36, no. 6, pp. 933–946, 2004.
- [37] L. Dimitriou, K. Stylianou, and M. A. Abdel-Aty, "Assessing rear-end crash potential in urban locations based on vehicle-by-vehicle interactions, geometric characteristics and operational conditions," *Accident Analysis & Prevention*, vol. 118, pp. 221–235, 2018.
- [38] W. Wang, R. Chen, R. Y. Chen, and P. Liu, "Conditional inference tree-based analysis of hazardous traffic conditions for rear-end and sideswipe collisions with implications for control strategies on freeways," *IET Intelligent Transport Systems*, vol. 8, no. 6, pp. 509–518, 2014.