The Influence of Urban Built Environment on the Spatial Heterogeneity of Road Traffic Accidents

——Taking Tianjin, China as an example

Keke Zhang[1], Chao Kong[1*], Wenhui Qin[1], Hubin Yan[1], Shaohua Wang[1]

[1] School of Automotive and Transportation, Tianjin University of Technology and Education, Tianjin, 300222, China;
kekezhang@tute.edu.cn(K.Z.); 0721231011@tute.edu.cn(W.Q.); 0721231012@tute.edu.cn(H.Y.); shaohuawang@tute.edu.cn(S.W.)
[*] Correspondence: unfollower@qq.com

Abstract. The urban built environment is closely related to the occurrence of traffic accidents. Accurately identifying the spatial heterogeneity characteristics of factors affecting traffic accidents is of great significance for accurate traffic safety management and control. In this study, 2,114 Traffic Analysis Zones (TAZ) in Tianjin, China were used as the research units, 25,869 road traffic accident data were collected as the dependent variable, and 12 built environment density data such as roads and Point of Interest (POI) were extracted as independent variables. After confirming the spatial dependence of the data based on the spatial autocorrelation test, three spatial econometric models of Ordinary Least Squares (OLS), Geographically Weighted Regression (GWR), and Multiscale Geographically Weighted Regression (MGWR) are constructed for optimization analysis. The research results show that MGWR is the optimal model to describe the spatial heterogeneity of traffic accidents in different built environment factors. There is a negative correlation between recreational density and intersection density with traffic accidents, while other influencing factors show a positive correlation. The regression coefficient for retail industry density is significantly higher than other influencing factors, while the regression coefficient for leisure and entertainment density is significantly lower than other influencing factors.

Keywords: urban built environment; road traffic accidents; spatial heterogeneity; multiscale geographically weighted regression;

1 INTRODUCTION

Traffic accidents have caused serious loss of life and property to people around the world. Road traffic accidents cause about 1.3 million deaths worldwide each year[1]. The urban built environment includes roads, transportation facilities, land use, and other aspects, which to varying degrees affect the accident rate[2]. Due to the vast area, dense population, high vehicle density, and complex urban environment, the frequency of

G. Chen et al. (eds.), *Proceedings of the 2024 International Conference on Rail Transit and Transportation (ICRTT 2024)*, Advances in Engineering Research 254,
https://doi.org/10.2991/978-94-6463-610-9_34

traffic accidents in megacities increases, and the severity is often more significant. Therefore, in-depth study of the relationship between road traffic accidents and urban built-up environmental factors, especially in megacities, is of great significance for formulating accurate traffic safety control policies.

Researchers have made a detailed analysis of the specific utility of the factors affecting the built environment on the accident. Zafri[3] and others found that factors positively correlated with the occurrence of accidents include entertainment land density, road density, and intersection density. Osama[4] and others found that accidents were positively correlated with the density of bus stations and commercial areas, and negatively correlated with the density of leisure and residential areas. Wang[5] and others' s modeling results show that the density of commercial POI has a significant negative impact on the accident. Most previous studies have adopted non-spatial global regression models, such as ordinary least squares method[6], Poisson and negative binomial regression models[7][8], neglecting spatial effects and failing to adequately explain spatial heterogeneity. For this reason, the GWR model considering spatial heterogeneity is applied[9]. Wang[10] and others used a GWPR model to explore the spatial effect of built environment on the collision density of drunk driving. The Fotheringham team further developed MGWR, which identifies and utilizes spatial scales of variables, offering more precise parameter estimates and model fits[11]. Liu[12] and others found that MGWR outperforms OLS and GWR in identifying the spatial relationships between influencing factors and traffic accidents.

At present, although the literature reveals the potential impact of the urban built environment on traffic accidents in some cities, it still needs to be improved. Although some studies have found a positive correlation between accidents and commercial area density[4], other research has indicated that commercial density has a significant negative impact on accidents[5]. The unique geographical, transportation, population, and behavioral patterns of mega-cities result in significant differences in the impact effects of different cities. Taking Tianjin, China as a case study, this paper conducts spatial autocorrelation tests to confirm the spatial dependence of the data. Then, three spatial econometric models including OLS, GWR, and MGWR are constructed for optimal analysis to explore the spatial heterogeneity characteristics of built environment factors at specific spatial scales. The research results are helpful to promote the improvement and development of the theoretical system of active traffic safety management and control, and provide empirical basis for formulating reasonable traffic management and control strategies in cities in the future.

2 RESEARCH DATA

2.1 Research Area

Tianjin is a municipality directly under the Central Government of the People's Republic of China. Tianjin City has a total of 16 districts, with a land area of 11966.45km2 and a permanent population of 13.64 million people, divided into 2114 TAZ based on land use attributes[13]. This study is based on Python programming to extract 25,869 road traffic accident data from judicial appraisal documents as the dependent variables.

The geographical location of the accidents is extracted through the Baidu Maps API, and built environment data such as intersection density and road network density provided by the Tianjin Municipal Institute of Traffic Planning and Design are collected. In addition, web crawler technology is used to extract the POI data of built environment related to traffic accidents, including 15 POI densities[4] for government agencies, railway stations and subway stations, bus stations, gas stations, financial services, commercial buildings, companies and enterprises, logistics express delivery, hotels and residential quarters. All data were collected and preprocessed based on TAZ as the basis for spatial division.

2.2 Data Preprocessing

Taking road traffic accident data as explanatory variables, 15 built environment density data were extracted as the dependent variable. Considering that when there is a high correlation between two or more explanatory variables, there will be a multicollinearity problem, which makes the estimation of the individual effects of the explanatory variables inaccurate[10]. Therefore, STATA software was used to program multicollinearity test and stepwise regression, and the influencing factors with VIF greater than 10 were removed. Retain 12 influencing factors through multicollinearity test. Secondly, in view of the large gap in data magnitude of traffic accident influencing factors, the data is logarithmic processing[14] before empirical analysis to slow down the influence of extreme values on empirical results.

3 RESEARCH METHODS

3.1 OLS model

The OLS model is the most classic regression analysis method, typically used to estimate the parameters of a linear regression model. The model expression is as follows:

$$y_i = \beta_0 + \sum_{i=1} \beta_k x_{ik} + \varepsilon_i \tag{1}$$

In the formula: i represents the $i\text{−}th$ sample point, y_i represents the accident data value of the $i\text{−}th$ sample point, and X_{ik} represents the $k\text{−}th$ traffic accident influencing factor of the $i\text{−}th$ sample point. ε_i is a random error term, β_0 represents the regression constant of the regression equation and β_k represents the regression coefficient of the $k\text{−}th$ traffic accident influencing factor.

3.2 MGWR Model

The MGWR model comprehensively considers the spatial characteristics of different explanatory variables, further refining the spatial relationship between explanatory variables and the explained variables[12]. The mathematical expression of the MGWR model is as follows:

$$y_i = \beta_0(\mu_i, v_i) + \sum_k \beta_{bwk}(\mu_i, v_i)X_{ik} + \varepsilon_i \tag{2}$$

In the formula: $y_i$represents the accident data value of the $i-th$ sample point, and X_{ik} represents the $k-th$ traffic accident influencing factor of the $i-th$ sample point, and X_{ik} represents the $k-th$ traffic accident influencing factor of the $i-th$ sample point. ε_i is a random error term, $\beta_0(\mu_i, v_i)$ represents the regression constant of the $i-th$ sample point, $\beta_{bwk}(\mu_i, v_i)$ represents the regression coefficient of the $k-th$ traffic accident influencing factor of the $i-th$ sample point.

4 MODEL RESULTS AND ANALYSIS

4.1 Model Calculation and Related Tests

(1) Selection of Spatial Econometric Models

Previous studies have shown that the higher the coefficient of determination (R^2), the smaller the Akaike Information Criterion corrected version ($AICc$), which is usually considered as an indicator of better model fit[12]. As shown in Table 1, the R^2 of the MGWR model is 0.073 higher than the GWR model and 0.136 higher than the OLS model; at the same time, the $AICc$ and RSS values of the MGWR model are both smaller than those of the GWR and OLS models, indicating that the MGWR model can better explain the spatial heterogeneity effects of independent variables on traffic accidents.

Table 1. Model comparison analysis results

Evaluation Metrics	OLS	GWR	MGWR
R^2	0.497	0.560	0.633
AICc	4565.773	4438.401	4314.125
RSS	1057.793	926.290	776.594

(2) Analysis of MGWR Model Results

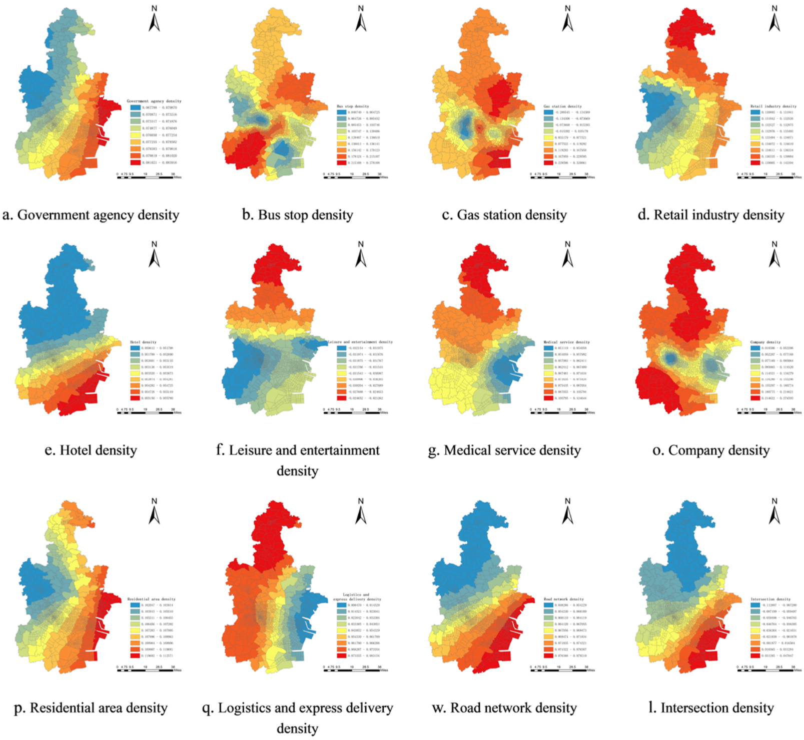
As shown in Table 2, the spatial heterogeneity regression coefficients of various influencing factors on traffic accidents were obtained through MGWR model calculation. The results indicate that there are significant differences in the regression coefficients of various influencing factors, suggesting significant variations in the impact of different built environment factors on traffic accidents. Specifically, there is a negative correlation between leisure and entertainment density and intersection density with traffic accidents, while the other variables show a positive correlation. The regression coefficient of retail industry density is significantly higher than other factors, indicating that compared to other influencing factors, the positive impact of retail industry density on traffic accidents is stronger; while the regression coefficient of leisure and entertainment density is significantly lower than other factors, indicating that the negative impact of leisure and entertainment density is stronger.

Table 2. MGWR Model Estimation Results

Variable	AVG	MEAN	MIN	MED	MAX
Government agency density	0.077	0.003	0.068	0.076	0.084
Bus stop density	0.123	0.043	0.041	0.118	0.278
Gas station density	0.057	0.118	-0.209	0.083	0.321
Retail industry density	0.134	0.002	0.131	0.133	0.143
Hotel density	0.054	0.001	0.051	0.054	0.056
Leisure and entertainment density	-0.032	0.002	-0.032	-0.032	-0.021
Medical service density	0.068	0.012	0.051	0.072	0.125
Company density	0.119	0.052	0.019	0.116	0.275
Residential area density	0.108	0.002	0.102	0.108	0.113
Logistics and express delivery density	0.048	0.023	0.006	0.062	0.083
Road network density	0.069	0.007	0.048	0.070	0.079
Intersection density	-0.023	0.038	-0.113	-0.036	0.048

4.2 Analysis of Spatial Heterogeneity of Influencing Factors

As shown in Figure 1, ArcGIS software visualizes the regression coefficients of the influencing factors, revealing the spatial heterogeneity of the influence of the influencing factors on the risk of traffic accidents.



a. Government agency density
b. Bus stop density
c. Gas station density
d. Retail industry density
e. Hotel density
f. Leisure and entertainment density
g. Medical service density
o. Company density
p. Residential area density
q. Logistics and express delivery density
w. Road network density
l. Intersection density

Fig. 1. Spatial distribution of regression coefficients of environmental impact factors

The regression coefficient of the density of government agencies is between 0.068 and 0.084, which is positive, indicating that as the density of government agencies increases, the risk of road traffic accidents increases, especially in the vicinity of Binhai New Area. This positive correlation may be related to the peak access of government agencies, the flow of people and the increase of vehicles, increasing the incidence of accidents. The regression coefficient of bus station density is between 0.041 and 0.215, and the coverage is large, reflecting strong spatial heterogeneity. There is a significant positive correlation in some areas of the outer suburbs. High-density bus stations mean higher traffic flow and increase the potential risk of traffic accidents.

The regression coefficients for gas station density cover a wide range, from -0.209 to 0.321, and the impact decreases successively from urban areas to near suburbs and far suburbs. Although the increase in fuel demand leads to higher traffic volume and may cause accidents, the urban area has more comprehensive traffic management and facilities to reduce the accident rate. The regression coefficient of retail industry density fluctuates slightly, overall positive. The regression coefficients for the density of hotels are between 0.051 and 0.067, all of which are positive values. The increased density in the retail industry and hotel sector means more foot traffic, leading to an increased risk of accidents.

The regression coefficient of leisure and entertainment density ranges from -0.032 to -0.021, indicating a negative correlation. Leisure and entertainment activities help people relieve fatigue and stress from work, thereby alleviating anxiety and reducing the risk of accidents. The regression coefficient of medical service density ranges from 0.051 to 0.125, indicating that an increase in density will lead to an increase in traffic flow, thereby increasing the risk of accidents. The regression coefficient of corporate density ranges from 0.019 to 0.275. Equivalent to the influence of suburban and exurban areas is relatively low, but overall, the mean density coefficient of companies is positive. Regions with high density of corporate enterprises are usually accompanied by traffic flow, leading to an increase in traffic accidents, especially during peak hours of commuting to and from work. The regression coefficients for residential neighborhood density are all positive, ranging from 0.102 to 0.113, indicating that as density increases, the risk of traffic accidents also increases. Similarly, the regression coefficient of logistics and express delivery density ranges from 0.007 to 0.083, showing spatial heterogeneity, with less impact near the far suburbs and Binhai new areas.

The regression coefficients of road network density range from 0.048 to 0.079, with an overall positive trend. The increase in road network density is usually accompanied by higher traffic volume and more complex traffic patterns, which may lead to driver misjudgments and thus increase the risk of accidents. The regression coefficients of intersection density range from -0.113 to 0.049, with a positive impact near the Binhai new area in the far suburbs on road traffic accidents. In Binhai New Area, there are multiple attractions such as Tianjin Polar Ocean Park and National Maritime Museum. The presence of intersections usually increases traffic volume and complexity, which may lead to an elevated risk of traffic accidents.

4.3 Managerial Implications

This paper discusses the spatial heterogeneity of built environment factors at specific spatial scales, which can help prevent traffic accidents from several aspects. First, in the process of urban construction, the comprehensive consideration of road network density and intersection density and other indicators help to rationally design and plan the road infrastructure to avoid the formation of traffic accident-prone road sections; second, in traffic management, the identification of spatial heterogeneity of influencing factors closely related to traffic accidents helps to implement different management and control measures in different spatial areas of the city to realize accurate accident prevention. accident prevention. Finally, in the management of traffic accident black spots, the study of identifying influencing factors closely related to the built environment can help to accurately find the root causes of accidents and carry out targeted management from the source, thus avoiding accidents.

4.4 Limitations

Although this article analyzes the spatial heterogeneity between the built environment and road traffic accidents based on TAZ traffic zones as the basic unit, there are still some shortcomings. Firstly, the choice of research scale may affect the results, as different grid scales can reveal completely different patterns of relationships. In the future, we will attempt to compare research results at different scales to evaluate the impact of scale effects on road traffic accidents, and determine the optimal analysis scale based on actual conditions. Secondly, the relationship between built environment and traffic accidents is extremely complex. In the future, nonlinear research methods will be considered to better identify the impact of built environment on road traffic accidents, providing more comprehensive theoretical support for the optimization of built environment in the future.

5 CONCLUSION

(1) Compared to OLS and GWR models, the MGWR model has better goodness of fit, can reflect the influence of different variables in different spatial ranges, highlight spatial heterogeneity, and provide a more detailed analysis of road traffic accidents.

(2) There are significant differences in the impact of built environment factors on traffic accidents. There is a negative correlation between the density of leisure and entertainment facilities and the density of intersections with traffic accidents, indicating that improving the density of both can enhance traffic safety. Other influencing factors have a positive correlation, meaning that as these factors increase, the incidence of traffic accidents also rises, especially the impact of retail industry density is the most significant. In addition, the impact of intersection density exhibits significant heterogeneity, especially in the Binhai new area.

ACKNOWLEDGEMENT

Funding: This research was funded by the Science and Technology Commissioner Project of Tianjin, grant number 22YDTPJC00570, the Scientific Research Project of Tianjin Education Commission, grant number 2021KJ017, and the industry-University Cooperative Education Project of the Ministry of Education, grant number 202102012013.

REFERENCE

1. World health statistics 2021: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.
2. Ding, C., Chen, P., & Jiao, J. (2018). Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: A machine learning approach. Accident Analysis & Prevention, 112, 116-126.
3. Zafri, N. M., & Khan, A. (2022). A spatial regression modeling framework for examining relationships between the built environment and pedestrian crash occurrences at macroscopic level: A study in a developing country context. Geography and sustainability, 3(4), 312-324.
4. Osama, A., & Sayed, T. (2017). Macro-spatial approach for evaluating the impact of socioeconomics, land use, built environment, and road facility on pedestrian safety. Canadian Journal of Civil Engineering, 44(12), 1036-1044.
5. Wang, J., Ji, L., Ma, S., Sun, X., & Wang, M. (2023). Analysis of Factors Influencing the Severity of Vehicle-to-Vehicle Accidents Considering the Built Environment: An Interpretable Machine Learning Model. Sustainability, 15(17), 12904.
6. Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., & Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. Accident Analysis & Prevention, 41(1), 137-145.
7. Ukkusuri, S., Miranda-Moreno, L. F., Ramadurai, G., & Isa-Tavarez, J. (2012). The role of built environment on pedestrian crash frequency. Safety science, 50(4), 1141-1151.
8. Hosseinpour, M., Prasetijo, J., Yahaya, A. S., & Ghadiri, S. M. R. (2013). A comparative study of count models: Application to pedestrian-vehicle crashes along Malaysia federal roads. Traffic injury prevention, 14(6), 630-638.
9. Huang, Y., Wang, X., & Patton, D. (2018). Examining spatial relationships between crashes and the built environment: A geographically weighted regression approach. Journal of transport geography, 69, 221-233.
10. Wang, S., Liu, J., Chen, N., Xiao, J., & Wei, P. (2023). How does the built environment affect drunk-driving crashes? A spatial heterogeneity analysis. Applied Sciences, 13(21), 11813.
11. Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). Annals of the American Association of Geographers, 107(6), 1247-1265.
12. Liu, J., Das, S., & Khan, M. N. (2024). Decoding the impacts of contributory factors and addressing social disparities in crash frequency analysis. Accident Analysis & Prevention, 194, 107375.
13. National Bureau of Statistics of the People’s Republic of China. China Statistical Yearbook; China Statistics Press: Beijing, China, 2022.

14. Zhang, K., Wang, S., Song, C., Zhang, S., & Liu, X. (2024). Spatiotemporal Heterogeneity Analysis of Provincial Road Traffic Accidents and Its Influencing Factors in China. Sustainability (2071-1050), 16(17).