Accident Analysis and Prevention 122 (2019) 189–198

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

A multivariate spatial approach to model crash counts by injury severity

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

[a] Department of Civil and Natural Resources Engineering, University of Canterbury, 20 Kirkwood Ave, Christchurch, 8041, New Zealand
[b] Department of Civil & Urban Engineering, Center for Urban Science and Progress (CUSP), C2SMART Center, New York University (NYU), 6 MetroTech Center, 4th Floor, Brooklyn, NY, 11201, USA
[c] Department of Modeling, Simulation & Visualization Engineering, Old Dominion University (ODU), 4700 Elkhorn Ave, Norfolk, VA, 23529, USA

ARTICLE INFO

Keywords:
Safety analysis
Spatial statistics
Multivariate conditional autoregressive model
Bayesian method

ABSTRACT

Conventional safety models rely on the assumption of independence of crash data, which is frequently violated. This study develops a novel multivariate conditional autoregressive (MVCAR) model to account for the spatial autocorrelation of neighboring sites and the inherent correlation across different crash types. Manhattan, which is the most densely populated urban area of New York City, is used as the study area. Census tracts are used as the basic geographic units to capture crash, transportation, land use, and demo-economic data. The specification of the proposed multivariate model allows for jointly modeling counts of various crash types that are classified according to injury severity. Results of Moran's *I* tests show the ability of the MVCAR model to capture the multivariate spatial autocorrelation among different crash types. The MVCAR model is found to outperform the others by presenting the lowest deviance information criterion (DIC) value. It is also found that the unobserved heterogeneity was mostly attributed to spatial factors instead of non-spatial ones and there is a strong shared geographical pattern of risk among different crash types.

1. Introduction

Crash frequency models, which are commonly developed to correlate traffic, geometric, and environmental characteristics with crash occurrence, are helpful in investigating risk factors, conducting before-after evaluation, and identifying crash hotspots. Poisson models (Jones et al., 1991; Miaou and Lum, 1993), Poisson-Gamma models (Poch and Mannering, 1996; Abdel-Aty and Radwan, 2000), and Poisson-lognormal models (Lord and Miranda-Moreno, 2008; El-Basyouny and Sayed, 2009a) are widely used in crash frequency modeling. Those crash frequency models rely on the assumption of independence of observed crashes. However, the independence assumption of different crashes is frequently violated. Firstly, spatial autocorrelation can exist in neighboring sites; namely, that observed or unobserved risk factors affect the likelihood of crashes at one site and its surrounding sites. For instance, crashes occurred at the downstream of a highway may lead to rear-end secondary crashes at the upstream due to the disrupted traffic (Yang et al., 2014, 2018). Also, traffic congestion in the central areas can result in traffic shifting to surrounding areas and thus change the likelihood of crashes not only in the central areas but also in the surrounding areas. Secondly, when modeling crashes of different types, there can be inherent correlation across crash types due to the existence of unobserved risk factors that are jointly associated with the frequencies of various crash types. For instance, poor road lighting would increase non-injury crashes as well as injury crashes, but data availability may limit the inclusion of road lighting as an explanatory variable and thus its effect on safety cannot be accounted for. Modeling crash frequencies without properly considering the intrinsic dependence of crash data can lead to biased inferences.

The main objective of this study is to develop a crash frequency model with multivariate responses that can jointly account for the spatial autocorrelation and inherent correlation among crash types. Manhattan, which is the most densely populated urban area of New York City, is used as the study area. This paper starts with introduction and literature review. It is followed by the section of data preparation and methodology where a novel multivariate spatial model is introduced to model multiple crash types classified by injury severity. Detailed discussion on addressing multivariate spatial autocorrelation is also presented. This paper ends with the summary and conclusions.

2. Literature review

To account for spatial autocorrelation of crash observations, generalized estimating equations (GEEs) have been frequently sought (Abdel-Aty and Wang, 2006). However, GEEs have the constraint of requiring the same correlation matrix for all groups (Xie et al., 2014).

* Corresponding author.
E-mail addresses: kun.xie@canterbury.ac.nz (K. Xie), kaan.ozbay@nyu.edu (K. Ozbay), hyang@odu.edu (H. Yang).

https://doi.org/10.1016/j.aap.2018.10.009
Received 21 January 2017; Received in revised form 15 September 2018; Accepted 16 October 2018
Available online 31 October 2018
0001-4575/

Alternatively, simultaneous autoregressive (SAR) models[1] developed by Whittle (1954) and conditional autoregressive (CAR) models by Besag (1974) have drawn increasing attention due to their flexibility in accounting for the spatial autocorrelation by specifying details of correlation matrices. The SAR models have two typical spatial specifications: 1) the spatial error specification that assumes the spatial autocorrelation is only due to spatial error correlation effects (unobserved risk factors at one site can affect the crash observations of itself and its neighboring sites), and 2) the spatial lag specification that allows spatial autocorrelation through both spatial error correlation effects and spatial spillover effects (observed risk factors at one site can affect the crash observations of itself and its neighboring sites) (Narayanamoorthy et al., 2013; Xie et al., 2014). SAR models have been applied by transportation researchers to anticipate land-use change (Wang et al., 2014b), to estimate incident durations (Xie et al., 2015a) and to conduct safety analysis (Quddus, 2008; Castro et al., 2012; Narayanamoorthy et al., 2013). The SAR models are generally estimated using the maximum likelihood method (Xie et al., 2015a) and the composite marginal likelihood (Castro et al., 2012; Narayanamoorthy et al., 2013). Likewise, the CAR models are mostly developed in the full Bayesian framework where the magnitude of correlation between observations can be specified with CAR priors. It should be noted that CAR models cannot accommodate spatial spillover effects but only spatial error correlation effects (Narayanamoorthy et al., 2013). CAR models have been used to analyze safety performance of various entities such as intersections (Xie et al., 2014), arterials (El-Basyouny and Sayed, 2009b), and census block groups (Saha et al., 2018). The CAR models are more commonly used to accommodate spatially correlated count data, whereas it is challenging to use the SAR models in a count-response setting, especially with large datasets (Wang and Kockelman, 2013). Moreover, the Bayesian framework of the CAR model enables a flexible selection of crash count distributions and can accommodate complicated model structures (Lan et al., 2009; Xie et al., 2013). Therefore, the CAR specification is adopted to analyze crash count data in this study.

When modeling crash frequencies of different types, it is likely that unobserved risk factors can affect all crash types simultaneously at each site (Lord and Mannering, 2010). Multivariate models (Ma and Kockelman, 2006; Park and Lord, 2007; Xie et al., 2015b) can address correlation among different crash types by incorporating shared error terms. Xie et al. (2015b) developed a multivariate model in Bayesian framework and showed the flexibility to specify the Bayesian multivariate model for accommodating complicated data structure. Copula-based approaches (Nashad et al., 2016) and fractional split modeling approaches (Yasmin et al., 2016) have also been used in the literature to account for correlation across crash types. Multivariate conditional autoregressive (MVCAR) models have been proposed (Song et al., 2006; Wang and Kockelman, 2013; Barua et al., 2014; Cheng et al., 2018) to jointly account for the spatial autocorrelation among neighboring sites and correlation among crash types, by including a multivariate conditional autoregressive effect term. Barua et al. (2016) extended the MVCAR model by allowing all the regression coefficients to vary randomly across sites and thus potential spatial heterogeneity can be addressed. The present study exploits the MVCAR approach to model crash counts by injury severity using data from Manhattan, New York. It is worth to mention that we also attempted to develop a MVCAR model with random parameters (similar to the approach by Barua et al. (2016)), allowing a selection of regression coefficients to vary randomly. However, this model didn't yield distinct improvement compared with the MVCAR model and thus was not reported in this study.

[1] Also referred to as spatial autoregressive models in literature.

3. Data preparation

The census tracts (n = 282) of Manhattan were used as the basic geographical units for data preparation and safety modeling. The census tracts could be easily connected to the demo-economic data provided by the U.S. Census Bureau. Crash, transportation, and land use data were also collected for each census tract.

Three-year crash record data (05/01/2008-04/30/2011) were obtained from the New York State Department of Transportation (NYSDOT[2]). Crashes were classified into five types by injury severity, i.e., property-damage-only, possible injury, non-incapacitating injury, incapacitating injury, and fatality. Considering the similarity of crash injuries, these five crash types were combined into three categories for model development: (1) property-damage-only (PDO); (2) minor injury (MI) including possible injury and non-incapacitating injury crashes; and (3) serious injury and fatality (SIF) including incapacitating injury and fatality crashes. The crash counts by injury severity for the analyzed census tracts are shown in Fig. 1. We can see from Fig. 1 that the zones with similar colors tend to be close to each other, which shows the spatial clustering of each individual crash type. Additionally, it is found that the zones with higher PDO frequencies are likely to be close to zones with more MI and SIF crashes and it implies a potential multivariate spatial correlation of different crash types.

Traffic AADT data were obtained from the Short Count Program (SCP) of NYSDOT.[3] In the SCP, approximately 12,000 statewide counts of 2–7 days' duration were taken every year and were used to calculate AADT after undergoing quality control procedures. Vehicle miles traveled (VMT) was computed for each census tract based on the AADT data. The output of the Best Practice Model (BPM) developed by the New York Metropolitan Transportation Council (NYMTC[4]) was used to estimate the ratio of truck flow to total flow for each census tract. The geographic information system (GIS) data of bus and subway stations were obtained from the Metropolitan Transportation Authority (MTA[5]). The number of bus and subway stations were calculated for each census tract using spatial tools of the software ArcGIS. The intersection number, length of different road types (freeway, avenue, and street), and road density were computed based on the road network of the regional planning model.

The land use data were obtained from the New York City Department of City Planning (NYCDCP[6]). The main zoning categories of interest include commercial, residential, mixed and park. The areas by zoning category were obtained for each census tract using a Visual Basic for Applications (VBA) program developed in ArcGIS and then the ratio of each zoning category to the total zone area was computed.

The demo-economic data for the studied census tracts were obtained from the 2011 census data provided by the U.S. Census Bureau[7]. The main categories of demo-economic data include demographic (e.g., total population, population under 14, and population over 65), economic (e.g., unemployment rate and median income), housing (e.g., median value and household average size), and commuting (e.g., the ratios of commuters by driving alone, carpooling, public transit, and walking) data. The description and descriptive statistics of crash, transportation, land use and demo-economic data are summarized in Table 1.

[2] Source: http://www.dmv.ny.gov/stats.htm.
[3] Source: https://gis.ny.gov/gisdata.
[4] Source: http://www.nymtc.org/project/bpm/bpmindex.html.
[5] Source: http://web.mta.info/developers/download.html.
[6] Source: http://www.nyc.gov/html/dcp/html/bytes/dwn_pluto_mappluto.shtml.
[7] Source: http://factfinder.census.gov.

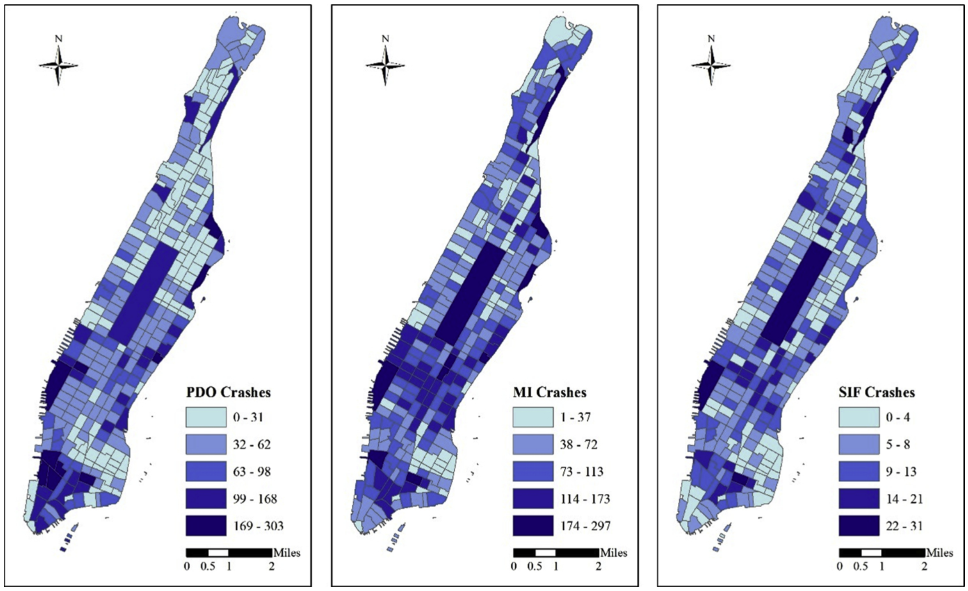

Fig. 1. Counts of PDO, MI and SIF crashes (05/01/2008-04/30/2011).

Table 1
Description and descriptive statistics of data (n = 282 census tracts).

Variable	Description	Mean	Standard Deviation
Crash			
PDO	Count of property-damage-only crashes	50.95	46.54
MI	Count of possible injury and non-incapacitating injury crashes	74.92	49.11
SIF	Count of incapacitating injury and fatality crashes	7.08	5.21
Transportation			
LogVMT	The logarithm of annual average daily vehicle miles traveled (veh.mile)	9.56	1.69
Truck ratio	The average ratio of truck flow to total flow	0.06	0.04
Freeway length	Total length of freeway segments (mile)	0.14	0.28
Avenue length	Total length of avenues (mile)	0.39	0.27
Street length	Total length of streets (mile)	0.27	0.34
Road density	Ratio of total road length to census tract area (mile/mile^2)	1.83	0.85
Intersection number	Count of intersection in each census tract	11.86	6.90
Subway station number	Count of subway stations in each census tract	0.52	0.92
Bus stop number	Count of bus stops in each census tract	8.09	5.92
Land use			
Commercial ratio	The ratio of commercial zone area to the whole area	0.30	0.33
Residential ratio	The ratio of residential zone area to the whole area	0.56	0.35
Mixed ratio	The ratio of mixed zone area to the whole area	0.07	0.17
Park ratio	The ratio of park area to the whole area	0.06	0.14
Demo-economic			
Population	Total population (10^3)	5.56	3.14
Population under 14	Population under 14 years (10^3)	0.70	0.54
Population over 65	Population 65 years and over (10^3)	0.74	0.56
Median age	Median age of population	37.36	7.55
Median income	Median income per household (10^3 $)	76.41	44.49
Median housing value	Median value of housing (10^3 $)	615.61	235.36
Household size	Number of people per household	2.08	0.57
Unemployment rate	Share of the labor force that is unemployed	0.09	0.05
Drive alone ratio	The ratio of commuters by driving alone	0.07	0.04
Carpool ratio	The ratio of commuters by carpooling	0.02	0.02
Public transit ratio	The ratio of commuters by public transit	0.56	0.17
Walk ratio	The ratio of commuters by walking	0.22	0.14

4. Methodology

4.1. Model specification

4.1.1. Model 1. Poisson-Gamma (PG) model

Let y_i^k denote the count of k^{th} type crashes at i^{th} sites ($i = 1, 2, ..., n$, n is the total number of sites; $k = 1, 2, ..., K$; and K is the total number of crash types) during the study period. It is commonly assumed that y_i^k follows Poisson distribution with the mean λ_i^k. The probability of observing y_i^k crashes at a site can be given by:

$$P(y_i^k|\lambda_i^k) = \frac{e^{-\lambda_i^k}\lambda_i^{k y_i^k}}{y_i^k!} \tag{1}$$

To specify the Poisson parameter λ_i^k, explanatory variables X_{pi}^k ($p = 1, ..., P_k$, where, P_k is the total number of explanatory variables for k^{th} crash type) are incorporated into the model:

$$\ln(\lambda_i^k) = \beta_0^k + \sum_{p=1}^{P_k} \beta_p^k X_{pi}^k + \varepsilon_i^k \tag{2}$$

where β_0^k and β_p^k are the regression coefficients to be estimated. Error term ε_i^k is included to address the over-dispersion issue, with $\exp(\varepsilon_i^k)$ assumed to be gamma-distributed with mean 1 and variance α_k^2. EqS. (1) and (2) constitute K independent Poisson-Gamma models that serves as a basis for modeling crash frequency.

4.1.2. Model 2. Univariate conditional autoregressive (UCAR) model

To capture the spatial autocorrelation of specific crash types, univariate conditional autoregressive (UCAR) model is given as follows:

$$\ln(\lambda_i^k) = \beta_0^k + \sum_{p=1}^{P_k} \beta_p^k X_{pi}^k + \varepsilon_i^k + S_i^k \tag{3}$$

where, S_i^k is a CAR effect term to account for the spatial autocorrelation of k^{th} crash type. It should be noted that the CAR effects for different crash types are assumed to be independent from each other. An intrinsic version of the CAR effect proposed by Besag et al. (1991) is used in this study. The full conditional distribution of S_i^k given S_{-i}^k turns out to be a normal distribution with mean $\sum_{j\neq i} \frac{\mathrm{w_{ij}}}{\mathrm{w}_{i+}} S_j^k$ and variance $\frac{\sigma_{Sk}^2}{\mathrm{w}_{i+}}$:

$$S_i^k | S_{-i}^k \sim \mathrm{N}\left(\sum_{j\neq i} \frac{\mathrm{w_{ij}}}{\mathrm{w}_{i+}} S_j^k, \frac{\sigma_{Sk}^2}{\mathrm{w}_{i+}}\right) \tag{4}$$

where, S_{-i}^k is the set of S_j^k for any $j \neq i$. w_{ij} indicates the spatial autocorrelation between sites i and j, with $\mathrm{w}_{ij} = 1$ if sites i and j are adjacent, and $\mathrm{w}_{ij} = 0$ otherwise. w_{i+} is the aggregation of weights for site i, with $\mathrm{w}_{i+} = \sum_{j=1}^{n} w_{ij}$. σ_{Sk}^2 is a parameter controlling the variance for spatial autocorrelation of k^{th} crash type. During the Bayesian procedure, Eq. (4) is used to assign CAR priors to S_i^k for k^{th} crash type independently.

4.1.3. Model 3. Multivariate conditional autoregressive (MVCAR) model

To account for the spatial autocorrelation of neighboring sites and inherent correlation among different crash types simultaneously, a K-dimensional multivariate autoregressive (MVCAR) model is proposed:

$$\ln(\lambda_i^k) = \beta_0^k + \sum_{p=1}^{P_k} \beta_p^k X_{pi}^k + \varepsilon_i^k + S_{ki} \tag{5}$$

where, S_{ki} is a multivariate CAR effect term. The difference between the multivariate CAR effect S_{ki} and the univariate CAR effect S_i^k in Eq. (3) is that S_{ki} accounts for the spatial correlation among different crash types. The full conditional distribution of $\mathbf{S_i} = (S_{1i}, S_{2i}, ..., S_{Ki})'$ follows a K-dimensional multivariate normal distribution (Thomas et al., 2004):

$$\mathbf{S_i} | \mathbf{S_{-i}} \sim MV\mathrm{N_K}\left(\sum_{i'\neq i} \frac{\mathrm{w}_{i'i}}{\mathrm{w}_{i+}} \mathbf{S}_{i'}, \frac{\mathbf{\Omega}}{\mathrm{w}_{i+}}\right) \tag{6}$$

Similar to S_{-i}^k in Eq. (4), $\mathbf{S_{-i}}$ is the set of $\mathbf{S}_{i'}$ for any $i' \neq i$. $\mathbf{\Omega}$ is the variance-covariance matrix for correlation:

$$\mathbf{\Omega} = \begin{pmatrix} \sigma_{S11}^2 & \sigma_{S12}^2 & \cdots & \sigma_{S1K}^2 \\ \sigma_{S21}^2 & \sigma_{S22}^2 & \cdots & \sigma_{S2K}^2 \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{SK1}^2 & \sigma_{SK2}^2 & \cdots & \sigma_{SKK}^2 \end{pmatrix} \tag{7}$$

Diagonal elements of $\mathbf{\Omega}$ (i.e., σ_{S11}^2, σ_{S22}^2, ...,σ_{SKK}^2) indicates the conditional variance of the spatial effects of individual crash types and the off-diagonal elements (i.e., σ_{S12}^2, σ_{S13}^2, ...,σ_{SKK-1}^2) represent the conditional with-in site covariance of the spatial effects of different crash types (Thomas et al., 2004).

4.2. Bayesian approach

4.2.1. Estimation of Bayesian models

All the aforementioned models are estimated in the full Bayesian framework. Bayesian method combines prior distributions with a likelihood function obtained to create posterior distributions as estimates. The theoretical framework for Bayesian inference can be expressed as:

$$p(\theta|\mathbf{y}) \propto L(\mathbf{y}|\theta)\pi(\theta) \tag{8}$$

where, $\mathbf{y}$ is the vector of observed data; θ is the vector of parameters required for the likelihood function (regarding the MVCAR model, θ contains β_0^k, β_p^k, α_k^2, $\mathbf{S_i}$, $\mathbf{\Omega}$); $p(\theta|\mathbf{y})$ is the posterior distribution of θ given $\mathbf{y}$; $L(\mathbf{y}|\theta)$ is the likelihood function (Eqs. (1) and (5) construct the likelihood function of MVCAR model); and $\pi(\theta)$ is the prior distribution of θ. Bayesian inference generally is performed using Markov Chain Monte Carlo (MCMC) algorithm (Gilks et al., 1998). 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 MCMC simulation (Spiegelhalter et al., 2002).

The priors of the univariate and multivariate CAR effect terms S_i^k and $\mathbf{S_i}$ were generated from Eqs. (4) and (6). Without credible prior information, uninformative priors were assumed for the other parameters. The chosen prior distributions are consistent to the previous studies such as El-Basyouny and Sayed (2009a), Guo et al. (2010), and Wang et al. (2014a). All regression coefficients were assumed to follow the Gaussian distribution $(0,10^5)$. The variance of the univariate CAR distribution σ_{Sk}^2 were assumed to follow the Inverse-Gamma distribution $(10^{-3}, 10^{-3})$. The logarithm of the variance of the Poisson-Gamma error term ($\ln\alpha_k^2$) was assumed to follow the Gaussian distribution $(0, 10^3)$. The variance-covariance matrix for correlation $\mathbf{\Omega}$ was assumed to follow a Wishart distribution (Thomas et al., 2004).

4.2.2. Autocorrelation and convergence of MCMC samples

Autocorrelation and convergence of MCMC samples were examined to make reliable statistical inferences. We used four variables (i.e, logVMT, avenue length, residential ratio and walk ratio) of the proposed MVCAR model as demonstration examples. The autocorrelation functions (ACFs) of these four variables are presented in Fig. 2. It was found that the autocorrelations for each MCMC chain were near zero when the lag was higher than or equal to one. It indicates that the MCMC samples could be regarded as independent and representative. Thus, technical approaches like thinning (Link and Eaton, 2012) to reduce autocorrelation are not necessary in this case.

We used relative standard deviation (RSD) of MCMC samples as an indicator of convergence for a single MCMC chain. RSD was defined as the standard deviation divided by the absolute value of the mean for MCMC samples. The plots of RSD are presented in Fig. 3. It was found that the after the initial fluctuation, overall tendency of RSD was decreasing as the iteration number increased. RSD plateaued after 60,000 iterations, which indicated the convergence of MCMC chains. Additionally, the potential scale reduction factor (PSRF) proposed by (Brooks and Gelman, 1998) was used to assess the convergence of multiple chains. The PSRF was obtained by dividing mixture variance by the average within-chain variance for each parameter. Convergence was assumed to occur when PSRF is less than 1.2. Fig. 4 presents the median and 97.5th percentile of PSRF during the MCMC process for the selected variables. All the estimates were less than 1.2 after the first 60,000 iterations. Considering convergence and time of updating, two MCMC chains of 100,000 iterations were run, and the first 60,000 samples were discarded as burn-in.

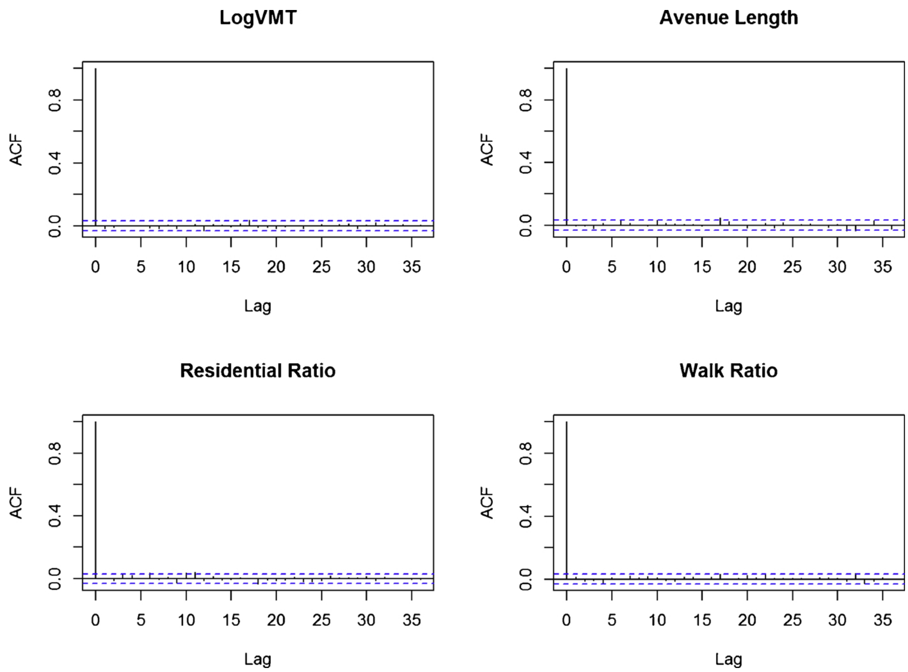

Fig. 2. Autocorrelation functions (ACFs) for the selected variables.

4.2.3. Assessment of Bayesian models

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

$$DIC = \overline{D(\theta)} + p_D \tag{9}$$

$D(\theta)$ is the Bayesian deviance of the estimated parameter θ, $D(\theta) = -2\log(L(\mathbf{y}|\theta)) + C$, where C is a constant than cancels out in all calculations that compare different models. $\overline{D(\theta)}$ 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 DIC difference of 5 or greater suggests that the model with a smaller DIC should be favored.

5. Modeling results

The PG, UCAR, and MVCAR models specified in the methodology section were developed in the Bayesian framework. A stepwise Akaike information criterion (AIC) method (Yamashita et al., 2007) was used for variable section. The variables kept by the stepwise AIC method were further examined according to their significance and contribution to the goodness of fit. Different explanatory variables were selected to model the PDO, MI, and SIF crashes. Variance inflation factors (VIFs) were used to diagnose multicollinearity in each model. A VIF greater than 5 indicates the existence of multicollinearity problem (O'brien, 2007). As presented in Table 2, all the VIFs of explanatory variables are far less than 5, and it indicates that no multicollinearity is detected.

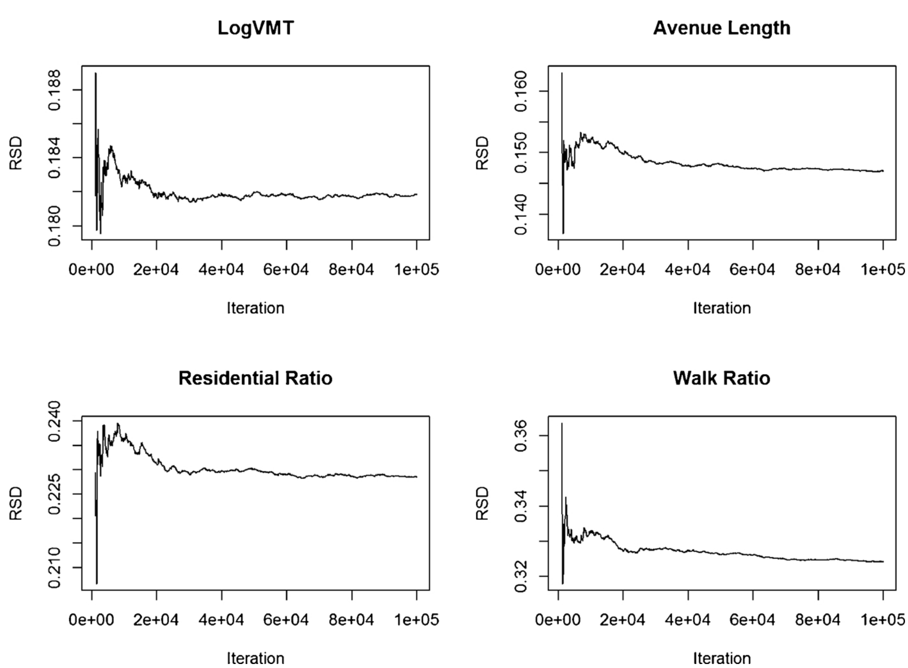

Fig. 3. Relative standard deviation (RSD) for the selected variables.

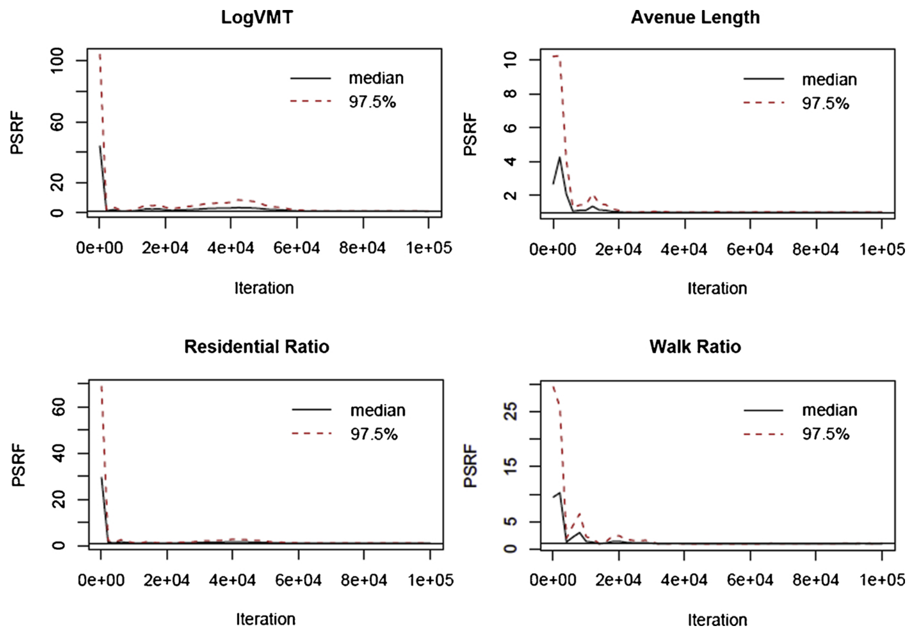

Fig. 4. Convergence diagnostics using potential scale reduction factor (PSRF) for the selected variables.

Table 2
Detection of multicollinearity using variance inflation factors (VIF).

Variables	Variance Inflation Factors (VIF)		
	PDO Crash Model	MI Crash Model	SIF Crash Model
Transportation			
LogVMT	1.607	1.380	1.413
Truck ratio	–	1.267	1.116
Road density	1.664	–	1.649
Avenue length	1.088	1.111	1.112
Street length	–	1.731	–
Intersection number	1.323	–	–
Land use			
Commercial ratio	1.575	–	–
Residential ratio	–	1.945	1.857
Mixed ratio	1.135	–	–
Park ratio	–	1.397	–
Demo-economic			
Population	1.372	–	–
Population under 14	–	1.555	1.629
Median income	–	1.355	1.320
Public transit ratio	1.418	–	–
Walk ratio	–	1.608	1.406

Table 3
Summary of DIC values.

Model	DIC
Model 1. Poisson-Gamma (PG) model	5699
Model 2. Univariate conditional autoregressive (UCAR) model	5685
Model 3. Multivariate conditional autoregressive (MVCAR) model	5644

The summary of DIC statistic is presented in Table 3. The DIC value of the MVCAR model (5644) is 55 and 41 less than that of the PG model (5699) and the UCAR model (5685), respectively. This suggests that the MVCAR model has the best performance. Compared to the PG model, the UCAR model that accounts for the spatial autocorrelation of crash data by including the CAR effect terms can be regarded as superior since it reduces DIC value by 14 (from 5699 for the PG model to 5685 for the UCAR model). In contrast with the UCAR model, the MVCAR model is further improved by including a multivariate CAR effect term to jointly address the spatial autocorrelation and the inherent correlation of different crash types.

Regarding the best performance of the MVCAR model, its Bayesian posterior estimates are used for variable interpretation, as presented in Table 4. The 95% Bayesian Credible Interval (95% BCI) is used to examine the significance of explanatory variables. Estimates can be regarded as significant at the 95% level if the BCIs do not cover 0 and vice versa (Gelman, 2004). Except road density and median income in the SIF crash model, all the other explanatory variables included in the MVCAR model are found to be statistically significant.

Vehicle miles traveled (VMT) is found to be positively related to the crash counts of all three severity levels. This finding is consistent with numerous previous studies (Abdel-Aty et al., 2013; Wang and Kockelman, 2013; Lee et al., 2015; Xie et al., 2017). The explanation is straight forward that vehicles traveled more miles are exposed to more crash risk. For this log-transformed variable VMT, its coefficients can be interpreted as: 1% increase in VMT is expected to raise the PDO, MI, and SIF crash counts by 0.272%, 0.182% and 0.180%, respectively. Additionally, previous studies (Ladrón De Guevara et al., 2004; Cottrill and Thakuriah, 2010; Ukkusuri et al., 2011; Wang and Kockelman, 2013; Lee et al., 2015) show the positive impact of road length and density on crash occurrence. In this study, avenue length is found to have positive impacts on PDO, MI, and SIF crash counts and one-unit increase in street length is predicted to result in 21.0% ($e^{0.191}$-1) more MI crashes. We also find that the increase of road density would lead to more PDO and SIF crashes and the increase of the intersection number could raise the likelihood of PDO crashes. A possible reason is that higher road density and more intersections are accompanied by shorter intersection distances, and it will limit the gap for making safe lane changes and result in more traffic conflicts (Xie et al., 2013). Another predictive variable is the truck ratio and it shows that a higher truck ratio promotes the likelihood of MI and SIF crashes, which is consistent with the findings in Martin (2002) and Xie et al. (2017). It is intuitive that trucks can disturb the traffic flow and cause more severe crashes because of their heavy weight.

Previous studies (Wier et al., 2009; Pulugurtha and Sambhara, 2011; Ukkusuri et al., 2011; Xie et al., 2017) show that land use patterns could influence the occurrence of crashes. In Table 4, the results show that both the ratio of commercial area and the ratio of mixed zone area have positive effect on PDO crash count. A possible reason is that more traffic attracted to the commercial areas imposes more exposure on crashes. Wang and Kockelman (2013) also found that areas with

Table 4
Modeling results of MVCAR model.

Variables	PDO Crash Model		MI Crash Model		SIF Crash Model	
	Mean	95% BCI	Mean	95% BCI	Mean	95% BCI
Intercept	0.422	(0.056, 0.778)	1.997	(1.639, 2.343)	−0.584	(−1.142, 0.137)
Transportation						
LogVMT	0.272	(0.234, 0.307)	0.182	(0.149, 0.217)	0.180	(0.118, 0.231)
Truck ratio	–	–	1.541	(0.549, 2.468)	2.519	(0.461, 4.491)
Road density	0.098	(0.035, 0.160)	–	–	0.038	(−0.046, 0.121)
Avenue length	0.399	(0.152, 0.605)	1.023	(0.851, 1.187)	1.051	(0.795, 1.297)
Street length	–	–	0.191	(0.089, 0.313)	–	–
Intersection number	0.020	(0.012, 0.028)	–	–	–	–
Land use						
Commercial ratio	0.446	(0.250, 0.666)	–	–	–	–
Residential ratio	–	–	−0.622	(−0.839, −0.461)	−0.565	(−0.843, −0.301)
Mixed ratio	0.540	(0.263, 0.848)	–	–	–	–
Park ratio	–	–	−0.467	(−0.776, −0.178)	–	–
Demo-economic						
Population	0.038	(0.020, 0.053)	–	–	–	–
Population under 14	–	–	0.262	(0.169, 0.350)	0.262	(0.112, 0.402)
Median income	–	–	−0.001	(−0.002, 0.000)	−0.001	(−0.003, 0.001)
Public transit ratio	−0.707	(−1.016, −0.437)	–	–	–	–
Walk ratio	–	–	0.433	(0.170, 0.693)	0.745	(0.283, 1.184)
Dispersion α_k	0.168	(0.105, 0.222)	0.023	(0.000, 0.063)	0.030	(0.000, 0.090)

mixed land use patterns were associated with higher crash frequencies. This finding can also be attributed to the mixed traffic in those areas, which is composed of pedestrians, cars, buses, and delivery trucks. One the contrary, the ratio of residential area has negative effect on MI and SIF crash counts and the ratio of park area has negative effect on MI crash count. A possible reason could be that both the residential and park areas have relatively lower traffic and smaller portions of heavy vehicles compared with the commercial and mixed areas.

Previous studies found the relationship between crash occurrence and demo-economic features including population (Ladrón De Guevara et al., 2004; Wier et al., 2009; Pulugurtha and Sambhara, 2011; Ukkusuri et al., 2011; Lee et al., 2015), age composition (Ladrón De Guevara et al., 2004; Wier et al., 2009; Lee et al., 2015) and income (Cottrill and Thakuriah, 2010; Lee et al., 2015). In Table 4, population is positively associated with PDO crashes, and in the census tracts with more population under 14 years old, the likelihoods of MI and SIF crashes are higher. Another finding is that the census tracts with high median income are associated with lower MI and SIF crash counts. In addition, it is an inspiring result that the ratio of commuters by public transit has a negative effect on SIF crash count. It is possible that if more commuters take public transit and fewer people drive, it could reduce the overall exposure to crashes. Census tracts with higher ratios of commuters by walking are associated with higher MI and SIF crashes because pedestrians are vulnerable road users and are more likely to be injured in crashes.

6. Discussion on spatial autocorrelation

This section first introduces Moran's *I* statistics to assess the significance of spatial autocorrelation of variables of interest. Then the ability of the PG, UCAR, and MVCAR models to adjust for spatial autocorrelation is examined by conducting Moran's *I* tests on their residuals.

6.1. Moran's I tests

To assess the spatial autocorrelation of variables in a quantitative way, global univariate and bivariate Moran's *I* statistics are computed. The univariate Moran's *I* is defined as (Moran, 1948):

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{S_0\sum_{i=1}^{n}(z_i - \bar{z})^2} \tag{10}$$

where, n is the total number of spatial units indexed by i and j; z is the variable of interest; $\bar{z}$ is the mean of z; w_{ij} represents the spatial weight between spatial units i and j, with $w_{ij} = 1$ if units i and j are adjacent, and $w_{ij} = 0$ otherwise; and S_0 is the aggregation of spatial weights with $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$.

As a generalization of univariate Moran's *I*, bivariate Moran's *I* can be used to measure the spatial correlation of two variables (Anselin et al., 2002). It indicates whether a variable is spatially correlated with the other variable. Bivariate Moran's *I* for two different variables is given by:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(z_i^A - \overline{z^A})(z_j^B - \overline{z^B})}{S_0\sum_{i=1}^{n}(z_i^A - \overline{z^A})^2} \tag{11}$$

where, z^A and z^B are the two variables of interest; and $\overline{z^A}$ and $\overline{z^B}$ are the means of z^A and z^B. Anselin et al. (2006) suggested a random permutation test to assess the significance of Moran's *I* using pseudo p-value. In the permutation test, the variables of interest were randomly reallocated to sites and Moran's statistics were computed repeatedly. The pseudo p-value is given by Eq. (12):

$$pseudop - value = \frac{M + 1}{S + 1} \tag{12}$$

where, S is the total number of permutations; and M is the number of instances with Moran's *I* equal to or greater than that obtained from the observed data. If pseudo p-value is less than 0.05, it suggests significant spatial correlation of observations.

6.2. Addressing spatial autocorrelation

GeoDa (Anselin, 2003) was used to compute the univariate Moran's *I* statistics for PDO, MI and SIF crashes separately and bivariate Moran's *I* statistics for each pair of them. A total of 9999 permutations were performed for each test. The outcomes of Moran's *I* statistics are presented in Table 5 (a). All the pseudo p-values are found to be far less than 0.05. The results not only confirm the spatial autocorrelation of each crash type at adjacent sites, but also indicate the multivariate spatial autocorrelation among different crash types. Ignoring the multivariate spatial autocorrelation of crash observations is likely to cause biased estimates and unreliable statistical inferences.

It is therefore essential to adjust for the spatial autocorrelation when modelling crash counts by injury severity to achieve more precise

Table 5
Summary of Moran's *I* tests on (a) crash counts, (b) residuals of the PG model, (c) residuals of the UCAR model, and (d) residuals of the MVCAR models.

	I	Pseudo p-value
(a)		
Univariate Moran's *I*		
PDO	0.3520	0.0001
MI	0.1895	0.0001
SIF	0.1264	0.0005
Bivariate Moran's *I*		
PDO vs MI	0.2050	0.0001
PDO vs SIF	0.1328	0.0002
MI vs SIF	0.1367	0.0001
(b)		
Univariate Moran's *I*		
PDO	0.2423	0.0001
MI	0.1508	0.0003
SIF	0.1778	0.0001
Bivariate Moran's *I*		
PDO vs MI	0.1123	0.0004
PDO vs SIF	0.0570	0.0298
MI vs SIF	0.1370	0.0002
(c)		
Univariate Moran's *I*		
PDO	−0.0630	0.0514
MI	−0.0334	0.1908
SIF	−0.0497	0.1281
Bivariate Moran's *I*		
PDO vs MI	−0.0359	0.1239
PDO vs SIF	−0.0585	0.0219
MI vs SIF	−0.0476	0.0481
(d)		
Univariate Moran's *I*		
PDO	−0.0327	0.2315
MI	−0.0354	0.2132
SIF	−0.0366	0.2064
Bivariate Moran's *I*		
PDO vs MI	0.0221	0.2662
PDO vs SIF	−0.0234	0.2079
MI vs SIF	0.0343	0.1456

estimation. Univariate and bivariate Moran's *I* tests were conducted on the residuals of the PG, UCAR and MVCAR models. If the outcome of univariate Moran's *I* tests is insignificant, it indicates that there is no spatial clustering of overestimation or underestimation and that the corresponding model can address the univariate spatial autocorrelation properly. Similarly, if the outcome of bivariate Moran's *I* tests is insignificant, it indicates that the corresponding model can capture the multivariate spatial autocorrelation. Univariate and bivariate Moran's *I* tests (9999 permutations for each test) were performed to quantify the spatial autocorrelation of residuals of a single crash type and each pair of them. The outcomes of the Moran's *I* tests are presented in Table 5(b)–(d). As shown in Table 5(b), all the Moran's *I* statistics of the residuals of the PG model were found to be significant, and it affirmed the fact that the PG model could not address the spatial autocorrelation. As shown in Table 5(c), the UCAR model was able to deal with univariate spatial autocorrelation by presenting insignificant outcomes in univariate Moran's *I* tests, whereas failed to deal with multivariate spatial autocorrelation by showing two significant outcomes (PDO vs. SIF and MI vs. SIF) in bivariate Moran's *I* tests. The result indicates that although UCAR could handle the spatial effect of each crash type separately, it tends to overestimate (or underestimate) one crash type for one tract and meanwhile overestimate (or underestimate) the other crash types for the neighbouring tracts. Regarding the MVCAR model, as shown in Table 5(d), all the pseudo p-values of the Moran's statistics were found to be greater than 0.05, indicating insignificant spatial autocorrelation of crash residuals. This finding shows that the MVCAR model accounted for not only the spatial autocorrelation of each individual crash type but also the multivariate spatial autocorrelation among different crash types. There is no spatial clustering of overestimation or underestimation of crashes by type.

Table 6
Estimates of the conditional standard deviation (SD) of the spatial effects of each individual crash type and the within-site conditional correlation of the spatial effects of each pair of crash types.

	Mean	95% BCI
Conditional standard deviation (SD)		
PDO: σ_{S11}	0.858	(0.748, 0.975)
MI: σ_{S22}	0.921	(0.834, 1.016)
SIF: σ_{S33}	0.998	(0.859, 1.148)
Within-site conditional correlation		
PDO vs MI: $\sigma_{S12}^2/(\sigma_{S11} \times \sigma_{S22})$	0.805	(0.726, 0.873)
PDO vs SIF: $\sigma_{S13}^2/(\sigma_{S11} \times \sigma_{S33})$	0.715	(0.596, 0.817)
MI vs SIF: $\sigma_{S23}^2/(\sigma_{S22} \times \sigma_{S33})$	0.898	(0.843, 0.938)

The results of the Moran's *I* tests in Table 5 confirm the presence of multivariate spatial autocorrelation of crash data and that it could be properly addressed by the proposed MVCAR model. Table 6 presents the estimates of the conditional standard deviation (SD) of the spatial effects of each individual crash type and the within-site conditional correlation of the spatial effects of different crash types in the MVCAR model. All the estimates in Table 6 were found to be statistically significant (95% BCIs do not cover 0) and the results provided further evidence for the multivariate spatial autocorrelation of different crash types.

The conditional SDs of the spatial effects of PDO crashes are 0.858, 0.921 and 0.998 and thus, one SD higher in spatial effect indicates 136% ($e^{0.858}$-1), 151% ($e^{0.921}$-1), and 171% ($e^{0.998}$-1) higher expectation in PDO, MI, and SIF crash counts, respectively. The great variation in crash occurrence could be caused by unobserved spatial factors such as the road geometric design, traffic control and enforcement that vary largely among downtown, midtown and uptown in Manhattan. Though the crash counts of PDO, MI, and SIF differ greatly in magnitude, the SDs of the spatial effects by crash type are of similar levels, because the spatial effects influence the crash counts proportionally. It is an expected outcome that the spatial effect of SIF crashes has a slightly higher SD than the other two crash types, because generally SIF crashes are rarer and more random and factors contributing to severe crashes such as speeding and non-use of seat belts (Zhang et al., 2000; Xie et al., 2018) are not included for modeling in this study. Moreover, the SDs of the spatial effects are found distinctly greater than those of the dispersion effects (0.168 for the PDO crashes, 0.023 for the MI crashes, and 0.030 for the SIF crashes) caused by non-spatial factors as presented in Table 4, which indicates that the unobserved heterogeneity is mostly attributed to spatial factors rather than non-spatial ones. Regarding the correlation between crash types, all the within-site conditional correlation coefficients of spatial effects are positive and greater than 0.7, suggesting a strong shared geographical pattern of risk among different crash types in Manhattan. Another finding is that the correlation between MI and SIF crashes is the highest (0.898) among all three pairs. A possible reason for that is the factors contributing to minor and severe crash injuries are similar.

7. Summary and conclusions

In this study, Manhattan, which is the most densely populated urban area of New York City, was used as the study area. Crash, transportation, land use and demo-economic data of Manhattan were collected for this study. Poisson-Gamma (PG) model, univariate conditional autoregressive (UCAR) model, and multivariate conditional autoregressive (MVCAR) model were developed and compared. The results of DIC suggest that the proposed MVCAR model outperforms the others by incorporating multivariate CAR effects. The proposed MVCAR model can improve the accuracy of coefficient estimates by properly accounting for the spatial dependence of crash observations. Results also show that vehicle miles traveled (VMT), truck ratio, road density, avenue length, street length, intersection number, ratio of commercial area, ratio of mixed land use area, population, population under 14, and ratio of commuters by walking have positive impact on crash occurrence; whereas ratio of residential area, ratio of park area, median income, and ratio of commuters by public transit are negatively associated with crash occurrence. By conducting Moran's *I* tests, residuals of the MVCAR model were found to be distributed randomly over space, which indicates that the MVCAR model can account for the multivariate spatial autocorrelation among different crash types. It is also found that unobserved heterogeneity was mostly attributed to spatial factors rather than non-spatial ones and there was a strong shared geographical pattern of risk among different crash types.

Modeling crash counts by injury severity using the MVCAR model in this study can help identify high-risk locations with consideration of injury severity. The MVCAR model has been used in a few previous studies but has not gain wide attention. Regarding the necessity to analyze and to model the occurrence of crashes by types (e.g., injury severity, time of day, travel mode, etc.) and the fact that crash data is generally spatially correlated, the MVCAR model can be helpful to avoid biased inferences and to gain insights into road safety management. This study also contributes to the literature by explicitly examining the univariate and multivariate spatial autocorrelation using Moran's I statistics. Approaches are presented to investigate how the unobserved heterogeneity is attributed to spatial or non-spatial factors and how strong the shared geographical pattern of risk among different crash types is. In addition, a detailed procedure of estimating Bayesian models is presented, including approaches to diagnose autocorrelation and convergence of MCMC.

One limitation of this study is that the boarders of census tracts coincide with major arterials and crashes occurring on those arterials are arbitrarily assigned to adjacent census tracts and this would lead to biased inferences. For future study, special consideration will be given to crashes on the boarders. Furthermore, all the Bayesian models developed in this study are Poisson-based. It is a future direction worth pursuing to compare Poisson-based models with other approaches such as Tobit models (Anastasopoulos et al., 2012a, b) that accommodate continuous response variable like crash rate and could potentially capture more risk factors and improve the mixing of MCMC chains.

Acknowledgments

The work is partially funded by the Connected Cities for Smart Mobility towards Accessible and Resilient Transportation (C2SMART) Center at New York University (NYU) and the Connected Traffic and Complex Systems (CTCS) Lab at University of Canterbury. The authors would like to thank the New York State Department of Transportation, New York City Department of City Planning, Metropolitan Transportation Authority, New York Metropolitan Transportation Council and U.S. Census Bureau for publicly providing data used in this study. The contents of this paper present views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents of the paper do not reflect the official views or policies of the agencies.

References

Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accid. Anal. Prev. 32 (5), 633–642.
Abdel-Aty, M., Wang, X., 2006. Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors. Trans. Res. Rec.: J. Trans. Res. Board (1953), 98–111.
Abdel-Aty, M., Lee, J., Siddiqui, C., Choi, K., 2013. Geographical unit based analysis in the context of transportation safety planning. Transp. Res. Part A Policy Pract. 49, 62–75.
Anastasopoulos, P.C., Mannering, F.L., Shankar, V.N., Haddock, J.E., 2012a. A study of factors affecting highway accident rates using the random-parameters tobit model. Accid. Anal. Prev. 45, 628–633.
Anastasopoulos, P.C., Shankar, V.N., Haddock, J.E., Mannering, F.L., 2012b. A multivariate tobit analysis of highway accident-injury-severity rates. Accid. Anal. Prev. 45, 110–119.
Anselin, L., 2003. Geoda™ 0.9 user's guide. Urbana 51, 61801.
Anselin, L., Syabri, I., Smirnov, O., 2002. Visualizing multivariate spatial correlation with dynamically linked windows. Urbana 51, 61801.
Anselin, L., Syabri, I., Kho, Y., 2006. Geoda: An introduction to spatial data analysis. Geogr. Anal. 38 (1), 5–22.
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., 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc. Ser. B 192–236.
Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Stat. Math. 43 (1), 1–20.
Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. 7 (4), 434–455.
Castro, M., Paleti, R., Bhat, C.R., 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: application to predicting crash frequency at intersections. Transp. Res. Part B Methodol. 46 (1), 253–272.
Cheng, W., Gill, G.S., Dasu, M., Jia, X., 2018. An empirical evaluation of multivariate spatial crash frequency models. Accid. Anal. Prev. 119, 290–306.
Cottrill, C.D., Thakuriah, P.V., 2010. Evaluating pedestrian crashes in areas with high low-income or minority populations. Accid. Anal. Prev. 42 (6), 1718–1728.
El-Basyouny, K., Sayed, T., 2009a. Accident prediction models with random corridor parameters. Accid. Anal. Prev. 41 (5), 1118–1123.
El-Basyouny, K., Sayed, T., 2009b. Urban arterial accident prediction models with spatial effects. Trans. Res. Rec.: J. Trans. Res. Board (2102), 27–33.
Gelman, A., 2004. Bayesian Data Analysis, 2nd ed. Chapman & Hall/CRC, Boca Raton, Florida.
Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. Pattern Analysis and Machine Intelligence. IEEE Trans. (6), 721–741.
Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1998. Markov Chain Monte Carlo in Practice. Chapman & Hall, Boca Raton, Florida.
Guo, F., Wang, X.S., Abdel-Aty, M.A., 2010. Modeling signalized intersection safety with corridor-level spatial correlations. Accid. Anal. Prev. 42 (1), 84–92.
Jones, B., Janssen, L., Mannering, F., 1991. Analysis of the frequency and duration of freeway accidents in seattle. Accid. Anal. Prev. 23 (4), 239–255.
Ladrón De Guevara, F., Washington, S., Oh, J., 2004. Forecasting crashes at the planning level: Simultaneous negative binomial crash model applied in tucson, arizona. Trans. Res. Rec.: J. Trans, Res, Board (1897), 191–199.
Lan, B., Persaud, B., Lyon, C., Bhim, R., 2009. Validation of a full bayes methodology for observational before-after road safety studies and application to evaluation of rural signal conversions. Accid. Anal. Prev. 41 (3), 574–580.
Lee, J., Abdel-Aty, M., Choi, K., Huang, H., 2015. Multi-level hot zone identification for pedestrian safety. Accid. Anal. Prev. 76 (0), 64–73.
Link, W.A., Eaton, M.J., 2012. On thinning of chains in mcmc. Methods Ecol. Evol. 3 (1), 112–115.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Trans. Res. Part a-Policy Pract. 44 (5), 291–305.
Lord, D., Miranda-Moreno, L.F., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of poisson-gamma models for modeling motor vehicle crashes: A bayesian perspective. Saf. Sci. 46 (5), 751–770.
Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. Winbugs-a bayesian modelling framework: Concepts, structure, and extensibility. Stat. Comput. 10 (4), 325–337.
Ma, J., Kockelman, K.M., 2006. Bayesian multivariate poisson regression for models of injury count, by severity. Trans. Res. Rec.: J. Trans. Res. Board 1950 (1), 24–34.
Martin, J.-L., 2002. Relationship between crash rate and hourly traffic flow on interurban motorways. Accid. Anal. Prev. 34 (5), 619–629.
Miaou, S.P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accid. Anal. Prev. 25 (6), 689–709.
Moran, P.A., 1948. The interpretation of statistical maps. J. R. Stat. Soc. Ser. 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. Transp. Res. Part B Methodol. 55, 245–264.
Nashad, T., Yasmin, S., Eluru, N., Lee, J., Abdel-Aty, M.A., 2016. Joint modeling of pedestrian and bicycle crashes: copula-based approach. Trans. Res. Rec.: J. Trans. Res. Board (2601), 119–127.
O'brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quant. 41 (5), 673–690.
Park, E.S., Lord, D., 2007. Multivariate poisson-lognormal models for jointly modeling crash frequency by severity. Trans. Res. Rec.: J. Trans. Res. Board 2019 (1), 1–6.
Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection-accident frequencies. J. Trans. Eng.-ASCE 122 (2), 105–113.
Pulugurtha, S.S., Sambhara, V.R., 2011. Pedestrian crash estimation models for signalized intersections. Accid. Anal. Prev. 43 (1), 439–446.
Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of london crash data. Accid. Anal. Prev. 40 (4), 1486–1497.

Saha, D., Alluri, P., Gan, A., Wu, W., 2018. Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models. Accid. Anal. Prev. 118, 166–177.
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., Best, N.G., Carlin, B.R., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B-Stat. Methodol. 64, 583–616.
Thomas, A., Best, N., Lunn, D., Arnold, R., Spiegelhalter, D., 2004. Geobugs User Manual. Medical Research Council Biostatistics Unit, Cambridge.
Ukkusuri, S., Hasan, S., Aziz, H., 2011. Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency. Trans. Res. Rec.: J. Trans. Res. Board (2237), 98–106.
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.
Wang, X., Xie, K., Abdel-Aty, M., Chen, X., Tremont, P., 2014a. Systematic approach to hazardous-intersection identification and countermeasure development. J. Trans. Eng.-ASCE 140 (6) 04014022.
Wang, Y., Kockelman, K.M., Damien, P., 2014b. A spatial autoregressive multinomial probit model for anticipating land-use change in austin, texas. Ann. Reg. Sci. 52 (1), 251–278.
Whittle, P., 1954. On stationary processes in the plane. Biometrika 434–449.
Wier, M., Weintraub, J., Humphreys, E.H., Seto, E., Bhatia, R., 2009. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. Accid. Anal. Prev. 41 (1), 137–145.
Xie, K., Wang, X., Huang, H., Chen, X., 2013. Corridor-level signalized intersection safety analysis in shanghai, china using bayesian hierarchical models. Accid. Anal. Prev. 50, 25–33.
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.
Xie, K., Ozbay, K., Yang, H., 2015a. Spatial analysis of highway incident durations in the context of hurricane sandy. Accid. Anal. Prev. 74, 77–86.
Xie, K., Ozbay, K., Yang, H., Holgurin-Veras, J., Morgul, E.F., 2015b. Modeling the safety impacts of off-hour delivery programs in urban areas. Trans. Res. Rec.: J. Trans. Res. Board 4784.
Xie, K., Ozbay, K., Kurkcu, A., Yang, H., 2017. Analysis of traffic crashes involving pedestrians using big data: investigation of contributing factors and identification of hotspots. Risk Anal. 37 (8), 1459–1476.
Xie, K., Ozbay, K., Yang, H., 2018. Secondary collisions and injury severity: a joint analysis using structural equation models. Traffic Inj. Prev. 19 (2), 189–194.
Yamashita, T., Yamashita, K., Kamimura, R., 2007. A stepwise aic method for variable selection in linear regression. Commun. Stat.—Theory Methods 36 (13), 2395–2403.
Yang, H., Ozbay, K., Xie, K., 2014. Assessing the risk of secondary crashes on highways. J. Safety Res. 49 (143), e1-149.
Yang, H., Wang, Z., Xie, K., Ozbay, K., Imprialou, M., 2018. Methodological evolution and frontiers of identifying, modeling and preventing secondary crashes on highways. Accid. Anal. Prev. 117, 40–54.
Yasmin, S., Eluru, N., Lee, J., Abdel-Aty, M., 2016. Ordered fractional split approach for aggregate injury severity modeling. Trans. Res. Rec.: J. Trans. Res. Board (2583), 119–126.
Zhang, J., Lindsay, J., Clarke, K., Robbins, G., Mao, Y., 2000. Factors affecting the severity of motor vehicle traffic crashes involving elderly drivers in ontario. Accid. Anal. Prev. 32 (1), 117–125.