

## Research Article

# Macrolevel Traffic Crash Analysis: A Spatial Econometric Model Approach

Shaohua Wang <sup>1,2</sup>, Yanyan Chen <sup>1</sup>, Jianling Huang,<sup>1,3</sup> Ning Chen <sup>1</sup> and Yao Lu<sup>1</sup>

<sup>1</sup>Beijing University of Technology, Beijing Key Laboratory of Traffic Engineering, Beijing 100124, China

<sup>2</sup>Tianjin University of Technology and Education, Tianjin Collaborative Innovation Center of Traffic Safety and Control, Tianjin 300222, China

<sup>3</sup>Beijing Transportation Information Center, Beijing 100161, China

Correspondence should be addressed to Yanyan Chen; [cdyan@bjut.edu.cn](mailto:cdyan@bjut.edu.cn)

Received 29 January 2019; Revised 27 March 2019; Accepted 21 April 2019; Published 6 May 2019

Academic Editor: Mahmoud Mesbah

Copyright © 2019 Shaohua Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study presents a spatial approach for the macrolevel traffic crashes analysis based on point-of-interest (POI) data and other related data from an open source. The spatial autoregression is explored by Moran's I Index with three spatial weight features (i.e., (a) Rook, (b) Queen, and (c) Euclidean distance). The traditional Ordinary Least Square (OLS) model, the Spatial Lag Model (SLM), the Spatial Error Model (SEM), and the Spatial Durbin Model (SDM) were developed to describe the spatial correlations among 2,114 Traffic Analysis Zones (TAZs) of Tianjin, one of the four municipalities in China. Results of the models indicated that the SDM with the Rook spatial weight feature is found to be the optimal spatial model to characterize the relationship of various variables and crashes. The results show that population density, consumption density, intersection density, and road density have significantly positive influence on traffic crashes, whereas company density, hotel density, and residential density have significant but negative effects in the local TAZ. The spillover effects coefficient of population density and road density are positive, indicating that the increase of these variables in the surrounding TAZs will lead to the increase of crashes in the target zone. The impacts of company density and hotel density are just the opposite. In general, the research findings can help transportation planners and managers better understand the general characteristics of traffic crashes and improve the situation of traffic security.

## 1. Introduction

In 2013, 1.25 million people were killed by the road traffic crashes worldwide and more than 50 million were injured [1]. Moreover, casualty rates caused by traffic crashes were significantly higher in the low or middle-income countries than that in the high-income countries. Taking China as an example, traffic crashes caused 58,523 deaths and 211,882 injuries in 2014[2]. In the same period, the crashes caused 32,744 deaths and 2,338 thousand injuries in 2014 in the United States [3]. With the rapid growth of economic development and autoownership, traffic crashes have become a leading cause of mortality in many developing countries, which attracted increasing attention from both the government and the public. Thus, it has become increasingly necessary for all the countries in the world to put considerable efforts to enhance the road safety, particularly in the developing countries.

The microlevel research focuses on the specific influencing factors of traffic crashes and casualties in the field of traffic safety research. The purpose of the microlevel research is to propose targeted measures to improve the vehicle, road, and environment. It is easy to understand the correlation between these direct contributing factors and traffic crashes. However, from another perspective, the macrolevel research focuses on the relationship between traffic crashes and society, economy, and environment. Compared with the microlevel safety research, the macrolevel safety analysis can identify safety problems more effectively in a larger area, which is more useful in helping establish a long-term planning policy to improve the traffic safety [4].

Though great progress had been made, the obtainment of data about traffic crashes and related influence factors is the main obstacle for crash analysis in under-developed countries [5]. In China, some scholars have used foreign

crashes data for analysis [6]. Other researchers focused on traffic violations such as drunk driving and speeding based on the traffic survey [7]. However, the situation is gradually changing. The road safety research platform (RSRP) was built to share traffic accident data by the Ministry of Public Security of People's Republic of China at 2015.

More importantly, with the continual development of data mining technology, open source data has raised more and more attention in recent years. The point-of-interest (POI) data are the more specific data of land use factors with exact information of location which are supposed to be highly related to the user characteristics and traffic crashes in both macro- and microaspects [5]. A POI database can be applied to describe the specific influence factors which are spatially correlated to the distribution of macrolevel traffic crashes. This study focuses on the spatial autocorrelation between the crashes and the impact of the different types of POI densities on the occurrences of crashes in the target units and adjacent units. The purpose of this paper is twofold: (a) to investigate the optimal spatial econometric model and (b) to evaluate the spatial direct effect and spillover effects of contributory factors that related to traffic crashes by using the POI dataset.

The remainder of the paper is organized as follows. In Section 2, a literature review of previous researches on traffic crashes and corresponding measurement methods are presented. Section 3 describes the POI data and crash data ( $N=26,121$ ) which are collected and processed within Traffic Analysis Zones (TAZs,  $N=2,114$ ) in Tianjin municipality of China. In Section 4, the author focuses on the spatial econometric model used from the following 3 aspects: (1) using Moran's I Index to check the spatial autocorrelation of the traffic crashes; (2) introducing the traditional Ordinary Least Square regression (OLS) model, the Spatial Lag Model (SLM), the Spatial Error Model (SEM), and testing the model by Lagrange Multiplier (LM); (3) furthermore, introducing the Spatial Durbin Model (SDM) to further estimate the spatial performance of the related factors. In Section 5, the empirical results are quantitative presented and analyzed. Section 6 provides the discussion, including policy implications and suggestions for further studies. Finally, the conclusions are presented and followed by references.

## 2. Literature Review

**2.1. Safety Covariates.** A wide variety of exposure variables were described in the traffic crash models in the previous studies. All the factors can be divided into five parts: (1) human related factors, including age, fatigue driving, and drunk driving [8, 9]; (2) vehicle factors, especially different types of vehicle and nonvehicle [10, 11]; (3) road factors, including geometric design features [12], number of lanes at road, number of intersections, and road density; (4) environment factors, including traffic characteristics, land use type [4], and weather condition. Meanwhile, socioeconomic variables, such as population, employment, and household income, were reported to be connected with the frequency of traffic crashes. Besides, the crash characteristics, such as collision type, were also pointed to affect the occurrence and severity of crashes. More often, many studies are based on a

comprehensive analysis of the factors above. As a new perspective for traffic crashes analysis, POI data are the specific data of land use factors with the exact information of location that are expected to be highly related to user characteristics and traffic crashes in both macro- and microaspects.

**2.2. Crash Modeling.** As collisions are believed to be discrete, nonnegative, and random, most of the previous literature related to the collision models is accountable for the Poisson regression models. Poisson model requires that the variance of data be equal to the mean, which is difficult to achieve in practice. Therefore, Poisson lognormal (PLN) model and Negative Binomial (NB) regression model are proposed to overcome these shortcomings [13]. However, the commonly used PLN model and NB model assume that the distribution of crashes is independent in space, while the crashes data have the spatial correlation characteristics in reality.

Most studies used spatial models to examine collision spatial correlation (spatial effects), since the cause of the crashes is often related to the specific influencing factors in the nearby areas. There are several ways to study the spatial effect in the models for count data, such as Bayesian hierarchical models and spatial econometric models [14]. The difference is that the former was designed to incorporate spatial autocorrelation in count models. Many different types of traffic accidents have been studied based on counting models [15, 16]. The spatial autocorrelation is realized by specifying a conditional autoregressive prior (CAR) model to the residual term of the link function in an ordinary Poisson regression [6]. However, there are also some scholars who point out that none of the transfer methods found broad reception so far [17]. The latter were originally designed for continuous data and count data must be transformed to meet the model's assumptions. The log transformation is widely used transformation to change a Poisson Log-linear Model to a Linear Log Model [14]. More importantly, spatial econometric models usually aim at estimating a parameter of spatial autocorrelation from the data and identifying spatial spillover effects. But the spatial spillover effects of explanatory and interpreted variables are not clearly described in the previous studies.

At last, spatial econometric models are mathematical models based on statistical theory that takes full account of population characteristics, social behavior, road traffic, and land attributes. Therefore, it is an important method to evaluate the safety levels in a certain area, and it is regarded as scientific and reasonable [5]. Meanwhile, variables used in previous studies were aggregated at different spatial levels, such as provinces, municipalities, specific regions, and TAZs [18]. The crash models based on macroscopic spatial level are beneficial to the implementation of corresponding traffic management before the crash emerge, so it is considered to be an active and effective crash prevention method.

**2.3. The Current Paper.** This study focuses on the macrolevel traffic crashes using the spatial econometric model in TAZs level. The POI data has been used to estimate the spatial spillover effects of traffic crashes influencing factors. In summary, we believe that by distinguishing the vital POI

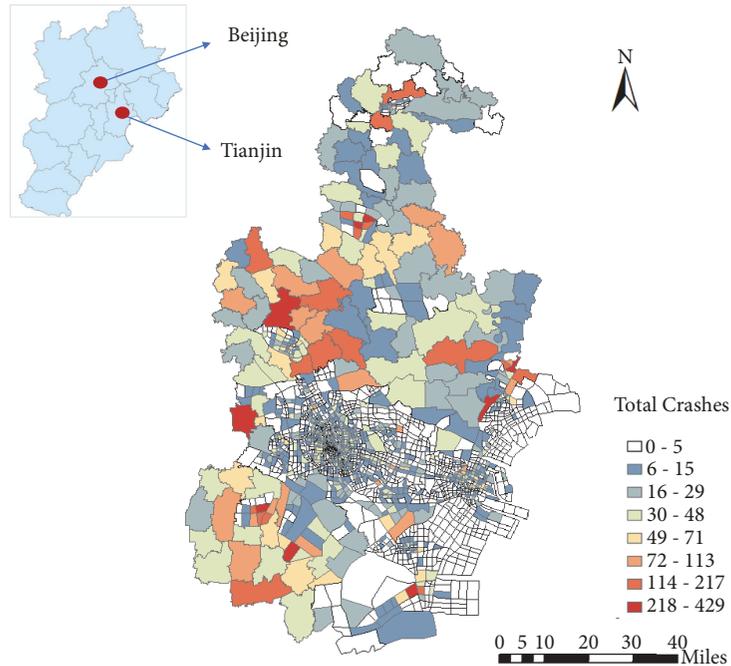


FIGURE 1: Spatial distribution of total crashes in the TAZ of Tianjin.

features and quantitative analyzing spatial spillover effects on the occurrence of traffic crashes we can contribute some recommendations to improve safety through traffic control policy management.

### 3. Data Description

**3.1. Data Preparation.** This study conducts an empirical research of Tianjin municipality of China. Tianjin is one of the four municipalities and is the industrial center in North China. It has a land area of 11 thousand square kilometers and a population of 15 million. By the end of 2017, Tianjin had 16 municipal districts with a total of 245 township-level districts. It is adjacent to Beijing, the capital of China, as shown in Figure 1.

The Traffic Analysis Zones (TAZs) scale has been proven to be reasonable and reliable in previous literature. With the aim of urban traffic planning and management, urban land is divided into several TAZs by Traffic Planning Department according to the principle of land attribute and so on [19]. In this study, the whole city is divided into 2,114 TAZs units according to the traditional transportation planning theory. As is shown in Figure 1, a total of 26,121 crashes records were obtained from the forensic institutions of the Justice Department of Tianjin for the period from 2011 to 2013. All the data collection behaviors are in conformity with the provisions of related laws and do not infringe upon personal privacy. With the aid of the data mining technology, different types of POI were collected based on web crawler. It contains nine categories of POI data, such as administration departments, schools, companies, hospitals, retail stores, restaurants, entertainments venues, hotels, and residential zones. The number of intersections and road densities were

obtained as the road factors. As another important related factor, population data were also collected from the Statistical Yearbook of the city and the Sixth National Township Census.

**3.2. Data Preprocessing.** Based on the Amap Application Programming Interface (API), the longitude and latitude coordinates of crashes location and POI location were extracted. As shown in Figure 2, the POI densities of each TAZ are all the corresponding POI within the zone's border divided by the area using zonal statistics. Firstly, Figure 2(a) shows that different types of POI point layers and spatial scale surface layers need to be joined. Secondly, Figure 2(b) shows that the total number of POI points in the corresponding space is calculated, then the area of every TAZs is obtained by "Computational Geometry" in the attribute field of the surface layer. Finally, point density can be obtained by dividing the total number of POI points by the area in "Field Calculator". The calculation method of traffic crash occurrence point density is similar. Geocoding was done with ArcGIS 10.2.

When there is a multiple collinearity problem between the explanatory variables, it may lead to an estimation deviation of the independent effect of the explanatory variables. Thus, correlation test and linear regression are used to diagnose any multicollinearity problems by SPSS 25 before model parameter estimation. There are strong correlations among retail store density, restaurant density, and entertainment venues density. Firstly, principal component analysis is used to reduce the dimensions of these three variables, and a unified consumption place density variable is obtained. Then all the data were processed through logarithm transforming to eliminate the heteroscedasticity or skewness, avoiding the model being too sensitive to the extreme value and reduce

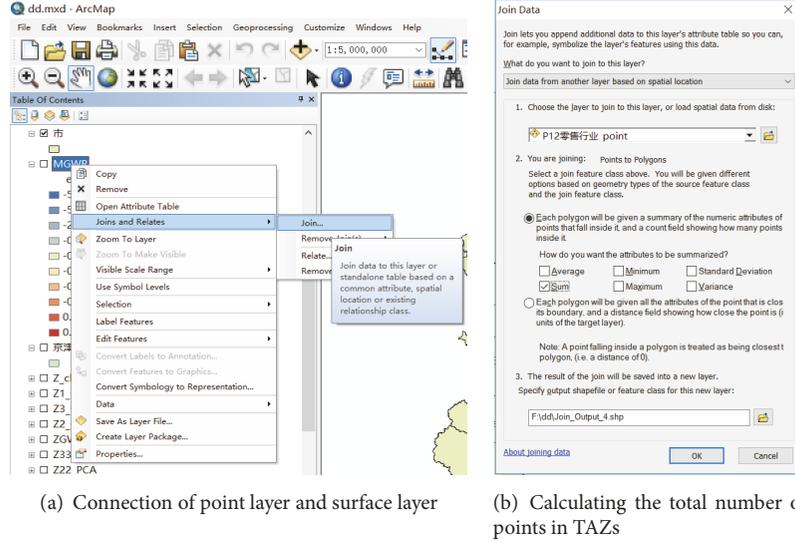


FIGURE 2: POI point density calculation.

the range of variables [20]. Finally, the Variance Inflation Factor (VIF) was used to determine the existence of any multicollinearity problems. Since all VIF values are less than 10, there is no multiple collinearity problem between the variables used in the model. The descriptive statistics of the explanatory variables are summarized in Table 1.

## 4. Methodologies

**4.1. Spatial Autocorrelation Analysis.** In this paper, the global Moran's I Index is applied to measure the spatial autocorrelation between each regional unit and the adjacent regional unit in the TAZ scale. The Moran's I Index is shown in formula (1). Only the Moran's Index passes the test; the spatial econometric model could be built to analyze the influence of the explanatory variables. The Moran's I value is between -1 and 1. If the Moran's I Index is less than zero, it means a negative correlation; while the value greater than 0, it means a positive correlation. A normalization statistics Z-score is usually used to verify the spatial autocorrelation of research units.

$$\text{Moran's } I = \frac{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_{i=1}^n \sum_{j=1}^n \omega_{ij}} \quad (1)$$

where  $x_i$  and  $x_j$  represent the alcohol-related road crashes density of zone  $i$  and zone  $j$ , respectively,  $n$  is the number of TAZ, and  $S$  denotes the standard deviation of the samples.  $\omega$  is the spatial weight matrix.

**4.2. Spatial Weight Feature.** The spatial weight matrix expresses the adjacency relationship between spatial units. It is pointed out that the choice of spatial weight matrix affects the degree of spatial autocorrelation observed in geographic studies.

The spatial weight matrix generally refers to an adjacency spatial weight matrix (0–1 swf) or a distance spatial weight

matrix (GCD swf). The 0–1 swf includes two types of "Rook" and "Queen". If there is a common boundary between  $\text{TAZ}_i$  and  $\text{TAZ}_j$ , "Rook" adjacency relation is defined as " $W_{ij}=1$ ". Otherwise, it is defined as " $W_{ij}=0$ ". The "Queen" adjacency relation includes a common boundary or a common point. The GCD swf means that distances (e.g., Euclidean distance and Manhattan distance) are used to reflect the correlation of different zones.  $W_{ij}$  equals the different types of weight between the zones. The "rook", the "queen" matrix, and "Euclidean distance" weight matrix are used to conduct the study. Every spatial weight matrix is normalized so that the factors of each row are summed up to unity before modeling process.

$$W = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1n} \\ W_{21} & W_{22} & \cdots & W_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ W_{n1} & W_{n2} & \cdots & W_{nn} \end{bmatrix} \quad (2)$$

**4.3. Spatial Regression Models.** The spatial econometric model is used to solve the spatial dependence problem. The general spatial econometric model with all types of interactions is shown as follows [21]:

$$Y = \rho WY + \alpha I_n + X\beta + WX\theta + u \quad (3)$$

$$u = \lambda W_u + \varepsilon \quad (4)$$

where  $Y$  denotes a dependent variable matrix,  $X$  denotes an explanatory variable matrix,  $WY$  denotes the endogenous interaction effects among the dependent variable,  $WX$  denotes the exogenous interaction effects among the independent variables, and  $W_u$  refers to the interaction effects among the disturbance term of the different units.  $I_n$  is a vector of ones associated with the constant term parameter  $\alpha$  to be estimated,  $\rho$  is called the spatial autoregressive

TABLE 1: Descriptive statistical table of variables.

Variables	Minimum	Maximum	Mean	S.D.	VIF
Ln_Population_Density	-0.298	4.959	3.050	1.067	1.814
Ln_Admin_Dept._Density	-1.921	2.410	0.171	0.581	2.669
Ln_School_Density	-1.672	2.594	0.206	0.613	3.518
Ln_Company_Density	-1.459	2.235	0.731	0.641	2.658
Ln_Hospital _Density	-2.107	2.167	0.244	0.608	4.899
Ln_Consumption _Density	-7.834	8.728	0.000	2.791	7.716
Ln_Hotel_Density	-2.015	1.799	0.096	0.452	2.020
Ln_Residential_Density	-2.107	2.241	0.337	0.669	4.110
Ln_Intersection _Density	-1.882	2.016	0.580	0.686	4.458
Ln_Road_Density	-2.930	1.345	0.350	0.521	3.656
Ln_Toal Crashes	0.000	2.632	0.652	0.603	

coefficient,  $\lambda$  is the spatial autocorrelation coefficient,  $\theta$ ,  $\beta$  are vectors with unknown parameters to be estimated, and  $\varepsilon$  is a vector of disturbance errors.

(1) When the parameter  $\theta = 0$  and  $\lambda = 0$ , the general model changes to the Spatial Autoregressive (SAR) model as shown in formula (5). The SAR model is mainly used to analyze the interaction between dependent variables. The magnitude of  $\rho$  reflects the degree of spatial diffusion and spillover. If the value of  $\rho$  is significant, it indicates that there is a spatial dependency between the interpreted variables.

$$Y = \rho WY + \alpha l_n + X\beta + \varepsilon \quad (5)$$

(2) When the parameter  $\rho = 0$  and  $\theta = 0$ , the Spatial Error Model (SEM) is defined as follows. The Spatial Error Model (SEM) mainly concentrates on the spatial interaction effect of the missing items in the modeling process.

$$\begin{aligned} Y &= \alpha l_n + X\beta + u \\ u &= \lambda W_u + \varepsilon \end{aligned} \quad (6)$$

(3) When the parameter  $\lambda = 0$ , the Spatial Durbin Model (SDM) can be expressed as formula (7). Not only is the

spatially lagged correlation of dependent variables considered, but also the explanatory variables are considered. The SDM will produce unbiased coefficient estimates for different data and is widely used [22].

$$Y = \rho WY + \alpha l_n + X\beta + WX\theta + u \quad (7)$$

More importantly, the space Durbin model also considers the impact of interpretation variables on traffic crashes in the adjacent zones. The SDM model can be rewritten as [23]

$$(I_n - \rho W) y = \alpha l_n + X\beta + WX\theta + u \quad (8)$$

$$y = \sum_{r=1}^k S_r(W) x_r + V(W) l_n \alpha + V(W) u$$

$$S_r(W) = V(W) (I_n \beta_r + W\theta_r) \quad (9)$$

$$\begin{aligned} V(W) &= (I_n - \rho W)^{-1} \\ &= I_n + \rho W + \rho^2 W^2 + \rho^3 W^3 + \dots \end{aligned}$$

Equation (8) can be rewritten as

$$\begin{aligned} &\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \\ &= \sum_{r=1}^k \begin{pmatrix} S_r(W)_{11} & S_r(W)_{12} & \cdots & S_r(W)_{1n} \\ S_r(W)_{n1} & S_r(W)_{22} & \cdots & S_r(W)_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ S_r(W)_{n1} & S_r(W)_{n2} & \cdots & S_r(W)_{nn} \end{pmatrix} \\ &\cdot \begin{pmatrix} x_{1r} \\ x_{2r} \\ \vdots \\ x_{nr} \end{pmatrix} + V(W) \alpha l_n + V(W) u \end{aligned} \quad (10)$$

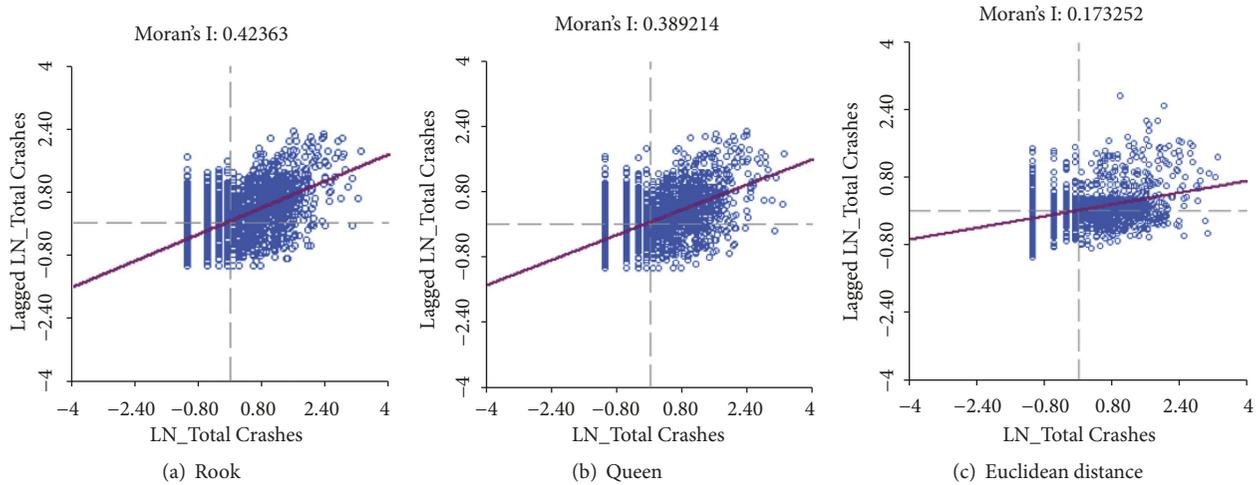


FIGURE 3: Moran's I Index under different spatial weighting matrices.

Among them, the mean value of the diagonal elements of the matrix represents the direct effect of the explanatory variables, and the mean of the diagonal elements represents the spatial spillover effect of the explanatory variables.

**4.4. Spatial Econometric Analysis Process.** After processing the related data, a series of related spatial econometric model are built to obtain the regression results. The main steps are shown as Steps 1–4.

**Step 1.** The Moran's Index was calculated to explore spatial autoregression of crashes based on different spatial features before the regression analysis.

**Step 2.** The traditional OLS model is built. The model is checked by the F-test and the variables are evaluated by the t-test. The Lagrange Multiplier test (LM) and the Robust Lagrange Multiplier test (Robust LM) are used to judge whether a spatial lag or spatial error terms exist in the model.

**Step 3.** If the test above indicates the existence of the spatial effect, the SAR and SEM model are built. The log-likelihood, Akaike's Information Criterion (AIC), and Schwarz information criterion (SIC) [24] are used to compare the OLS, SAR and SEM models.

**Step 4.** The SDM will further describe the direct effect, indirect effect, and total effect of the alcohol outlets densities on road crashes.

In this study, parameter estimates were obtained by the maximum likelihood (ML) method based on the above tests. All spatial regression models and tests were conducted by using Elhorst's spatial econometrics MATLAB toolbox and the software Geoda 1.12.

## 5. Results

**5.1. Results of Spatial Autoregression.** The spatial autocorrelation of all the crashes was analyzed based on different

TABLE 2: Results of Lagrange Multiplier tests.

LM test	Value	p value
LM (lag)	72.093***	0.001
Robust LM (lag)	37.998***	0.001
LM (error)	39.047***	0.001
Robust LM (error)	4.952**	0.026

Note: \*\*\* and \*\* indicate significance at the 1% and 5% levels, respectively.

spatial weighting matrices. The Moran's Index is calculated by Geoda1.12 and is shown in Figure 3. All the Moran's Indexes are positive and significant. The Z score is far greater than 1.96 and the P values are 0.001, which all passed the 5% level of significance test. It means the road crashes have significantly positive spatial relationship under all the three different spatial weights. It also indicates the maximum value of Moran's I (0.424) under the "Rook" spatial weight matrix. Thus, the empirical research is based on the "Rook" matrix in this study.

**5.2. Conventional Results of the Spatial Regression Models.** As shown in Tables 2 and 3, the OLS, SAR, and SEM model estimation were built and analyzed based on the cross-section data.

Firstly, the results show that the OLS model passed through F-test. The Ln\_Admin Dept.\_Density, Ln\_School\_Density, and Ln\_Hospital\_Density were removed because they did not pass the t-test. Other explanatory variables were significant at 5% level.

Secondly, the results of LM-lag and LM-error reject the null hypothesis at 1% level, Robust LM-lag passes the 1% level of significance test, and Robust LM-error passes the 5% level of significance test. The diagnostic tests indicated clear autocorrelation problems.

Thirdly, the adjust  $R^2$  of OLS model is 0.34 which shows a relatively strong explanatory power for the occurrence of the crashes. However, unlike some studies, it is important to note

TABLE 3: Estimation results of OLS, SAR, and SEM model.

Variables	OLS	SAR	SEM
Ln_Population_Density	0.184* * *	0.161* * *	0.171* * *
Ln_Company_Density	-0.667* * *	-0.619* * *	-0.642* * *
Ln_Consumption_Density	0.089* * *	0.086* * *	0.087* * *
Ln_Hotel_Density	-0.095* * *	-0.091* * *	-0.090* * *
Ln_Residential_Density	-0.071* *	-0.077* *	-0.071* *
Ln_Intersection_Density	0.066* *	0.072* *	0.073* *
Ln_Road_Density	0.239* * *	0.216* * *	0.219* * *
$\rho/\lambda$	/	0.219* * *	0.195* * *
Adjust R <sup>2</sup>	0.342	0.436	0.434
Log-likelihood	-1485.7	-718.872	-733.734
AIC	2987.690	2740.370	2765.440
SIC	3032.940	2791.280	2810.690

Note: \* \* \* and \* \* indicate significance at the 1% and 5 % levels, respectively.

TABLE 4: Estimation results of SDM model.

Variables	Coefficient	Variables	Coefficient
Ln_Population_Density	0.147*	$w \times \text{Ln\_Population\_Density}$	0.058* * *
Ln_Company_Density	-0.612* * *	$w \times \text{Ln\_Company\_Density}$	-0.056
Ln_Consumption_Density	0.086* *	$w \times \text{Ln\_Consumption\_Density}$	0.002
Ln_Hotel_Density	-0.089* * *	$w \times \text{Ln\_Hotel\_Density}$	-0.027
Ln_Residential_Density	-0.084* *	$w \times \text{Ln\_Residential\_Density}$	0.009
Ln_Intersection_Density	0.075* *	$w \times \text{Ln\_Intersection\_Density}$	-0.080
Ln_Road_Density	0.208* * *	$w \times \text{Ln\_Road\_Density}$	0.124*
$\rho$		0.183* * *	
Log-likelihood		-713.783	

Note: \* \* \*, \* \*, and \* indicate significance at the 1%, 5%, and 10% levels, respectively.

that the value of goodness-of-fit cannot be used as basis for spatial model comparison and selection [21].

At the same time, the outcomes indicate that the coefficient  $\rho/\lambda$  connected with spatial autocorrelation is positive and significantly different from zero for all spatial estimations. It shows that the dependent variable in neighboring TAZs has significantly influenced the crash frequency. More importantly, the spatial model has been found to have a better interpretation than the traditional OLS model. The SAR model has the highest value of log-likelihood and the lowest value of AIC and SIC. Hence, the results indicate that it is necessary to consider spatial factors while building a crash model, and it has been proved statically that the SAR model would be a better approach than the SEM model.

As is shown in Table 4, the SDM is introduced to further estimate the interaction effects and exogenous interaction effects. Meanwhile, the average value of these variables from neighboring zones is also included in the explanatory variables since this is a Spatial Durbin Model. The SDM has the highest log-likelihood (-713.783). The log-likelihood function value indicates that the SDM outperforms other models. Finally, the population density, consumption density, intersection density, and road density have a significantly positive influence on crashes. At the same time, the company

density, hotel density, and residential density also have a significantly negative effect on the crash risks.

In addition, direct effects, spillover effects, and total effects of explanatory variables are further studied. As is shown in Table 5, the increase of the population density, consumption density, intersection density, and road density will increase the possibility of traffic crashes in the target TAZs. Especially, for every 1% increase in the average of population density and road density, the crashes of local zone increase around 0.15% and 0.21%. Meanwhile, the company density, hotel density, and residential density have a significant negative effect on the local zone. It is an important observation that a percent increase in company decreases the traffic crashes level by 0.62.

The spillover effect coefficient of population density and road density is positive, indicating that the increase of these variables in the surrounding areas will lead to the increase of crashes in the target areas. The impact of company density and hotel density is just the opposite. The result of consumption density is same as other documents [25]; it does not have the indirect effects on the crashes of neighboring zones. Overall, the total effects of variables are like the spillover effect. The increase of population density and road density will lead to the corresponding increase of traffic crashes. The

TABLE 5: Direct, indirect, and total effects of SDM model.

Variables	Direct Effect		Indirect Effect		Total Effect	
	Coefficient	T Statistic	Coefficient	T Statistic	Coefficient	T Statistic
Ln_Population_Density	0.154*	1.951*	0.102***	4.387***	0.256**	2.544**
Ln_Company_Density	-0.620***	-16.946***	-0.199**	-2.485**	-0.819***	-8.503***
Ln_Consumption_Density	0.089**	2.297**	0.023	0.252	0.112	1.102
Ln_Hotel_Density	-0.091***	-5.641***	-0.052*	-1.736*	-0.143***	-4.285***
Ln_Residential_density	-0.084**	-2.509**	-0.004	-0.072	-0.088	-1.402
Ln_Intersection_Density	0.070**	2.248**	-0.079	-1.145	-0.009	-0.121
Ln_Road_Density	0.214***	6.667***	0.187**	2.408**	0.401***	4.631***

Note: \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% levels, respectively.

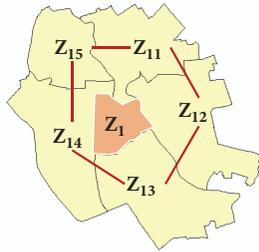


FIGURE 4: Impact analysis diagram of traffic network conditions.

higher the densities of companies and hotels are, the lower the frequency of road crashes is.

As is shown in Table 5, through the analysis of road density and intersection density, it is found that the direct effects are both positive. This is due to the complexity of road network and the increase of traffic conflicts caused by higher road density and intersection density in target region lead to more crashes, which is similar to the existing traffic safety literature. But interestingly, the spillover effect of intersection density is negative, while the spillover effect of road network density is positive, which is different from people's general understanding. The reason maybe is the spatial interaction effect between the areas. As is shown in Figure 4, it is assumed that traffic zone  $Z_i$  is adjacent to five traffic zones. If there are many intersections in the adjacent zones, drivers will face more traffic conflicts and risks in the process of going through the surrounding traffic zones to  $Z_i$ . Therefore, the increase of intersections in adjacent areas will lead to a reduction in crashes in the region. However, the indirect effect and total effect for intersection density are not significant. It means that the negative effect of intersection density should not be considered.

Although the intersection density and road density are all represent road complexity, the latter is slightly different. The density of the road network in the adjacent area represents the accessibility to the  $Z_i$  area. The greater the road density in the adjacent areas, the easier the driver will reach the  $Z_i$  area, which will lead to crash in the area. Therefore, the density of adjacent intersections inhibits the arrival of vehicles in the region, so the spillover effect is negative, while the density of road network in adjacent areas promotes the arrival of vehicles in the region, so the spillover effect is positive.

## 6. Discussion

This study proposed a spatial econometric model to evaluate the spatial direct effect, spillover effects and total effects based on open source POI data and related data. Our main findings are as follows.

**6.1. Spatial Effect of Variables.** By using multivariate regression and spatial analyses, the results confirm that a clear spatial association exists among the traffic crashes in Tianjin of China. A high population density and road density in the region and adjacent areas will lead to an increase of crashes. More interestingly, this study can serve as a proof-of-concept that the direct effect and spillover effect of the company density and hotel density are negative, which indicates that the increase of the hotel density will lead to the decrease of crash rate in the target area and adjacent area. It is partly because that most companies have strict rules in working. Another reason to explain this might be that a free shuttle bus operates between the company location and the homeplace; the professional driver of the bus will reduce the crashes. The reason about hotel density may be that most of people living in hotels are strangers and most of them use bus or taxis rather than self-driving to hotels, which results in a lower probability of crashes.

**6.2. Policy Implications.** Based on the findings of this empirical study, there are some important and direct implications for transport policies. Firstly, the results will be helpful to make traffic safety planning based on the POI data. Traffic safety planning requires that traffic safety be taken into account at all levels of traffic planning, emphasizing the prediction and planning of safety level from macro to microlevel. The macrolevel crash prediction model may aid transportation agencies in more proactively incorporating safety consideration into the long-term transportation planning process [26]. Based on the spatial econometric model described in this paper, traffic planners can predict the level of traffic safety at the transportation planning zones, select more reasonable planning schemes, and establish more useful crash prevention facilities. Secondly, the research also has some practical significance to improve safety through traffic control policy management. For example, it is very important to choose a suitable distance when doing traffic crashes hotspot analysis or density analysis. The appropriate

distance threshold or radius can be determined by examining the spatial correlation of different types of traffic crashes at different distances. The selected distance or radius is helpful for the traffic policeman to determine the scope for implementing traffic management. Finally, more publicity and education should be carried out on the zones to enhance the citizens' awareness of traffic safety based on the determined POIs related to traffic crashes. All of these applications will contribute to the improvement of traffic safety level.

**6.3. Limitations and Future Directions.** Though some important discoveries have been revealed in this study, there are still some research limitations. First, the POI densities only reflect the number of different points of mixed land use, but not the size. Thus, further research needs to be carried out through more detailed investigations to get more specific data about different types of POI. Second, different types of traffic crashes have different influence factors. For example, the occurrence of drunk driving may be more likely to relate to restaurants density. It is necessary to study on typical traffic crashes in the future. Third, spatial heterogeneity of crashes should not be ignored.

## 7. Conclusions

This empirical research investigated different techniques to estimate the correlation of POI data and other related data with traffic crashes at a macrolevel. Four types of models are discussed in the paper, i.e., OLS, SAR, SEM, and SDM. Data from 2,114 TAZs in the city of Tianjin were applied to develop the macrolevel crash models which incorporate the original covariates related to land use and environment. The results indicate that spatial effect, especially the spillover effect, should be considered to build the crash model and the spatial model can be much better than the traditional OLS model. In contrast with the safety performance of spatial regression models, it has shown that the Spatial Durbin Model enjoys the highest degree of fitting.

## Data Availability

(1) Previously reported crashes and POI data were used to support this study and are available at <https://doi.org/10.1016/j.aap.2018.09.018>. The prior study is cited at relevant places within the text as references. (2) The crash records analyzed during the current study are confidential and cannot be used on personal computers due to the personal data privacy policy of the Justice Department and Public Security Bureau of Tianjin but other related data (POI data, etc.) are available from the corresponding author upon reasonable request. (3) The POI data can be easy to obtain from Google Map Application Programming Interface (API) and Baidu (China's biggest search engine) Map API. Of course, all the POI data, population data, and road data are available from the corresponding author upon request.

## Conflicts of Interest

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

## Acknowledgments

This research was funded by the National Key Research and Development Program of China, grant number 2017YFC0803903, the Key Project of the Natural Science Foundation of Tianjin city, grant number 16JCZDJC38200, and the Major Projects in Artificial Intelligence Science and Technology of Tianjin city, grant number 17ZXRGGX00070.

## References

- [1] WHO, "Road traffic injuries. world health organization," 2018.
- [2] *Traffic Management Bureau, Ministry of Public Security, Annual Report of Road Traffic Crashes in People's Republic of China*, 2014.
- [3] T. A. Board and M. Wigan, "National highway traffic safety administration," *Transportation Research Record*, pp. 1–9, 2015.
- [4] X. Wang, J. Yang, C. Lee, Z. Ji, and S. You, "Macro-level safety analysis of pedestrian crashes in Shanghai, China," *Accident Analysis & Prevention*, vol. 96, pp. 12–21, 2016.
- [5] R. Jia, A. Khadka, and I. Kim, "Traffic crash analysis with point-of-interest spatial clustering," *Accident Analysis & Prevention*, vol. 121, pp. 223–230, 2018.
- [6] H. Huang, B. Song, P. Xu, Q. Zeng, J. Lee, and M. Abdel-Aty, "Macro and micro models for zonal crash prediction with application in hot zones identification," *Journal of Transport Geography*, vol. 54, pp. 248–256, 2016.
- [7] K. Jia et al., "Drunk driving offenders knowledge and behavior in relation to alcohol-involved driving in Yinchuan and a comparison with Guangzhou," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 38, pp. 182–193, 2016.
- [8] A. Kuerbis and P. Sacco, "The impact of retirement on the drinking patterns of older adults: A review," *Addictive Behaviors*, vol. 37, no. 5, pp. 587–595, 2012.
- [9] W. Roberts and M. T. Fillmore, "Curbing the DUI offender's self-efficacy to drink and drive: A laboratory study," *Drug and Alcohol Dependence*, vol. 172, pp. 73–79, 2017.
- [10] Y. Peng, X. Wang, S. Peng, H. Huang, G. Tian, and H. Jia, "Investigation on the injuries of drivers and copilots in rear-end crashes between trucks based on real world accident data in China," *Future Generation Computer Systems*, vol. 86, pp. 1251–1258, 2018.
- [11] M. Moomen, M. Rezapour, and K. Ksaibati, "An investigation of influential factors of downgrade truck crashes: A logistic regression approach," *Journal of Traffic and Transportation Engineering (English Edition)*, 2018.
- [12] X. Zhao, X. Zhang, J. Rong, and J. Ma, "Identifying method of drunk driving based on driving behavior," *International Journal of Computational Intelligence Systems*, vol. 4, no. 3, pp. 361–369, 2011.
- [13] M. Mohammadi, G. Shafabakhsh, and A. Naderan, "Effects of modal shares on crash frequencies at aggregate level," *Accident Analysis & Prevention*, vol. 120, pp. 295–303, 2018.
- [14] P. Simões, I. Natário, and Spatial., "Econometric approaches for count data: an overview and new directions," *International Scholarly and Scientific Research & Innovation*, vol. 10, no. 1, pp. 348–357, 2016.
- [15] Y. Guo, A. Osama, and T. Sayed, "A cross-comparison of different techniques for modeling macro-level cyclist crashes," *Accident Analysis & Prevention*, vol. 113, pp. 38–46, 2018.

- [16] T. Bhowmik, S. Yasmin, and N. Eluru, "A joint econometric approach for modeling crash counts by collision type," *Analytic Methods in Accident Research*, vol. 19, pp. 16–32, 2018.
- [17] Stephanie Glaser, "A review of spatial econometric models for count data," *HOHENHEIM Discussion Papers in Business, Economics and Social Sciences*, pp. 1–31, 2017.
- [18] X. Jiang, M. Abdel-Aty, J. Hu, and J. Lee, "Investigating macro-level hotzone identification and variable importance using big data: A random forest models approach," *Neurocomputing*, vol. 181, pp. 53–63, 2016.
- [19] J. Lee, M. Abdel-Aty, and X. Jiang, "Development of zone system for macro-level traffic safety analysis," *Journal of Transport Geography*, vol. 38, pp. 13–21, 2014.
- [20] Woolridge, *Introductory Econometrics - A Modern Approach*, Cengage Learning Press, Boston, USA, 2015.
- [21] J. P. Elhorst, *Spatial Econometrics: From Cross Sectional Data to Spatial Panels*, Springer Press, Groningen, Netherlands, 2010.
- [22] J. LeSage and R. K. Pace, *Introduction to Spatial Econometrics*, CRC Press, Boca Raton, USA, 2009.
- [23] A. Tientao, D. Legros, and M. C. Pichery, "Technology spillover and TFP growth: A spatial Durbin model," *International Economics*, vol. 145, pp. 21–31, 2016.
- [24] C. Liu and A. Sharma, "Exploring spatio-temporal effects in traffic crash trend analysis," *Analytic Methods in Accident Research*, vol. 16, pp. 104–116, 2017.
- [25] D. M. Gorman, P. W. Speer, P. J. Gruenewald, and E. W. Labouvie, "Spatial dynamics of alcohol availability, neighborhood structure and violent crime," *Journal of Studies on Alcohol*, vol. 62, no. 5, pp. 628–636, 2001.
- [26] B. Xie, Z. An, Y. Zheng, and Z. Li, "Incorporating transportation safety into land use planning: Pre-assessment of land use conversion effects on severe crashes in urban China," *Applied Geography*, vol. 103, pp. 1–11, 2019.

