

Research Article

Developing a New Spatial Unit for Macroscopic Safety Evaluation Based on Traffic Density Homogeneity

Chen Wang , Lin Liu, and Chengcheng Xu 

School of Transportation, Southeast University, Nanjing 210096, China

Correspondence should be addressed to Chengcheng Xu; xuchengcheng@seu.edu.cn

Received 13 November 2019; Accepted 6 December 2019; Published 22 January 2020

Guest Editor: Feng Chen

Copyright © 2020 Chen 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.

Macrolevel crash modeling has been extensively applied to investigate the safety effects of demographic, socioeconomic, and land use factors, in order to add safety knowledge into traffic planning and policy-making. In recent years, with the increasing attention to regional traffic management and control, the safety effects of macrolevel traffic flow parameters may also be of interest, in order to provide useful safety knowledge for regional traffic operation. In this paper, a new spatial unit was developed using a recursive half-cut partitioning procedure based on a normalized cut (NC) minimization method and traffic density homogeneity. Two Bayesian lognormal models with different conditional autoregressive (CAR) priors were applied to examine the safety effects of traffic flow characteristics at the NC level. It was found that safety effects of traffic flow exist at such macrolevel, indicating the necessity of considering safety for regional traffic control and management. Furthermore, traffic flow effects were also examined for another two spatial units: Traffic Analysis Zone (TAZ) and Census Tract (CT). It was found that ecological fallacy and atomic fallacy could exist without considering traffic flow parameters at those planning-based levels. In general, safety needs to be considered for regional traffic operation and the effects of traffic flow need to be considered for spatial crash modeling at various spatial levels.

1. Introduction

Macroscopic safety evaluation was often conducted, with the purpose of finding factors that could be improved or controlled at the planning stage or during policy-making process. Traditional macroscopic crash models (e.g., Poisson lognormal models) rely on the assumption of independence across observations. However, in recent years, spatial crash models have gained a lot of attention, by adding spatial dependence into macroscopic crash models. Socioeconomic, land use, demographic, and traffic network characteristics were of interest, and they were often aggregated at varying levels of spatial units [1], including Traffic Analysis Zones (TAZ), Census Tracts (CT), census wards, statistical area levels, block groups, counties, and states. In general, spatial crash models have shown their superiority over conventional macroscopic crash models.

However, an important issue of spatial crash models is the choice of a certain level of spatial unit, which is also

called Modifiable Aerial Unit Problem (MAUP). Wang et al. [2] discussed the possible ecological fallacy of spatial aggregation (Davis) that modeling results from spatially aggregated data may not be fully applied to disaggregated data. They also argued the possible atomistic fallacy caused by disaggregated data, which is unable to take “system-wide effects” captured by spatially aggregated data [3, 4]. Abdel-Aty et al. [5] compared spatial crash models based on three different aggregate-level spatial units (i.e., TAZs, CT, and block groups) and found the effects of different spatial units on the significance of model estimates. Xu et al. [6] conducted a sensitive analysis on the effects of different aggregate-level spatial TAZ units. They found that more aggregated TAZs tended to have better model performance but fewer variables. This result was consistent with Wang et al. [2], as spatially aggregated data tend to decrease atomistic fallacy while increase ecological fallacy if not properly aggregated. Gyimah et al. [1] examined the effect of six different aggregate-level spatial units on unobserved

spatial heterogeneity. According to the results, the aggregation level could significantly affect spatial heterogeneity, as well as model performance.

Both Xu et al. [6] and Gyimah et al. [1] pointed out the necessity of defining a new spatial unit from the perspective of macroscopic safety evaluation, instead of using long-term planning-based TAZs. To note, most spatial units like TAZ were not initially defined for safety analysis but for traffic planning purposes. Thus, they are possibly delineated based on the homogeneity of demographic, socioeconomic, or land use factors, regardless of traffic flow characteristics. Traffic flow characteristics were largely found to be associated with crash occurrence and safety in the previous literature. Thus, an aggregate-level spatial unit, like TAZs at any aggregate level, may contain roadways with significantly distinct traffic flow characteristics. As a result, the relationship between traffic flow characteristics and crash occurrence could be weakened or nonsignificant (i.e., an example of ecological fallacy for a spatial unit due to the inhomogeneity in traffic), leading to biased model estimates. To overcome possible ecological fallacy and atomistic fallacy (caused by disaggregated data), Wang et al. [2] claimed the necessity of a better mathematical partitioning scheme based on spatial homogeneity in both demographic and traffic flow characteristics.

There are many different spatial partitioning methods in the literature to define spatial units, such as k -means algorithm, AZP [7], and REDCAP (regionalization with dynamically constrained agglomerative clustering and partitioning) [8]. Some have also been introduced into spatial crash modeling. In order to study the effect of enforcement on road crashes in Greece, Yannis et al. [9] formed spatial units based on spatial homogeneity in traffic characteristics and road safety parameters, with a direct k -means algorithm. Yannis concluded the statistical results might be more reliable if spatial units are more homogeneous. Xu et al. [6] proposed a zoning scheme of aggregating similar TAZs into a spatial unit, based on REDCAP. The homogeneity of crash risk was considered as the clustering criteria. However, although considering spatial homogeneity in various factors, these partitioning methods still rely on predefined spatial units for planning, such as TAZ.

Ji and Geroliminis [10] introduced a graph cut minimization method to divide urban traffic networks into multiple spatial units. With such method, a traffic network was partitioned into spatial units with homogeneous traffic flow characteristics, and macroscopic fundamental diagram (MFD) of each region was successfully identified. Such method is very flexible, by not depending on any sort of predefined spatial unit (e.g., TAZs). Although initially proposed for controlling traffic and improving congestion, the idea of the graph cut minimization method could also be beneficial for macroscopic safety evaluation. As traffic flow characteristics have been shown strong relationships with crash, it is reasonable to believe that a spatial unit aggregating roadway with similar traffic flow characteristics would be more suitable for spatial crash modeling. Based on such method, the underlying relationship between traffic flow and

crash may be better explored. Thus, active regional traffic control and management strategies [11–15] could be expected to improve macroscopic safety, by managing traffic flow in some certain state with lower crash risk.

Thus, in this paper, we will define a new spatial unit for macroscopic safety evaluation, by considering the homogeneity of traffic densities. A graph cut method will be introduced, based on which a spatial partitioning procedure is proposed. Two Bayesian spatial modeling techniques are employed to analyze crash data at the new level, in order to identify possible traffic flow effects on safety. The remainder of the paper is organized as follows. In Section 2, we present the detail of the graph cut minimization method for spatial partitioning as well as spatial modeling techniques. Section 3 gives a brief description of the data. Section 4 summarizes the spatial partition results and modeling results and includes a discussion. The last part concludes the paper and recommends future research directions.

2. Spatial Partitioning

2.1. A Normalized Cut (NC) Minimization Method. In order to divide an area into multiple regions, a normalized cut minimization (NC) method is introduced, considering intersections as nodes and roadways as edges. NC method has been used in the previous literature for spatial partitioning [10].

Suppose the node set V in an undirected graph $G = (V, E)$ where E indicates the set of edges in G . Assume that each edge between two vertices v_i and v_j carries a non-negative weight $w_{ij} = w_{ji} \geq 0$. The weight adjacency matrix of the graph can be defined as $W = (w_{ij})_{i,j=1,\dots,n}$. When $w_{ij} = 0$, it indicates that the two vertices are not connected. The degree of a vertex $v_i \in V$ is defined as $d_i = \sum_{j=1}^n w_{ij}$. The degree matrix D is defined as the diagonal matrix with degree d_1, d_2, \dots, d_n on the diagonal.

Consider edge to be a measure of the similarity between nodes. We want to find a partition of the graph such that edges between different groups have a very low weight, indicating that points in different clusters are dissimilar from each other. Moreover, the edges within a group need to have high weights, implying that points within the same cluster are similar.

For two disjoint subsets A and B ,

$$\text{cut}(A, B) = \sum_{i \in A, j \in B} w_{ij}. \quad (1)$$

In 2000, Shi and Malik proposed a 2-way normalized cut functions: Ncut and Nassoc, which indicate the homogeneity and heterogeneity of two clusters:

$$\begin{aligned} \text{Ncut}(A, B) &= \frac{\text{cut}(A, B)}{\text{cut}(A, V)} + \frac{\text{cut}(A, B)}{\text{cut}(B, V)}, \\ \text{Nassoc}(A, B) &= \frac{\text{cut}(A, A)}{\text{cut}(A, V)} + \frac{\text{cut}(B, B)}{\text{cut}(B, V)}, \end{aligned} \quad (2)$$

$$\text{cut}(A, A) + \text{cut}(A, B) = \text{cut}(A, V),$$

$$\text{Ncut}(A, B) = 2 - \text{Nassoc}(A, B).$$

Thus, the objective can be determined as the minimization of Ncut:

$$\begin{aligned} \min_x \quad & \text{Ncut}(x). \\ \text{s.t.} \quad & x \in V. \end{aligned} \quad (3)$$

Minimizing Ncut value exactly is NP-complete. The discrete solution can be approximated efficiently by solving an eigenvalue system in the real value domain. A common way is to convert the normalized cut into the unnormalized graph Laplacian. Set

$$w(i, j) = \begin{cases} \exp\left(-(d_i - d_j)^2\right), & a_{ij} = 1, \\ 0, & a_{ij} = 0. \end{cases} \quad (4)$$

Then, we have

$$\begin{aligned} \min_x \quad & \text{Ncut}(x) = \min_y \frac{y^T(D - W)y}{y^T D y}, \\ \text{s.t.} \quad & x_i = 1 \text{ if } x_i \in A, \\ & x_i = 1 \text{ if } x_i \in B, \\ & y_i \in \{-1, b\}, \\ & y^T D 1 = 0, \end{aligned} \quad (5)$$

where D is an $N * N$ diagonal matrix with $d_{ii} = \sum_j \omega(i, j)$ on its diagonal. $D - W$ is the Laplacian matrix, known to be positive semidefinite. Based on Rayleigh–Ritz theorem, the solution is to solve the generalized eigenvalue system:

$$(D - W)y = \lambda D y, \quad (6)$$

where λ is the eigenvalue and y is the eigenvector. The second smallest eigenvector is called the Fiedler vector, which is the real-valued solution to normalized cut problem.

2.2. A Recursive Half-Cut Partitioning Procedure. To apply the above method for spatial partitioning, we propose a recursive half-cut procedure:

Step 1: set up a weighted graph $G = (V, E)$ based on the topology of traffic network. Intersections are treated as nodes while roadways are considered as edges.

Step 2: set the weight on the edge connecting two nodes, using a measure of similarity between two notes (i.e., traffic density).

Step 3: solve $(D - W)y = \lambda D y$ for eigenvectors with the smallest eigenvalues.

Step 4: cut the graph into two clusters based on the second smallest eigenvector (i.e., Fiedler vector).

Step 5: decide if the current partition should be further divided.

Step 6: repeat the first five steps until certain criteria were met.

When partitioning a graph with the Fiedler vector, different strategies can be used. There are three general ways: (1) partition the graph with the median value; (2) cut the graph with value 0 (negative versus positive); and (3) cut the graph based on the largest interval between every two elements. In this study, the third approach was utilized.

2.3. Spatial Model Configuration. Crash modeling includes severity modeling [16–18] and crash frequency modeling [19, 20]. Spatial crash modeling belongs to crash frequency modeling, which contains multiple model structures [21–26], including Poisson lognormal model, negative binomial spatial model, Poisson lognormal spatial model, geographic weighted Poisson regression model, and Bayesian spatial varying-coefficient model. Since the purpose of this study is to examine the effect of traffic flow characteristics, two Bayesian lognormal models with different CAR priors were applied, since they have been widely applied in many different research fields such as epidemiology.

A generalized Bayesian lognormal model with CAR prior can be presented as

$$\begin{aligned} Y_i & \sim \text{Poisson}(\lambda_i), \\ \ln(\lambda_i) & = \ln(E) + \beta_0 + \beta_k X_{ik} + \theta_i + \phi_i, \end{aligned} \quad (7)$$

where λ_i is the expected mean of crash occurrence for observation i ; E is the exposure/expectation for observation i ; β_k is the parameter coefficient of k th variable; X_{ik} is the k th variable for i th observation; θ_i is the unstructured error, often assumed as a prior normal distribution; and ϕ_i is the spatial correlation.

For the spatial correlation term ϕ_i , the intrinsic conditional autoregressive prior (CAR prior) can be defined as follows [18]:

$$\phi_i | \phi_{-i}, W, \tau^2 \sim N\left(\frac{\sum_{i \neq j} \phi_j m_{ij}}{\sum_{i \neq j} m_{ij}}, \frac{\tau_c^2}{\sum_{i \neq j} m_{ij}}\right), \quad (8)$$

where m_{ij} denotes the binary entries of proximity matrix (1 represents adjacency while 0 indicates nonadjacency). τ_c is the precision parameter, assumed as a prior gamma distribution. In essence, the conditional expectation ϕ_i is the average of spatial correlations of adjacent areas; conditional variance τ^2 is inversely proportional to the number of adjacent areas.

A Cressie autoregressive prior can be written as follows:

$$\phi_i | \phi_{-i}, W, \tau^2 \sim N\left(\rho \frac{\sum_{i \neq j} \phi_j m_{ij}}{\sum_{i \neq j} m_{ij}} + (1 - \rho) \frac{\sum_{j=1}^n \phi_j}{n}, \frac{\tau_c^2}{\sum_{i \neq j} m_{ij}}\right). \quad (9)$$

Different from IAC priors, the conditional expectation of ϕ_i is modified into the weighted average of the average of adjacency area and the average of the entire area. Weight parameter ρ indicates the intensity of spatial autocorrelation. When $\rho = 0$, it indicates a complete spatial independence, and with the increase of ρ , spatial autocorrelation

increases. When $\rho = 1$, the Cressie model degenerates to an intrinsic CAR model.

Based on a fitted model, the relative risk (RR) of a subregion can be calculated as follows:

$$RR = e^{(\beta_0 + \beta_k X_{ik} + \theta_i + \phi_i)} \quad (10)$$

3. Data Preparation

Crash data for the central area of Kunshan City (within the Kunshan Middle Ring Road) in 2015 were acquired from the Kunshan Police Department. A total of 5538 crashes were collected. The crash data contain detailed information on drivers, roadways, and vehicles. For each crash record, there is a unique coordinate, which can be further used for locating it on the map. In order to conduct macroscopic crash modeling, spatially aggregated features also need to be collected.

Planning-based data were extracted from the planning department of Kunshan City. More importantly, detailed traffic data are necessary for spatial partition and spatial modeling. Thus, traffic data between 2015.8 and 2015.9 were extracted from microwave detectors with the 30 s interval, including density, speed, and counts. The average traffic density was considered as the measure of traffic homogeneity, used in spatial partitioning. Figure 1 shows the roadway network, the location of microwave detectors, and the land use sketch of the studied area.

As for Bayesian modeling, the total number of crashes was used as the dependent variable. Unlike previous studies, we calculated the expected crash number as exposure. The expected crash number of a spatial unit can be calculated as the total crash number times the proportion of the exposure (daily traffic volume * total population * area size) of the area. The calculation assumed that the expected average risk of each area is comparable. However, it is also reasonable that the relative exposure risk among different areas can largely vary. Many factors could contribute to it, including traffic flow parameters (used as explanatory variables), planning, and land use factors. Four aggregated traffic variables (average flow, average density, average speed, and speed variance) were calculated for each area, which were also used as explanatory variables for spatial modeling. Other explanatory variables include most planning-based aggregated variables, commonly used in the previous literature.

4. Results

4.1. Partitioning Results. The NC partitioning method was applied to partition the graph with 99 nodes (i.e., 99 intersections). Initially, by solving the Laplacian matrix, all eigenvalues were calculated. Based on the similarity of eigenvectors and their spatial adjacency, 8 clusters were determined. However, NC 8 has the area size of 10.98 m², which is much larger than other NCs. In order to obtain all NCs with comparable size, further division efforts were conducted in NC 8, which has 23 intersections within the

area. Based on another two rounds of partitioning on NC 8, 13 NCs were finally delineated, as shown in Figure 2.

The descriptive statistics of planning, roadway, and traffic flow parameters of 13 NCs are summarized in Table 1.

4.2. Model Results. Intrinsic and Cressie Bayesian CAR models were developed to examine the relationship between various variables with crash risk. First, multicollinearity needs to be examined for those variables. VIF tests and stepwise methods were applied to eliminate those variables with high multicollinearity (VIF >10).

Then, for each Bayesian CAR model, 100000 iterations were conducted with 5000 iterations as burn-in period. All three models appeared to reach convergence within the simulation period. Figure 3 gives an illustration of model convergence.

Both intrinsic and Cressie CAR models showed significant traffic flow effects on crash risk. The results were comparable for the two models. Detailed results are shown in Table 2. A region having higher speed variance tends to have higher crash risk. This is reasonable. With higher traffic density, there is a slight increase in crash risk. Previous literature studies suggest controversial findings on the relationship between density and crash risk. Some claimed a positive linear relationship while others suggested a quadratic function. It is reasonable to expect that, at first, the higher density creates more interaction and thus more crashes. While it reaches some certain point, the traffic becomes congested and the speed significantly drops down. In this case, crashes could possibly decrease. Since we only consider an average effect without regarding spatial and temporal heterogeneity, the detailed density effect of each zone needs to be further explored. Average daily traffic volume (ADT) was not found to be significant as the explanatory variable, implying that traffic volume is not significantly connected with average crash risk. Thus, the effect of ADT on crash risk could be considered as the only exposure effect (Figure 4).

As for the effects of planning/roadway factors on crash risk, there are slight differences between the two models. According to intrinsic CAR models, the increase of major/minor arterial density will increase crash risk. This was consistent with the previous literature. There is no significant relationship between local road density and crash risk. It was expected that the increase of local roads would lower crash risk, according to the previous literature. A possible reason could be the inclusion of significant traffic flow variables adjusts the effects of local roads density. In other words, traffic flow variables could be endogenous variables that cannot be ignored. According to the Cressie CAR model, with the increase of industrial land use, crash risk decreases by 2.5%. More school land use and residential land use are associated with higher crash risk. This was reasonable in China. During morning and evening peak hours, parents pick children to increase the disorder of traffic. Thus, it is critical to deal with school land use. Those low-income residential areas are older district. Buildings are too old so that there is enough parking space. Oftentimes, roadways



FIGURE 1: (a) Roadway network and location of microwave detectors and (b) land use sketch of the subject area.

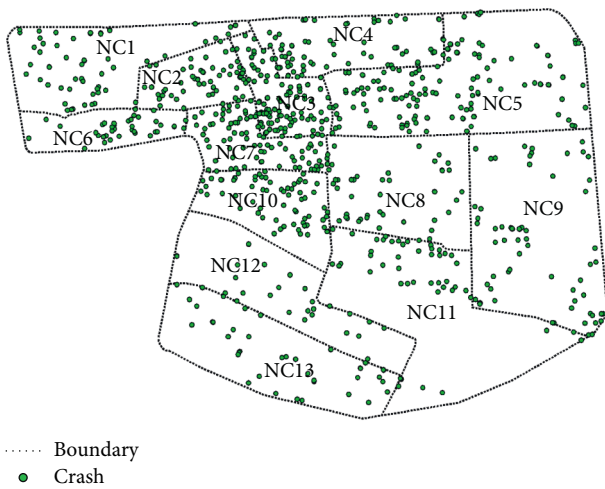


FIGURE 2: Final NC partitioning results.

TABLE 1: Descriptive analysis of 13 NCs.

Roadway characteristics	Min	Max	Mean	SD
Area (km ²)	1.016	4.269	2.329	1.237
Roadway density (km/km ²)	3.181	10.919	6.923	2.423
Major arterials (km/km ²)	0.392	4.805	2.648	1.396
Minor arterials (km/km ²)	0.022	2.386	1.432	0.660
Land use characteristics	Average percentage (%)			
School land use	2.391			
Public land use	5.761			
Commercial land use	11.822			
Industrial land use	9.816			
Low-income residentials	1.015			
High-income residentials	28.265			

were occupied by vehicles. These could be potential safety issues. Thus, certain traffic management could be considered.

The relative risk of each spatial unit was also calculated, based on the two models. In general, two models estimated the similar results. Comparing with crash frequency, it should be noted that spatial models can effectively identify

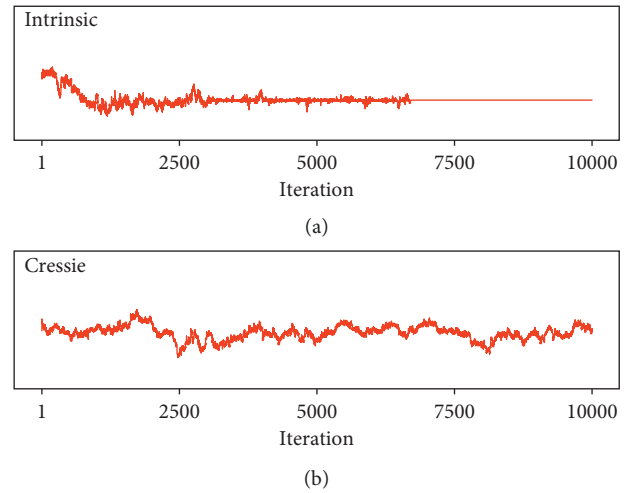


FIGURE 3: Illustration of Bayesian model convergence: intrinsic (a) and Cressie (b).

actual relative risk, by accounting for exposure and spatial autocorrelation.

4.3. Ecological Fallacy and Atomic Fallacy. To further discuss ecological fallacy and atomic fallacy, spatial modeling was also conducted for another two spatial units: TAZ and Census Tract (CT) zone. For each spatial unit, two models were developed, one considering traffic flow parameters (model 2) and another without considering them (model 1). Detailed results can be found in Table 3. Note that both original TAZ and CT were decided by the Kunshan Traffic Planning Department (as shown in Figure 5).

For CT model 2, average daily traffic volume, average speed, business, administrative, and public service land use were found as significant variables. For TAZ model, only land use variables were found as significant, including the public management and public service, business, and residence. Using TAZ as spatial units, there were no significant traffic flow effects on crash risk. However, for CT units, traffic flow effects emerge.

TABLE 2: Two bayesian modeling results of NC spatial units.

Varibale	BYM intrinsic CAR			BYM cressie CAR		
	Mean	s.d.	95% CI	Mean	s.d.	95% CI
Roadway density	-0.136	0.067	(-0.239, -0.012)	-0.081	0.091	(-0.206, 0.096)
Major arterial density	0.089	0.199	(-0.526, 0.309)	0.222	0.252	(-0.158, 0.596)
Minor arterial density	0.121	0.313	(-0.304, 1.012)	-0.601	0.242	(-1.083, -0.193)
School land use	-0.224	0.150	(-0.841, 0.015)	0.047	0.111	(0.034, 0.228)
Public land use	0.032	0.106	(0.121, 0.700)	0.066	0.058	(-0.049, 0.176)
Commercia land use	-0.066	0.033	(-0.099, 0.008)	0.040	0.018	(-0.074, 0.085)
Industrial land use	-0.037	0.019	(-0.064, 0.006)	-0.025	0.025	(-0.067, -0.012)
Low-income residence	0.222	0.194	(-0.257, 0.531)	0.145	0.201	(-0.185, -0.510)
High-income residence	-0.026	0.032	(-0.076, 0.011)	-0.013	0.019	(-0.046, 0.015)
Daily traffic density	0.021	0.010	(0.009, 0.033)	0.014	0.012	(0.007, 0.028)
Average traffic speed	0.013	0.024	(-0.023, 0.052)	0.011	0.020	(-0.017, 0.041)
ADT	0.223	0.141	(-0.019, 0.788)	0.213	0.131	(-0.011, 0.634)
Speed variance	0.082	0.012	(0.053, 0.114)	0.077	0.011	(0.049, 0.103)

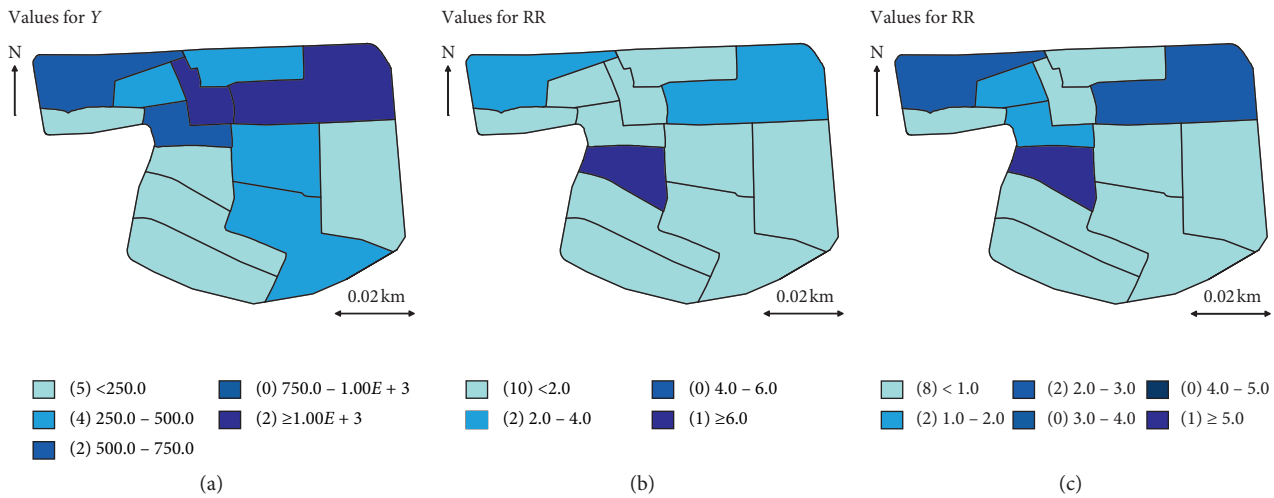


FIGURE 4: Crash count, relative risk of intrinsic and Cressie model estimation. (a) Crash Count. (b) Intrinsic model estimation. (c) Cressie model estimation.

TABLE 3: Bayesian modeling results of CT and TAZ model.

Variables	CT model 2*		CT model 1		TAZ model 2*		TAZ model 1	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
Intercept	-3.034	0.665	-2.993	0.385	-2.603	0.438	-2.702	0.425
<i>Planning parameters</i>								
Minor arterials								
Local roads					-0.001	0.005	-0.001	0.004
Public service land use	0.040	0.017	0.057	0.021	0.025	0.009	0.027	0.010
Commercial land use	0.053	0.016	0.078	0.017	0.046	0.011	0.048	0.010
Residential land use	—	—	0.027	0.009	0.029	0.007	0.031	0.009
<i>Traffic parameters</i>								
Daily traffic density	-0.1264	0.0318			0.02459	0.02557		
Average traffic speed	0.0216	0.0062	—	—	0.00161	0.00535		
ADT	-0.0415	0.0090			-0.0109	0.00674		
Speed variance	0.0432	0.0076			-0.0321	0.00339		

*The modeling efforts of adding traffic flow parameters to the corresponding spatial levels.

It is known that TAZ is delineated by land use, socio-economics, and demographics. From the results, it can be concluded that the major issue of TAZ-based spatial

modeling is the risk of atomic fallacy. In other words, the actual traffic flow effects were not detected based on TAZs. Moreover, in practice, any active interventions or policies



FIGURE 5: (a) TAZ and (b) CT spatial units of the studied area.

cannot be proposed on regional traffic flow control or management to improve traffic safety, as only land use parameters were found as significant.

For the CT-based model, average daily traffic volume (ADT) was found as a significant variable, with a negative sign indicating that higher traffic volume would result in lower crash rates. This finding could be challenged because traffic volume was considered positively associated with crash risk as the crash exposure, per the previous literature studies. In addition, average speed was found to be positively associated with regional crash risk. However, according to Abdel-Aty et al. [27], relatively low-speed area tended to have higher crash risk. Moreover, previous literature largely suggested the insignificance of possible linear relationship between average speed and crash risk [28]. Thus, the significant traffic flow effects from CT appear to be ecological fallacy [29]. As known, CT was also defined based on nontraffic parameters.

5. Conclusion

Crash occurrence was believed and found to be associated with traffic flow characteristics. Macroscopic spatial crash modeling was initially conducted with the purposes of adding safety consideration into long-term traffic planning and policy-making, as spatial units used were mainly defined from the planning perspective. However, it could also be expected to propose some effective control and management strategies to improve not only efficiency but also regional traffic safety. Thus, exploring the possible linkage among traffic flows and crash risk at regional level appears to be necessary. In addition, previous literature argued the major flaw of TAZ-based spatial modeling of both atomic fallacy and ecological fallacy. Thus, conducting macroscopic spatial crash modeling using a better-defined spatial unit is of interest in this paper.

Based on a normalized cut minimization method, we defined a spatial unit for regional safety evaluation. Microwave

data were used to partition the subject area into multiple parts, according to the homogeneity of traffic density. Crash data, planning data, and traffic data were all collected for spatial modeling purpose. In order to account for the spatial dependency among each unit and the potential overfitting issue (caused by the availability of detectors), two Bayesian CAR models were employed.

The results proved the existence of traffic flow effects at macrolevels. Note that this level was often used to study regional traffic control and management strategies. Thus, it indicates the necessity of deeply examining the relationship among traffic flow characteristics (e.g., MFD) and crash risk at regional level, in order to enlighten traffic professionals to propose time-dependent active regional traffic control and management strategies for safety improvement.

Admittedly, the current study still has some limitations. First, most traffic data were collected for arterial roads, where microwave detectors are installed. Only those roads can be used as edges in graph cut minimization. Thus, limited by the sparsity of microwave detectors and the size of the subject area, only 13 NC zones were finally defined. This could possibly cause overfitting issue. Bayesian CAR models were introduced to deal with the issue, and the coefficients were assumed to follow prior normal distributions (similar with L2 regularization). In our future study, we will obtain more detailed traffic data through different ways and expand the study area. Second, spatial and temporal heterogeneities were not considered in this model. Spatial heterogeneity had been discussed in the previous literature. Since this study mainly focuses on defining new spatial unit and comparing it with other units, spatial heterogeneity could be examined in the future, especially for the effect of traffic characteristics. Moreover, since NC is defined based on traffic flow data instead of planning data, temporal heterogeneity can also be discussed in the future. Third, only three simple aggregated traffic parameters were considered in the study. It is interesting to extract other features of traffic flow (e.g.,

macroscopic fundamental diagram (MFD) and examine their possible effects on safety. Last, it appeared that macrolevel traffic flow parameters were not significant at TAZ levels. It deserves a deeper investigation and possible endogenous factors need to be examined in the future.

Data Availability

Crash data are confidential and cannot be shared without permission from the local government in China.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research has been supported by the National Key R&D Program of China (2018YFE0102700).

References

- [1] R. Gyimah, M. Saberi, and M. Sarvi, "The effect of variations in spatial units on unobserved heterogeneity in macroscopic crash models," *Analytic Methods in Accident Research*, vol. 13, pp. 28–51, 2017.
- [2] C. Wang, M. A. Quddus, T. Ryley, M. Enoch, and L. Davison, "Spatial models in transport: a review and assessment of methodological issues," in *Proceedings of the 91st Annual Meeting of the Transportation Research Board*, Washington, DC, USA, 2012.
- [3] R. B. Noland and M. A. Quddus, "A spatially disaggregate analysis of road casualties in England," *Accident Analysis & Prevention*, vol. 36, no. 6, pp. 973–984, 2004.
- [4] R. Haynes, A. Jones, V. Kennedy, I. Harvey, and T. Jewell, "District variations in road curvature in England and Wales and their association with road-traffic crashes," *Environment and Planning A: Economy and Space*, vol. 39, no. 5, pp. 1222–1237, 2007.
- [5] M. Abdel-Aty, J. Lee, C. Siddiqui, and K. Choi, "Geographical unit based analysis in the context of transportation safety planning," *Transportation Research Part A: Policy and Practice*, vol. 49, pp. 62–75, 2013.
- [6] P. Xu, H. Huang, N. Dong, and S. C. Wong, "Sensitivity analysis in the context of regional safety modeling: identifying and assessing the modifiable areal unit problem," *Accident Analysis & Prevention*, vol. 70, pp. 10–120, 2014.
- [7] D. Martin, "Extending the automated zoning procedure to reconcile incompatible zoning systems," *International Journal of Geographical Information Science*, vol. 17, no. 2, pp. 181–196, 2003.
- [8] D. Guo and H. Wang, "Automatic region building for spatial analysis," *Transactions in GIS*, vol. 15, no. 1, pp. 29–45, 2011.
- [9] G. Yannis, E. Papadimitriou, and C. Antoniou, "Multilevel modelling for the regional effect of enforcement on road accidents," *Accident Analysis & Prevention*, vol. 39, no. 4, pp. 818–825, 2007.
- [10] Y. Ji and N. Geroliminis, "On the spatial partitioning of urban transportation networks," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1639–1656, 2012.
- [11] A. Kouvelas, M. Saeedmanesh, and N. Geroliminis, "Enhancing model-based feedback perimeter control with data-driven online adaptive optimization," *Transportation Research Part B: Methodological*, vol. 96, pp. 26–45, 2017.
- [12] N. Geroliminis, J. Haddad, and M. Ramezani, "Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: a model predictive approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 348–359, 2013.
- [13] H. Zhao, Q. Chen, W. Shi, T. Gu, and W. Li, "Stability analysis of an improved car-following model accounting for the driver's characteristics and automation," *Physica A: Statistical Mechanics and Its Applications*, vol. 526, Article ID 120990, 2019.
- [14] D. Sun, H. Zhao, H. Yue, M. Zhao, S. Cheng, and W. Han, "ST TD outlier detection," *IET Intelligent Transport Systems*, vol. 11, no. 4, pp. 203–211, 2017.
- [15] H. Zhao, H. Yue, T. Gu, and W. Li, "CPS-based reliability enhancement mechanism for vehicular emergency warning system," *International Journal of Intelligent Transportation Systems Research*, vol. 17, no. 3, pp. 232–241, 2019.
- [16] F. Chen, M. Song, and X. Ma, "Investigation on the injury severity of drivers in rear-end collisions between cars using a random parameters bivariate ordered probit model," *International Journal of Environmental Research and Public Health*, vol. 16, no. 14, p. 2632, 2019.
- [17] F. Chen and S. Chen, "Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1677–1688, 2011.
- [18] C. Wang, C. Xu, J. Xia, and Z. Qian, "Modeling faults among E-bike-related fatal crashes in China," *Traffic Injury Prevention*, vol. 18, no. 2, pp. 175–181, 2017.
- [19] C. Wang, C. Xu, J. Xia, Z. Qian, and L. Lu, "A combined use of microscopic traffic simulation and extreme value methods for traffic safety evaluation," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 281–291, 2018.
- [20] C. Wang, C. Xu, and Y. Dai, "A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data," *Accident Analysis and Prevention*, vol. 123, pp. 365–373, 2019.
- [21] J. Besag, J. York, and A. Molli, "Bayesian image restoration, with two applications in spatial statistics," *Annals of the Institute of Statistical Mathematics*, vol. 43, no. 1, pp. 1–20, 1991.
- [22] W. Cheng, G. S. Gill, R. Dasu, M. Xie, X. Jia, and J. Zhou, "Comparison of Multivariate Poisson lognormal spatial and temporal crash models to identify hot spots of intersections based on crash types," *Accident Analysis & Prevention*, vol. 99, pp. 330–341, 2017.
- [23] 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.
- [24] T. Wang, "Incorporating safety into transportation planning for small and medium-sized communities," Master thesis, Iowa State University, Ames, IA, USA, 2011.
- [25] T. Wang, R. R. Souleyrette, and K. Gkritza, "Incorporating safety into transportation planning at smaller agencies," in *Proceedings of the TRB 92nd Annual Meeting*, Washington, DC, USA, January 2013.
- [26] 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.
- [27] M. Abdel-Aty and A. Pande, "Identifying crash propensity using specific traffic speed conditions," *Journal of Safety Research*, vol. 36, no. 1, pp. 97–108, 2005.

- [28] P. Marchesini and W. Weijermars, *The Relationship between Road Safety and Congestion on Motorways*, SWOV Publication R-2010-12, Leidschendam, Netherlands, 2010.
- [29] G. A. Davis, "Is the claim that "variance kills" an ecological fallacy?," *Accident Analysis & Prevention*, vol. 34, no. 3, pp. 343–346, 2002.

