Accident Analysis and Prevention 132 (2019) 105268

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

# Freeway single and multi-vehicle crash safety analysis: Influencing factors and hotspots

Xuesong Wang[a,b,*], Mingjie Feng[a,b]

[a] The Key Laboratory of Road and Traffic Engineering, Ministry of Education, China
[b] School of Transportation Engineering, Tongji University, Shanghai, 201804, China

ARTICLE INFO

*Keywords:*
Freeway safety
Hotspot identification
Single-vehicle crashes
Multi-vehicle crashes
Bivariate negative binomial spatial CAR model

ABSTRACT

Single-vehicle (SV) and multi-vehicle (MV) crashes have been recognized as differing in spatial distribution and influencing factors, but little consideration has been given to these differences as related to hotspot identification. For the purpose of better hotspot identification, this study aims to analyze influencing factors of SV and MV crashes and to explore the consistency between SV and MV hotspots. Crash data, roadway geometric design features, and traffic characteristics were collected along the two directions of a 45-km freeway section in Shanghai, China. Univariate negative binomial conditional autoregressive (NB-CAR) and bivariate negative binomial spatial conditional autoregressive (BNB-CAR) models were developed to analyze the influencing factors and specifically address (1) site correlation between SV and MV crashes within the same freeway segment, and (2) spatial correlation among different freeway segments within the same direction. The modeling results showed substantial differences in the significant factors that influence SV and MV crashes, including both roadway geometric features and traffic operational factors. A non-negligible site correlation was found between SV and MV crashes. Taking into account the site correlation, the BNB-CAR model outperformed the NB-CAR model in terms of parameter estimation and model fitting. For hotspot identification, potential for safety improvement based on the empirical Bayes method was adopted to handle the crash fluctuation problem. Substantial inconsistency was found between SV and MV hotspots despite the site correlation: in the top ten hotspots, no hotspot was shared by the two crash types. This result highlights the importance of differentiating SV and MV crashes when identifying hotspots, providing insight into freeway safety analysis.

## 1. Introduction

Single-vehicle (SV) and multi-vehicle (MV) crashes have been associated with different vehicle maneuvers and causality, or influencing, factors (Ivan et al., 1999; Martensen and Dupont, 2013; Ma et al., 2016). The Annual Statistical Report on Roadway Traffic Accidents (2016) shows that in China, 93.86% of SV crashes on freeways involve hitting fixed objects or rollover crashes, and 93.81% of MV crashes are rear-end or sideswipe crashes. As suggested by a previous study, using different control strategies according to SV and MV hazard factors may be more effective for reducing their crash risk (Yu and Abdel-Aty, 2013). For instance, creating more forgiving roadsides may be more appropriate for hotspots of SV crashes, while attention to traffic flow control strategies seem more suitable for MV hotspot improvement. Given limited government resources, implementation of these different countermeasures depends on separate accurate identification of SV and MV hotspots, which therefore deserves further consideration. Although it has been noted that the two crash types' spatial distributions differ (Kilamanu et al., 2011; Schneider et al., 2017), existing studies have concentrated on analyzing distinct influencing factors for SV and MV crashes (Ivan et al., 1999; Abdel-Aty et al., 2006; Dong et al., 2018), but have not taken the next step of using the knowledge of the differing factors as rationale for identifying separate SV and MV hotspots.

SV crashes have been found to be less likely to occur than MV crashes. In 2015, SV crashes only accounted for 21.3% of the total crashes in New Mexico (NMDOT, 2017). In a study by Feng et al. (2019) that investigated the transferability of freeway safety performance functions between China and the United States, the proportion of SV crashes on the studied U.S. freeways was only 15%, but 37% on Chinese. The high SV crash proportion in China, as well as evidence for influencing factors (Yu and Abdel-Aty, 2013) and spatial distributions (Kilamanu et al., 2011) that differ from MV crashes, motivates the present study of SV and MV crashes.

Driver loss of control, a usually suspected distinction between SV

* Corresponding author.
*E-mail address:* wangxs@tongji.edu.cn (X. Wang).

https://doi.org/10.1016/j.aap.2019.105268
Received 9 May 2018; Received in revised form 5 August 2019; Accepted 11 August 2019
Available online 26 August 2019
0001-4575/ 

and MV crashes, is not addressed in this study. A substantial proportion of SV crashes have been found to be caused by driver loss of control, and in only a somewhat random portion of these crashes is there engagement with other vehicles that happen to be in the wrong place at the wrong time, creating an MV crash; likewise, an SV crash might result from interactions among several vehicles (Martensen and Dupont, 2013). Thus, differentiating crashes according to loss of control or no loss of control, rather than by SV or MV, would seem a reasonable way to proceed. However, over 60% of freeway crashes in the Chinese police crash database have no clear records on their causation, making loss-of-control distinctions impossible. The in-depth crash investigation needed to gather sufficient causation detail requires substantial commitment of resources, funds, and professional and interdisciplinary teams, when it can even be done after the fact. Due to high costs, only small samples of road crashes can be this thoroughly investigated, samples too small to be statistical representative (Usami et al., 2017), and, in fact, the examples of SV-MV overlap mentioned above have been determined to be the exception rather than the rule (Lee and Abdel-Aty, 2008; Martensen and Dupont, 2013). In contrast to the problems involved in loss-of-control investigation, the number of vehicles involved in a crash is always recorded in the daily work of traffic police, making the SV-MV distinction readily observable.

The crash prediction model is one of the most widely used methods for analyzing the relationships between traffic crashes and their influencing factors. Separate crash prediction models have been suggested for SV and MV crashes (Ivan et al., 1999; Lord et al., 2005a; Geedipally and Lord, 2010b; Ma et al., 2016), but a significant correlation between SV and MV crashes was identified (Geedipally and Lord, 2010b), so, as recommended, will be considered in modeling. Site correlation, as it is called in this study, refers to the likelihood that factors specific to a certain road segment will impact both SV and MV crashes. A second type of correlation, known as spatial correlation, refers to the effects of adjacent road segments (Aguero-Valverde and Jovanis, 2008). Adjacent road segments have similarities in geometric design and traffic operational characteristics, which cause correlation over space. Spatial correlation among various freeway segments has been confirmed, and addressed in several papers by use of a spatial conditional autoregressive (CAR) model (Ma et al., 2017a, b; Wen et al., 2018). In this study, crashes on freeway segments of the same direction have been confirmed to be spatially correlated. Both site correlation and spatial correlation may adversely affect the parameter estimate precision, and so will be accounted for (Lord and Mannering, 2010).

For hotspot identification, specific crash related characteristics have been modeled to obtain more accurate results. For example, hotspots were found to differ in crash patterns (Wang et al., 2014; Kim et al., 2006) and severities (Dezman et al., 2016). Various methods have been employed, including a simple evaluation of crash frequency or rate, a proportion method, and a model-based method. Among these, the best performance has come from the model-based method (Huang et al., 2009; Montella, 2010; Cafiso and Di Silvestro, 2011), in which the estimates obtained according to crash prediction models are then used to identify hotspots. Based on the crash predictions, the empirical Bayes (EB) method and potential for safety improvement (PSI) have frequently been used to handle the random fluctuation of crashes.

The overall aim of this study is to identify the distinct influencing factors for SV and MV crashes, and to use that information to determine if hotspots can be better pinpointed by considering SV and MV crashes separately. To achieve that objective, this study: (1) developed crash prediction models to identify the influencing factors while accounting for site and spatial correlation, and (2) used the models to identify SV and MV hotspots, and determine their degree of overlap by investigating their spatial distribution consistency.

The balance of this study is divided into six sections. First, previous studies related to crash modeling and hotspot identification for SV and MV crashes are discussed. The next section provides a brief explanation of data collection procedures and statistical descriptions of the data used in this study. Then, the modeling methodology and hotspot identification methods are introduced. The fourth section shows the results of parameter estimation and hotspot identification, and finally, discussion and conclusions from this study are provided.

## 2. Literature review

### *2.1. Contributing factors for SV and MV crashes*

Previous studies have found that potential contributing factors differ between SV and MV crashes (Ivan et al., 1999; Kockelman and Kweon, 2002; Savolainen and Mannering, 2007; Wang et al., 2017). In 1999, Ivan et al. found that the right (passenger side) roadway shoulder width had opposite effects on the two crash groups; and that significant contributors to SV crashes, including sight distance, level of service, and speed limit, were not found to be related to MV crashes, whereas road type and number of signals showed significant impact only on MV crashes. Similarly, Yu and Abdel-Aty (2013) found that significant factors influencing MV crashes, such as degree of curvature and curve length ratio, were not significant for SV crashes, while speed limit and longitudinal grade only have impact on SV crashes. Ma et al. (2016) determined, using a random-effects bivariate Poisson lognormal model, that inside shoulder length, temperature, and chemically wet-surface percentage only influenced SV crash count, while number of lanes related only to MV crashes. Researchers have not yet reached consistent conclusions regarding the impact of traffic volume (Persaud and Mucsi, 1995; Ivan et al., 1999; Lord et al., 2005a; Pande and Abdel-Aty, 2009). Studies conducted by Persaud and Mucsi (1995) and Geedipally and Lord (2010a, b) showed that SV and MV crashes were both positively influenced by traffic volume, whereas Lord et al. (2005a) had found earlier that traffic volume significantly contributed to SV and MV crashes, but with opposite effects; that is, traffic volume was positively related to MV crash count, but negatively to SV crash count. In another contrast to the above studies, Yu and Abdel-Aty (2013) and Ma et al. (2016) found that traffic volume had a significant impact on MV crashes only.

### *2.2. Crash prediction models*

#### *2.2.1. Correlation between SV and MV crashes*

When Öström and Eriksson (1993) investigated the impact of drivers' blood alcohol content on crashes, they were among the first to use separate SV and MV models. Since then, many researchers have developed separate predictive models for SV and MV crashes to analyze crash influencing factors or to predict crash frequency (Mensah and Hauer, 1998; Ivan et al., 1999; Lord et al., 2005a; Geedipally and Lord, 2010a, b). Mensah and Hauer (1998) developed separate prediction models for SV and MV crashes, plus one aggregated model for total crashes, and found that the aggregated model predicted fewer crashes than combining the results of the two separate models. Likewise, Lord et al. (2005a) advocated developing separate predictive models for SV and MV crashes, noting that the single aggregate model for all crashes was not adequate for prediction. Geedipally and Lord (2010b) confirmed that crash predictions based on separate models were much closer to the true crash count values, but at the same time developed a bivariate negative binomial model that found SV and MV crash counts to be significantly correlated. In 2013, Yu and Abdel-Aty (2013) corroborated the correlation with a bivariate Poisson-lognormal model that analyzed SV and MV crashes simultaneously. Based on the data collected from a mountainous freeway, they found that the bivariate model performed better than the hierarchical Poisson-lognormal model, which did not consider the correlation. More recently, Ma et al. (2016) have extended the bivariate Poisson-lognormal model with independent and correlated random effects in order to address the spatial correlation between SV and MV crashes. According to these studies, it can be concluded that it is best to model SV and MV crashes separately, as well as to use bivariate models to consider the correlation between the two crash types.

### 2.2.2. Spatial correlation

Road segments along a given freeway have similarities in geometric design, traffic operational characteristics, and other unobserved factors that can cause correlation over space. This spatial correlation can result in losses in parameter estimation efficiency for crash prediction models (Lord and Mannering, 2010). To handle spatial correlation, the conditional autoregressive (CAR) model is one of the most widely used methods, and has been applied at the macro level (e.g. canton and census tract) (Aguero-Valverde, 2013; Islam et al., 2016; Wang and Kockelman, 2013) and the micro level (e.g., intersection and roadway segment) (Barua et al., 2014, 2016; Huang et al., 2017). When Aguero-Valverde and Jovanis (2008) investigated spatial correlation at the segment level using a Poisson lognormal CAR model, they found the CAR model could improve the goodness of fit and reduce the bias associated with model misspecification. Ma et al. (2017a, b) and Wen et al. (2018) developed multivariable CAR models to analyze different severity crashes on freeways, and the results have justified the applicability of the CAR model in addressing the spatial correlation effects.

## 2.3. Spatial distribution and hotspot identification for SV and MV crashes

Previous studies have identified hotspots using total crashes, and have distinguished among crash types such as rear-end, left-turn and sideswipe crashes, but few hotspot identification studies have considered the differences between SV and MV crashes (Miranda-Moreno et al., 2005; Jiang et al., 2014; Park et al., 2014; Wang et al., 2014). The different spatial distributions between SV and MV crashes have been recognized, however. By using spider graphs and kernel density estimation, Kilamanu et al. (2011) found that the spatial distribution patterns of SV and MV crashes in Western Australia differ. Schneider et al. (2017) compared alcohol-related SV and MV crash hotspots at different times of day, and concluded that incidence of the two crash types coinciding at the same locations in the same time periods was very limited. Separate prediction models for SV and MV crashes were developed by Geedipally and Lord (2010a) in order to identify hotspots, which were determined as sites at which the combined crash prediction, that is, the sum of the predicted SV and MV crash counts, was greater than the total mean crash count of all sites on the studied roads. They found that the combined estimate predicted fewer false positives and negatives than a single aggregated model of total crash frequency, so recommended using separate SV and MV crash prediction models for identifying hotspots. The researchers did not identify separate SV and MV hotspots.

Most of these studies are limited in similar ways. That is, they have developed crash prediction models distinguishing SV and MV crashes in order to investigate differing influencing factors and to predict crash counts, and a few have modeled SV and MV separately. Fewer still have attempted to identify separate SV and MV crash hotspots. This study investigates the influencing factors that may contribute to separate hotspots, and assesses the consistency between SV and MV hotspots in order to determine the need for separate identification.

## 3. Data description

A 45-km freeway section in Shanghai was analyzed in this study. Roadway geometric design features, traffic characteristics, and 2012–2013 crash data along the two directions of the freeway section were collected.

## 3.1. Roadway geometric design features

The roadway geometric design features included horizontal, vertical and cross-section alignments. Horizontal and vertical alignment features were extracted from the freeway's blueprint, while the cross-sectional features such as median width and number of lanes were collected using Google Earth®. Horizontal alignment variables included mainline type, curve type, curvature, curve length, transition curve length, and curve proportion. Vertical alignment variables were section type, maximum gradient, vertical curvature, change of grade, longitudinal slope length and vertical curve proportion.

To keep alignment features consistent within each road segment in order to handle any disparities in roadway design, the freeway section was split into homogenous segments. As previous research has indicated that short segments may result in low exposure and lead to statistical uncertainty (Ahmed et al., 2011), the U.S. Highway Safety Manual (HSM) suggests that roadway segments should be no shorter than 0.1 miles (160.9 m) (Bonneson, 2010). Accordingly, the short segments in this study were merged with adjacent segments based on their alignment features. Due to subtle differences in geometric design between the roadway's two directions, the north-to-south direction was finally divided into 174 study segments, and the south-to-north direction into 186 segments. In the end, the 45-km freeway section was split into 360 homogeneous segments with an average length of 224.315 m. Table 1 shows the descriptive statistics of the geometric variables.

## 3.2. Traffic characteristics

Traffic operational characteristics were extracted from videos collected by 76 traffic cameras installed along both directions of the

**Table 1**
Descriptive Statistics for Road Segments' Design Features.

| Explanatory Variables | Description | Summary Statistics |
|---|---|---|
| Mainline type | 1: segments away from ramps | Observations: 324 |
| | 2: segments near ramps | Observations: 36 |
| Horizontal curve type | 1: straight segment | Observations: 198 |
| | 2: curve segment | Observations: 140 |
| | 3: tangent-curve segment | Observations: 22 |
| Horizontal curvature | Reciprocal of curve radius ($10^{-4}\,m^{-1}$) | Mean: 1.91<br>Std dev: 2.477 |
| Horizontal curve length | Length of whole curve alignment(km) | Mean: 2.504<br>Std dev: 1.866 |
| Transition curve length | Length of transition curve between straight line and curve (km) | Mean: 0.173<br>Std dev: 0.127 |
| Horizontal curve proportion | Ratio of curve and tangent alignment (%) | Mean: 42.064<br>Std dev: 48.185 |
| Vertical section type | 1: uphill | Observations: 63 |
| | 2: downhill | Observations: 61 |
| | 3: convex vertical curve | Observations: 87 |
| | 4: concave vertical curve | Observations: 76 |
| | 5: combined segments with slope and curve | Observations: 73 |
| Maximum gradient | Maximum grade of slope in segment (%) | Mean: 0.876<br>Std dev: 0.883 |
| Vertical curvature | Reciprocal of vertical curve radius ($10^{-4}\,m^{-1}$) | Mean: 0.631<br>Std dev: 0.652 |
| Change of grade | Difference between maximum and minimum grade within segment (%) | Mean: 1.002<br>Std dev: 1.229 |
| Longitudinal slope length | Total length of vertical slope (km) | Mean: 0.270<br>Std dev: 0.141 |
| Vertical curve proportion | Ratio of vertical curve length in segment (%) | Mean: 56.9<br>Std dev: 44.8 |
| Barrier-protected median width | 1: median width is 2m | Observations: 45 |
| | 2: median width is more than 2m | Observations: 315 |
| Number of lanes (in one direction) | 1: two lanes | Observations: 146 |
| | 2: more than two lanes | Observations: 214 |
| Length | Length of road segment (m) | Mean: 224.315<br>Std dev: 95.047 |

*Notes*: Each set of observations, e.g., number of lanes, totals 360, on which the statistical parameters are based.
Std dev. is the standard deviation.

**Table 2**
Descriptive Statistics for Traffic Operational Data.

| Explanatory Variables | Statistical Description | Mean | Std dev. | Min | Max |
|---|---|---|---|---|---|
| Traffic volume | Average hourly traffic volume per lane (vehicle/h) | 456.342 | 155.923 | 184 | 943 |
| Truck proportion | Proportion of trucks in total volume (%) | 28.841 | 5.803 | 19.54 | 45.65 |
| Average speed | Average speed of vehicles passing through road segment (km/h) | 82.893 | 12.001 | 60.02 | 103.82 |
| Speed variance | Speed variance of vehicles passing through road segment | 48.769 | 11.798 | 13.27 | 74.83 |

*Note:* Std dev. is the standard deviation.

studied freeway section. Average hourly traffic volume per lane, proportion of trucks, average speed, and speed variance were extracted using video recognition software. Detailed information about traffic operational data is given in Table 2.

### 3.3. Crash data

A crash is defined as an incident that causes casualties or property damage due to an accident or fault of a vehicle on the road. In China, traffic police use two procedures to handle the recording of crashes: simple and normal. The simple procedure mainly deals with property damage only crashes (PDO), in which they use portable devices to provide them with standardized options to record information. The normal procedure collects more detailed data, and is mainly used to deal with injury and fatal crashes. In these cases, the officer is required to sketch the scene of the crash spot and record additional investigative information such as pavement condition. PDO and injury (including fatal) crash data are thus entered and stored in two different tables in the Integrated Traffic Management Database (ITMD), a uniform system used throughout China.

For this study, 2012–2013 crash data from a 45-km section of the G15 freeway was extracted from the Shanghai ITMD, for a total of 1293 crashes. Of these, 473 were SV and 820 were MV crashes. All the crashes were located using a linear referencing method based on the freeway mile point and the location record of each crash. The crash frequencies of SV and MV crashes were then computed for each of the 360 road segments. The statistical descriptions for SV, MV and total crashes are shown in Table 3.

Spatial distributions of SV and MV crashes for the two directions of the studied freeway section are shown in Fig. 1. The horizontal coordinate is the roadway segment ID, and the vertical coordinates are the SV and MV crash proportions, calculated as each segment's SV and MV crash counts divided by the respective SV and MV totals for the freeway section. For example, 1.5% of the 473 total SV crashes occurred on segment 66. Crash proportions are used in preference to absolute numbers because dangerous SV sites, for example, are obtained by comparison with the numbers of SV crashes (but not MV crashes) on other freeway segments. The color of the Fig. 1 stripes indicates which of the SV or MV crash proportions is greater. It can be seen that SV and MV crashes tend to aggregate on different road segments. For example, SV crash proportions of segments #66, #165 and #179 exceed their MV crash proportions substantially, while the SV proportion of segment #155 is much lower than the segment's MV proportion.

The total 360 freeway segments were ranked in descending order of SV and MV crash counts separately. The two sets of segment rankings, in which a ranking of 1 indicates the highest crash count, are compared in Fig. 2. Because the SV and MV crash counts for many segments are the same, those segments share the same rank, which is indicated by the size of the point. That is, the larger the point, the more segments share the rank. It can be noted that many of the larger points are distributed off the diagonal: as points along the diagonal indicate segments with similar ranks of both SV and MV crashes, the off-diagonal dispersion shows that the SV and MV segment crash rankings differ. Spearman's rho was calculated to measure the statistical dependence between the two sets of segment ranking. The resulting value of 0.532 implies a moderate correlation. If highway safety improvement funding were more available, the moderate correlation would support continuing to identify hotspots by total crashes. However, due to limited financial resources, only the highest-ranked segments receive funding, so Spearman's rho was recalculated for the top 10%-ranked segments. The low negative correlation of -0.219 reveals inconsistent spatial distribution of the top-ranked segments and encourages separate hotspot identification for the two crash types.

**Table 3**
Summary Statistics for Crash Count Frequency.

| Crash type | Frequency | Mean | Std dev. | Min | Max |
|---|---|---|---|---|---|
| SV crashes | 473 | 1.31 | 1.80 | 0 | 13 |
| MV crashes | 820 | 2.28 | 3.24 | 0 | 22 |
| Total crashes | 1293 | 3.59 | 4.52 | 0 | 28 |

*Note:* Std dev. is the standard deviation.

## 4. Methodology

### 4.1. Negative binomial regression model

Poisson regression is a prediction model widely and traditionally used for crash analysis, but the problem of over-dispersion of crash data has limited its application (Lord and Mannering, 2010), and has often led researchers to use a negative binomial model (NB) (Anastasopoulos and Mannering, 2009; Geedipally and Lord, 2010b; Ma et al., 2017a, b).

The graphs in Fig. 3 show the empirical distribution frequencies of SV, MV and total crashes, overlaid with best-fit negative binomial curves. Three parameters are noted for each of the graphs: *size* is the dispersion parameter of the negative binomial distribution, *mu* is the crash mean, and *var* is the variance. For all crash types, the variance is larger than the mean, indicating the crash data used in this study is over-dispersed. The consistently positive *size* parameter demonstrates that the negative binomial model is appropriate.

This study therefore assumes that the crashes across the 360 segments follow the negative binomial distribution. The standard negative binomial probability density function and the negative binomial regression model are expressed as (Chou and Steenhard, 2009):

$$f(y_i \mid \lambda_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})}\left(\frac{\lambda_i}{\lambda_i + \alpha^{-1}}\right)^{y_i}\left(\frac{\alpha^{-1}}{\lambda_i + \alpha^{-1}}\right)^{\alpha^{-1}} \quad (1)$$

$$\log(\lambda_i) = \beta_0 + \boldsymbol{\beta}\mathbf{X_i} \quad (2)$$

where $y_i$ is the number of crashes occurring on segment $i$ during a period of time, and $\lambda_i$ is the mean predicted crash frequency for segment $i$ and is assumed to be a function of the explanatory vector $\mathbf{X_i}$. $\beta_0$ and $\boldsymbol{\beta}$ are regression coefficients and $\alpha$ is the over-dispersion parameter. When $\alpha = 0$, the NB model reduces to a Poisson regression model.

### 4.2. Negative binomial spatial CAR model

The studied data was collected from a freeway section where the two directions are separated by a median. Crashes on freeway segments of the same direction were confirmed to be positively spatially correlated by Moran's $I$; due to the spatial correlation, crashes across the 360

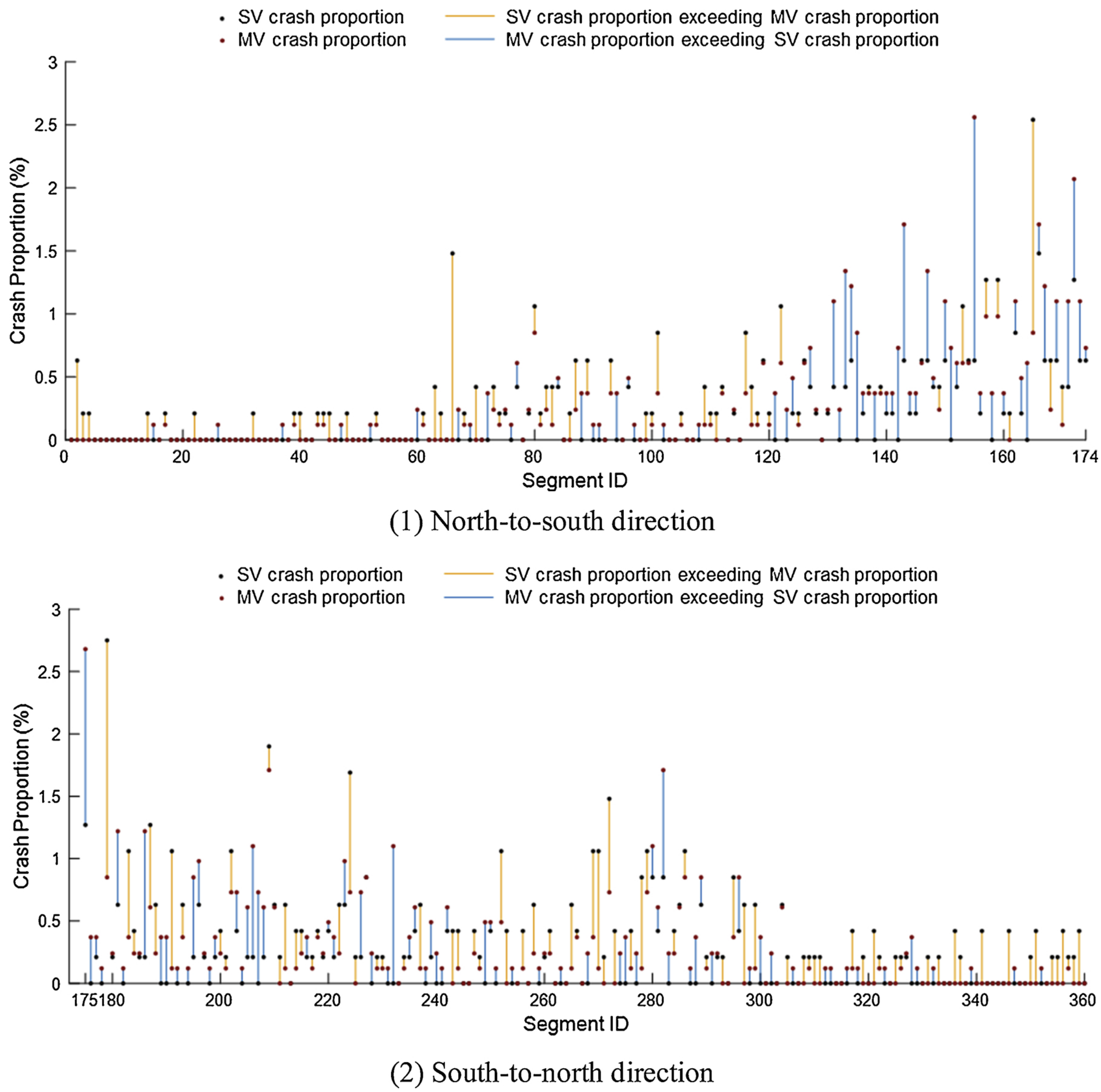



**Fig. 1.** Spatial Distribution of SV and MV Crashes.
(1) North-to-south direction.
(2) South-to-north direction.

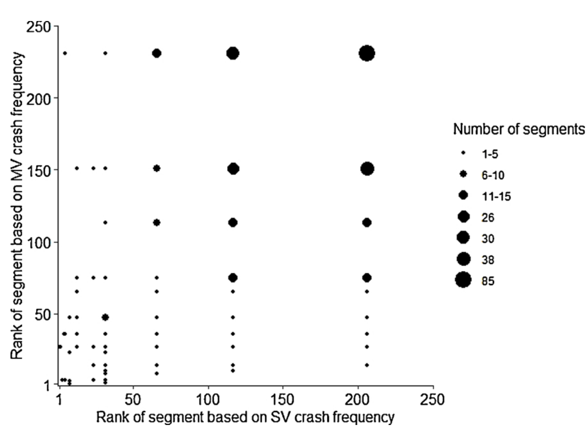



**Fig. 2.** Segment Rank Based on SV and MV Crash Frequency.

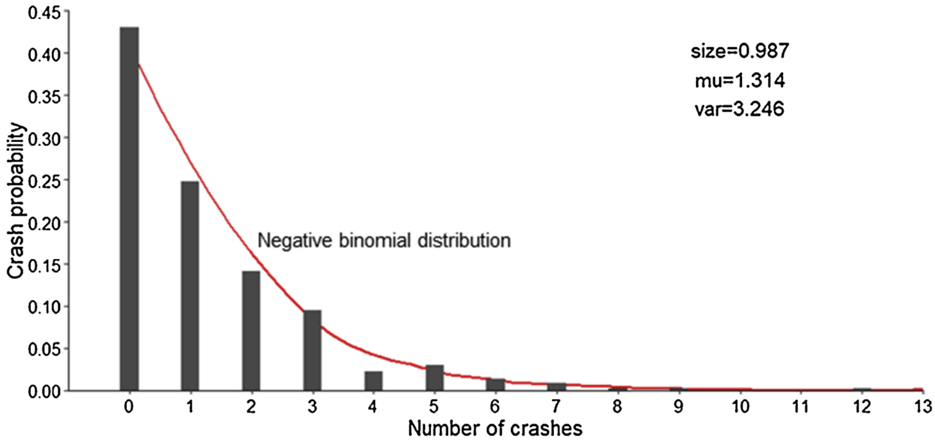
segments cannot be considered as purely randomly distributed. The conditional autoregressive (CAR) model has flexibility in measuring the magnitude of correlation and has been suggested in many previous studies (El-Basyouny and Sayed, 2009; Guo et al., 2010; Huang and Abdel-Aty, 2010; Song et al., 2006; Xie et al., 2014). In the CAR model, the random effect term $\varphi_i$ is used to capture the spatial correlation. The negative binomial spatial CAR model (NB-CAR) is defined as:

$$y_i \sim NB(\lambda_i, \alpha) \tag{3}$$

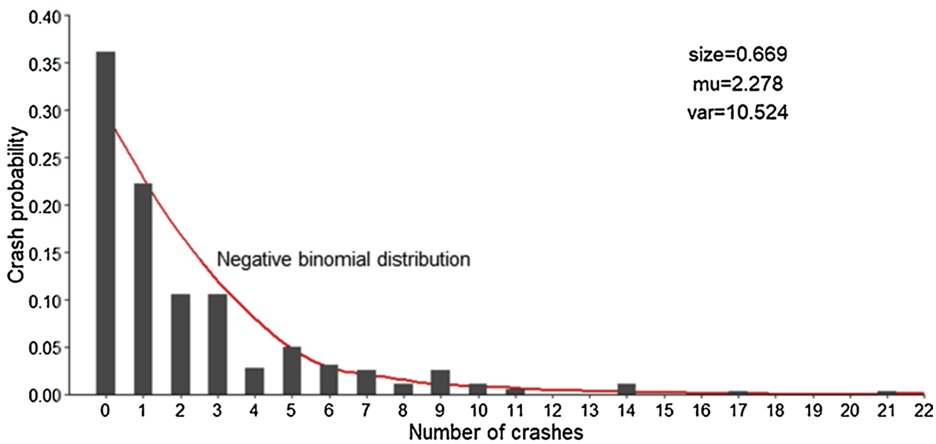
$$\log(\lambda_i) = \beta_0 + \boldsymbol{\beta}\mathbf{X_i} + \varphi_i \tag{4}$$

where $\varphi_i$ is set to follow a CAR prior, which can represent the spatial relationship among road segments by a proximity matrix $\mathbf{W}$ with entry $w_{i,j}$ reflecting the spatial association between segments $i$ and $j$. When segments $i$ and $j$ run in the same direction, the $w_{i,j}$ is defined as the reciprocal of the distance between the midpoints of segment $i$ and segment $j$; otherwise $w_{i,j}$ is set to be 0 representing the independence of freeway segments running in different directions (Li and Wang, 2017). Therefore the full conditional distribution of $\varphi_i$ follows (Miaou and Song, 2005):

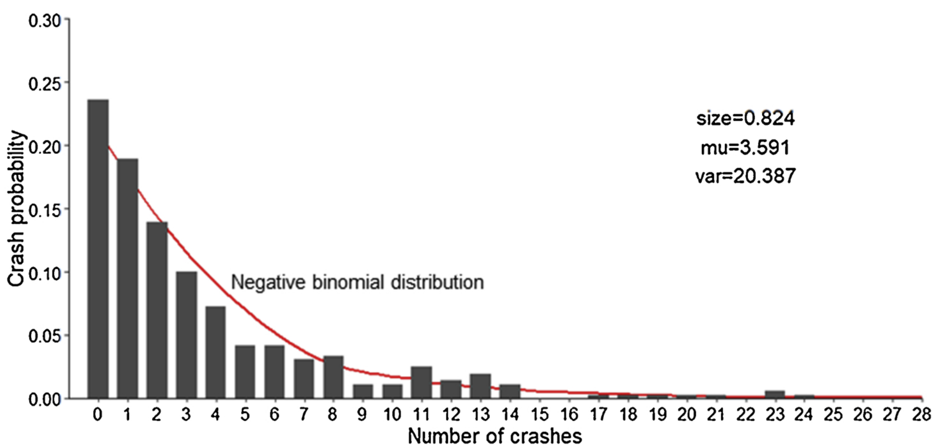
$$P(\varphi_i \mid \varphi_{-i}) \propto \exp[-\frac{w_{i+}}{2\sigma_\varphi^2}(\varphi_i - \delta\sum_{i\neq j}\frac{w_{i,j}}{w_{i+}}\varphi_j)^2] \tag{5}$$




(a) Probability distribution histogram of SV crashes




(b) Probability distribution histogram of MV crashes




(c) Probability distribution histogram of total crashes

**Fig. 3.** Probability Distributions and Best-fit Negative Binomial Curves for SV, MV and Total Crashes.
(a) Probability distribution histogram of SV crashes.
(b) Probability distribution histogram of MV crashes.
(c) Probability distribution histogram of total crashes.

where $\varphi_{-i}$ is the set of $\varphi_i$ for any $i \neq j$ and $w_{i+} = \sum_{i=1}^{n} w_{i,j}$ representing the sum of weights for the given freeway direction.

An intrinsic version of the CAR model in Besag et al. (1991) was used in this study. $\delta$ in Eq. (5) was set as 1. The conditional distribution of $\varphi_i$ becomes:

$$\varphi_i \mid \varphi_{-i} \sim N\left(\sum_{i \neq j} \frac{w_{i,j}}{w_{i+}} \varphi_i, \frac{\sigma_\varphi^2}{w_{i+}}\right) \tag{6}$$

### 4.3. Bivariate negative binomial spatial CAR model

To consider the site correlation between SV and MV crashes, a bivariate negative binomial spatial CAR model (BNB-CAR) was developed. Two jointly normally distributed random error variables, $\boldsymbol{\varepsilon}_1$, $\boldsymbol{\varepsilon}_2$, were used to measure the site correlation. The model is expressed as:

$$y_{i1} \sim NB(\lambda_{i1}, \alpha_1) \tag{7}$$

$$y_{i2} \sim NB(\lambda_{i2}, \alpha_2) \tag{8}$$

$$\log(\lambda_{i1}) = \beta_{01} + \boldsymbol{\beta}_1 \mathbf{X}_{i1} + \varepsilon_{i1} + \varphi_i \tag{9}$$

$$\log(\lambda_{i2}) = \beta_{02} + \boldsymbol{\beta}_2 \mathbf{X}_{i2} + \varepsilon_{i2} + \varphi_i \tag{10}$$

$$(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2) \sim N\{(\mathbf{0}, \mathbf{0}), (\sigma_1^2, \rho\sigma_1\sigma_2, \sigma_2^2)\} \tag{11}$$

where $y_{i1}$ and $y_{i2}$ are SV and MV crash counts occurring on segment $i$. $\lambda_{i1}$ and $\lambda_{i2}$ are crash expectations. $\mathbf{X}_{i1}$ and $\mathbf{X}_{i2}$ are independent variables, and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors of regression coefficients. $\varphi_i$ is the CAR effect term and $\rho$ is the site correlation coefficient.

### 4.4. Modeling process

The crash prediction models were developed in WinBUGS®. For each model, two chains of 20,000 iterations were set up, with the first 5000 iterations used in the burn-in step. Methods including monitoring the trace plots and checking for autocorrelations were used to confirm convergences. The DIC (deviance information criterion) was used to evaluate the models' performance (Spiegelhalter et al., 2003): a lower DIC value means a better model fitting, and in a comparison between two models, a difference greater than 10 indicates better performance in the model with the lower DIC.

Correlation tests for variables were conducted before model estimation to avoid multicollinearity issues. Results showed that horizontal curve type was significantly correlated with horizontal curvature and horizontal curve proportion; vertical section type was highly correlated with vertical curvature, and average speed was highly correlated with median width. Model construction therefore avoided including these highly correlated variables in one model simultaneously.

A BNB-CAR model was developed to analyze SV and MV crashes, and three NB-CAR models (for SV, MV and total crashes separately) were developed for comparison. The mean standardized absolute deviance (MSAD) was calculated to compare the crash frequency prediction precision, and results are shown in Table 4. The 360 freeway segments had been divided into two groups: 80% of the segments were used to develop, or train, the models, and 20% were used to test their prediction performance. For both training and testing, the BNB-CAR model performed better than the NB-CAR models. For MV crashes, the BNB-CAR model showed slight over-fitting (i.e., the MSAD of the testing dataset was larger than that of the training dataset), but its prediction precision still outperformed the NB-CAR model. For SV crashes, the BNB-CAR model did not show an over-fitting problem.

**Table 4**
Model Performance Measured by MSAD Values.

| Crash | Model | Training Dataset | Testing Dataset |
|---|---|---|---|
| SV crash | NB-CAR | 0.646 | 0.652 |
| | BNB-CAR | 0.622 | 0.601 |
| MV crash | NB-CAR | 0.667 | 0.739 |
| | BNB-CAR | 0.534 | 0.633 |

These were, however, preliminary results. Because the sample was relatively small and all 360 road segments came from the same freeway in Shanghai, the 20% of freeway segments used for model testing, which also contain important information, were reincorporated to rebuild the final BNB-CAR and NB-CAR models to obtain more efficient parameter estimates. In other words, the final BNB-CAR and NB-CAR models are based on the entire 360-segment sample.

### 4.5. Hotspot identification

Potential for safety improvement (PSI) was used to identify hotspots. PSI is an index that identifies roadway sites that can be made less risky by undertaking safety measures. It is calculated as the difference between the expected crash frequency and the predicted crash count of a crash prediction model (Montella, 2010). The greater a site's PSI, the more likely improvement will reduce the number of vehicle crashes for that site. PSI thus has been used as the evaluation index in many studies for ranking hotspots (Persaud et al., 1999; El-Basyouny and Sayed, 2006). Based on the crash predictions, the empirical Bayes (EB) method was used to calculate the PSI of each freeway segment:

$$EB_i = \frac{1}{1 + (\lambda_i/k)} \times \lambda_i + \left(1 - \frac{1}{1 + (\lambda_i/k)}\right) \times y_i \tag{12}$$

$$PSI_i = EB_i - \lambda_i \tag{13}$$

where $EB_i$ is the empirical Bayes expected crash frequency of segment $i$, and $k$ is the over-dispersion parameter of the negative binomial model.

The freeway segments were ranked in descending order of PSI to identify SV, MV and total crash hotspots. Hotspot identification process is shown in Fig. 4, where $Crash_{pre,i}$ is the crash prediction for segment $i$ based on the BNB-CAR and NB-CAR models, and $Crash_{obs,i}$ is the crash observation. The SV, MV and total crash hotspots were then tested for consistency to determine the degree of overlapping.

The method consistency test (MCT) was used to test the hotspot consistency. It is a quantitative evaluation method used to compare the effects of different hotspot identification methods by measuring the number of identical hotspots identified in two time periods (Cheng and Washington, 2008; Montella, 2010; Jiang et al., 2014). In this study, the MCT is repurposed to measure the number of hotspots shared by SV, MV and total crashes. The test statistic is given as:

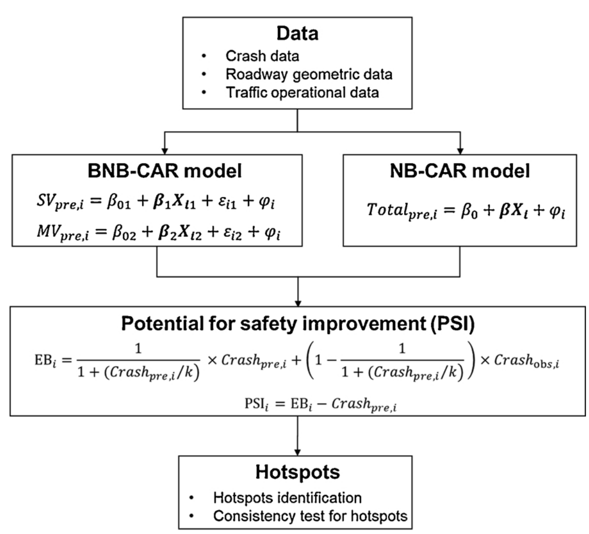



**Fig. 4.** Hotspot identification process.

**Table 5**
Modeling Results for SV, MV and Total Crashes.

| Variables | | NB-CAR models | | | | | | BNB-CAR model | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Total Crashes | | SV Crashes | | MV Crashes | | SV Crashes | | MV Crashes | |
| | | Mean | Std. | Mean | Std. | Mean | Std. | Mean | Std. | Mean | Std. |
| Intercept | | −5.229[a] | 1.413 | −2.838[b] | 1.481 | −8.075[a] | 1.496 | −3.339[a] | 1.189 | −8.952[a] | 1.681 |
| Mainline type (reference:1) | 2 | 0.523[a] | 0.182 | | | 0.890[a] | 0.288 | | | 0.870[a] | 0.224 |
| Horizontal curve proportion | | | | 0.481[a] | 0.197 | 0.405[a] | 0.190 | 0.440[a] | 0.162 | 0.418[a] | 0.157 |
| Longitudinal slope length | | 0.002[a] | 0.001 | | | 0.002[a] | 0.001 | | | 0.002[a] | 0.001 |
| Number of lanes (reference:1) | 2 | 1.022[a] | 0.145 | | | | | | | | |
| Median width (reference:1) | 2 | | | −0.566[b] | 0.288 | −0.425 | 0.296 | −0.521[a] | 0.236 | −0.552[a] | 0.224 |
| Vertical section type (reference:1) | 2 | 0.149 | 0.192 | | | 0.226 | 0.314 | | | 0.241 | 0.252 |
| | 3 | 0.484[a] | 0.206 | | | 0.654[a] | 0.325 | | | 0.638[a] | 0.262 |
| | 4 | 0.219 | 0.232 | | | 0.287 | 0.362 | | | 0.359 | 0.288 |
| | 5 | 0.682[a] | 0.267 | | | 0.863[a] | 0.390 | | | 0.948[a] | 0.334 |
| Speed variance | | | | | | 0.011 | 0.008 | | | 0.013[b] | 0.007 |
| Truck proportion | | −0.034[a] | 0.014 | −0.037[b] | 0.020 | −0.086 [a] | 0.023 | −0.042[a] | 0.017 | −0.061[a] | 0.017 |
| Ln(Volume) | | 0.434[a] | 0.150 | | | 1.029[a] | 0.283 | | | 0.973[a] | 0.243 |
| Ln(Length) | | 0.558[a] | 0.218 | 0.816[a] | 0.237 | 0.637[a] | 0.276 | 0.902[a] | 0.196 | 0.701[a] | 0.251 |
| $\sigma_{11}(\sigma_{22})$ | | | | | | | | 0.29[a] | 0.117 | 0.302[a] | 0.122 |
| $\sigma_{12}$ | | | | | | | | 0.263[a] | 0.100 | | |
| Dispersion | | 0.896 | | 0.784 | | 0.781 | | 0.687 | | 0.688 | |
| DIC | | 1576.86 | | 1106.96 | | 1324.14 | | 1077.9 | | 1305.12 | |

*Note:* Std. is the standard error of the corresponding variable.
[a] Variables statistically significant at 95% confidence interval.
[b] Variables statistically significant at 90% confidence interval.

$$MCT_{i,j} = \{k_1, \ldots, k_m, \ldots, k_{n\delta}\}_i \cap \{k_1, \ldots, k_m, \ldots, k_{n\delta}\}_j \tag{14}$$

where $i$, $j$ represent any combination of SV crashes, MV crashes and total crashes, and $i \neq j$. $k_m$ is the hotspot ranked $m$, $n$ is the total number of freeway segments, and $\delta$ is the proportion of segments defined as hotspots.

## 5. Results

This section shows the modeling results of the crash prediction models, followed by the hotspot identification results for SV, MV, and total crashes.

### 5.1. Modeling results

The parameter estimation results are shown in Table 5. In the BNB-CAR model, the dependent variables are SV crash counts and MV crash counts, and the independent variables include mainline type, horizontal curve proportion, longitudinal slope length, median width, vertical section type, speed variance, truck proportion, traffic volume, and segment length. The site correlation coefficient $\rho$ of the BNB-CAR model was 0.89, indicating a non-negligible correlation effect between SV and MV crashes.

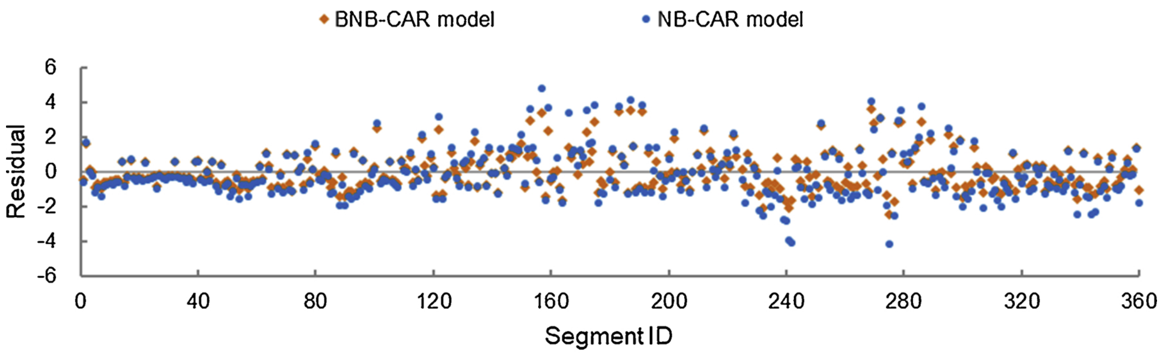
As can be seen in the modeling results, the BNB-CAR model outperformed the NB-CAR models by three measures. First, the BNB-CAR model provided more efficient parameter estimates; i.e., its regression coefficients have smaller standard errors than the NB-CAR models (e.g., 0.224 vs. 0.288 for mainline type in the MV models). Second, the DIC values of the BNB-CAR model are lower than those of the NB-CAR models, the large differences in DIC values indicating that the bivariate model had better model fitting. Third, the MSADs of the BNB-CAR model (0.497 for SV, 0.421 for MV) are smaller, and therefore more precise, than those of the NB-CAR models, for which the MSAD was 0.641 for SV crashes and 0.579 for MV crashes. The regression residuals of the BNB-CAR model and the NB-CAR models are compared in Fig. 5, which shows that residuals of the BNB-CAR model are more concentrated toward zero than the NB-CAR models, especially for MV crashes.

After taking account of the site correlation between SV and MV crashes, some previously insignificant variables became significant. For SV crashes, median width and truck proportion, which were significant at the 90% confidence interval in the univariate NB-CAR model, became significant at the 95% confidence interval in the BNB-CAR model. For MV crashes, median width and speed variance were not significant in the univariate model, but became significant in the bivariate model. The BNB-CAR model is thus thought to provide more efficient parameter estimates. Therefore, in the following discussion, the BNB-CAR modeling results were used to analyze the crash influencing factors and to identify hotspots for SV and MV crashes, while the NB-CAR results were used to analyze total crashes.

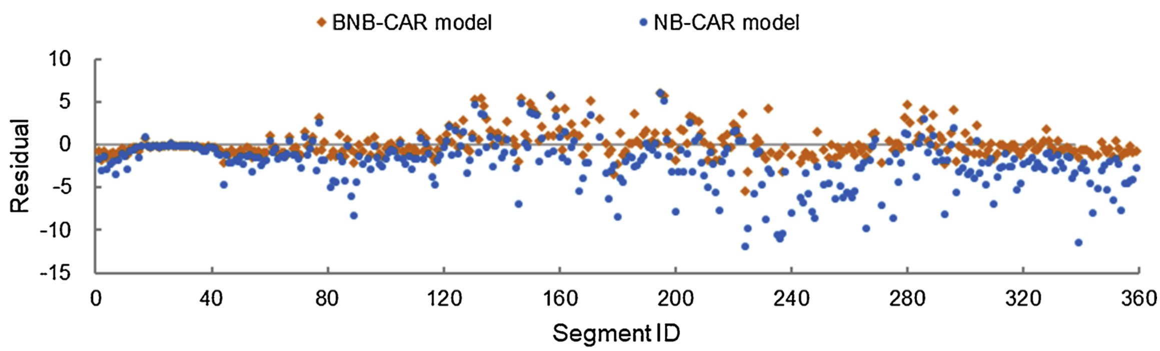
Only truck proportion and segment length were found to be significant across the SV, MV and total crash models. Results show that the proportion of trucks was negatively related to the three crash groups, similar to results that have been reported in previous studies (Hiselius, 2004; Lord et al., 2005b). Hiselius (2004) speculated that a decrease in average speed might be responsible for reducing these crash counts; this seems likely when considering the results of Roh et al.'s (2017) recent study, which demonstrated that the average traffic flow speed tends to decrease as the percentage of trucks increases. Segment length was significant with a positive sign in this study's SV, MV and total crash models, indicating that longer segments were likely to have more crashes. Similar results have been reported in previous studies (Yu and Abdel-Aty, 2013; Ma et al., 2016).

Significant variables for SV crashes also included horizontal curve proportion and barrier-protected median width. Horizontal curve proportion was positively associated with SV crashes, indicating that as the proportion of horizontal curves increased, the SV crash count increased. The negative sign for median width means that segments with a median width over 2 m had fewer SV crashes than segments with 2 m-wide medians, a result consistent with Yu and Abdel-Aty (2013). The reason might be that the wider medians could reduce the impact of glare on drivers.

Horizontal curve proportion and median width were also significant for MV crashes, and with the same sign as in the SV model. However, unlike SV crashes, MV crashes were also significantly influenced by longitudinal slope length, mainline type, vertical section type, speed variance and traffic volume. The positive parameter estimation of longitudinal slope length means that longer longitudinal slopes




(1) Regression Residual Plot of SV Crashes




(2) Regression Residual Plot of MV Crashes

**Fig. 5.** Regression Residual Plots of BNB-CAR and NB-CAR Models.
(1) Regression Residual Plot of SV Crashes.
(2) Regression Residual Plot of MV Crashes.

increased the likelihood of MV crashes. The positive coefficient of mainline type indicates that freeway segments near ramps were prone to more MV crashes than those away from ramps. This is reasonable because weaving traffic flow near a ramp tends to increase traffic conflicts, which may result in a high inter-vehicle crash risk (Fazio et al., 1993). The positive signs in vertical section type indicate that, compared with uphill sections, convex vertical curves (vertical section type 3) and combined segments with slope and curve (type 5) were more likely to have MV crashes (vertical section types 2 and 4 also show positive signs, but the results were not significant). These results may be due to vertical curves and combined alignments limiting drivers' sight distance and driving comfort, leading to unsuitable driving operation. MV crash occurrence increased with speed variance, that is, as the standard deviation of driving speed increased, a result consistent with a study by Garber and Ehrhart (2000). Large traffic volume also appeared to increase MV crash frequency, a result similar to those noted in previous studies (Yu and Abdel-Aty, 2013; Ma et al., 2016; Yu et al., 2018).

In the NB-CAR model for total crashes, longitudinal slope length, mainline type, vertical section type and traffic volume were significant, and with the same sign, as in the MV model. Additionally, number of lanes was found to be significant with a positive coefficient, indicating that the total crash frequency on three-lane or four-lane segments was higher than on two-lane segments. This result may be attributed to an increased possibility of conflict associated with a greater number of lanes (Kononov et al., 2008). However, variables such as horizontal curve proportion, median width, and speed variance, which are shown to have a significant impact on SV and/or MV crash, are not significant in the total crash model. Such discrepancies demonstrate that the crash prediction model of total crashes does not appear adequate for either SV or MV crashes.

### 5.2. Hotspot identification results

Hotspots were identified using PSI values, which were calculated based on the crash prediction models. Specifically, SV and MV crash hotspots were identified based on the results of the BNB-CAR model, and total crash hotspots were identified based on the results of the total crash NB-CAR model.

As shown in Table 6, segments with high SV crash proportions in Fig. 1, such as #66, #165 and #179, were indeed identified as SV hotspots (along with segment #175 and #155 for MV crashes). This means the spatial distribution of the crashes shown in Fig. 1 can to some extent reflect the spatial distribution of SV and MV hotspots. In Table 6, each crash type shows quite different hotspots for its top 10 rankings. No segment is shared by SV and MV crashes, and segment 159,

**Table 6**
Top 10 Freeway Hotspots for SV, MV and Total Crashes.

| Hotspot Rank | Segment ID | | |
|---|---|---|---|
| | SV Crashes | MV Crashes | Total Crashes |
| 1 | 179 | 175 | 175 |
| 2 | 165 | 155 | 179 |
| 3 | 209 | 172 | 166 |
| 4 | 224 | 166 | 209 |
| 5 | 66 | 143 | 165 |
| 6 | 187 | 282 | 155 |
| 7 | 157 | 173 | 157 |
| 8 | 272 | 147 | 172 |
| 9 | 269 | 133 | 143 |
| 10 | 191 | 131 | 159 |

**Table 7**
Hotspot Consistency Test Results for SV, MV and Total Crashes.

| Proportion of Hotspots | Shared by MV & SV | Shared by MV & Total | Shared by SV & Total |
|---|---|---|---|
| Top 1% (top 4) | 0 | 2 | 2 |
| Top 2.5% (top 9) | 0 | 5 | 4 |
| Top 5% (top 18) | 2 | 11 | 8 |
| Top 10% (top 36) | 10 | 24 | 20 |
| Top 15% (top 54) | 19 | 37 | 33 |
| Top 20% (top 72) | 28 | 50 | 43 |

identified as a hotspot for total crashes, is not included in either SV or MV hotspots. To determine the degree of overlapping hotspots for SV, MV and total crashes, segments with PSIs in the top 20% (top 72 segments) were selected for a detailed consistency test.

The top 1%, 2.5%, 5%, 10%, 15% and 20% of freeway segments ranked in PSI descending order were used to conduct the hotspot consistency test, as shown in Table 7. The consistency test results indicate that SV crashes have few hotspots in common with MV crashes. In the top 2.5% of hotspots, none are mutually identified for SV and MV crashes, and of the 5%, 10%, 15% and 20%, fewer than 40% are shared by SV and MV crashes. These results indicate that hotspot consistency between the two crash types is weak. Total crash hotspots show a slightly higher consistency with MV crashes than with SV crashes, but the consistency in both cases is low. Nearly half of SV and MV hotspots, 55.6% of SV and 44.4% of MV, are not included in total crash hotspots.

## 6. Discussion

A highly correlated effect between SV and MV crashes was captured by using the bivariate model. The site correlation coefficient of 0.89 is similar to that of studies by Yu and Abdel-Aty (2013) and Ma et al. (2016), in which the coefficient was 0.68 and 0.72, respectively. This site correlation, measured by random error terms, is believed to be caused by unobserved factors affecting crash frequencies across crash types simultaneously (Ma and Kockelman, 2006; Lee et al., 2015). Because it incorporates site correlation, the BNB-CAR model can facilitate more reliable parameter estimates and more accurately fit the data. Two MV crash influencing factors, median width and speed variance, became newly significant in the BNB-CAR model. This suggests that the univariate model's ignoring of site correlation may have caused imprecise parameter estimation and even inaccurate analysis, which highlights the importance of considering site correlation when analyzing SV and MV crashes simultaneously.

Similar to results of previous research, this study's modeling results showed substantial differences in the factors that influence SV and MV crashes. It was found, for example, that SV crash occurrence was not significantly related to some of the traffic operation characteristics that had significant impact on MV crashes. Traffic volume and speed variance were shown to have no significant influence on SV crashes, which may indicate that SV crashes are as likely to occur on roadways with high-traffic, high-service level, low-traffic, or low-service level, but further confirmation is needed. The study of Yu and Abdel-Aty (2013) reached a similar conclusion, but results from other previous studies have been inconsistent (Persaud and Mucsi, 1995; Geedipally and Lord, 2010b). The MV crash count in this study, however, increased as traffic volume and speed variance increased, consistent with previous studies (Persaud and Mucsi, 1995; Geedipally and Lord, 2010b; Yu and Abdel-Aty, 2013). Also important to note is that the prediction model based on total crashes was shown to be inadequate. Variables such as horizontal curve proportion, median width, and speed variance, significant in this study's SV and MV models, were not identified as significant factors in the total crash model. These findings indicate the importance of developing separate crash prediction models for SV and MV crashes, so as to avoid missing important influencing factors.

According to the potential for safety improvement (PSI) comparison results, SV and MV crashes tend to aggregate at different locations, and are characterized by the varying influencing factors noted above. None of the top ten SV and MV hotspots are mutually shared. Total crash hotspots have a slightly higher consistency with MV than with SV hotspots: the top 5% of hotspots (top 18 segments) identified using the total crash prediction model contain 61% of MV hotspots, and 44% of SV hotspots. This phenomenon might be attributed to the overall data characteristic that MV crashes account for 63.42% of total crashes, while the proportion of SV crashes is around 36.58%. This parallel reminds one of the vital role of the original data composition. The MV-Total and SV-Total hotspot consistency test results also indicate that nearly half of SV and MV hotspots might be overlooked by use of the total crash prediction model to identify hotspots. Further, using the total crash model may produce false positives; that is, it may identify sites that are in the top 10 hotspots for neither SV nor MV crashes, sites that should receive lower priority for improvement. Just as using the total crash model to determine SV and MV influencing factors has shortcomings, using it to identify hotspots may lead to biased and inaccurate results.

The high degree of site correlation between SV and MV crashes seems to be in contradiction with the inconsistency between SV and MV hotspots. One reason, though not yet very clear, might be that the unobserved factors that are believed to cause the site correlation make, nonetheless, only small contributions to traffic crashes; whereas the disparities of SV and MV hotspots spatial distribution are due to the large differences in influence of the observed variables in this study. Another possible reason is that the site correlation represents the degree of correlation across the entire studied freeway section (i.e., all 360 segments), while the overall smaller number of SV and MV crashes on segments with top-ranking PSI, that is, the hotspots, are less correlated, as determined by the Spearman's rho calculation for segment ranking based on crash frequency. Confirming the reason for this disparity is a task for future research.

## 7. Conclusions

With the aim of identifying crash hotspots more accurately, this study used roadway geometric design features, traffic operation characteristics and crash data collected along the two directions of a 45-km freeway section in Shanghai to build upon previous studies that have found differing influencing factors and spatial distributions for single and multiple-vehicle crashes. A BNB-CAR model was applied to address (1) site correlation between SV and MV crashes within the same freeway segment, and (2) spatial correlation among different freeway segments running in the same direction. For comparison, NB-CAR models were developed separately for SV, MV and total crashes. Based on the models' crash estimates, influencing factors and hotspots of SV, MV and total crashes were identified and compared.

Two important results were found: first, SV and MV crashes differ in both influencing factors and hotspots; second, analysis and hotspot identification based on results of the total crash prediction model proved inadequate. The total crash model missed quite a few significant factors that influenced only SV or MV crashes, such as speed variance; and not only did it miss several important hotspots, but it also showed a tendency to produce false positives. While future research is suggested to determine if total crash models are adequate for identifying hotspots in certain circumstances, this study has made clear that separate identification can be essential. Developing separate crash prediction models for SV and MV crashes simultaneously, as well as considering site correlation and spatial correlation, should generate more accurate detection of influencing factors, and can therefore provide the improved hotspot identification results that are vital for roadway safety management.

As this study was conducted based on data from a particular freeway section in Shanghai, China, the results are somewhat unique to the studied roadway and should be confirmed by further research with different data and road facilities. Attention should also be given to important factors that were not included in this study due to limited availability of sufficient and detailed data. Such factors include speed limit, roadway lighting, roadside slopes, and weather conditions, and, as noted earlier, better data would allow crashes to be differentiated by driver's loss of control. Additionally, Wang et al. (2014) have noted that hotspots at intersection approaches differ in crash pattern. Indeed, using crash pattern, and more specific crash types such as rear-end, sideswipe, and run-off-road crashes, may well improve identification of hotspots; differences in crash severity and between day and night hours also warrant study and verification with more comprehensive factors and new crash data. Merging crash data with naturalistic driving data is one way that future research can conduct the more advantageous in-depth crash investigations needed to better classify crashes, pinpoint causation, and enhance roadway safety.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This study was sponsored by the International Science & Technology Cooperation Program of China (2017YFE0134500), the Science and Technology Commission of Shanghai Municipality (18DZ1200200), and the 111 Project (B17032).

## References

Abdel-Aty, M., Pemmanaboina, R., Hsia, L., 2006. Assessing crash occurrence on urban freeways by applying a system of interrelated equations. Transp. Res. Rec. 1953 (1), 1–9.
Ahmed, M., Huang, H., Abdel-Aty, M., Guevara, B., 2011. Exploring a bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. Accid. Anal. Prev. 43 (4), 1581–1589.
Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of road crash frequency with spatial models. Transp. Res. Rec. 2061 (1), 55–63.
Aguero-Valverde, J., 2013. Multivariate spatial models of excess crash frequency at area level: case of Costa Rica. Accid. Anal. Prev. 59, 365–373.
Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accid. Anal. Prev. 41 (1), 153–159.
Barua, S., El-Basyouny, K., Islam, M.T., 2014. A full Bayesian multivariate count data model of collision severity with spatial correlation. Anal. Methods Accid. Res. 3, 28–43.
Barua, S., El-Basyouny, K., Islam, M.T., 2016. Multivariate random parameters collision count data models with spatial heterogeneity. Anal. Methods Accid. Res. 9, 1–15.
Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Stat. Math. 43 (1), 1–20.
Bonneson, J.A., 2010. Highway Safety Manual. American Association of State Highway and Transportation Officials, Washington, DC.
Cafiso, S., Di Silvestro, G., 2011. Performance of safety indicators in identification of black spots on two-lane rural roads. Transp. Res. Rec. (2237), 78–87.
Cheng, W., Washington, S., 2008. New criteria for evaluating methods of identifying hot spots. Transp. Res. Rec. (2083), 76–85.
Chou, N.-T., Steenhard, D., 2009. A flexible count data regression model using sas® proc nlmixed. Proceedings of the SAS Global Forum: Statistics and Data Analysis 1–14.
Dezman, Z., De Andrade, L., Vissoci, J.R., El-Gabri, D., Johnson, A., Hirshon, J.M., Staton, C.A., 2016. Hotspots and causes of motor vehicle crashes in Baltimore, Maryland: a geospatial analysis of five years of police crash and census data. Injury, S0020138316304326.
Dong, B., Ma, X., Chen, F., Chen, S., 2018. Investigating the differences of single-vehicle and multivehicle accident probability using mixed logit model. J. Adv. Transp., 2702360.
El-Basyouny, K., Sayed, T., 2006. Comparison of two negative binomial regression techniques in developing accident prediction models. Transp. Res. Rec. (1950), 9–16.
El-Basyouny, K., Sayed, T., 2009. Urban arterial accident prediction models with spatial effects. Transp. Res. Rec. (2102), 27–33.
Fazio, J., Holden, J., Rouphail, N.M., 1993. Use of freeway conflict rates as an alternative to crash rates in weaving section safety analyses. Transp. Res. Rec. (1401), 61–69.
Feng, M., Wang, X., Lee, J., Abdel-Aty, M., 2019. Transferability of safety performance functions and hotspot identification for freeways of the United States and China. Proceedings of the Transportation Research Board 98th Annual Meeting.
Garber, N.J., Ehrhart, A.A., 2000. Final Report: the Effect of Speed, Flow, and Geometric Characteristics on Crash Rates for Different Types of Virginia Highways. Virginia Transportation Research Council, Charlottesville. Virginia, pp. 10–15 1.
Geedipally, S., Lord, D., 2010a. Identifying hot spots by modeling single-vehicle and multivehicle crashes separately. Transp. Res. Rec. (2147), 97–104.
Geedipally, S.R., Lord, D., 2010b. Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of poisson–gamma models. Accid. Anal. Prev. 42 (4), 1273–1282.
Guo, F., Wang, X., Abdel-Aty, M.A., 2010. Modeling signalized intersection safety with corridor-level spatial correlations. Accid. Anal. Prev. 42 (1), 84–92.
Hiselius, L.W., 2004. Estimating the relationship between accident frequency and homogeneous and inhomogeneous traffic flows. Accid. Anal. Prev. 36 (6), 985–992.
Huang, H., Abdel-Aty, M., 2010. Multilevel data and Bayesian analysis in traffic safety. Accid. Anal. Prev. 42 (6), 1556–1565.
Huang, H., Chin, H., Haque, M., 2009. Empirical evaluation of alternative approaches in identifying crash hot spots: Naive ranking, empirical bayes, and full bayes methods. Transp. Res. Rec. (2103), 32–41.
Huang, H., Zhou, H., Wang, J., Chang, F., Ma, M., 2017. A multivariate spatial model of crash frequency by transportation modes for urban intersections. Anal. Methods Accid. Res. 14, 10–21.
Islam, M.T., El-Basyouny, K., Ibrahim, S.E., Sayed, K., 2016. Before–After safety evaluation using full bayesian macroscopic multivariate and spatial models. Transp. Res. Rec. 2601 (1), 128–137.
Ivan, J.N., Pasupathy, R.K., Ossenbruggen, P.J., 1999. Differences in causality factors for single and multi-vehicle crashes on two-lane roads. Accid. Anal. Prev. 31 (6), 695–704.
Jiang, X., Abdel-Aty, M., Alamili, S., 2014. Application of poisson random effect models for highway network screening. Accid. Anal. Prev. 63, 74–82.
Kilamanu, W., Xia, J.C., Caulfield, C., 2011. Analysis of spatial and temporal distribution of single and multiple vehicle crash in western australia: a comparison study. Proceedings of International Congress on Modelling and Simulation 683–689.
Kim, D.G., Washington, S., Oh, J., 2006. Modeling crash types: new sights into the effects of covariates on crashes at rural intersections. J. Transp. Eng. 4 (282), 282–292.
Kockelman, K.M., Kweon, Y.-J., 2002. Driver injury severity: An application of ordered probit models. Accid. Anal. Prev. 34 (3), 313–321.
Kononov, J., Bailey, B.A., Allery, B.K., 2008. Exploratory analysis of relationship between number of lanes and safety on urban freeways. Proceedings of the Transportation Research Board 87th Annual Meeting.
Lee, C., Abdel-Aty, M., 2008. Presence of passengers: does it increase or reduce driver's crash potential? Accid. Anal. Prev. 40, 1703–1712.
Lee, J., Abdel-Aty, M., Jiang, X., 2015. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. Accid. Anal. Prev. 78, 146–154.
Li, J., Wang, X., 2017. Safety analysis of urban arterials at the meso level. Accid. Anal. Prev. 108, 100–111.
Lord, D., Manar, A., Vizioli, A., 2005a. Modeling crash-flow-density and crash-flow-v/c ratio relationships for rural and urban freeway segments. Accid. Anal. Prev. 37 (1), 185–199.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transp. Res. Part A Policy Pract. 44 (5), 291–305.
Lord, D., Middleton, D., Whitacre, J., 2005b. Does separating trucks from other traffic improve overall safety? Transp. Res. Rec. (1922), 156–166.
Ma, J., Kockelman, K.M., 2006. Bayesian multivariate Poisson regression for models of injury count, by severity. Transp. Res. Rec. 1950 (1), 24–34.
Ma, X., Chen, S., Chen, F., 2016. Correlated random-effects bivariate poisson lognormal model to study single-vehicle and multivehicle crashes. J. Transp. Eng. 142 (11), 04016049.
Ma, X., Chen, S., Chen, F., 2017a. Multivariate space-time modeling of crash frequencies by injury severity levels. Anal. Methods Accid. Res. 15, 29–40.
Ma, Z., Zhang, H., Steven, I., Chien, J., Wang, J., Dong, C., 2017b. Predicting expressway crash frequency using a random effect negative binomial model: a case study in China. Accid. Anal. Prev. 98, 214–222.
Martensen, H., Dupont, E., 2013. Comparing single vehicle and multivehicle fatal road crashes: a joint analysis of road conditions, time variables and driver characteristics. Accid. Anal. Prev. 60, 466–471.
Mensah, A., Hauer, E., 1998. Two problems of averaging arising in the estimation of the relationship between accidents and traffic flow. Transp. Res. Rec. (1635), 37–43.
Miaou, S.P., Song, J.J., 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. Accid. Anal. Prev. 37 (4), 699–720.
Miranda-Moreno, L., Fu, L., Saccomanno, F., Labbe, A., 2005. Alternative risk models for ranking locations for safety improvement. Transp. Res. Rec. (1908), 1–8.
Montella, A., 2010. A comparative analysis of hotspot identification methods. Accid. Anal. Prev. 42 (2), 571–581.
New Mexico Department of Transportation (NMDOT), 2017. New Mexico Traffic Crash Annual Report 2015. NMDOT.
Öström, M., Eriksson, A., 1993. Single-vehicle crashes and alcohol: A retrospective study of passenger car fatalities in northern sweden. Accid. Anal. Prev. 25 (2), 171–176.
Pande, A., Abdel-Aty, M., 2009. A novel approach for analyzing severe crash patterns on multilane highways. Accid. Anal. Prev. 41 (5), 985–994.
Park, B.-J., Lord, D., Lee, C., 2014. Finite mixture modeling for vehicle crash data with application to hotspot identification. Accid. Anal. Prev. 71, 319–326.
Persaud, B., Lyon, C., Nguyen, T., 1999. Empirical bayes procedure for ranking sites for safety investigation by potential for safety improvement. Transp. Res. Rec. (1665), 7–12.
Persaud, B.N., Mucsi, K., 1995. Microscopic accident potential models for two-lane rural roads. Transp. Res. Rec. (1485), 134–139.

Roh, C.-G., Park, B.-J., Kim, J., 2017. Impact of heavy vehicles on highway traffic flows: case study in the seoul metropolitan area. J. Transp. Eng. Part A Syst. 143 (9), 05017008.
Savolainen, P., Mannering, F., 2007. Probabilistic models of motorcyclists' injury severities in single-and multi-vehicle crashes. Accid. Anal. Prev. 39 (5), 955–963.
Schneider, I., William, H., Stakleff, B., Buser, L., 2017. Alcohol-related Hot-spot Analysis and Prediction.
Song, J.J., Ghosh, M., Miaou, S., Mallick, B., 2006. Bayesian multivariate spatial models for roadway traffic crash mapping. J. Multivar. Anal. 97 (1), 246–273.
Spiegelhalter, D.J., Thomas, A., Best, N.G., Lunn, D., 2003. WinBUGS Version 1.4.1 User Manual. MRC Biostatistics Unit, Cambridge, UK.
Trafc Management Bureau of the Public Security Ministry of the People's Republic of China (TMBPSM), 2016. Annual Statistical Report on Roadway Traffic Accidents 2015. TMBPSM.
Usami, D.S., Giustiniani, G., Persia, L., Gigli, R., 2017. Aggregated analysis of in-depth accident causation data. Int. J. Inj. Contr. Saf. Promot. 24 (2), 165–173.
Wang, K., Ivan, J.N., Ravishanker, N., Jackson, E., 2017. Multivariate poisson lognormal modeling of crashes by type and severity on rural two lane highways. Accid. Anal. Prev. 99, 6–19.
Wang, X., Xie, K., Abdel-Aty, M., Chen, X., Tremont, P.J., 2014. Systematic approach to hazardous-intersection identification and countermeasure development. J. Transp. Eng. 140 (6), 04014022.
Wang, Y., Kockelman, K.M., 2013. A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. Accid. Anal. Prev. 60, 71–84.
Wen, H., Sun, J., Zeng, Q., Zhang, X., Yuan, Q., 2018. The effects of traffic composition on freeway crash frequency by injury severity: a Bayesian multivariate spatial modeling approach. J. Adv. Transp Article ID 6964828.
Xie, K., Wang, X., Ozbay, K., Yang, H., 2014. Crash frequency modeling for signalized intersections in a high-density urban road network. Anal. Methods Accid. Res. 2, 39–51.
Yu, R., Abdel-Aty, M., 2013. Multi-level bayesian analyses for single-and multi-vehicle freeway crashes. Accid. Anal. Prev. 58, 97–105.
Yu, R., Quddus, M., Wang, X., Yang, K., 2018. Impact of data aggregation approaches on the relationships between operating speed and traffic safety. Accid. Anal. Prev. 120, 304–310.