Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

CrossMark

Crash frequency modeling for signalized intersections in a high-density urban road network

Kun Xie [a,b,c,*], Xuesong Wang [d], Kaan Ozbay [a,b,c], Hong Yang [a,b,c]

[a] Department of Civil and Urban Engineering, New York University, Brooklyn, NY 11201, USA
[b] Center for Urban Science and Progress (CUSP), New York University, Brooklyn, NY 11201, USA
[c] Urban Mobility and Intelligent Transportation Systems (UrbanMITS) Laboratory, New York University, Brooklyn, NY 11201, USA
[d] Department of Transportation Engineering, Tongji University, Shanghai, China

ARTICLE INFO

ABSTRACT

Conventional crash frequency models rely on an assumption of independence among observed crashes. However, this assumption is frequently proved false by spatially related crash observations, particularly for intersection crashes observed in high-density road networks. Crash frequency models that ignore the hierarchy and spatial correlation of closely spaced intersections can lead to biased estimations. As a follow-up to our previous paper (Xie et al., 2013), this study aims to address this issue by introducing an improved crash frequency model. Data for 195 signalized intersections along 22 corridors in the urban areas of Shanghai was collected. Moran's I statistic of the crash data confirmed the spatial dependence of crash occurrence among the neighboring intersections. Moreover, Lagrange Multiplier test was performed and it suggested that the spatial dependence should be captured in the model error term. A hierarchical model incorporating a conditional autoregressive (CAR) effect term for the spatial correlation was developed in the Bayesian framework. A deviance information criterion (DIC) and cross-validation test were used for model selection and comparison. The results showed that the proposed model outperformed traditional models in terms of the overall goodness of fit and predictive performance. In addition, the significance of the corridor-specific random effect and CAR effect revealed strong evidence for the presence of heterogeneity across corridors and spatial correlation among intersections.

1. Introduction

Statistical modeling of the inter-relationship between crash frequencies and contributing factors associated with intersections is of great interest in intersection safety studies. Traditional crash frequency models mainly assume that crash observations are independent from each other. However, this assumption is often proved false by crash observations occurring at signalized intersections in high-density urban road networks. Take the urban areas of Shanghai as an example, where the average intersection spacing is only about 200 m. Such short spacing leads to potential, and sometimes unavoidable, dependence among crash observations at adjacent intersections. This dependence can be attributed to two factors: (a) hierarchy – signalized intersections located along the same corridor share similar traffic flow, geometric design

and land use, and thus characteristics of the intersection can be described by a two-level hierarchical data structure consisting of the intersection-level information and corridor-level information; and (b) spatial correlation – neighboring intersections typically interact with one another in operation, for example, signals of adjacent intersections are usually coordinated and queue spillovers may frequently occur in dense road networks. Modeling aggregated crash frequencies without properly addressing the issues of hierarchy or spatial correlation can lead to unreliable findings.

This study is a follow-up to our recent paper (Xie et al., 2013) which focuses on the analysis of contributing factors to crashes in view of the unique traffic characteristics of Chinese cities. Different from the previous one, the main objective of this study is to develop an improved crash frequency model for closely spaced signalized intersections by accounting for the hierarchy and spatial correlation of the crash observations.

2. Literature review

Poisson models (Jones et al., 1991; Miaou and Lum, 1993) and negative binomial (NB) models (Miaou, 1994; Poch and Mannering, 1996; Abdel-Aty and Radwan, 2000) have been widely used to capture the relationship between traffic crashes and contributing factors. It is widely recognized that Poisson models outperform the standard regression models in handling the nonnegative, random and discrete features of crash counts (Joshua and Garber, 1990; Maher and Summersgill, 1996). Despite the improved performance, however, the constraint of the mean being equal to the variance in Poisson models is often violated by over-dispersed crash data. Alternatively, NB models are used to accommodate this over-dispersion issue by incorporating an independently distributed error term. However, with the assumption of independent observations, neither the Poisson models nor the NB models address any inherent correlation of crash data.

To complement the Poisson models and NB models, random effect models have been proposed in previous studies to account for the potential heterogeneity across homogeneous groups (Shankar et al., 1998; Chin and Quddus, 2003; Wang et al., 2014). In addition, random parameter models which can be viewed as extensions of random effect models are developed to incorporate the variability of both the intercept and the variable coefficients across observations and thus provide more flexibility for handling the heterogeneity (Anastasopoulos and Mannering, 2009; El-Basyouny and Sayed, 2009a; Venkataraman et al., 2011, 2013, 2014). More recently, hierarchical models have become the preferred method to accommodate a multilevel data structure (Jones and Jorgensen, 2003; Lenguerrand et al., 2006; Kim et al., 2007; Huang and Abdel-Aty, 2010; Ahmed et al., 2011; Xie et al., 2013; Chen and Persaud, 2014). Hierarchical models can accommodate the heterogeneity among different groups and have the ability to incorporate variables at the specific levels where impacts of specific variables occur (Gelman and Hill, 2007).

Spatial dependence is primarily modeled in two ways: using a spatially lagged dependent variable and using an error term. (Anselin, 1988b). The former way is denoted the spatial lag specification which allows spatial dependence through both spatial spillover effects (observed variables at one location can affect the dependent variable of itself and its neighboring locations) and spatial error correlation effects (omitted variables at one location can affect the dependent variable of itself and its neighboring locations) (Narayanamoorthy et al., 2013; Chiou et al., 2014). The latter is referred to as the spatial error specification that assumes the spatial dependence is only due to spatial error correlation effects. To model spatially correlated observations, generalized estimating equations (GEEs) are often the go-to method (Lord and Persaud, 2000; Hutchings et al., 2003; Abdel-Aty and Wang, 2006; Wang and Abdel-Aty, 2006). However, GEEs have the limitation of setting the same correlation matrix for all intersection groups, and thus cannot reflect the discrepancies in correlations among different groups of intersections. Conditional autoregressive (CAR) models can provide more flexibility in specifying the magnitude of correlation and have been recommended in many recent studies (Song et al., 2006; Aguero-Valverde and Jovanils, 2008; El-Basyouny and Sayed, 2009b; Guo et al., 2010). The CAR models capture the spatial dependence using the spatial error specification (Narayanamoorthy et al., 2013). To address the hierarchy and spatial correlation of crash data, a hierarchical model incorporating CAR effect terms is proposed in this study.

3. Data description

For this study, a total of 195 signalized intersections located along 22 corridors was selected in the urban areas of Shanghai. These intersections were limited to either 3-legged or 4-legged designs so that the analysis of geometric features could be simplified. To ensure the independence of intersections across corridors, efforts were made to avoid choosing intersecting corridors.

Variables at both intersection and corridor levels were extracted from different sources. First, the crash data for the year of 2009 were geocoded on a GIS map using the crash location description, and then linked to selected intersections. Second, traffic volumes for each intersection were acquired from loop detectors. Third, the Sydney coordinated adaptive traffic system (SCATS) provided the geometric design and traffic control data. In addition, for each intersection, the presence or absence of an elevated road over the intersection was observed and distances among adjacent intersections were measured from the projected coordinates of intersections on the GIS map.

In addition, floating car data (FCD) was obtained from over 40,000 GPS equipped taxis in Shanghai, which provided a continuous position data of each taxi operating in the city every ten seconds. The locations of the taxis were matched with the GIS road network, and only taxi samples that passed through the study corridors were chosen. The mean speed and speed variance of each corridor were acquired from the FCD of taxis traveling along the corridor.

Table 1
Intersection related variables with their description and descriptive statistics.
Source: Xie et al. (2013).

Variables	Description	Min	Max	Mean	S.D.
Intersection type	0 for 4-leg, 1 for 3-leg	0	1	0.17	0.38
Minimum angle	Minimum angle of intersecting roadways	30.00	90.00	84.25	13.12
Maximum angle	Maximum angle of intersecting roadways	90.00	150.00	95.66	12.25
Total number of lanes	Total number of lanes for all entering approaches	3	25	10.74	4.34
Number of through lanes	Number of through lanes for all entering approaches	0	14	4.87	3.35
Number of right-turn lanes	Number of right-turn lanes for all entering approaches	0	6	1.12	1.34
Number of left-turn lanes	Number of left-turn lanes for all entering approaches	0	9	2.23	2.42
Number of through-right lanes	Number of through-right mixed use lanes for all entering approaches	0	4	1.24	1.21
Number of through-left lanes	Number of through-left mixed use lanes for all entering approaches	0	4	0.51	0.94
Ratio of turning lanes	The ratio of the total number of turning lanes to the total number of lanes	0	0.92	0.21	0.24
Under elevated roads or not	1 for intersection under elevated roads, 0 for not	0	1	0.23	0.42
Proximity of intersections	Distance to the nearest intersection along corridors (km)	0.09	1.17	0.35	0.23
Total number of phases	0 for 2–3 phases, 1 for 4–6 phases	0	1	0.69	0.46
Cycle length	Traffic signal cycle length (s)	40.00	240.00	171.11	42.23
Saturation degree	Saturation degree of an intersection	0.00	166.00	85.71	26.22
Intersection ADT	Average daily traffic entering the entire intersection (in 10^4 vehicles)	0.77	14.03	4.62	2.35
Intersection ADTPL	Average daily traffic per lane of entire intersection (in 10^4 vehicles)	0.10	1.36	0.51	0.22

Table 2
Corridor related variables with their description and descriptive statistics.
Source: Xie et al. (2013).

Variables	Description	Min	Max	Mean	S.D.
Presence of a median	0 for without median, 1 for with median	0	1	0.50	0.51
One-way road or not	0 for two-way road, 1 for one-way road	0	1	0.09	0.29
ADT along corridor	Average daily traffic volume along corridors in two directions (in 10^4 vehicles)	0.75	5.64	2.62	1.21
ADTPL along corridor	Average daily traffic volume per lane in two directions along corridors (in 10^4 vehicles)	0.21	0.82	0.54	0.23
Mean speed	Mean speed of taxi samples on each studied corridor (km/h)	10.78	42.01	22.61	6.94
Speed variance	Speed variance of taxi samples on each studied corridor	5.18	98.01	46.12	29.41

For a detailed description of the dataset, please refer to Xie et al. (2013). Listed in Tables 1 and 2 are the 17 intersection-level variables and 6 corridor-level variables, along with brief descriptions and statistics for each variable.

4. Diagnostic tests for spatial dependence

4.1. Moran's I test

Global Moran's I statistic proposed by Moran (1948) has been widely used to measure the spatial autocorrelation of observations (Goodchild, 1986; Anselin, 1988b; Getis and Ord, 1992). The Moran's I statistic is defined as

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} z_i z_j}{S_0 \sum_{i=1}^{n} z_i^2} \tag{1}$$

where z_i and z_j are deviations of observations i and j from the mean, w_{ij} represents the spatial weight between observations i and j, n is the total number of observations, and S_0 is the aggregation of spatial weights $\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$. The z-score of Moran's I can be computed by the following equation:

$$z_I = \frac{I - E[I]}{\sqrt{V[I]}} \tag{2}$$

where $E[I]$ and $V[I]$ are the expectation and variance of I

$$E[I] = -\frac{1}{n-1} \tag{3}$$

$$V[I] = E[I^2] - E[I]^2 \tag{4}$$

A statistically significant positive z-score suggests that the distribution of the observations are spatially clustered, whereas a negative z-score implies that the observations tend to be more dissimilar than one might expect. A z-score close

Table 3
Summary of Moran's *I* statistics.

Threshold distance (m)	*I*	*E*[*I*]	*V*[*I*]	*z*-Score	*p*-Value	Pseudo *p*-value
909	0.1182	−0.0052	0.0022	2.6046	0.0092	0.0086
1000	0.1152	−0.0052	0.0020	2.6715	0.0076	0.0078
1500	0.1099	−0.0052	0.0013	3.1372	0.0017	0.0027
2000	0.0928	−0.0052	0.0010	3.0796	0.0021	0.0029

Table 4
Summary of Lagrange Multiplier statistics.

Test	Specification	Degree of freedom	LM statistic	*p*-Value
Lagrange Multiplier	Spatial lag	1	18.2212	< 0.0001
	Spatial error	1	22.2308	< 0.0001
Adjusted Lagrange Multiplier	Spatial lag	1	0.3621	0.5474
	Spatial error	1	4.3709	0.0366

to zero indicates that observations are likely to be randomly and independently distributed in space (Goodchild, 1986). By assuming *z*-score is from a standard normal distribution, its associated *p*-value can be obtained.

If the *z*-score follows a skewed distribution instead of standard normal distribution, the *p*-value obtained can be unreliable. Anselin et al. (2006) suggested a more robust metric called pseudo p-value based on a random permutation test. The pseudo *p*-value is computed as $(M+1)/(S+1)$, where *M* is the number of instances with Moran's *I* equal to or greater than that of the observed data and *S* is the total number of permutations.

The spatial statistics tool in *ArcGIS* was used to test whether or not intersection crashes are spatially correlated (Mitchell, 2005). *GeoDa* was used to perform 9999 permutations to compute the pseudo *p*-value (Anselin, 2003). The spatial weight w_{ij} was defined by the inverse distance between intersections *i* and *j* if they were within the threshold distance. To perform the test, the minimum threshold distance should be long enough to ensure all intersections would have at least one neighbor (Mitchell, 2005). The minimum threshold distance (909 m) and additional longer threshold distances (1000 m, 1500 m and 2000 m) were used to test the spatial autocorrelation of crashes.

The summary of Moran's *I* statistics is reported in Table 3. According to Table 3, all the *z*-scores are significantly positive at the confidence level of 95% (*p*-values < 0.05). Moreover, the small pseudo *p*-values suggest highly significant Moran's *I* statistic. These results confirm the clustering characteristics of crashes at adjacent intersections. Ignoring the spatial correlation of crash data would reduce the precision of the estimates.

4.2. Lagrange Multiplier test

Spatial dependence can be caused by different kinds of spatial correlation effects (i.e., spatial error and spatial lag). Lagrange Multiplier (LM) test provides a good basis for the indication of selecting an appropriate spatial regression models (Anselin and Florax, 1995). The LM test was used by Burridge (1980) to diagnose spatial error dependence and later used for spatial lag dependence by Anselin (1988a). An adjusted LM test later proposed by Bera and Yoon (1993) has good finite sample properties and is proved to be more suitable for the identification of the source of dependence than the original LM test (Anselin et al., 1996). The results from the adjusted LM test are robust to the presence of a nuisance parameter representing spatial autocorrelation (Anselin et al., 1996).

GeoDa was used to perform LM and adjusted LM tests (Anselin, 2003). All estimations were conducted using the spatial weight matrix generated based on the minimum threshold distance. The summary of LM statistics is presented in Table 4. For both the spatial lag and spatial error specifications, their LM statistics are found to be highly significant with their *p*-values < 0.0001. When both LM statistics reject the null hypothesis, the adjusted LM statistics should be considered (Anselin, 2003). The adjusted LM statistic for the spatial error specification is found to be significant (*p*-value=0.0366) but not for the spatial lag specification (*p*-value=0.5474). The results suggest that a spatial error model should be developed to account for the spatial dependence of crash data.

5. Statistical modeling methodology

5.1. Model specification

Let Y_{ij} denote the crash frequency at *i*th intersection on *j*th corridor ($i=1,…,n_j$, n_j is the total number of intersections on *j*th corridor; $j=1,…J$, *J* is the total number of corridors) in a given time period. Assuming Y_{ij} follows Poisson distribution with

the mean λ_{ij}, the probability of observing y_{ij} crashes in the intersection can be given by

$$P(Y_{ij} = y_{ij}) = \frac{e^{-\lambda_{ij}}\lambda_{ij}^{y_{ij}}}{y_{ij}!} \tag{5}$$

To specify the Poisson parameter λ_{ij}, site-specific explanatory variables X_{pij} ($p=1,\ldots,P$, P is the total number of explanatory variables) are incorporated into the model

$$\ln(\lambda_{ij}) = \beta_0 + \sum_{p=1}^{P} \beta_p X_{pij} \tag{6}$$

where β_0 and β_p ($p=1,\ldots,P$) are the regression coefficients to be estimated. Eqs. (5) and (6) constitute the Poisson model which serves as a basis for modeling crash frequency. In this section, five types of crash frequency models derived from the Poisson model are proposed to identify the most appropriate one which can accommodate the specific data structure used in this study.

Model 1. Negative binomial (NB) model
To address the over-dispersion issue, the negative binomial model is obtained by introducing an error term ε_{ij} into Eq. (6)

$$\ln(\lambda_{ij}) = \beta_0 + \sum_{p=1}^{P} \beta_p X_{pij} + \varepsilon_{ij} \tag{7}$$

where $\exp(\varepsilon_{ij})$ is assumed to be gamma-distributed with mean 1 and variance α.
Model 2. Random effect negative binomial (RENB) model
Random effect models have the ability to deal with the hierarchy of data. To account for unobserved heterogeneity across corridors, a corridor-specific random effect term ν_j is incorporated into Eq. (7)

$$\ln(\lambda_{ij}) = \beta_0 + \sum_{p=1}^{P} \beta_p X_{pij} + \varepsilon_{ij} + \nu_j \tag{8}$$

where ν_j is a normally distributed term with mean 0 and variance σ_c^2. The variance component σ_c^2 represents the variation across corridors.
Model 3. Random parameter (RPNB) model
Random parameter models are capable of allowing some or all estimated parameters to vary across corridors. Random parameter models are equivalent to random effect models, if only the constant term is allowed to vary. The specification of the random parameter models is given in Eqs. (9) and (10) (Anastasopoulos and Mannering, 2009).

$$\ln(\lambda_{ij}) = \beta_{0j} + \sum_{p=1}^{P} \beta_{pj} X_{pij} + \varepsilon_{ij} \tag{9}$$

$$\beta_{pj} = \beta_p + \kappa_j \tag{10}$$

where κ_j is a normally distributed term with mean 0 and variance σ_c^2.
Model 4. Hierarchical negative binomial (HNB) model
As mentioned above, the data structure used in this study consists of a two-level hierarchy, namely the intersection level and the corridor level. In the hierarchical model proposed, parameters at the intersection level are expressed by probability models at the corridor level. The HNB model, which was also used in our previous study (Xie et al., 2013), can be specified as follows:
Intersection-level model:

$$\ln(\lambda_{ij}) = \beta_{0j} + \sum_{p=1}^{P} \beta_{pj} X_{pij} + \varepsilon_{ij} + \nu_j \tag{11}$$

Corridor-level model:

$$\beta_{0j} = \gamma_{00} + \sum_{q=1}^{Q} \gamma_{0q} Z_{qj} \tag{12}$$

$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$
$$\ldots$$
$$\beta_{Pj} = \gamma_{P0}$$

where Z_{qj} are the corridor-level variables ($q=1,\ldots,Q$, Q is the total number of corridor-level variables), γ_{00}, γ_{10}, γ_{20}, … γ_{P0} and γ_{0q} are the regression coefficients to be estimated. Like in Eq. (8), ν_j is the random effect at the corridor level where $\nu_j \sim N(0, \sigma_c^2)$.

Model 5. Hierarchical conditional autoregressive (HCAR) model

The Moran's *I* statistic confirmed the clustering characteristics of crashes in the neighboring intersections and results of the Lagrange Multiplier test indicated the spatial dependence should be captured using spatial error specification. The proposed hierarchical conditional autoregressive (HCAR) model cannot only account for the heterogeneity across different corridors but also can capture the spatial dependence among different intersections in the error term. The HCAR model assumes that within a corridor, intersections are spatially correlated and the correlation of intersections close to each other is expected to be greater than those far apart. We derive the HCAR model from the HNB model, with Eq. (11) modified as

$$\ln(\lambda_{ij}) = \beta_{0j} + \sum_{p=1}^{P} \beta_{pj} X_{pij} + \varepsilon_{ij} + \nu_j + \phi_{ij} \tag{13}$$

In Eq. (13), a conditional autoregressive (CAR) effect term ϕ_{ij} was introduced to capture the spatial correlation among intersections located along the same corridor. The joint probability distribution of ϕ under the Gaussian CAR model is as follows (Ghosh et al., 1999; Miaou and Song, 2005):

$$p(\phi_{11}, \phi_{12}, \dots \phi_{nJ}) \propto \exp\left[-\frac{1}{2\sigma_\phi^2} \Sigma_{ij \neq i'j'} w_{ij,i'j'} (\phi_{ij} - \rho \phi_{i'j'})^2 \right] \tag{14}$$

where ρ represents the direction and magnitude of the spatial autocorrelation, and σ_ϕ^2 is a parameter controlling the variance of the Gaussian CAR distribution. A proximity matrix W was used to describe the spatial relationship, and the correlation $w_{ij,i'j'}$ between intersections ij and $i'j'$ was defined as the inverse function of their distance $d_{ij,i'j'}$:

$$w_{ij,i'j'} = \begin{cases} \frac{1}{d_{ij,i'j'}} & \text{if } j = j' \\ 0 & \text{if } j \neq j' \end{cases} \tag{15}$$

When the corridor index $j \neq j'$, $w_{ij,i'j'} = 0$ represents that intersections located along different corridors are independent from each other. Based on Eq. (15), the full conditional distribution of ϕ_{ij} can be obtained as follows (Miaou and Song, 2005):

$$p(\phi_{ij}|\phi_{-ij}) \propto \exp\left[-\frac{w_{j+}}{2\sigma_\phi^2} \left(\phi_{ij} - \rho \Sigma_{i' \neq i} \frac{w_{i'j,ij}}{w_{j+}} \phi_{i'j} \right)^2 \right] \tag{16}$$

where ϕ_{-ij} is the set of $\phi_{i'j}$ for any $ij \neq i'j$ and $w_{j+} = \Sigma_{i'=1}^{n_j} w_{ij,i'j}$ representing the sum of weights on jth corridor.

An intrinsic version of the CAR model proposed by Besag et al. (1991) is adopted in this study. The parameter ρ is set to be 1 and the conditional distribution of ϕ_{ij} given ϕ_{-ij} turns out to be a normal distribution with mean $\Sigma_{i' \neq i} (w_{i'j,ij}/w_{j+}) \phi_{i'j}$ and variance σ_ϕ^2/w_{j+}

$$\phi_{ij}|\phi_{-ij} \sim N\left(\Sigma_{i' \neq i} \frac{w_{i'j,ij}}{w_{j+}} \phi_{i'j}, \frac{\sigma_\phi^2}{w_{j+}} \right) \tag{17}$$

In the Bayesian procedure, Eq. (17) is used to assign CAR priors to ϕ_{ij}. The Eqs. (5), (12), (13) and (17) represent the specification of the HCAR model.

5.2. Bayesian procedure

All model parameters were estimated using the Bayesian method that combines prior distributions with a likelihood function obtained from the observed data to estimate posterior distributions. Several researchers have shown the advantages of Bayesian methods in accommodating complex model structures when compared to classical statistical methods (Mitra and Washington, 2007; Persaud et al., 2010; Xie et al., 2013; Yang et al., 2013). Bayesian inference is usually implemented by a Markov Chain Monte Carlo (MCMC) algorithm (Gilks et al., 1998). MCMC is a classic method that utilizes independent and identically distributed simulations of a random process to approximate the desired distribution. The primary technique of MCMC is Gibbs sampling (Geman and Geman, 1984), each iteration of which draws a new value for each unobserved stochastic node from its full conditional distribution given the current values of all the other quantities in the model (Lunn et al., 2000). The WinBUGS statistical software package was used to provide a computing approach for the calibration of Bayesian models using Gibbs sampling (Spiegelhalter et al., 2002).

As mentioned above, the priors of the CAR effect term ϕ_{ij} are generated from Eq. (17). Without credible prior information for other parameters, uninformative priors were assumed. Uninformative priors express vague and general information about parameters. All regression coefficients were assumed with the Normal distribution $(0, 10^6)$; the variance of the randomly distributed terms σ_ε^2 and the variance of the Gaussian CAR distribution σ_ϕ^2 were assumed to follow the Inverse-Gamma distribution $(10^{-3}, 10^{-3})$. The logarithm of the variance of the NB error term ($\ln \alpha$) was assumed with Normal distribution $(0, 10^3)$. The Brooks–Gelman–Rubin (BGR) diagnostic proposed by Brooks and Gelman (1998) was used to assess the convergence of multiple chains. Convergence was assumed to occur when the BGR statistic is less than 1.2. Considering

convergence and time of updating, two MCMC chains of 150,000 iterations were run, and the first 50,000 samples were discarded as burn-in.

5.3. Model assessment

5.3.1. Deviance information criterion

The deviance information criterion (DIC) is widely used as a Bayesian measure of model fitting and complexity (Spiegelhalter et al., 2002). Specifically, DIC is calculated as follows:

$$\text{DIC} = \overline{D(\theta)} + p_D \tag{18}$$

where $D(\theta)$ is the Bayesian deviance of the estimated parameter θ. $\overline{D(\theta)}$ denotes the posterior mean of $D(\theta)$ and can be used to indicate how well the model fits the data. p_D defines the effective number of parameters and can be taken as a measure of model complexity. A difference in DIC that is larger than 5 suggests that the model with a smaller DIC should be favored.

5.3.2. Cross-validation test

To compare the predictive performance of different models, the cross-validation procedure proposed by Gelfand (1996) was used. A recent study by Yang et al. (2013) showed the practical use of such a procedure in traffic safety modeling. Unlike the original procedure that leaves out a single observation at each step, the modified procedure was designed to leave out all observations for each corridor at the same time at each step. A data set of n intersections was repeatedly split into a training set of size $n - n_j$ and a test set of size of n_j, where n_j denotes the number of intersections at jth corridor. Based on the remaining training data set, the conditional predictive ordinate (CPO) of each intersection in the test data set was calculated by using the following equation:

$$\text{CPO}_{ij} = f(y_{ij,obs} | Y_{-j,obs}, M_a) \tag{19}$$

where $Y_{-j,obs}$ means that all observations other than that of jth corridor are used to estimate a model M_a, and CPO_{ij} is the predictive density for observation $y_{ij,obs}$ for ith intersection at jth corridor by applying the model M_a. Using the λ_{ij} estimated from specific models, CPO_{ij} can be computed from Eq. (5). To compare the results of different models, the CPO ratio (CR) for intersection ij is calculated by using

$$\text{CR}_{ij} = \frac{f(y_{ij,obs} | Y_{-j,obs}, M_a)}{f(y_{ij,obs} | Y_{-j,obs}, M_b)} \tag{20}$$

$\text{CR}_{ij} > 1$ indicates that the observation $y_{ij,obs}$ supports model M_a, whereas $\text{CR}_{ij} > 1$ indicates that the observation $y_{ij,obs}$ supports model M_b. The model supported by more observations is preferred.

The CR for all the observations is the cumulative product of the CR_{ij} of all associated intersections. The cumulative CR (CCR) can be calculated by Eq. (21)

$$\text{CCR} = \prod_{i=1}^{n_j} \prod_{j=1}^{J} \text{CR}_{ij} \tag{21}$$

$\text{CCR} > 1$ indicates that overall the observations support model M_a, whereas $\text{CCR} < 1$ indicates the observations support model M_b.

6. Results and discussion

The NB, RENB, RPNB, HNB and HCAR models for intersection crash frequency were developed based on both the intersection and the corridor related variables listed in Tables 1 and 2. To conduct comparisons, all the explanatory variables included in the five models were kept the same. It should be noted that the HNB and HCAR models included corridor related variables into the corridor-level models, while the NB, RENB and RPNB treated corridor related variables in the same way as did the intersection related variables. In the RPNB model, if the estimated standard deviation of a parameter distribution was not statistically different from 0, the parameter was fixed across the corridors. Consequently, the intercept and the parameter of intersection ADT were allowed to vary randomly from one corridor to another in the RPNB model. Bayesian posterior estimations of the five models were presented in Table 5. The 95% Bayesian Credible Interval (95% BCI) was used to examine the significance of estimations. Estimations can be regarded as significant at the 95% level if the BCIs do not cover 0 and vice versa (Gelman, 2004). The ratio of turning lanes in the HCAR model was the only variable not found to be significant.

6.1. Discussion on explanatory variables

According to the variable coefficients of the HCAR model presented in Table 5, the intersections associated with higher crash frequencies were found to be under elevated roads, in close proximity to each other, with larger total number of phases, with higher ratios of turning lanes, with greater ADT, and along corridors with higher mean speeds. The proximity of intersections indicates the road network density. Box-plots of the intersection crash rate for the 22 corridors, ordered by

Table 5
Posterior summary of Bayesian model fitting.

Variable	NB		RENB		RPNB		HNB		HCAR	
	Mean	95% BCI	Mean	95% BCI	Mean	95% BCI	Mean	95% BCI	Mean	95% BCI
Intercept	1.163	(0.668,1.750)	1.232	(0.721,1.779)	1.527	(0.921,2.671)	1.347	(0.675,2.103)	0.720	(0.025,1.607)
Intersection level										
Under elevated roads or not	0.492	(0.155,0.848)	0.488	(0.051,0.981)	0.470	(0.066,0.819)	0.476	(0.076,0.873)	0.602	(0.023,1.232)
Proximity of intersections	−0.719	(−1.531,−0.049)	−0.722	(−1.478,−0.013)	−0.739	(−1.437,−0.005)	−0.700	(−1.503,−0.011)	−0.861	(−1.646,−0.045)
Ratio of turning lanes	1.000	(0.292,1.749)	0.906	(0.188,1.595)	0.968	(0.169,1.686)	0.902	(0.129,1.709)	0.720	(−0.009,1.591)
Total number of pases	0.368	(0.061,0.670)	0.386	(0.107,0.685)	0.369	(0.048,0.697)	0.398	(0.085,0.761)	0.448	(0.161,0.815)
Intersection ADT	0.054	(0.003,0.113)	0.051	(0.005,0.112)	0.060	(0.001,0.115)	0.057	(0.005,0.127)	0.055	(0.006,0.121)
(SD of parameter distribution)	–	–	–	–	0.040	(0.019,0.080)	–	–	–	–
Corridor level										
Mean speed	0.090	(0.067,0.110)	0.087	(0.061,0.112)	0.072	(0.022,0.100)	0.093	(0.059,0.116)	0.112	(0.072,0.148)
Dispersion α	0.743	(0.598,0.915)	0.663	(0.524,0.837)	0.670	(0.517,0.850)	0.656	(0.517,0.829)	0.668	(0.524,0.843)
Random effect σ_c^2	–	–	0.098	(0.018,0.253)	0.071	(0.010,0.231)	0.098	(0.012,0.261)	0.155	(0.011,0.581)
CAR effect var(ϕ_{ij})	–	–	–	–	–	–	–	–	0.498	(0.240,0.959)
Assessment										
DIC	1825		1817		1812		1810		1795	
CCR[a]	2.662		1.719		1.459		1.225		–	

[a] CCR is calculated using Eqs. (20) and (21) where the HCAR model is used as M_a while the other models are used as M_b. The model with lower CCR has better predictive performance.
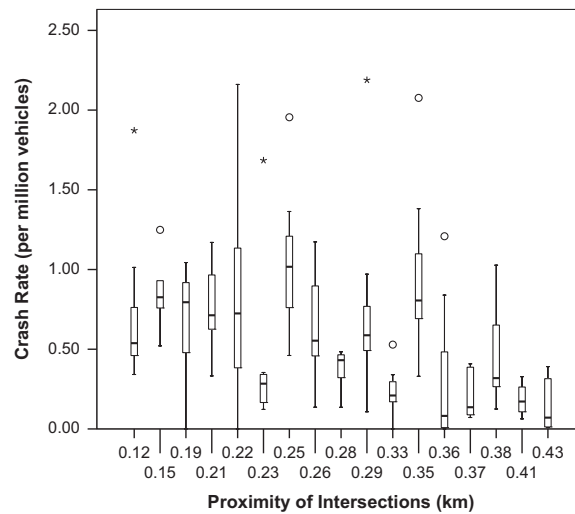
Fig. 1. The box-plot of crash rate for each corridor ordered by average proximity of intersections.

proximity of intersections, are presented in Fig. 1. In the box-plot, crash rate values more than three interquartile ranges from the end of a box are denoted with asterisks (∗), and values more than 1.5 interquartile ranges but less than 3 interquartile ranges are shown in circles (o). Overall, increases in the proximity of intersections are associated with decreases in the crash rates. Marginal effects represent an expected change in crash frequency given a unit change in the explanatory variables (Mitra and Washington, 2012). The estimated marginal effect of the proximity of intersections was −7.69 for the HCAR model with all other variables set at their means. This indicates that a one-kilometer increase in the distance between two intersections is expected to reduce the crash frequency by 7.69, annually. This appears to be counterintuitive for saturated arterials, where the traffic slows down due to the queue spillback between close intersections. When two intersections are separated by a larger distance, cars speed up and cause more crashes. The possible reason for this finding is that the shorter distance between intersections limits the gaps to implement safe lane changes, and ultimately results in more traffic conflicts. This finding of the impact of the distances between two intersections is important for designing safe road networks in urban areas.

According to the marginal effect of mean speed, one unit increase of mean speed is associated with an expected increase in crash frequency by 1.00, annually. Somewhat unexpectedly, we did not find speed variance to be significantly related to the frequency of crashes. The possible reason for this finding is that the speed features extracted from the taxi FCD data could be different from those of other vehicles due to the taxis' stop-and-go behavior as they pick up or drop off clients. This difference may introduce the issue of selectivity bias which can cause the inaccurate statistical inferences on speed features. The results regarding the speed features should thus be used with caution. Additionally, it should be noted that some factors affecting crash frequency such as bus traffic, pedestrian volume, median width and pavement conditions were not available and thus were unable to be included in the models. Due to these limitations of data, the number of explanatory variables that were found to be statistically significant is relatively small. The omission of other contributing factors may introduce unobserved heterogeneity and has risk of leading to biased estimations of explanatory variables in the models (Mannering and Bhat, 2014).

6.2. Discussion on model structure and performance

The estimated dispersion values (0.743 in the NB model, 0.663 in the RENB model, 0.670 in the RPNB model, 0.656 in the HNB model and 0.668 in the HCAR model) validate the assumption of over-dispersion. The random effect variances (0.098 in the RENB model, 0.071 in the RPNB model, 0.098 in the HNB model and 0.155 in the HCAR model) provide strong evidence for the presence of heterogeneity across corridors. The CAR effect variance (0.498 in the HCAR model) is found to be statistically significant and confirms the spatial correlation of intersections along the same corridor. Ignoring the spatial correlation would have resulted in the biased estimations of parameters in the models without the CAR effect term.

Table 5 presents the DIC values for the three Bayesian models. The DIC value of the HCAR model (1795) is 30, 22, 17 and 15 less than that of the NB model (1825), the RENB model (1817), the RPNB model (1812) and the HNB model (1810) respectively. The HCAR model is shown to be superior to the other four models in terms of the overall goodness of fit with the lowest DIC value. By including the random effect term, the RENB model is found to outperform to the NB model. The model performance is further improved in RPNB by allowing the effects of intersection ADT to vary across corridors. The HNB model has a DIC value close to that of the RPNB model and shows substantial improvement compared to the NB and RENB models. It should be noted that the only difference between the RENB and HNB models is that the HNB model includes the variable mean speed into the corridor-level model instead of mixing it with other intersection related variables. The HNB

model shows its advantage over the RENB model by incorporating the mean speed at the level where its actual impact occurs. The HCAR model yields an improved result compared with the HNB model by incorporating the CAR effect term. The DIC value is reduced significantly by 15 (from 1810 for the HNB model to 1795 for the HCAR model) when the spatial correlation among intersections is properly addressed.

A recent study by Geedipally et al. (2014) suggests not using DIC as the sole model selection criterion if the likelihood does not remain the same across all the models under consideration. Therefore, the cross-validation test described in the previous section was conducted to further compare the HCAR model with the other four crash frequency models alternatively. By applying Eqs. (20) and (21), the CR for each intersection and the CCR for all the observations were estimated. The CCRs of the HCAR model versus the NB, RENB, RPNB, HNB are reported in Table 5. Referring that CCRs of the NB model (2.662), the RENB model (1.719), the RPNB model (1.459) and the HNB model (1.225) are all greater than 1, the HCAR model is confirmed to have better predictive performance than any other models. Moreover, according to CRs of the HCAR model versus the NB model shown in Fig. 2a, the HCAR model is preferred in 166 (denoted with green circles) out of the 195 cases, whereas the NB model is supported by only 29 intersection observations (denoted with red triangles). Similarly, as shown in Fig. 2b–d, the HCAR model performs better than the RENB, RPNB and HNB models because 149, 109 and 103 out of 195 CRs respectively are greater than 1.

The aforementioned Moran's I statistic confirms the necessity of considering spatial dependence of crash data and the results from the Lagrange Multiplier test advocates the needs of developing a model with a spatial error specification. The proposed HCAR model can capture the spatial dependence in the CAR effect term ϕ_{ij}. As discussed in the model specification, the conditional distribution of ϕ_{ij} is assumed to follow a normal distribution with the weighted average of CAR effects of neighboring intersections as its mean. The advantage of including CAR effects over random effects is that the
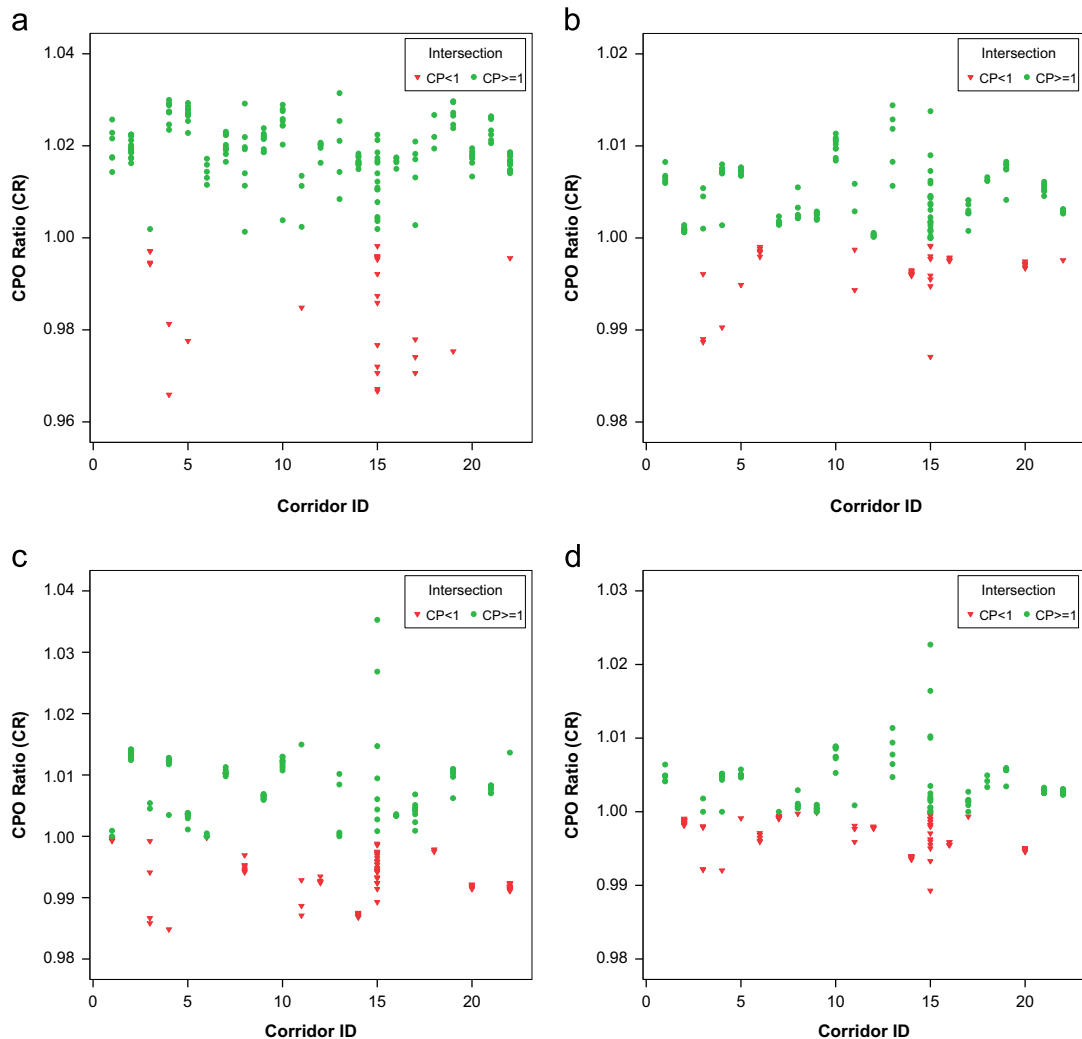

Fig. 2. Model comparison based on CPO ratio (CR), (a) HCAR vs NB, (b) HCAR vs RENB, (c) HCAR vs RPNB and (d) HCAR vs HNB.

distribution mean can vary from one entity to another and is related to the weight matrix which can be specified flexibly. In the HCAR model, the variation of the crash frequency is accommodated by the combination of dispersion parameter, the corridor-specific random effect term and the CAR effect term. According to the DIC and cross-validation test, the HCAR model performs better than other models in terms of the overall goodness of fit and predictive performance, by taking corridor-level heterogeneity and distance-related spatial dependence into account.

7. Conclusions

As a follow-up to our previous paper (Xie et al., 2013), this study proposed an advanced crash frequency model that can account for the hierarchy and spatial correlation of crash data of signalized intersections in a high-density urban road network. The proposed model can serve as a useful complement to the previous safety analysis methods in terms of the clustering characteristics of crash data in high-density road networks.

Closely spaced signalized intersections in the urban areas of Shanghai were used as a case study. Detailed crash data as well as related traffic and geometric data were collected. The Moran's *I* statistic was computed for crash observations and the results confirmed the clustering characteristics of crashes in the neighboring intersections. This finding indicates the necessity of accounting for spatial effects when modeling crash observations. Moreover, we conducted the Lagrange Multiplier tests and found that the spatial dependence should be captured in the model error term. To address the spatial correlations of crashes among different intersections, a hierarchical conditional autoregressive (HCAR) model incorporating both intersection-level and corridor-level variables and CAR effects was developed in the Bayesian framework. To examine the performance of the proposed model, a negative binomial (NB) model, a random effect negative binomial (RENB) model and a random parameter negative binomial model (RPNB) and a hierarchical negative binomial (HNB) model were also estimated for comparison purposes.

The significance of the corridor-specific random effects in the RENB, RPNB, HNB and HCAR models confirmed the hierarchy of the crash data. In addition, the significance of the CAR effects in the HCAR model indicated that intersections along the same corridor were spatially correlated. These findings affirm the needs to capture the unique structure of crash data for intersections in high-density networks. Besides, the safety effect of intersection proximity as associated with road network density was investigated. A negative relation was found between the crash frequency and the distance of two adjacent intersections. This finding is important for planning, designing, and managing high-density urban road networks.

The deviance information criterion (DIC) and cross-validation test were used for model comparison. Considering the drawbacks of using DIC as the sole model criterion (Geedipally et al., 2014), we emphasized the application of less frequently used cross-validation test to measure the predictive performance of models. Overall, the proposed HCAR model was found to outperform the NB, RENB, RPNB and HNB models. These results should be attributed to the two advantages of the HCAR model: first, the HCAR model is able to construct multilevel (i.e., intersection-level and corridor-level) model structure and to account for the safety effects of intersection related and corridor related variables at specific levels of models; second, the HCAR model can capture the spatial dependence of intersection crashes by specifying the correlation matrix for intersections along each corridor. However, an advantage of the RPNB model over the HCAR model is that it can account for the unobserved heterogeneity and thus the effects of explanatory variables can be adjusted. The preference of the HCAR and RPNB models can be related to the site specific characteristics of the data used in this study. To further understand the performance of those models, additional studies based on other datasets are suggested.

Despite the performance of the proposed approach, we should mention the limitation of missing other potential explanatory variables due to their unavailability in this study. This may alter the estimations of certain parameters and result in unreliable inferences. To further improve the performance of the HCAR model, additional procedures, for instance, incorporation of random parameters, are suggested to further account for the unobserved heterogeneity. Moreover, the temporal correlations and correlations among different crash types at the same site can be considered simultaneously as well as spatial correlations among different sites.

References

Abdel-Aty, M., Wang, X., 2006. Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors. Transportation Research Record 1953, 98–111.
Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention 32 (5), 633–642.
Aguero-Valverde, J., Jovanils, P.P., 2008. Analysis of road crash frequency with spatial models. Transportation Research Record 2061, 55–63.

Ahmed, M., Huang, H., Abdel-Aty, M., Guevara, B., 2011. Exploring a bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. Accident Analysis and Prevention 43 (4), 1581–1589.

Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis and Prevention 41 (1), 153–159.

Anselin, L., 1988a. Lagrange Multiplier test diagnostics for spatial dependence and spatial heterogeneity. Geographical Analysis 20 (1), 1–17.

Anselin, L., 1988b. Spatial Econometrics: Methods and Models. Kluwer Academic, Dordrecht.

Anselin, L., 2003. Geoda™ 0.9 user's guide. Urbana 51, 61801.

Anselin, L., Bera, A.K., Florax, R., Yoon, M.J., 1996. Simple diagnostic tests for spatial dependence. Regional Science and Urban Economics 26 (1), 77–104.

Anselin, L., Florax, R.J., 1995. New Directions in Spatial Econometrics. Springer Verlag, Berlin.

Anselin, L., Syabri, I., Kho, Y., 2006. Geoda: an introduction to spatial data analysis. Geographical Analysis 38 (1), 5–22.

Bera, A.K., Yoon, M.J., 1993. Specification testing with locally misspecified alternatives. Econometric Theory 9 (4), 649–658.

Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics 43 (1), 1–20.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7 (4), 434–455.

Burridge, P., 1980. On the cliff-ord test for spatial correlation. Journal of the Royal Statistical Society. Series B (Methodological), 107–108.

Chen, Y., Persaud, B., 2014. Methodology to develop crash modification functions for road safety treatments with fully specified and hierarchical models. Accident Analysis & Prevention 70, 131–139.

Chin, H.C., Quddus, M.A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accident Analysis and Prevention 35 (2), 253–259.

Chiou, Y.-C., Fu, C., Chih-Wei, H., 2014. Incorporating spatial dependence in simultaneously modeling crash frequency and severity. Analytic Methods in Accident Research 2, 1–11.

El-Basyouny, K., Sayed, T., 2009a. Accident prediction models with random corridor parameters. Accident Analysis and Prevention 41 (5), 1118–1123.

El-Basyouny, K., Sayed, T., 2009b. Urban arterial accident prediction models with spatial effects. Transportation Research Record 2102, 27–33.

Geedipally, S.R., Lord, D., Dhavala, S.S., 2014. A caution about using deviance information criterion while modeling traffic crashes. Safety Science 62, 495–498.

Gelfand, A.E., 1996. Model determination using sampling based methods. In: Gilks, W., Richardson, S., Spiegelhalter, D. (Eds.), Markov Chain Monte Carlo in Practice, Chapman and Hall, London, pp. 145–161.

Gelman, A., 2004. Bayesian Data Analysis, 2nd ed. Chapman & Hall/CRC, Boca Raton, Fla.

Gelman, A., Hill, J., 2007. Data Analysis using Regression and Multilevel/Hierarchical Models. Cambridge University Press, Cambridge; New York.

Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741.

Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geographical Analysis 24 (3), 189–206.

Ghosh, M., Natarajan, K., Waller, L.A., Kim, D., 1999. Hierarchical bayes glms for the analysis of spatial data: an application to disease mapping. Journal of Statistical Planning and Inference 75 (2), 305–318.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1998. Markov Chain Monte Carlo in Practice. Chapman & Hall, Boca Raton, Fla.

Goodchild, M.F., 1986. Spatial Autocorrelation. Geo Books, Norwich.

Guo, F., Wang, X.S., Abdel-Aty, M.A., 2010. Modeling signalized intersection safety with corridor-level spatial correlations. Accident Analysis and Prevention 42 (1), 84–92.

Huang, H., Abdel-Aty, M., 2010. Multilevel data and bayesian analysis in traffic safety. Accident Analysis and Prevention 42 (6), 1556–1565.

Hutchings, C.B., Knight, S., Reading, J.C., 2003. The use of generalized estimating equations in the analysis of motor vehicle crash data. Accident Analysis and Prevention 35 (1), 3–8.

Jones, A.P., Jorgensen, S.H., 2003. The use of multilevel models for the prediction of road accident outcomes. Accident Analysis and Prevention 35 (1), 59–69.

Jones, B., Janssen, L., Mannering, F., 1991. Analysis of the frequency and duration of freeway accidents in seattle. Accident Analysis and Prevention 23 (4), 239–255.

Joshua, S.C., Garber, N.J., 1990. Estimating truck accident rate and involvements using linear and poisson regression models. Transportation Planning and Technology 15 (1), 41–58.

Kim, D.G., Lee, Y., Washington, S., Choi, K., 2007. Modeling crash outcome probabilities at rural intersections: application of hierarchical binomial logistic models. Accident Analysis and Prevention 39 (1), 125–134.

Lenguerrand, E., Martin, J.L., Laumon, B., 2006. Modelling the hierarchical structure of road crash data – application to severity analysis. Accident Analysis and Prevention 38 (1), 43–53.

Lord, D., Persaud, B.N., 2000. Accident prediction models with and without trend: application of the generalized estimating equations procedure. Transportation Research Record 1717, 102–108.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. Winbugs—a bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10 (4), 325–337.

Maher, M.J., Summersgill, I., 1996. A comprehensive methodology for the fitting of predictive accident models. Accident Analysis and Prevention 28 (3), 281–296.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: methodological frontier and future directions. Analytic Methods in Accident Research 1, 1–22.

Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road sections – poisson versus negative binomial regressions. Accident Analysis and Prevention 26 (4), 471–482.

Miaou, S.P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accident Analysis and Prevention 25 (6), 689–709.

Miaou, S.P., Song, J.J., 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. Accident Analysis and Prevention 37 (4), 699–720.

Mitchell, A., 2005. The Esri Guide to Gis Analysis, Volume 2: Spatial Mesurements and Statistics. ESRI Press.

Mitra, S., Washington, S., 2007. On the nature of over-dispersion in motor vehicle crash prediction models. Accident Analysis and Prevention 39 (3), 459–468.

Mitra, S., Washington, S., 2012. On the significance of omitted variables in intersection crash modeling. Accident Analysis and Prevention 49, 439–448.

Moran, P.A., 1948. The interpretation of statistical maps. Journal of the Royal Statistical Society. Series B (Methodological) 10 (2), 243–251.

Narayanamoorthy, S., Paleti, R., Bhat, C.R., 2013. On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. Transportation Research Part B 55, 245–264.

Persaud, B., Lan, B., Lyon, C., Bhim, R., 2010. Comparison of empirical bayes and full bayes approaches for before-after road safety evaluations. Accident Analysis and Prevention 42 (1), 38–43.

Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection-accident frequencies. Journal of Transportation Engineering 122 (2), 105–113.

Shankar, V., Albin, R., Milton, J., Mannering, F., 1998. Evaluating median crossover likelihoods with clustered accident counts: an empirical inquiry using the random effects negative binomial model. Transportation Research Record 1635, 44–48.

Song, J.J., Ghosh, A., Miaou, S., Mallick, B., 2006. Bayesian multivariate spatial models for roadway traffic crash mapping. Journal of Multivariate Analysis 97 (1), 246–273.

Spiegelhalter, D.J., Best, N.G., Carlin, B.R., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B—Statistical Methodology 64, 583–616.

Venkataraman, N., Shankar, V., Ulfarsson, G.F., Deptuch, D., 2014. A heterogeneity-in-means count model for evaluating the effects of interchange type on heterogeneous influences of interstate geometrics on crash frequencies. Analytic Methods in Accident Research 2, 12–20.

Venkataraman, N., Ulfarsson, G.F., Shankar, V.N., 2013. Random parameter models of interstate crash frequencies by severity, number of vehicles involved, collision and location type. Accident Analysis and Prevention 59, 309–318.

Venkataraman, N.S., Ulfarsson, G.F., Shankar, V., Oh, J., Park, M., 2011. Model of relationship between interstate crash occurrence and geometrics. Transportation Research Record 2236, 41–48.

Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis and Prevention 38 (6), 1137–1150.

Wang, X., Xie, K., Abdel-Aty, M., Tremont, P.J., Chen, X., 2014. Systematic approach for hazardous-intersection identification and countermeasure development. Journal of Transportation Engineering 140 (6), 04014022.

Xie, K., Wang, X., Huang, H., Chen, X., 2013. Corridor-level signalized intersection safety analysis in Shanghai, China using Bayesian hierarchical models. Accident Analysis and Prevention 50, 25–33.

Yang, H., Ozbay, K., Ozturk, O., Yildirimoglu, M., 2013. Modeling work zone crash frequency by quantifying measurement errors in work zone length. Accident Analysis and Prevention 55, 192–201.