Accident Analysis and Prevention 131 (2019) 297–307

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Taxicab crashes modeling with informative spatial autocorrelation

Qingyu Ma[a], Hong Yang[a,*], Kun Xie[b], Zhenyu Wang[a], Xianbiao Hu[c]

[a] Department of Computational Modeling and Simulation Engineering, Old Dominion University, Norfolk, VA 23529, United States
[b] Department of Civil and Environmental Engineering, Old Dominion University, Norfolk, VA 23529, United States
[c] Department of Civil, Architectural & Environmental Engineering, Missouri University of Science and Technology, Rolla, MO 65409-0030, United States

ARTICLE INFO

Keywords:
Taxi crashes
Taxi trips
Spatial weight
Spatial autocorrelation
Conditional autoregressive model
Bayesian estimation

ABSTRACT

Maintaining taxi safety is one of the important goals of operating urban transportation systems. Taxicabs are often prone to higher crash risk due to their long-time exposure to the complicated and dynamic traffic environments in urban areas. Despite existing efforts in understanding the safety issues associated with these vehicles, there were still few attempts that have specifically examined the relationship between taxi-involved crashes and other multifaceted contributing factors. To this end, this paper aims to develop crash frequency models for analyzing taxi-involved crashes. In particular, the spatial autocorrelations between variables were explored and the Poisson conditional autoregressive (Poisson-CAR) models for taxi-involved crashes were proposed. Unlike previous safety studies that mainly consider distance as the key indicator of spatial correlation, the present paper introduced the use of massive taxi trip data for constructing a more informative spatial weight matrix. The developed models with the taxi trip-based weight matrix were tested by using the 2016 taxi trip data collected in Washington D.C. The modeling results highlight the key explanatory factors such as road density, taxi activity, number of bus stops, and land use. More importantly, it demonstrates that the proposed Poisson-CAR models with the taxi trip-based weight matrix outperformed both the non-spatial Poisson model and the Poisson-CAR models using conventional distance-based weight matrix. Moran's I tests further indicate that our proposed models have sufficiently accounted for the spatial autocorrelation of the residuals. Thus, it deserves to consider informative spatial weight matrices when applying spatial models in traffic safety studies.

1. Introduction

Taxicabs are one of the most commonly used transportation modes in urban areas. These vehicles typically account for 6 to 12 percent of all trips in major cities (LTA, 2011). For example, according to the estimation by the New York City Taxi (NYC) and Limousine Commission (TLC) (Joshi, 2018), nearly 780 million trips were completed from January 2016 to June 2018. Likewise, in Washington D.C., the monthly number of taxi trips is between 0.6 and 0.8 million from May 2015 to July 2017 according to the Open Data DC (http://opendata.dc.gov). Meanwhile, with the rise of e-hailing companies such as Uber and Lyft, millions of trips are also served by the ridesharing vehicles. Arguably, taxi drivers are at a greater risk of being involved in a crash due to their occupational exposure to hazardous conditions such as fatigue and stress daily (Lam, 2004), which similarly applies to many e-hailing drivers. The large number of taxis as well as ridesharing vehicles cruising along streets is deemed to cause great challenges to transportation safety in urban areas.

An array of studies has already examined the behavioral characteristics of taxi drivers involved in crashes. For example, taxi drivers with an age below 30 were prone to be involved and injured in crashes than the elder group (Maag et al., 1997). However, given the significant effect of exposure to hazardous environments, it deserves to mention that only few studies have explored the external factors associated with taxi crashes (Yang et al., 2015). Unlike many commuters driving personal vehicles, taxis often provide their services according to passengers' requests. Thus, there are no fixed schedules and routes. Even experienced taxi drivers may need to drive in an unfamiliar community to pick up or drop off a passenger. The new environment in the community may cause safety challenges to taxi drivers because of changed traffic patterns, road network structures, speed limits, etc. In addition, urban areas often consist of a number of central business districts, tourist attractions, shopping centers, public transport terminals, airports, etc. Many taxi drivers often run business among these sites because of the strong demand of services. This naturally leads to distinguishable taxi mobility patterns in terms of the spatial and temporal characteristics among different zones in a city (Ma et al., 2019). Consequently, the safety risk is also expected to be different among the zones. Nonetheless, the relationship between taxi activities and the

* Corresponding author.
E-mail addresses: qma002@odu.edu (Q. Ma), hyang@odu.edu (H. Yang), kxie@odu.edu (K. Xie), zwang002@odu.edu (Z. Wang), xbhu@mst.edu (X. Hu).

https://doi.org/10.1016/j.aap.2019.07.016
Received 14 February 2019; Received in revised form 19 June 2019; Accepted 18 July 2019
Available online 24 July 2019

safety issues across these zones are largely unknown. This should be mainly attributable to the fact that it was almost impossible to determine precisely the spatiotemporal taxi activities in early days.

Taking advantages of the equipment with the global positioning systems (GPS), massive detailed taxi trajectory records can be gathered. These records provide a reliable probe for taxi activities within and between different zones. As a result, it also offers opportunities to quantitatively examine the relationship between these activities and taxi-involved crashes. Thus, this paper aims to examine the safety issues of taxicabs in urban areas by leveraging the massive taxi trip data derived from GPS devices. In particular, it is focused on the spatial interactions between taxi activities and taxi-involved crashes. Specifically, it introduces a unique way of constructing spatial weights using aggregated taxi trips. The derived spatial weights are then incorporated into proposed spatial models for analyzing taxi crash counts. The modeling results highlight the augmented performance of the developed spatial models with better characterization of the interactions between spatial units.

2. Literature review

Previous safety studies have investigated the occurrence of taxi-involved crashes from different perspectives, such as taxi driver behavior (Machin and De Souza, 2004; Ma et al., 2010), injury severity levels (Maag et al., 1997; Zhao et al., 2015), and causal relationships (Wang et al., 2018; Xie and Wang, 2018). Notably, more existing research were focused on behavioral analysis and individual crash characteristics. They have explored the unsafe driving maneuvers observed from vehicle acceleration and/or braking events (Machin and De Souza, 2004; Ma et al., 2010). Detailed crash characteristics such as the number of casualties, the number of involved cars, and crash types were also often analyzed (Zhao et al., 2015). The premise of such studies is the available detailed records of driving events (e.g., distraction, speeding, etc.) and individual crashes. However, collecting such precise information are often challenging. Instead of investigating individual taxi crashes, there has been a growing interest in probing taxi crash mechanisms from a macroscopic level to understand interactions between taxi activities, the environment, and crash risk (Zhao et al., 2015; Wang et al., 2018; Xie and Wang, 2018).

Despite the scarcity of the macroscopic analysis of taxi crashes, many other existing safety studies have provided extensive illustrations of crash modeling practices. For example, a number of studies have investigated traffic crash propensity at different spatial levels, such as census block (Levine et al., 1995; Loo, 2006), TAZ (Siddiqui et al., 2012; Chen, 2015; Dong et al., 2015), census tract (LaScala et al., 2000; Aguero-Valverde and Jovanis, 2006), zip code (Meliker et al., 2004; Treno et al., 2007), and others (Quddus, 2008). In macroscopic safety studies, as the spatial unit among these analyses becomes smaller, the number of crash count observed in each sampled unit decreases and the histogram becomes skewed with more zero-count units. Therefore, regarding the scale of taxi-involved crash data, small units such as census blocks may not be the suitable units. On the other hand, it's suggested that demographic factors (such as population, age, racial, etc.) affect traffic causalities. With the easy accessibility of demographic data, census tract is often the preferred spatial unit as compared to other levels. Additionally, like other crashes, taxi-involved crashes may also occur on roadways that are the boundaries of zones. Arbitrarily assigning the crashes to one zone can be biased. The literature suggested that geographic boundaries such as census tracts and TAZs may use the means of demarcation (Bureau, 2007). However, Siddiqui and Abdel-Aty (2016) argued that solely depending on the characteristics of spatial entities may not be a prudent way to allocate, analyze, and develop macroscopic safety models.

In order to capture the impact of factors in neighboring zones, a number of safety studies have also proposed the use of spatial models. In general, conventional crash frequency models (e.g., Poisson) are extended to account for the potential impact of factors from other neighboring zones (Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Huang et al., 2010). For example, Xie et al. (2014) modeled crash frequencies at sampled signalized intersections using conditional autoregressive (CAR) models. Siddiqui et al. (2012) introduced the Bayesian Poisson-lognormal model accounting for spatial correlation for bicycle and pedestrian crashes in the TAZs. Li et al. (2013) used the geographically weighted Poisson regression (GWPR) model to capture these spatially varying relationships in the county-level crash data. When considering the spatial relationship, existing studies typically considered the physical adjacency (Aguero-Valverde and Jovanis, 2006; Quddus, 2008) or distance (Li et al., 2013; Pirdavani et al., 2014; Xie et al., 2014) of different spatial units. The distance-based weight matrix is often defined as the inverse of physical distance between different spatial entities. Accounting for the impact of different zones, the spatial autocorrelation in residuals can be reduced, which in turn improves the modeling performance. Thus, it is expected that any macroscopic study on taxi crashes should also address the spatial effects.

Lately, there have been some attempts to leverage the massive taxi mobility data in supporting the understanding of taxi crashes. In general, recent studies used these data mainly in two ways: (i) heatmap visualization. For example, Xie and Wang (2018) visually described the relationship between taxi O-Ds and crash distributions using heatmaps. According to the visualization results, both originations and destinations were found to be spatially correlated with the crash distributions; and (ii) exposure variable; O-D trip information is commonly aggregated by the same zonal units as the crash frequencies, serving as an exposure variable in crash models. Besides simply counting the numbers, categorical method can also be applied. In a most recent paper by Bao et al. (2018), taxi trips are categorized into different temporal patterns indicating the activity status in each zone. The present paper notices that there can be another new way of using the taxi data to support crash modeling: describing the spatial connectivity. As mentioned earlier, many studies determined their spatial weight matrices by considering the distance between zones in different spatial models such as negative binomial model (Xu and Huang, 2015), Poisson-lognormal CAR model (Wang and Kockelman, 2013), and GWPR (Bao et al., 2018). However, distance does not necessarily warrant the strong connections between neighboring zones. For example, some geographical neighbors can be blocked by natural barriers (e.g., rivers and gullies) (Andris and Bettencourt, 2014). On the other hand, some zones (e.g., areas with many hotels) and the ones with major transportation terminals (e.g., railway stations, airports, etc.) are often expected to have strong connections in terms of taxi activities (Veloso et al., 2011; Yuan et al., 2012; Tang et al., 2015). Thus, to better describe the spatial relationship between zones, taxi O-D trips can be accumulated to replace the role of the distance between zones.

Built upon the lessons from existing safety studies, this paper intends to focus on the spatial modeling of taxi-involved crashes considering multiple contributing factors. In particular, during the process of constructing spatial weight matrices, both distance-based and taxi O-D trip-based measures are involved. This will help answer two questions. Firstly, whether spatially weighted models have a better performance than that of non-spatial models for taxi crash modeling. Secondly, whether models with weights constructed based on taxi O-D trips have improved performance comparing to the ones with traditional distance-based weights.

3. Data description

3.1. Taxi-involved crashes

The spatial units in this study are based on the census tracts in Washington D.C. area. According to the US Census 2010, there are 179 tracts in this area with the average population of 3361 per census tract. Since previous research suggest that census factors (e.g. population, age, racial) affecting traffic casualty (Noland and Quddus, 2004), crash records were aggregated by census tracts using the geographic

information system (GIS) techniques. The crash data are published and maintained by the District Department of Transportation (DDOT). Three-year (2015–2017) data were acquired and taxi-involved crashes were extracted for subsequent analysis.

Fig. 1 shows the spatial distribution of taxi-involved crashes with a total number of 5417 occurred during the 3-year period. It is noticeable that these crashes are clustered in the central areas, which represent downtown areas of D.C. Obviously, the spatial pattern illustrates a strong spatial correlation between neighboring tracts.

3.2. Taxi trip data and processing

Taxi trips data were obtained from the Open Data DC. The data portal stores hourly trip information of taxis from May 2015 to July 2017. As shown in Fig. 2, there are on average 0.837 million trips per month in Washington D.C. areas and there are notable differences between different months. This notable change reflects the temporal variations in human activities across the study area. In order to better describe the spatial distributions, spatial connection matrices can be derived by the following procedure.

3.2.1. Taxi O-D matrix

Taxi O-D matrix is calculated based on the geographical information of census tracts in D.C. Fig. 3 shows an example on how the taxi O-D matrix is generated. As shown in Fig. 3(a), A, B, and C are census tract polygons. There are four taxi trips $O_i - D_i$ ($i = 1, 2, 3, 4$) recorded with O-D information. The directional trip count can be described with the matrix M in Fig. 3(b). After adding itself with its transpose matrix M^T (except for values on diagonal), the symmetric trip count matrix $M_{Trip} = M + M^T$ in Fig. 3(c) can be derived to describe the activities associated with two zones. M_{Trip} will be used to describe the spatial autocorrelation. We will discuss it in more detail in the next section.

3.2.2. Taxi VMT

Taxi VMT can be an important explanatory variable for potential taxi crashes. In Fig. 3, if each row of the taxi trip matrix M_{Trip} in Fig. 3(c) is summed up, the taxi activity can be derived in Fig. 3(d). In this example, census tract A has the highest taxi activity with 3 trips by considering both the pickups and drop-offs. Then taxi VMT can be calculated by aggregating trip distances in similar way. Since the detailed trajectory data are unavailable, the taxi trip distance (VMT) is estimated using Euclidean distance between pick-up spots and drop-off spots. One should note that the actual trip distances typically will be longer than our estimated distances due to taxi drivers' different route selections between the same ODs.

3.3. Moran's I and Bivariate Moran's I

To test the spatial autocorrelation of taxi-involved crashes and taxi activity, global univariate and bivariate Moran's I statistics were computed. The univariate Moran's I was initially suggested by (Moran, 1948). Essentially, it is a cross-product statistic between a variable and its spatial lag, with the expression of $z_i = x_i - \bar{x}$, where $\bar{x}$ is the mean of variable x at census tract i. The univariate Moran's I statistic is then calculated as:

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

where, n is the number of spatial entities indexed by i and j; w_{ij} represents their spatial weight. If the entity i and j are adjacent, then $w_{ij} = 1$; otherwise, $w_{ij} = 0$; and S_0 is the aggregation of all the spatial weights.

The bivariate Moran's I is used to measure the spatial correlation of two variables (Anselin et al., 2002), which is a generalization of univariate Moran's I. For example, Xie et al. (2019) examined whether the spatial autocorrelation can be well captured between different types of crashes.

$$I = \frac{n}{S_0}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(z_i^A)(z_j^B)}{\sum_{i=1}^{n}(z_i^A)^2} \quad (2)$$

where, z^A and z^B are the deviations from the mean for variable A and B, respectiely.

Since the taxi data were not fully available in 2015 and 2017, the subsequent analysis and modeling are based on the crash data and taxi data collected in the year of 2016. As shown in Fig. 4, for 179 census tracts in Washington D.C, taxi crash frequency and taxi VMT individually show significant spatial autocorrelation with Moran's I values of 0.5425 and 0.3876 at 99.9% confidential level, respectively. Besides, the two variables are also spatially correlated with bivariate Moran's I of 0.5104 and p-value of 0.001. Therefore, it is important to incorporate spatial components in the model to account for the spatial correlation.

3.4. Explanatory variables for taxi-involved crashes

Other than taxi trip data, we have also explored other three major categories of explanatory variables that may be associated with taxi crashes. These include variables related to transportation, land use, and socio-demographic factors.

3.4.1. Transportation environment

Transportation environment has been frequently shown to fundamentally affect crash occurrences (Quddus, 2008; Li et al., 2013; Cai et al., 2017). Thus, a number of factors should be explored. In this paper, based on the data availability, we have examined variables related to public transit, intersections, road density, driving behavior. In general, it deserves to consider the number of transit stops that are key factors associated with citizens' travelling activities. The bus stops data were obtained from the Open Data DC. The number of bus stops were calculated for each census tract with spatial tools of the ArcGIS. In addition, dangerous intersections should be considered. The intersection data from the D.C. Crash Intersections Summary[1] were extracted to capture the intersections with historical crashes. The road density was also calculated based on the D.C. road network information. According to Cooper (1997), excessive speed conviction is positively correlated to crash occurrences. Thus, unsafe driving behavior and high speeding (15mph above speed limit) events were extracted and aggregated from the 2016 monthly moving violations reports archived on the Open Data DC. Resorting to the spatial join tool, above factors were counted or aggregated based on the unit of census tracts.

In addition to above factors, it is also important to account for the exposure in crash modeling. Thus, annual average daily traffic (AADT) data were obtained from the National Capital Region Transportation Planning Board[2], from where the exposure variable vehicle miles travelled (VMT) can be derived. For each census tract, the VMT is calculated using the function below:

$$VMT = \sum Length*AADT \quad (3)$$

where, "Length" indicates the length in miles of each road segment. The AADT for all segments located within the tract are aggregated and weighted by its length.

Other than VMT, taxi activity was also used as an exposure factor in the modeling process. Firstly, the monthly taxi activity counts in 2016 for each census tracts were calculated. Then the average numbers for each census tract were used in our model. To differ the patterns of taxi

[1] Source: https://www.arcgis.com/home/item.html?id=522992c64d8444618-5395d04d23f8d21.

[2] Source: http://rtdc-mwcog.opendata.arcgis.com/datasets/traffic-counts-annual-average-2009-2016-by-netowrk-link.

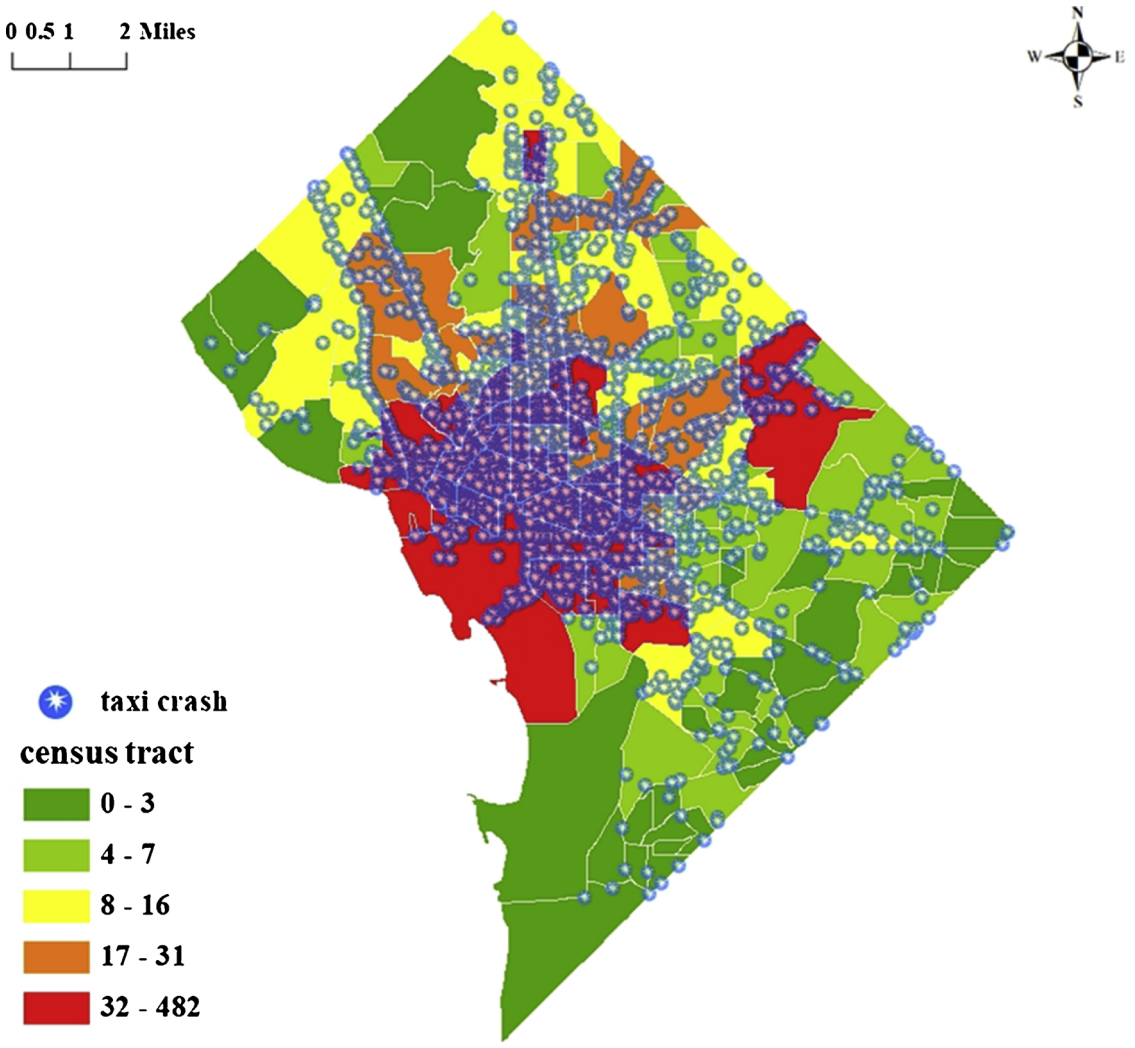

Fig. 1. The Spatial Distribution of Taxi-involved Crashes in Census Tracts.

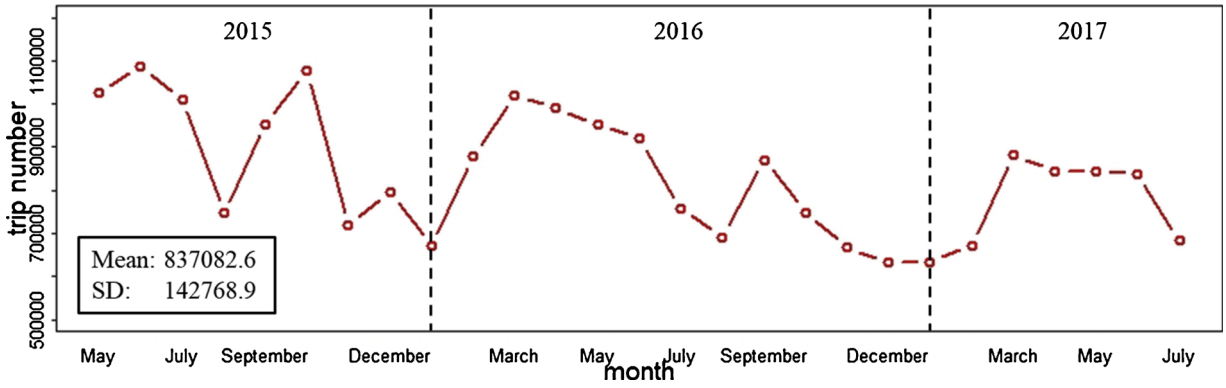

Fig. 2. Monthly Count of Taxi Trips in Washington D.C.

activity, taxi pickup ratio is also considered. It is calculated by dividing the number of pickups by the sum of pickups and drop-offs in each census tract. This ratio describes the role of the zone: with high ratio, a census tract has more taxi demand; and with low ratio, a census tract has more attractions.

3.4.2. Land use

The land use data were obtained from the Open Data DC[4]. The existing land use data were categorized as Residential, Institutional, Open Space, and Other. As shown in Fig. 5, by using the zonal statistics tool, each census tract was assigned a land use category based on its majority of land use.

3.4.3. Socio-demographic factors

Previous research suggest that various socio-demographic variables such as population, poverty, age, and racial affect traffic causalities (e.g., Graham and Glaister (2006); Noland and Quddus (2004)). The socio-demographic data for the studied census tracts were obtained

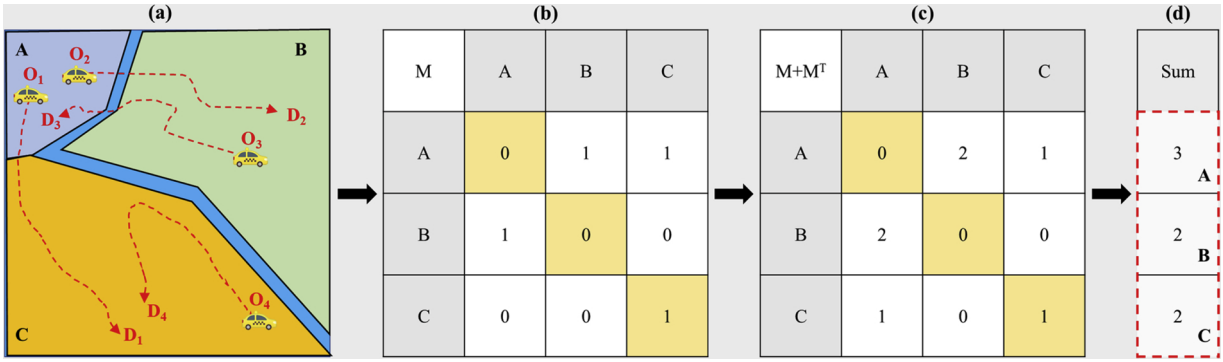

Fig. 3. Example of Trip Matrix Generation.

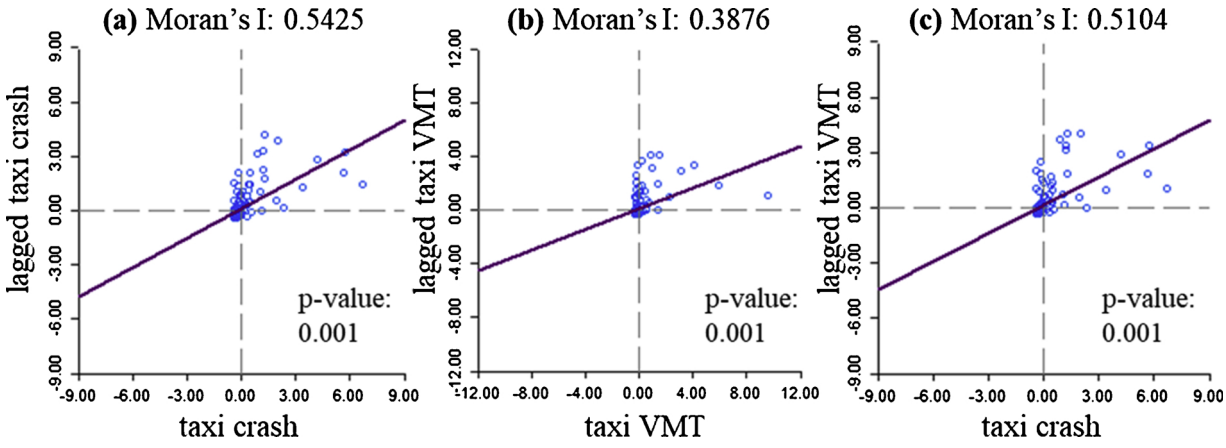

Fig. 4. Moran's I Scatter Plots. [(a) univariate Moran's I for taxi crash; (b) univariate Moran's I for taxi VMT; and (c) bivariate Moran's I of taxi crash and taxi VMT.

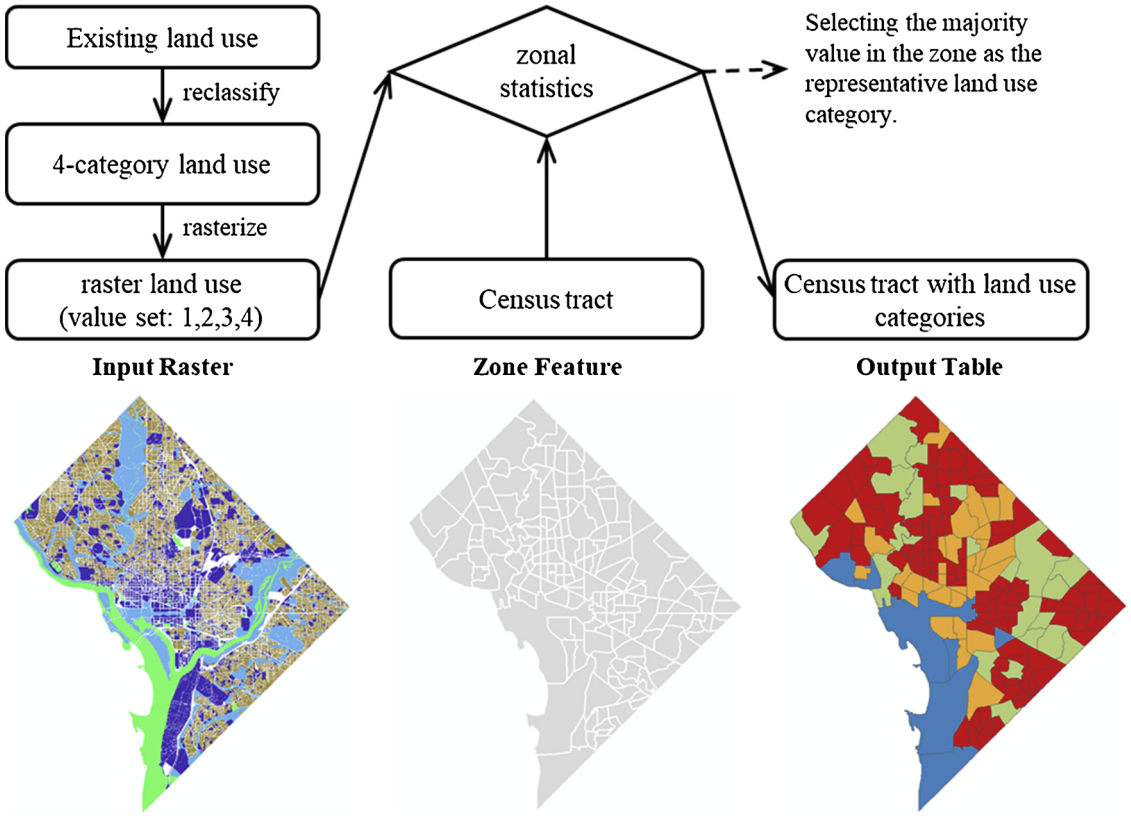

Fig. 5. The Process of Assigning Land Use Categories to Census Tracts.

Table 1
Descriptive Statistics of Prepared Data.

Variable	Description	Mean	Min	Max	Std.
Crash					
Taxi crash count	Count of taxi-related crashes in 2016	30.49	0	482	68.78
Transportation					
log(taxiVMT)	The logarithm of annual taxi miles traveled	9	6.38	14.07	1.56
Taxi pickup ratio	The ratio of taxi pickups to the sum of pickups and drop-offs	0.61	0.13	4.15	0.39
log(VMT)	The logarithm of annual average daily vehicle miles traveled	9.61	6.55	13.37	1.25
Intersection number	Count of intersections in each census tract	141.30	28	678	86.14
Road density	Ratio of total road length to census tract area (mile/mile^2)	20.67	0.25	40	8.33
Moving violations (> 15mph)	Count of moving violations in each census tract (10^3)	2.77	0	114.86	12.81
Bus stop number	Count of bus stops in each census tract	17.46	2	84	11.69
Land use (categorical)					
Residential (42.53%)	Census tract whose majority land use is residential.	—	—	—	—
Institutional (17.37%)	Census tract whose majority land use is institutional.	—	—	—	—
Open space (22.11%)	Census tract whose majority land use is open space.	—	—	—	—
Other (17.99%)	Census tract whose land use is mixed.	—	—	—	—
Demo-economic					
Population	Total population (10^3)	3.36	0.03	7.44	1.29
Median household income	Median income per household (10^3 $)	54.06	0	197.16	30.28
Total housing units	The number of housing units in each census tract (10^3)	1.66	0	5.38	0.80
POI number	Count of POI in each census tract	48.93	1	1332	109.83
Public school number	Count of public schools in each census tract	0.68	0	4	0.87

from the 2010 census tract data from the Open Data DC[4]. It mainly includes demographic (e.g. population, age, gender), housing (e.g. housing units, vacant houses), and economic (e.g. median household income). The GIS data of points of interest (POI) and public school were also obtained from the Open Data DC. Their numbers in each census tract were calculated using the spatial tools of ArcGIS. Table 1 provides the statistical description of all variables considered in this study.

4. Methodology

In this study, three types of models were developed and compared in modeling taxi-involved crash frequencies: (1) Poisson model; (2) Conditional Autoregressive model using distance matrix for explanatory purpose of spatial correlations; and (3) Conditional Autoregressive model using trip count matrix as explanatory factor of spatial components. The performance of above three models were evaluated in the context of Bayesian approaches conducted with the software package of WinBUGS (Spiegelhalter et al., 2003).

4.1. Model specification

4.1.1. Model 1. Poisson model

Poisson model is one of the frequently used model to analyze traffic crash count data. This model has been widely used because it can deal with non-negative integers. It is often shown to well capture the distribution of randomly occurred crashes. Thus, this model is considered as a benchmark in analyzing taxi crashes.

Based on the Poisson model, the probability of i^{th} zone entity (census tract) having y_i crashes is given by:

$$P(y_i|\lambda_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!} \tag{4}$$

where λ_i denotes the Poisson distribution mean parameter, which is specified by a function of explanatory variables X_{pi} ($p = 1, ..., P$ and P is the total number of explanatory variables):

$$\ln(\lambda_i) = \beta_0 + \sum_{p=1}^{P} \beta_p X_{pi} \tag{5}$$

where β_0 and β_p ($p = 1, ..., P$) are the regression parameters to be estimated. Eqs. (4) and (5) construct the Poisson model that serves as the basis for taxi crash frequency modeling.

4.1.2. Model 2. Conditional Autoregressive (CAR) model with distance matrix

According to the first law of Geography that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). It is often observed that both dependent variable and explanatory variables are clustered across the space. Without dealing with spatial components, the residuals of non-spatial model are often found to be spatially autocorrelated. Therefore, CAR model is introduced to capture the spatial autocorrelation. The equation is given as follows:

$$\ln(\lambda_i) = \beta_0 + \sum_{p=1}^{P} \beta_p X_{pi} + S_i \tag{6}$$

where, all components are the same with Eq. (2) except for the spatial autocorrelation S_i. The CAR effect used in this study was proposed by Besag et al. (1991), where the full conditional distribution for S_i is originally defined with the equation below:

$$S_i|S_{-i} \sim \mathrm{N}\left(\sum_{j\neq i} \frac{w_{ij}}{w_{i+}} S_j, \frac{\sigma_S^2}{w_{i+}}\right) \tag{7}$$

where, S_{-i} is the set of S_j for any $j \neq i$. w_{ij} determines the spatial autocorrelation between sites i and j, with maximum value of one if sites i and j are neighbored, and minimum value of zero. w_{i+} and σ_S^2 are the aggregation of weights and the variance for the set of S_i respectively.

In the CAR model, each site is weighted by its distance from the regression sites. The Gaussian and bi-square functions are commonly used to produce the weighting scheme as follow (Gill et al., 2017):

$$Gaussian: w_{ij} = \exp\left(-\frac{1}{2} \times \frac{dist_{ij}}{G_d}\right) \tag{8}$$

$$Bi-square: w_{ij} = \begin{cases} [1 - (dist_{ij}/G_d)^2]^2 if\ (dist_{ij} < G_d) \\ 0 \qquad\qquad otherwise \end{cases} \tag{9}$$

where, $dist$ is the distance matrix calculated by the centroid locations of the census tracts in D.C. $dist_{ij}$ indicates the Euclidean distance between census tracts i and j. The parameter G_d is a positive quantity known as the bandwidth. When G_d approaches infinity, w_{ij} approaches 1 and the CAR becomes a global model. In our study, we assume that G_d equals 3.1 miles (= 5 km), which is the approximate radius of the maximum inscribed circle in our study area.

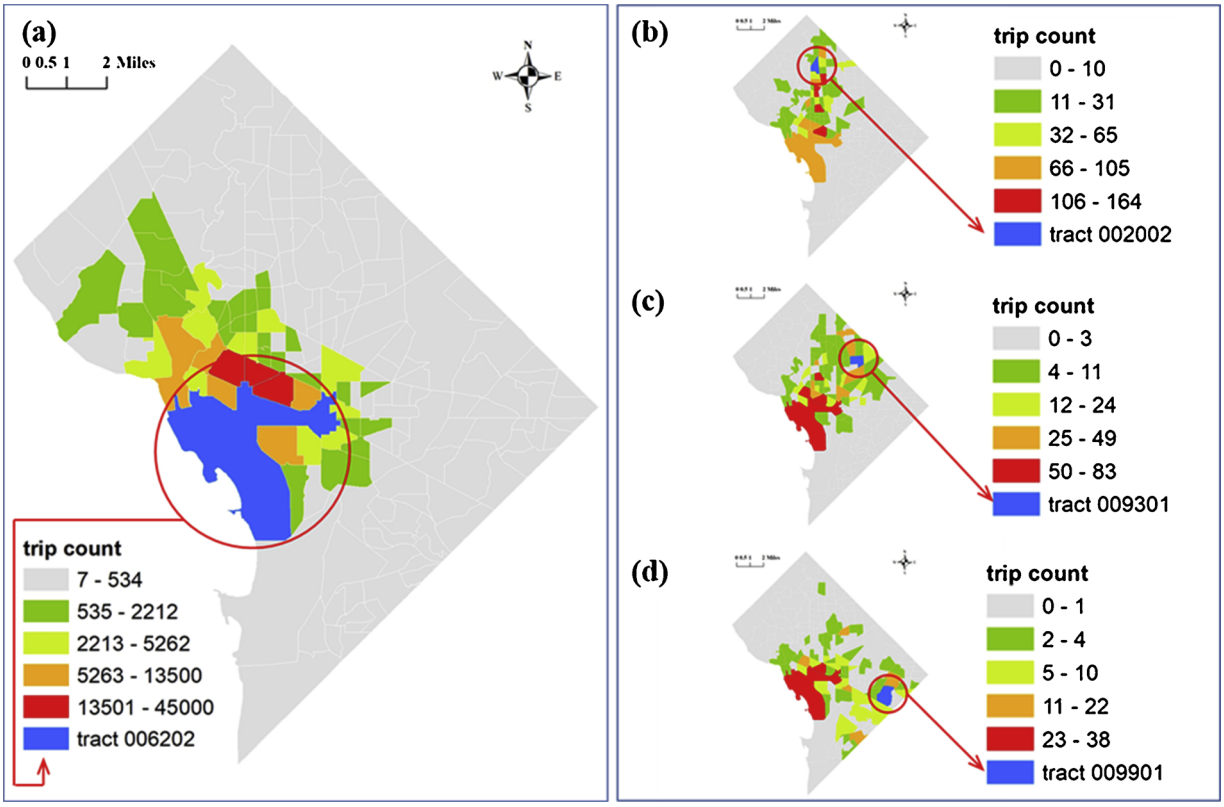

Fig. 6. Trip Aggregation between a Target Census Tract and Others in March 2016.

4.1.3. Model 3. Conditional Autoregressive model with taxi trip matrix

In previous CAR model, the first law of Geography may raise concerns about the spatial autocorrelation. Euclidean distance is the most commonly used quantitative metric in modeling crash frequencies (Li et al., 2013; Xu and Huang, 2015; Xie et al., 2019). However, in some fields of study, with the fast growth of informative data, there are alternative metrics to better construct the spatial autocorrelation matrix. For example, Buchin et al. (2014) created distorted map using travel times for the purpose of improving driving navigation. The maps in Fig. 6 show the trip counts between four selected census tracts with other zones during March 2016. The higher trip count indicates that a census tract is more intensively connected to the selected ones. For example, the pattern of tract (id = 006202) in Fig. 6(a) is consistent with the rule of spatial autocorrelation that nearer census tracts have higher connections. However, in Fig. 6(b)–(d), this rule no longer holds as we can see that taxis commuted more frequently between selected census tracts and the census tracts in downtown areas instead of nearer ones. In these cases, the Euclidean distance cannot reflect the taxi activities between census tracts as good as the trip count. Therefore, this paper proposes to use the weight matrix constructed by using the taxi O-D matrix introduced in the Section of Data Processing.

$$Gaussian: \mathrm{w_{ij}} = \exp\left(-\frac{1}{2} \times \frac{G_t}{\log(count_{ij})}\right) \tag{10}$$

$$Bi-square: \mathrm{w_{ij}} = \begin{cases} [1 - (G_t/\log(count_{ij}))^2]^2 & if\ (\log(count_{ij}) > G_t) \\ 0 & otherwise \end{cases} \tag{11}$$

As shown in the Eqs. (10) and (11), the variable *count* is obtained from the developed taxi O-D matrix. The parameter G_t is the customized bandwidth equaling the mean value of $\log(count_{ij})$. The role of the *count* matrix is different from the role of the *dist* matrix in the weight equations in Model 2. The range of the weight is from 0 to 1 and with higher count, the weight gets closer to 1.

4.2. Bayesian approach

4.2.1. Estimation of Bayesian models

All introduced models are estimated within the full Bayesian framework. The Bayesian method estimates models' parameters using posterior distributions, which is theoretically approximated by the following equation.

$$p(\theta|y) \propto L(y|\theta)\pi(\theta) \tag{12}$$

where, y is the vector of observed data; θ is the vector of parameters required for the likelihood function; $L(y|\theta)$ is the likelihood function; and $\pi(\theta)$ is the prior distribution of θ. The Bayesian inference is implemented by using the Markov Chain Monte Carlo (MCMC) algorithm (Gilks et al., 1998). Gibbs sampling (Geman and Geman, 1984) plays a primary role in the Markov chain Monte Carlo (MCMC) algorithm. In each iteration, unobserved stochastic values are drawn from their full conditional distribution given the current values of all the other quantities in the model (Lunn et al., 2000). The WinBUGS software package was used to provide an efficient computing tool for the calibration of Bayesian models using MCMC simulation (Spiegelhalter et al., 2002).

In the Poisson and CAR models, the prior distributions of the coefficients $\beta_i (i = 0, 1, ..., P)$ were set to follow normal distribution (0, 10^{-5}) because no prior information could be assumed. The CAR effect term S_i is determined by the following function.

$$S_i \sim car(adj, weight, num, tau) \tag{13}$$

where, *car* is the modeling function provided by WinBUGS with parameters *adj*, *weight*, and *num* that determine the weight matrix and neighboring relationships. *tau* is assumed to follow the distribution of $Gamma(0.5,\ 0.0005)$, which will allow the sampling of *tau* value in a wide range.

4.2.2. Assessment of Bayesian models

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

$$DIC = \overline{D(\theta)} + p_D \tag{14}$$

where, $D(\theta)$ is the Bayesian deviance of the estimated parameter θ, with $D(\theta) = -2\log(L(y|\theta)) + C$ and C is a constant. $\overline{D(\theta)}$ 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 five or greater suggests that the model with a smaller DIC should be favored.

5. Modeling results

5.1. Variable selection and estimation

To build converged models, no multi-collinearity should be detected among explanatory variables and suitable transformations should also be applied to certain variables. We conducted the correlation test using R studio. If two explanatory variables are highly correlated, the one with higher correlation with the dependent variable will be selected. On the other hand, the coefficients of the variables should be significantly different from zero to be included in the final model. This feature can be checked in WinBUGS during the simulation process. Based on these considerations, the final selected variables and modeling results are shown in Table 2. Based on the 95% Bayesian credible interval (BCI), road density, log(taxiVMT), and taxi pickup ratio are significant variables. Other than the non-spatial model, different land uses do not show distinctive impacts on taxi-involved crashes in spatial models 2–5.

It is worthy of mentioning that the taxi pickup ratio and log (taxiVMT) are significant in the modeling result. The negative coefficient of taxi pickup ratio indicates that taxi-involved crashes are more likely to occur in census tracts with more drop-off events. On the other hand, the positive coefficient of log(taxiVMT) shows that the occurrences of taxi-involved crash are closely related to taxi activities.

Our modeling results of taxi-crash indicate similarities and differences comparing to many other studies focusing on general vehicle crash modeling. In our results, for example, the significant variable road density has also been found to affect overall crash frequency in previous studies such as Xie et al. (2019). The positive coefficients of these variables indicate that the crashes are more likely to occur in zones with dense transportation facilities. On the other hand, unlike the significant variables of socio-demographic factor (e.g., population density, age, education) observed in the other studies (Lee et al., 2015), the occurrences of taxi-involved crash are found to be not sensitive to these factors. In addition, the selected land use variables (residential, open space, institution) are binary factors comparing to the "other" category. Since zeros are included within their credential intervals, comparing to "other" type of land use, the selected land use variables are not significant in models 2–5.

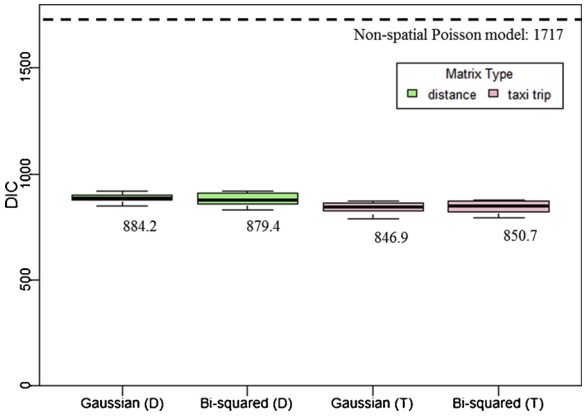

Fig. 7. DIC Evaluation Results with 15 Runs of Sampling Process.

5.2. Performance evaluation

The boxplots of DIC statistic is presented in Fig. 7. Each boxplot reflects the distribution of the simulation results of 15 sampling runs. Each model has been iterated for more than 200,000 times so that the model is converged before any data collection. The model convergence has been evaluated based on the with calculated Brooks-Gelman-Rubin (BGR) diagnostic plots (Spiegelhalter et al., 1996). We can observe that the non-spatial Poisson model has a stable DIC value of 1717. All other models considering spatial components have lower DICs compared to the non-spatial model, which indicates the need of spatial models for better performance. Specifically, among the four spatial models, the models using taxi trip-based weight matrix have lower DIC compared to those using distance-based weight matrix. Thus, the proposed methods

Table 2
Selected Variables and Modeling Results.

Variable		Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	Mean (Std.):	**−3.947** (0.134)	**−4.653** (0.441)	**−4.028** (0.504)	**−4.722** (0.409)	**−4.57** (0.373)
	95% BCI:	(−4.209, −3.684)	(−5.502, −3.754)	(−5.018, −3.04)	(−5.526, −3.946)	(−5.327, −3.82)
Taxi pickup ratio	Mean (Std.):	**−0.550** (0.054)	**−0.501** (0.120)	**−0.373** (0.126)	**−0.529** (0.122)	**−0.509** (0.133)
	95% BCI:	(−0.655, −0.445)	(−0.736, −0.269)	(−0.619, −0.125)	(−0.774, −0.29)	(-0.77, −0.247)
log(taxiVMT)	Mean (Std.):	**0.704** (0.017)	**0.665** (0.050)	**0.581** (0.054)	**0.693** (0.048)	**0.682** (0.046)
	95% Std.:	(0.671, 0.738)	(0.568, 0.763)	(0.475, 0.687)	(0.599, 0.788)	(0.592, 0.773)
log(POI)	Mean (Std.):	**0.064** (0.030)	**0.168** (0.072)	**0.217** (0.071)	**0.141** (0.071)	**0.168** (0.075)
	95% BCI:	(0.004, 0.124)	(0.029, 0.311)	(0.078, 0.357)	(0.002, 0.279)	(0.025, 0.318)
Residential	Mean (Std.):	**−0.533** (0.079)	**−0.052** (0.315)	**0.027** (0.333)	**−0206** (0.270)	**−0.410** (0.231)
	95% BCI:	(−0.685, −0.38)	(−0.67, 0.567)	(−0.617, 0.682)	(−0.723, 0.323)	(−0.858, 0.047)
Open space	Mean (Std.):	**−0.068** (0.078)	**0.489** (0.323)	**0.403** (0.342)	**0.409** (0.282)	**0.140** (0.244)
	95% BCI:	(−0.219, 0.084)	(−0.147, 1.108)	(−0.259, 1.077)	(−0.124, 0.973)	(−0.346, 0.603)
Institutional	Mean (Std.):	**−0.006** (0.061)	**0.417** (0.302)	**0.376** (0.324)	**0.316** (0.252)	**0.116** (0.208)
	95% BCI:	(−0.124, 0.113)	(-0.177, 1.010)	(−0.255, 1.014)	(−0.164, 0.819)	(−0.303, 0.52)
Road density	Mean (Std.):	**0.031** (0.002)	**0.035** (0.007)	**0.027** (0.008)	**0.038** (0.007)	**0.037** (0.007)
	95% BCI:	(0.027, 0.036)	(0.021, 0.05)	(0.011, 0.044)	(0.023, 0.052)	(0.024, 0.05)
Sigma	Mean (Std.):	–	**4.070** (0.312)	**1.915** (0.150)	**4.933** (0.373)	**2.291** (0.163)
	95% BCI:	–	(3.495, 4.709)	(1.641, 2.227)	(4.244, 4.917)	(1.992, 2.626)

Model 1: Poisson model; Model 2: Poisson-CAR model + Gaussian distance-based weight; Model 3: Poisson-CAR model + Bi-squared distance-based weight; Model 4: Poisson-CAR model + Gaussian taxi trip-based weight; Model 5: Poisson-CAR model + Bi-squared taxi trip-based weight; and Sigma = sqrt(1/tau): standard deviation.

Table 3
Moran's I Test of Modeled Residuals.

Model	Moran's I	p-value	Description
Model 1	0.307	0.005	Spatial autocorrelated at CI of 99.5%
Model 2	−0.032	0.200	No autocorrelation at CI of 99.5%
Model 3	0.020	0.155	No autocorrelation at CI of 99.5%
Model 4	0.037	0.750	No autocorrelation at CI of 99.5%
Model 5	0.056	0.055	No autocorrelation at CI of 99.5%

Model 1: Poisson model; Model 2: Poisson-CAR model + Gaussian distance-based weight; Model 3: Poisson-CAR model + Bi-squared distance-based weight; Model 4: Poisson-CAR model + Gaussian taxi trip-based weight; Model 5: Poisson-CAR model + Bi-squared taxi trip-based weight.

for taxi crash modeling using the weight matrix developed based on taxi trips have shown better performance than other models.

6. Discussion

6.1. Autocorrelation test of residuals

As mentioned previously, Moran's I is a method for testing spatial autocorrelation. As shown in Table 3, the residuals of non-spatial Poisson model are spatially autocorrelated. Recall that the distribution of taxi crash frequencies were tested to be spatially autocorrelated, this model fails to address the spatial effects and the residuals are still spatial-nonstationary. In contrast, the Models 2–5 have eliminated the residuals' spatial autocorrelation by adding spatial weights in the models. Therefore, the spatial models are preferred in modeling taxi-involved crashes based on the studied scenario.

6.2. The role of log(VMT) and log(taxiVMT)

The log(VMT) is the most frequently used variable in modeling crash frequencies. It reflects vehicles' activity within spatial entities. In general, log(VMT) has been demonstrated to be positively correlated to crash frequency. For example, in the modeling results of Mitra and Washington (2012); Li et al. (2013); Lee et al. (2015), and Cai et al. (2017), the coefficients of log(VMT) are positive and significantly different from zero. However, log(VMT) is no longer significant when modeling taxi-involved crashes in this study. One can anticipate that there can be few taxis passing by areas with large log(VMT) dominated by other types of vehicles. If taxi activity is replaced by log(VMT), the taxi crash model was not able to converge. On the other hand, taxi activity reflects the frequencies of taxi trips, which is more closely related to taxi-involved crashes. As shown in Fig. 8, the maps in Fig. 8(a)–(c) demonstrate the distributions of log(crash), log(taxiVMT), and log(VMT). We can see that log(crash) and log(taxiVMT) share similar spatial patterns by clustering at the central areas in D.C. Fig. 8(d)–(f) are the scatter plots of the three variables, where a clear positive linear relationship can be observed between the log(taxiVMT) and log(crash). On the other hand, the relationship between log(VMT) and log(crash) does not show strong linear relationship. Therefore, when modeling crashes involving a specific type of transportation mode, the variable reflecting its corresponding activities is highly recommended as a key explanatory variable (e.g., bike activities should be a good explanatory variable for modeling bike-involved crashes).

7. Conclusion

In this study, taxi-involved crashes in Washington, D.C. were analyzed and modeled by considering a set of factors related to transportation environment, land-use, and socio-demographic data. For this specific type of crashes, we proposed to use the historical trip data, instead of Euclidean distance, to generate the spatial weight matrix. Our proposed method is based on the conditional autoregressive model. Both Gaussian and Bi-squared functions were included in the computation of the spatial weight matrix. The non-spatial Poisson model, spatial Poisson-CAR model with distance-based weight matrix, and spatial Poisson-CAR model with the taxi trip-based weight matrix were compared. The modeling results suggest that the Poisson-CAR model outperforms the non-spatial model by successfully accounting for the spatial dependencies among variables. More specifically, our proposed Poisson-CAR models using the weight matrix constructed by taxi trips have better performance than Poisson-CAR models using the distance-based weight matrix. The results also show that taxi-VMT and road density are positively related to taxi-involved crash occurrence in a spatial unit. Meanwhile, the taxi-involved crashes are more likely to occur in census tracts with comparably higher taxi drop-off events. By conducting Moran's I tests, residuals of non-spatial Poisson model were found to be positively autocorrelated. After introducing CAR component to the Poisson model, the residuals were found to be randomly distributed over space, which has implied that the spatial Poisson-CAR models can account for spatial autocorrelation in modeling taxi-involved crashes.

Modeling taxi-involved crashes can help identify risky locations for taxi drivers considering up-to-date information. One of the key explanatory variables "log(taxiVMT)" allows models to leverage the value of massive taxi trip records. Its temporal scale can be flexible by using yearly, seasonally, monthly, or daily taxi trips, depending on the availability of historical data. Therefore, if given up-to-date information of taxis, such as taxis' distribution in most recent periods, the discussed models can be extended to provide timely taxi crash risk estimates.

Currently, this study is limited by the availability of taxi-involved crash data in Washington, D.C. The proposed model can be further tested at other places, where taxi and crashes data are both available. Besides, it is a worthy direction to explore and compare with other approaches for analyzing taxi crashes. For example, taxi crash rate can be calculated as the number of crashes per acreage in each census tract. This continuous response variable can be accommodated by Tobit models (Anastasopoulos et al., 2012a, b; Chen et al., 2014) and more risky variables would be potentially captured with MCMC methods. Meanwhile, zero-inflated structure can be added if the corresponding variable contains too many zeros (Anastasopoulos, 2016; Mannering et al., 2016; Fountas and Anastasopoulos, 2018). Also, if real-time taxi trajectory information and other data were available, it also possible to consider short-term taxi crash risk estimation like the ones introduced in other crash studies (Chen et al., 2018; Xie et al., 2018). The possibility of using such models for taxi crash analysis was scarce and deserves more experimenting.

Taxi safety is often affected by many factors such as individual and household specific characteristics, as well as spatiotemporal features. It should be noted that not all explanatory factors have been included in this paper due to the unavailability of the relevant data. The omitted factors will undoubtedly induce additional bias to the modeling results. In addition, unobserved heterogeneity is also likely to present due to other missing information, changeable effects of the same parameters on the response variables observed at different time/locations, etc. Thus, selection of appropriate functional forms and addressing misspecification issues, like in other crash modeling research, is strongly suggested. For example, one may consider introducing random parameters as a useful way to allow spatial unobserved heterogeneities in models (Anastasopoulos and Mannering, 2009). Also, modeling efforts such as the use of latent variables and random effect models worth exploring in future work. What's more, taking advantages of the informative spatial weight matrices constructed with taxi trip records or data alike, similar spatial modeling approaches can be extended to other traffic safety modeling practices in the future.

Declaration of Competing Interest

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

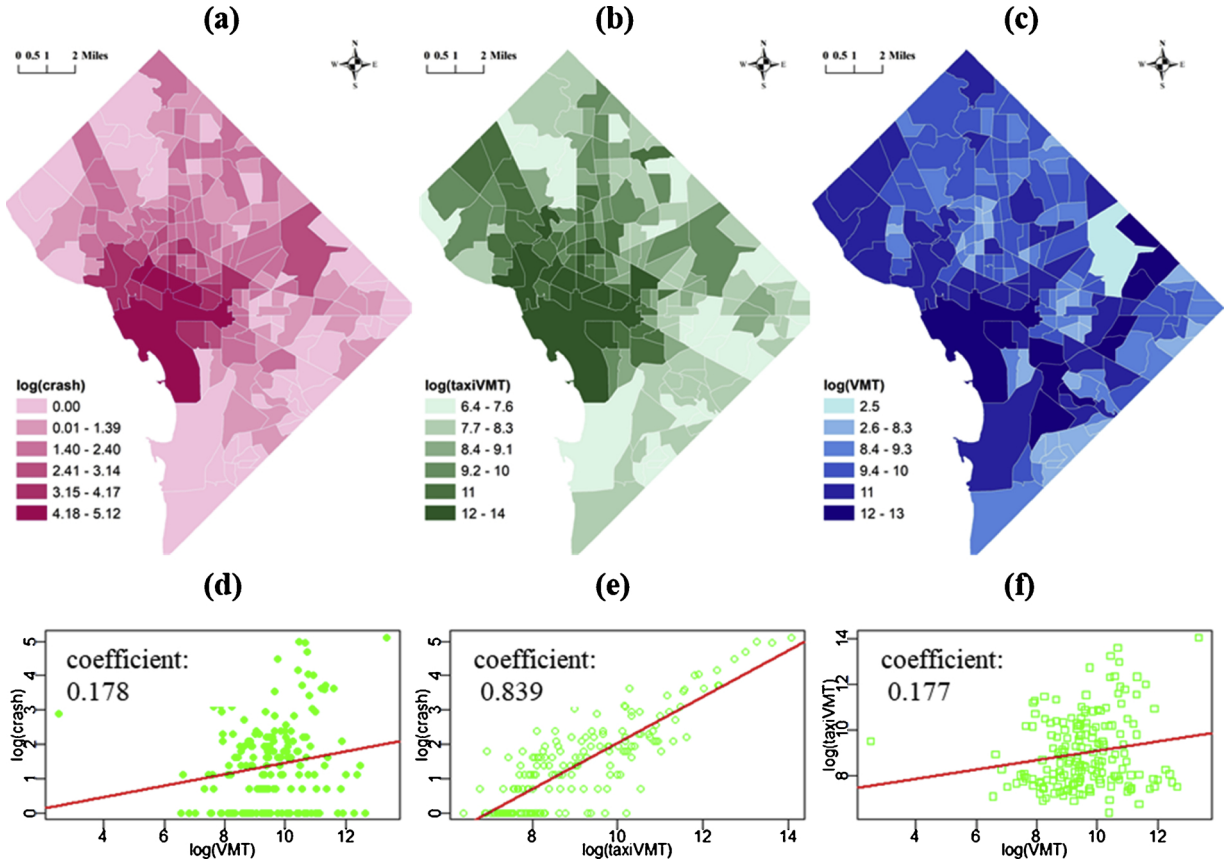

Fig. 8. Maps and Correlation Scatter Plots for log(crash), log(taxiVMT), and log(VMT).

Acknowledgments

The contents of this paper only reflect research views of the authors who are responsible for the facts and accuracy of the data and results presented herein. The contents and information derived from the paper do not necessarily reflect official views or policies of any sponsoring agencies. The author appreciate the related transportation agencies that have made the valuable data open access to the public. We also thank the reviewers for their sharing their great insights and suggestions that helped improve this paper.

References

Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accid. Anal. Prev. 38 (3), 618–625.
Anastasopoulos, P.C., 2016. Random parameters multivariate tobit and zero-inflated count data models: addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. Anal. Methods Accid. Res. 11, 17–32.
Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accid. Anal. Prev. 41 (1), 153–159.
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.
Andris, C., Bettencourt, L.M., 2014. Development, information and social connectivity in côte d'ivoire. Infrastruct. Complex. 1 (1), 1.
Anselin, L., Syabri, I., Smirnov, O., 2002. Visualizing multivariate spatial correlation with dynamically linked windows. Urbana 51, 61801.
Bao, J., Liu, P., Qin, X., Zhou, H., 2018. Understanding the effects of trip patterns on spatially aggregated crashes with large-scale taxi GPS data. Accid. Anal. Prev. 120, 281–294.
Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Stat. Math. 43 (1), 1–20.
Buchin, K., Van Goethem, A., Hoffmann, M., Van Kreveld, M., Speckmann, B., 2014. Travel-time maps: linear cartograms with fixed vertex locations. Proceedings of the International Conference on Geographic Information Science. pp. 18–33.
Bureau, C., 2007. Summary File 3. 2000 Census of Population and Housing. U.S. Department of Commerce.
Cai, Q., Abdel-Aty, M., Lee, J., Eluru, N., 2017. Comparative analysis of zonal systems for macro-level crash modeling. J. Saf. Res. 61, 157–166.
Chen, F., Chen, S., Ma, X., 2018. Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data. J. Saf. Res. 65, 153–159.
Chen, F., Ma, X., Chen, S., 2014. Refined-scale panel data crash rate analysis using random-effects tobit model. Accid. Anal. Prev. 73, 323–332.
Chen, P., 2015. Built environment factors in explaining the automobile-involved bicycle crash frequencies: a spatial statistic approach. Saf. Sci. 79, 336–343.
Cooper, P.J., 1997. The relationship between speeding behaviour (as measured by violation convictions) and crash involvement. J. Saf. Res. 28 (2), 83–95.
Dong, N., Huang, H., Zheng, L., 2015. Support vector machine in crash prediction at the level of traffic analysis zones: assessing the spatial proximity effects. Accid. Anal. Prev. 82, 192–198.
Fountas, G., Anastasopoulos, P.C., 2018. Analysis of accident injury-severity outcomes: the zero-inflated hierarchical ordered probit model with correlated disturbances. Anal. Methods Accid. Res. 20, 30–45.
Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. Pattern Anal. Mach. Intell. 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.
Gill, G.S., Cheng, W., Xie, M., Vo, T., Jia, X., Zhou, J., 2017. Evaluating influence of neighboring structures on spatial crash frequency modeling and site-ranking performance. Transp. Res. Rec. 2659 (1), 117–126.
Graham, D.J., Glaister, S., 2006. Spatial implications of transport pricing. J. Transp. Econ. Policy (JTEP) 40 (2), 173–201.
Huang, H., Abdel-Aty, M., Darwiche, A., 2010. County-level crash risk analysis in Florida: Bayesian spatial modeling. Transp. Res. Rec. J. Transp. Res. Board (2148), 27–37.
Joshi, B.D.B.M., 2018. 2018 Fact Book. New York City Taxi & Limousine Commission (TLC).
Lam, L.T., 2004. Environmental factors associated with crash-related mortality and injury among taxi drivers in New South Wales, Australia. Accid. Anal. Prev. 36 (5), 905–908.
Lascala, E.A., Gerber, D., Gruenewald, P.J., 2000. Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. Accid. Anal. Prev. 32 (5), 651–658.
Lee, J., Abdel-Aty, M., Jiang, X., 2015. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. Accid. Anal. Prev. 78, 146–154.
Levine, N., Kim, K.E., Nitz, L.H., 1995. Spatial analysis of honolulu motor vehicle crashes: Li. Zonal generators. Accid. Anal. Prev. 27 (5), 675–685.

Li, Z., Wang, W., Liu, P., Bigham, J.M., Ragland, D.R., 2013. Using geographically weighted poisson regression for county-level crash modeling in California. Saf. Sci. 58, 89–97.
Loo, B.P., 2006. Validating crash locations for quantitative spatial analysis: a gis-based approach. Accid. Anal. Prev. 38 (5), 879–886.
Lta, 2011. Passenger transport mode shares in world cities. Journeys. Land Transport Authority, Singapore.
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, M., Yan, X., Huang, H., Abdel-Aty, M., 2010. Safety of public transportation occupational drivers: risk perception, attitudes, and driving behavior. Transp. Res. Rec. 2145 (1), 72–79.
Ma, Q., Yang, H., Zhang, H., Xie, K., Wang, Z., 2019. Modeling and analysis of daily driving patterns of taxis in reshuffled ride-hailing service market. J. Transp. Eng. Part A Syst. https://doi.org/10.1061/JTEPBS.0000266.
Maag, U., Vanasse, C., Dionne, G., Laberge-Nadeau, C., 1997. Taxi drivers' accidents: how binocular vision problems are related to their rate and severity in terms of the number of victims. Accid. Anal. Prev. 29 (2), 217–224.
Machin, M.A., De Souza, J.M., 2004. Predicting health outcomes and safety behaviour in taxi drivers. Transp. Res. Part F Traffic Psychol. Behav. 7 (4–5), 257–270.
Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal. Methods Accid. Res. 11, 1–16.
Meliker, J.R., Maio, R.F., Zimmerman, M.A., Kim, H.M., Smith, S.C., Wilson, M.L., 2004. Spatial analysis of alcohol-related motor vehicle crash injuries in southeastern michigan. Accid. Anal. Prev. 36 (6), 1129–1135.
Mitra, S., Washington, S., 2012. On the significance of omitted variables in intersection crash modeling. Accid. Anal. Prev. 49, 439–448.
Moran, P.A., 1948. The interpretation of statistical maps. J. R. Stat. Soc. Ser. B (Methodological) 10 (2), 243–251.
Noland, R.B., Quddus, M.A., 2004. A spatially disaggregate analysis of road casualties in England. Accid. Anal. Prev. 36 (6), 973–984.
Pirdavani, A., Bellemans, T., Brijs, T., Wets, G., 2014. Application of geographically weighted regression technique in spatial analysis of fatal and injury crashes. J. Transp. Eng. 140 (8), 04014032.
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.
Siddiqui, C., Abdel-Aty, M., 2016. Geographical boundary dependency versus roadway hierarchy in macroscopic safety modeling: analysis with motor vehicle crash data. Transp. Res. Rec. J. Transp. Res. Board (2601), 59–71.
Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accid. Anal. Prev. 45, 382–391.
Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., 1996. Bayesian Inference Using Gibbs Sampling Manual (version Ii). BUGS 0.5. MRC Biostatistics Unit, Institute of Public Health, Cambridge, pp. 59.
Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. Winbugs User Manual. Version.
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.
Tang, J., Liu, F., Wang, Y., Wang, H., 2015. Uncovering urban human mobility from large scale taxi GPS data. Phys. A Stat. Mech. Appl. 438, 140–153.
Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. Econ. Geogr. 46 (sup1), 234–240.
Treno, A.J., Johnson, F.W., Remer, L.G., Gruenewald, P.J., 2007. The impact of outlet densities on alcohol-related crashes: a spatial panel approach. Accid. Anal. Prev. 39 (5), 894–901.
Veloso, M., Phithakkitnukoon, S., Bento, C., 2011. Urban mobility study using taxi traces. Proceedings of the Proceedings of the 2011 International Workshop on Trajectory Data Mining and Analysis. pp. 23–30.
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, Y., Li, L., Prato, C.G., 2018. The relation between working conditions, aberrant driving behaviour and crash propensity among taxi drivers in china. Accid. Anal. Prev.
Xie, K., Ozbay, K., Yang, H., 2019. A multivariate spatial approach to model crash counts by injury severity. Accid. Anal. Prev. 122, 189–198.
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., Yang, D., Ozbay, K., Yang, H., 2018. Use of real-world connected vehicle data in identifying high-risk locations based on a new surrogate safety measure. Accid. Anal. Prev. 125, 311–319.
Xie, X.-F., Wang, Z.J., 2018. Uncovering Urban Mobility and City Dynamics from Large-scale Taxi Origin-destination (od) Trips: Case Study in Washington Dc Area. Technical Report: WIO-TR-18-00. Wiomax LLC.
Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: random parameter versus geographically weighting. Accid. Anal. Prev. 75, 16–25.
Yang, H., Ozbay, K., Xie, K., 2015. Mining taxicab crashes in high-density urban areas. 2015 Road Safety & Simulation International Conference 119–129.
Yuan, J., Zheng, Y., Xie, X., 2012. Discovering regions of different functions in a city using human mobility and pois. Proceedings of the Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 186–194.
Zhao, Y., Zhang, J., He, X., 2015. Risk factors contributing to taxi involved crashes: a case study in Xi'an, China. Period. Polytech. Transp. Eng. 43 (4), 189–198.