Accident Analysis and Prevention 194 (2024) 107375

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Decoding the impacts of contributory factors and addressing social disparities in crash frequency analysis

Jinli Liu *, Subasish Das, Md Nasim Khan

Texas State University, 601 University Drive, San Marcos, Texas 78666, United States

ARTICLE INFO

Keywords:
Traffic safety
Social disparities
Spatial analysis
Geographically weighted regression

ABSTRACT

Understanding the relationship between social disparities and traffic crash frequency is essential for long-term transportation planning and policymaking. Few studies have systemically examined the influence of socioeconomic and infrastructure-related disparities in macro-level traffic crash frequency. This study provides a framework to spatially examine the relationships between crash rates and demographic and socioeconomic characteristics, as well as roadway infrastructure and traffic characteristics at the Census Block Groups (CBGs) level. Spatial autocorrelation analysis was first performed on the residual of the Ordinary Least Squares (OLS) model to identify whether non-stationarity exists. Then, the Geographically Weighted Regression (GWR) model and the Multiscale Geographically Weighted Regression (MGWR) model were applied to assess the impacts of these factors on crash rates spatially and statistically. Our findings indicate that MGWR outperforms both OLS and GWR in uncovering the spatial relationships between contributing factors and both fatal and injury (FI) crashes as well as property damage only (PDO) crashes. A thorough examination of local coefficient maps highlighted six pivotal variables that significantly influenced a majority of CBGs. Improving infrastructure, including pedestrian pathways and public transit facilities, in low-income areas can offer significant benefits. These findings and recommendations can inform the development of effective strategies for reducing crashes and guide the appropriate selection of modeling techniques for macro-level crash analysis.

1. Introduction

Ensuring safety has emerged as a critical factor in developing long-term transportation planning strategies. However, despite advances in transportation safety measures, the upward trend in crash and fatality risks in recent years is a cause for concern. Furthermore, these risks are not distributed evenly among different communities. Multiple earlier studies have provided evidence that members of ethnic or racial minority groups in the U.S. continue to experience a disproportionate risk of severe or fatal pedestrian traffic crashes (Lee, et al., 2019); (Apardian and Smirnov, 2020). Additionally, studies have highlighted a strong association between minority status and car crashes, with socioeconomic status being closely correlated with fatal crashes (Li, et al., 2022).

Recently, macroscopic traffic safety analysis has become popular among scholars to alleviate safety-related problems. In macro-level analysis, crash frequencies are aggregated at various spatial units such as counties, census tracts, traffic analysis zones (TAZ), block groups, intersections, and segments, among others, to investigate their correlation with several crash-inducing factors (Zafri and Khan, 2022). This type of analysis can identify the contributing built environment factors influencing crashes, explain spatial variation in crash occurrence, and inform the development of long-term safety improvement policies (Osama and Sayed, 2017).

Previous studies that have examined macro-level traffic frequency analysis have primarily concentrated on examining the influence of the built environment on pedestrian crashes in various settings (Amoh-Gyimah et al., 2016); (Chimba et al., 2018); (Apardian and Smirnov, 2020). A notable aspect of this research has been the investigation of equity concerns regarding pedestrian safety. However, the exploration of car crashes within this framework has received insufficient attention, with only a limited number of studies addressing this issue (Pulugurtha et al., 2013a); (Noland and Laham, 2018). Moreover, most of the existing research in this area has employed non-spatial global regression models to assess the association. However, the utilization of these models fails to adequately account for the crucial aspect of spatial autocorrelation, potentially introducing bias into the findings and compromising the accuracy of the analysis. Hence, there exists a critical need for further investigation utilizing spatially aware methodologies to

* Corresponding author.
E-mail address: j_l848@txstate.edu (J. Liu).

https://doi.org/10.1016/j.aap.2023.107375
Received 12 June 2023; Received in revised form 27 September 2023; Accepted 5 November 2023

obtain more precise insights into the relationship between the built environment and overall crash occurrences.

To address the existing research gap, this study aimed to examine the correlation between traffic crash frequency and contributory-built environment factors at a macro-level. A novel and comprehensive spatial regression modeling framework was developed, incorporating both the Geographically Weighted Regression model (GWR) and the Multiscale Geographically Weighted Regression model (MGWR). By applying this framework, the study focused on modeling the occurrence of car crashes at the Census Block Group (CBG) level in Texas, USA, to determine the most reliable and accurate modeling technique. It is important to note that CBG represents a geographic entity specifically employed by the United States Census Bureau. The CBG represents the most minimal spatial division for which the bureau disseminates sample data, referring to information gathered exclusively from a subset of households. In general, CBG are characterized by a population size ranging from 600 to 3,000 individuals (United States Census Bureau, 2012). Initially, the crash frequency for each CBG in Texas was averaged over a span of five years (2017——2021). To assess the spatial relationship, Global and Local Bivariate Moran's I tests were conducted. Subsequently, the GWR and MGWR models were employed to capture the spatially nonstationary influences of various variables on crash risks across Texas. Specifically, this study investigated the impact of socioeconomic, demographic, and infrastructure-related factors on the occurrence of crashes.

2. Literature review

A growing number of studies have been conducted from a social justice perspective to explore the relationship between roadway safety with different socioeconomic and demographic factors, such as income, housing, population structure, investments in urban infrastructures, and living environments at various scales (Chimba, Musinguzi and Kidando, 2018; (Noland and Laham, 2018). Chimba, Musinguzi and Kidando, (2018) explored the relationship between socioeconomic and demographic factors in pedestrian crashes. They applied the negative binomial model to assess seven selected factors' influences on pedestrian crash counts at the block group level. Their results indicated that the number of pedestrian crashes is positively correlated with some sociodemographic factors, including population density, population commuting to work by walking, the population aged 15 to 64, and households without vehicles.

Another similar study related to pedestrian safety by (Guerra et al., 2019) explored the relationship between population density and pedestrian safety in the Philadelphia region. The findings indicated a weak correlation between population density and traffic safety, with denser neighborhoods having a slightly lower rate of traffic crashes. The study also suggests that other factors, such as the design of streets and the presence of public transportation, may have a greater impact on traffic safety than population density.

Additionally, there are other studies that focus on exploring the relationship between pedestrian safety and socioeconomic and demographic factors. In a study conducted by (González et al., 2019), the impact of gentrification on pedestrian and cyclist safety was examined. Using a multivariate regression analysis, the researchers analyzed data from 213 rail stations in Los Angeles County and the Bay Area spanning the years 1997 to 2015. The findings of their study indicated that in Los Angeles County, pedestrians and cyclists face a greater risk of collision in areas around commercially gentrified stations compared to non-gentrified ones. However, this trend was not observed in the Bay Area.

In a separate investigation, (Wang, et al., 2016) conducted a study that aimed to examine the connection between pedestrian crash frequency and various predictor variables, including roadway, socioeconomic, and land-use features. The research utilized data from the urban area of Shanghai, which is the largest city in China. To capture the spatial correlations among TAZs, Bayesian Conditional Autoregressive (CAR) models were developed, incorporating seven different spatial weight features. The findings of the study indicated that TAZs with medium land use intensity had higher rates of pedestrian crashes compared to TAZs with low and high land use intensity. As a result, it is recommended to prioritize TAZs with medium land use intensity in order to enhance pedestrian safety.

The study conducted by Amoh-Gyimah, Saberi and Sarvi (2016) delved into the analysis of macro-level safety factors that impact the frequency of pedestrian and bicycle crashes. The research aimed to provide additional empirical evidence and methodological insights on this topic. Random parameter negative binomial models were employed to examine the relationship between various factors and the occurrence of pedestrian and bicycle crashes. The findings revealed significant and positive correlations between the number of pedestrian and bicycle crashes and factors such as vehicle kilometers traveled, population density, the percentage of commuters cycling or walking to work, and the percentage of households without motor vehicles. Additionally, the study identified a positive association between mixed land use and the frequency of pedestrian and bicycle crashes. Similarly, (Zafri and Khan, 2022) developed a comprehensive spatial regression modeling framework to examine the relationships between pedestrian crash occurrences and the built environment at the macroscopic level in a megacity.

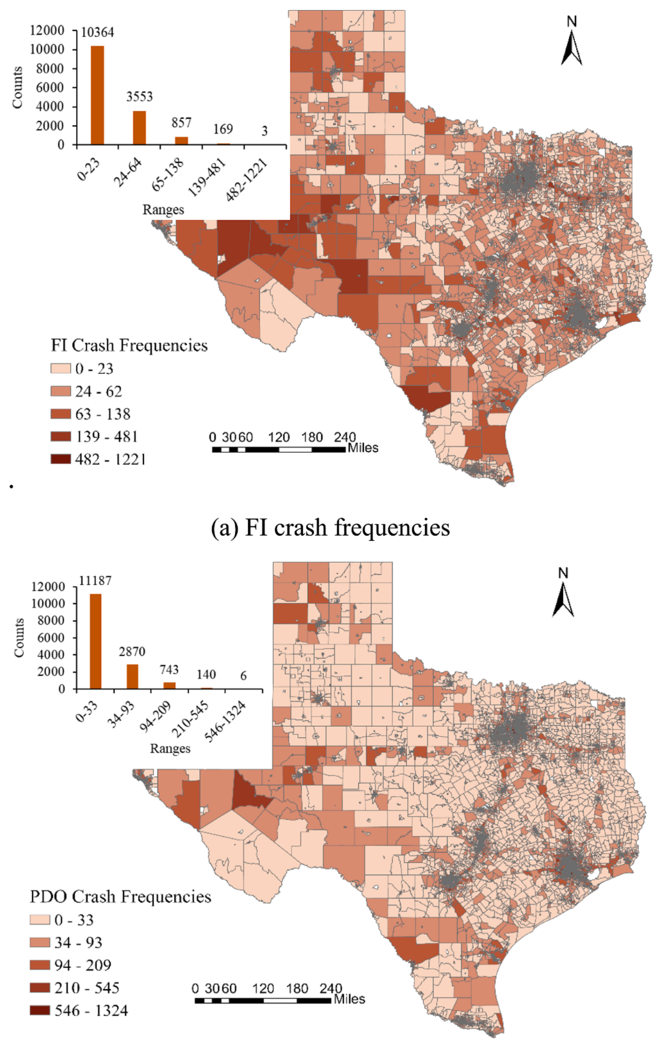
In addition to the aforementioned studies concerning pedestrian and bicycle safety regarding social disparities, there exists a body of research that focuses on vehicle crashes. For instance, a multivariate Poisson lognormal conditional autoregressive model at the macroscopic level was created by Lee, Abdel-Aty, and Jiang (2015) to analyze crashes involving various transportation modes, including motor vehicle, bicycle, and pedestrian crashes. (Cai, et al., 2017) performed a comparative analysis of various geographic units (census tracts (CTs), state-wide traffic analysis zones (STAZs), and a newly developed traffic-related zone system labeled traffic analysis districts (TADs)) for the purpose of macroscopic total, severe, and non-motorized mode crashes modeling analysis. The results of comparing these models for all types of crashes consistently showed that the models based on TADs consistently outperformed the others. Additionally, the models that took into account spatial autocorrelation exhibited better performance than those that did not consider it.

(Qu, 2021) presented a systematic method based on MGWR to explore the influence of reclassified points-of-interest (POI) on traffic crashes occurring on weekdays and weekends. Another similar study by (Li, et al., 2022) assessed the inequity in transportation safety by spatially examining the relationships between crash risks and the social vulnerability index (SVI) established by the Centers for Disease Control and Prevention (CDC). In Honolulu, Hawaii, (Kim et al., 2010) conducted a comprehensive study to explore the intricate relationship between demographic factors, land use characteristics, roadway accessibility variables, and different types of crashes. Their investigation involved constructing eight models to examine the associations between various crash categories, including total crashes, injury, fatality, pedestrian, bicycle, moped, motorcycle, and motor vehicle to motor vehicle crashes, and factors such as population, land use patterns, and accessibility measures such as road length, bus stops, length of bus route, number of intersections, and dead ends.

In a separate study by (Pulugurtha et al., 2013b), the focus was on developing crash estimation models at the Traffic Analysis Zone (TAZ) level, taking into consideration various land use characteristics. The findings revealed a strong association and statistically significant role of land use characteristics, such as mixed-use development, urban residential areas, single-family residential areas, multi-family residential areas, business districts, and office districts, in estimating crashes at the TAZ level. Notably, the coefficient for single-family residential areas indicated a negative relationship, suggesting a decrease in the number of crashes with an increase in the extent of single-family residential areas. Additionally, (Xu, et al., 2020) conducted a study investigating the relationship between land-use patterns and crash frequency at the TAZ level, utilizing K-means clustering and GWPR models. The analysis

Table 1
Yearly breakdown of crash numbers by severity level.

Severities	K (Fatal Injury)	A (Incapacitating crashes)	B (Non-Incapacitating)	C (Possible Injury)	O (No Apparent Injury)	Total
2017	3,317	13,317	54,409	90,859	335,595	497,497
2018	3,273	11,310	50,242	93,808	371,148	529,781
2019	3,253	12,453	52,447	102,292	370,293	540,738
2020	3,520	11,821	45,219	80,935	322,415	463,910
2021	4,033	15,431	59,286	85,302	376,000	540,052



(a) FI crash frequencies

(b) PDO crash frequencies

Fig. 1. Spatial distribution of FI and PDO level crash frequencies.

showed that the land-use pattern characterized by community and life services exhibited the highest crash risk, 1.295 times that of the residential-oriented land-use pattern when considering factors such as traffic, network, and road infrastructure characteristics. Conversely, the land-use pattern emphasizing nature and ecology had the lowest crash risk, with a rate of 0.754 times that of the residential-oriented land-use pattern.

The literature review summarizes a growing body of research exploring the relationship between roadway safety and socioeconomic and demographic factors at various scales from a social justice perspective. Studies have focused on pedestrian and vehicle crashes and have employed a variety of analytical approaches, including negative binomial models, spatial regression analysis, and GWR. The findings suggest that sociodemographic factors such as population density, poverty, street connectivity, and land use patterns influence roadway safety. Additionally, several studies have directed their attention toward vehicle crashes, further broadening the scope of investigation. These studies emphasize the importance of considering both individual and environmental factors in understanding the causes of vehicle crashes and developing effective safety policies. The review highlights the importance of considering these factors in transportation planning and designing safe and equitable transportation systems. The literature reviews shows that there is limited exploration of the broader regional or macro-scale factors that may influence roadway safety from a social justice perspective.

3. Data collection and processing

3.1. Study area

Texas is a state located in the South-Central region of the U.S., with a population density of 29.1 million residents in 2020. It is the second-largest U.S. state by both area and population. The state is characterized by a vast and diverse landscape, with large urban areas, sprawling suburbs, and vast expanses of rural land. To facilitate transportation across such a large and varied terrain, Texas has invested in developing one of the largest and most extensive highway and railway systems in the country. Houston, San Antonio, and Dallas are the most populous metropolitan areas in the state. In order to investigate the factors associated with crash rates at the marcrolevel, five years of crash data (2017——2021) along with their location were retrieved from the Crash Records Information System (CRIS), owned by the Texas Department of Transportation (TxDOT). CRIS is an automated database that collects and tracks statewide traffic crash records; it contains all the data received from the Texas Peace Officer's Crash Report (Texas Department of Transportation, 2023). There are over 5 million crash records each year. Table 1 provides a comprehensive breakdown of the number of yearly crashes across the entire state by different severity types. The original dataset classifies crashes into five distinct levels based on the severity of injuries sustained during the crash. These levels are represented using the acronym KABCO. Over the years, the aggregate number of crashes has seen fluctuations, and the distribution across the different severity levels exhibits distinct variations.

3.2. Crash frequency at CBG level

In this study, crash data will be aggregated by CBG. A CBG, as defined by the (United States Census Bureau, 1994), and it is a fundamental element in the hierarchical structure of census geographic entities. The state of Texas is divided into 254 counties, each of which is further segmented into census tracts. These tracts are subsequently subdivided into one or more block groups. The specific count of these divisions can vary, from different census years. In this study, we employed the block group boundaries from the 2018 Census, which represents the latest boundary version adopted in the SLD. As of 2018, there are a total of 14,949 CBGs. This aggregation is carried out using the intersect function in the ArcGIS Pro software. Many studies have focused on bifurcating crashes into two major categories (Bauer and Harwood, 2013); (Ahmad et al., 2023). Following this approach, rather than examining all five levels of severity, this study classified crashes into two primary categories: Fatal and Injury-level (FI) crashes, which

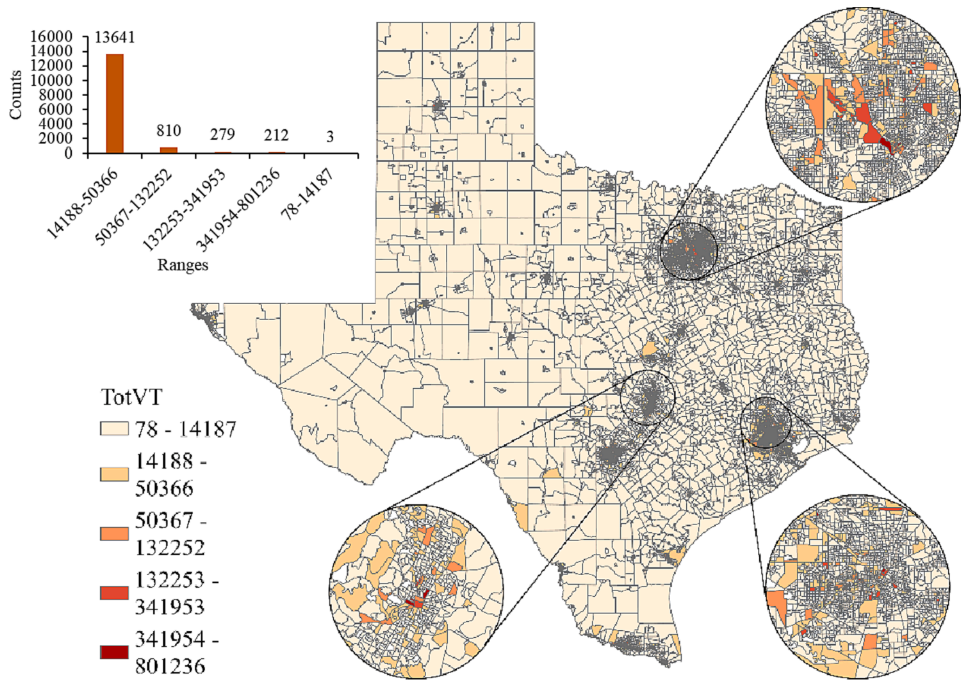


Fig. 2. Spatial distribution of total vehicle trip.

include Types K, A, B, and C; and Property Damage Only (PDO) crashes, encompassing Types O (Hall and Tarko, 2019). The cumulative crash frequency for each CBG is then averaged over a span of five years. With a total of 14,949 CBGs, there are correspondingly 14,949 FI crash frequencies and PDO crash frequencies. Fig. 1 illustrates the geographical distribution of these two crash frequency types at the CBG level. For the FI crashes, the five-year average spans from 0 to 1,221. Notably, 254 out of the 14,949 CBGs reported no crashes, constituting approximately 1.7 %. The majority of CBGs have crash frequencies within the 0–23 range. For the PDO crashes, the range is between 1 and 1,324, with 11,187 CBGs falling within the 0–33 range. Notably, no CBG reported an absence of PDO crashes.

3.3. Traffic exposure and independent variables

After examining various data sources, it was found that traffic exposure data is not directly available at the CBG level. However, this information can be sourced from the Smart Location Database (SLD). The SLD is a public data product jointly provided by the U.S. Environmental Protection Agency (EPA) and the U.S. General Services Administration (GSA). It compiles various demographic, employment, and built environment metrics for every CBG in the U.S. and is accessible as a file geodatabase. It is worth noting that there are over 120 variables in the original database. The variable $D2c_{TrpMx1}$, representing employment and household entropy, is availabe in SLD. Its mathematical formulation can be found in Equation (1).

$$\begin{aligned} D2c_{TrpMx1} = {} & (HH*11/TotVT)*\ln\left(HH*\frac{11}{TotVT}\right) + (E5\_Ret*22/TotVT) \\ & *\ln\left(E5\_Ret*\frac{22}{TotVT}\right) + (E5\_Off*3/TotVT)*\ln\left(E5\_Off*\frac{3}{TotVT}\right) \\ & + (E5\_Ind*2/TotVT)*\ln\left(E5\_Ind*\frac{2}{TotVT}\right) + (E5\_Svc*31/TotVT) \\ & *\ln\left(E5\_Svc*\frac{31}{TotVT}\right) + (E5\_Ent*43/TotVT)*\ln\left(E5\_Ent*\frac{43}{TotVT}\right) \end{aligned} \tag{1}$$

In this equation, *TotVT* denote the total trips generated, encompassing both production and attraction, for all activity categories in the CBG. The variable *HH* denotes the number of households in the CBG. Employment within a five-tier classification scheme is represented by *E5_Ret*, *E5_Off*, *E5_Ind*, *E5_Svc* , *E5_Ent* which denote retail jobs, office jobs, industrial jobs, service jobs, and entertainment jobs, respectively. All these variables are available in the SLD dataset, enabling the calculation of *TotVT*. Fig. 2 presents the spatial distribution of *TotVT*, which ranges from 42 to 801,236. As illustrated in the figure, metropolitan areas, including Dallas, Houston, Austin, and San Antonio, have the highest number of vehicle trips.

After *TotVT* values are obtained, both FI crash rates and PDO crash rates can be calculated. Two levels of crash rates per 100 trips at *i*th CBG were derived using the formulas: $\boldsymbol{FI\_Rate_i} = \boldsymbol{FICrash\ Counts_i}*100/\boldsymbol{TotVT_i}$; $\boldsymbol{PDO\_Rate_i} = \boldsymbol{PDOCrashCounts_i}*100/\boldsymbol{TotVT_i}$. In terms of crash predictor variables, previous studies have explored all kinds of variables, including demographic factors, socio-economic factors, land use, and land cover characteristics. Through a thorough review, the independent variables used in this study can be grouped into two categories: demographic and socio-economic characteristics and roadway and traffic characteristics. These variables were obtained from the SLD. After a careful review, 15 variables were selected. Table 2 describes and summarizes all the independent and dependent variables. Out of the original 14,949 CBGs, 4 had invalid values and were consequently excluded from the analysis dataset.

4. Methodology

This study proposed a spatial modeling framework to explore the associations of the demographic, socio-economic, and built environment factors on two severity levels of vehicle crash rates, incorporating spatial autocorrelation and spatial heterogeneity. The study flowchart is presented in Fig. 3. This framework was developed based on previous literature (Zafri and Khan, 2022). The analysis began with variable selection. Initially, 15 variables were chosen based on their categories and inherent data qualities. Following this, any variables exhibiting multicollinearity were identified and eliminated. The Variance Inflation Factor (VIF) served as the primary metric for assessing multicollinearity among these variables. Once the variables were refined, the study

Table 2
Variable description and descriptive statistics.

Variables	Description	Min	Max	Mean	STD
Crash Characteristics					
FI_Rate	Rate of Fatal and Injury (FI) crashes per 100 trips	0.000	53.741	0.564	1.008
PDO_Rate	Rate of Property Damaged Only (PDO) crashes per 100 trips	0.002	37.143	0.678	1.011
Demographic and Socio-economic Characteristics					
ResDens	Gross residential density (HU/acre) on unprotected land	0	188.84	2.54	3.51
PopDens	Gross population density (people/acre) on unprotected land	0	283.75	6.33	7.48
EmpDens	Gross employment density (jobs/acre) on unprotected land	0	309.61	2.37	8.61
WorkAgeRate	Ratio of population that is working aged 18 to 64 years, 2018	0	1.00	0.59	0.10
ZeroCarRate	Ratio of zero-car households in CBG, 2018	0	0.84	0.06	0.08
OneCarRate	Ratio of one-car households in CBG, 2018	0	1.00	0.33	0.15
MultiCarRate	Ratio of two-plus-car households in CBG, 2018	0	1.00	0.60	0.19
LowWageRate	Ratio of low wage workers in a CBG (home location), 2017	0	0.50	0.23	0.06
EmpHHEnt	Employment and household entropy	0	1.00	0.50	0.23
EmpEnt	Static eight-tier employment entropy	0	0.95	0.52	0.22
ActDens	Gross activity density (employment + HUs) on unprotected land	0	320.30	4.91	9.86
Roadway infrastructure and Traffic Characteristics					
AONetDens	Network density in terms of facility miles of auto-oriented links per square mile	0	40.59	1.47	2.78
MMNetDens	Network density in terms of facility miles of multi-modal links per square mile	0	26.82	2.55	2.55
PONetDens	Network density in terms of facility miles of pedestrian- oriented links per square mile	0	53.88	12.49	8.10
IntDens	Street intersection density per square mile	0	864.31	77.46	67.34

proceeded with the application of the OLS global regression model, followed by spatial autocorrelation analysis and spatial regression analysis. The OLS model is a linear regression approach that employs the Ordinary Least Squares method for parameter estimation (Shabrina et al., 2021). Notably, the OLS model is non-spatial, meaning it does not account for spatial non-stationarity in its modeling process.

4.1. Spatial autocorrelation analysis

To assess if spatially heterogeneous situation exists in the model, one of the common processes is to evaluate if the residual of the OLS model is spatially correlated (Huang et al., 2018). Then, the Global Moran's I test was performed to assess the spatial relationship between the crash frequency and independent variables. The Global Moran's I test is a

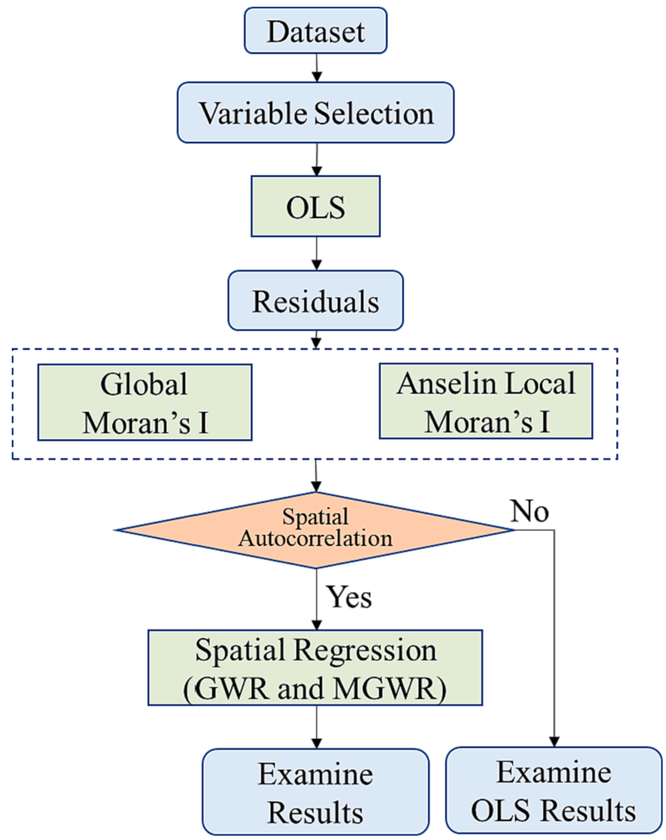


Fig. 3. Study flow chart.

Table 3
Comparation among OLS, GWR and MGWR (Shabrina, Buyuklieva and Ng, 2021).

Model	Description	Key Features
OLS (Ordinary Least Squares)	Global regression model that minimizes the sum of squared residuals	- Assumes constant relationships across geographical space, known as spatial non-stationary - Provides a single global regression equation
GWR (Geographically Weighted Regression)	Extends global regression models that allows relationships between variables to vary across space	- Uses spatial weights to give importance to nearby observations. - Assume univariate scale for all predictors
MGWR (Multiscale Geographically Weighted Regression)	Extends GWR by allowing different spatial scales for each variable	- Allows for varying spatial scales for each predictor. - Capture local to regional effects for different variables.

statistical measure used to assess spatial autocorrelation in a dataset (Getis and Ord, 1992). While the Global Moran's I test provides an overall assessment of spatial autocorrelation, the Local Moran's I allows for the detection of spatial clusters or patterns at a local scale. Therefore, Anselin Local Moran's I, a widely adopted local spatial autocorrelation test method (Anselin, 1995), was used to identify and map the clusters of significant associations between crash frequency and independent variables at local levels. This study developed these two models in ArcGIS Pro software.

Table 4
Variance Inflation Factor (VIF) values.

Variables	VIF value
WorkAgeRate	1.21
ZeroCarRate	3.53
OneCarRate	**9.25**
MultiCarRate	13.23
LowWageRate	1.42
ResDens	**> 1000**
PopDens	8.59
EmpDens	**> 1000**
ActDens	**> 1000**
EmpHHEnt	1.34
EmpEnt	1.52
AONetDens	1.13
MMNetDens	1.41
PONetDens	6.39
IntDens	6.35
TotVT	1.82

Note: VIF values (>7.5) indicate redundancy among explanatory.

4.2. Spatial regression analysis

If a spatial correlation is found to exist, the spatially nonstationary relationships between crash frequency and selected variables will be modeled using GWR and MGWR. Table 3 provides an overview of OLS, GWR, and MGWR. Compared to OLS, GWR, and MGWR, by their very nature, attempt to model spatially varying relationships, which can help account for spatial non-stationarity. In addition, MGWR is an extension of GWR, which allows for different spatial scales for each variable. These two models were conducted using the open-source Python package mgwr (Oshan, et al., 2019).

5. Results and analysis

5.1. Variable selection

This study first examined the multicollinearity problem. Table 4 presents the VIF results of all independent variables presented in Section 3.1. Six variables were found to have VIF values over 7.5 (O'brien, 2007). As the VIF value greater than 7.5 represents multicollinearity, these variables need to be removed. First, ResDens, EmpDens, and ActDens were removed since their VIF value was higher than 7.5. Then, after the experiment, it was found that a higher R square score could be acquired by removing OneCarRat instead of MultiCarRat and PopDens. Therefore, OneCarRate was removed as well.

With the final selected variables, a new OLS regression model was fitted. All the variables passed the multicollinearity test. The final modeling results are presented Table 5. The model had an R^2 value of 0.0847 for FI level crash and 0.1128 for PDO level crash, indicating that the eleven independent variables could explain an 8.47 % variation in the FI level crash frequency and an 11.28 % variation in the PDO level crash frequency. While both WorkAgeRate and ZeroCarRate are not statistically significant in either model, the other nine variables are significant at the 0.05 level in both the FI and PDO level models. The sign of the coefficient remains consistent across both models; however, the relationship between the same variables and the two dependent variables exhibits variation. For instance, MultiCarRate bears a positive sign in both models, but it has a larger value in the FI model compared to the PDO model. This suggests a stronger association with FI crashes.

In order to assess the validity of the results obtained from the OLS model, it was essential to investigate the presence of spatial autocorrelation or spatial non-stationarity. The Moran's I diagnostics of residual for the FI model was found to be statistically significant, with Moran's Index equal to 0.032132 (Table 6). This result suggested the presence of spatial autocorrelation in the model. Spatial autocorrelation in the residuals is often interpreted as an underlying spatial process that induces spatial autocorrelation in some of the variables is missing from the model. Therefore, this result highlights that the traditional non-spatial regression technique was not the best approach for modeling the FI level crash frequency. Similar results were obtained for the PDO level crash.

In addition to the application of Moran's I diagnostics, this study also conducted the Anselin Local Moran's I test to analyze two distinct categories of crashes. The analysis revealed the presence of four distinct cluster types: areas with high crash values surrounded by neighboring areas with high values (referred to as 'High-High' clusters), regions with low crash values surrounded by neighboring areas with low values ('Low-Low' clusters), areas with low crash values surrounded by neighboring areas with high values ('Low-High' clusters), and areas with high crash values surrounded by neighboring areas with low values ('High-Low' clusters). These significant clusters are visually represented in Fig. 4. Upon examining the spatial clustering patterns depicted in Fig. 4 (a) and Fig. 4 (b), it became evident that they exhibited notably different characteristics. Moreover, this analysis revealed that the

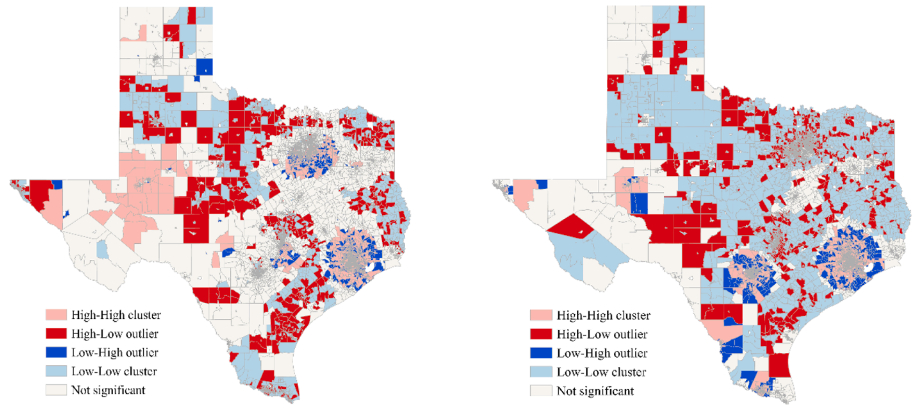
Table 6
Global Moran's I.

Models	FI level crash	PDO level crash
Moran's Index	0.032132	0.043912
Expected Index	−0.000067	−0.000067
Variance	0.000001	0.000001
z-score	32.561642	44.578859
p-value	0.000000*	0.000000*

Table 5
The results of the Possion regression model.

Models	FI level crash				PDO level crash			
Variable	Coefficient	STD	P-value	VIF	Coefficient	STD	P-value	VIF
Intercept	0.5637	0.0079	0.0000	–	0.6782	0.0078	0.0000	–
WorkAgeRate	−0.0011	0.0084	0.8935	1.1	−0.0175	0.0083	0.0347	1.1
ZeroCarRate	−0.0039	0.0101	0.6994	1.6	0.0033	0.0100	0.7424	1.6
MultiCarRate	0.0283	0.0106	0.0073	1.8	0.0375	0.0104	0.0003	1.8
LowWageRate	0.0787	0.0092	<0.0001	1.3	0.1378	0.0090	<0.0001	1.3
PopDens	−0.0564	0.0103	<0.0001	1.7	−0.0690	0.0102	<0.0001	1.7
EmpEnt	0.1008	0.0097	<0.0001	1.5	0.1208	0.0096	<0.0001	1.5
EmpHHEnt	0.1260	0.0084	<0.0001	1.1	0.2391	0.0082	<0.0001	1.1
AONetDens	0.0194	0.0093	0.0359	1.4	0.0489	0.0091	<0.0001	1.4
MMNetDens	−0.2957	0.0197	<0.0001	6.3	−0.1296	0.0195	<0.0001	6.3
PONetDens	0.1421	0.0197	<0.0001	6.3	0.0789	0.0195	0.0001	6.3
IntDens	0.5637	0.0079	0.0000	1.0	0.6782	0.0078	0.0000	1.0
AICc	41345.3843				40955.2124			
R-squared	0.0847				0.1128			

Note: AICc is short for Akaike's Information Criterion



(a) FI crash (b) PDO crash

Fig. 4. Spatial clusters identified by Anselin Local Moran's I.

Table 7
The goodness of fit results for the 6 models.

Models		Evaluation Indexes		
		AICc	R^2	Log-likelihood
OLS	FI	41345.38	0.085	−20662.68
———	PDO	40955.21	0.11	−20467.60
GWR	FI	34797.20	0.64	−13588.01
	PDO	36552.52	0.58	−14768.28
MGWR	FI	32976.74	0.71	−11892.87
	PDO	35487.23	0.64	−12798.26

clustering patterns observed in crashes at the FI level were less significant compared to those observed at the PDO level. This finding suggests that FI crashes exhibited lower levels of spatial clustering compared to PDO level crashes. Specifically, Fig. 4 (a) highlights the prevalence of High-High clusters concentrated in Houston, Dallas, and Fort Worth, providing strong evidence for the existence of spatial autocorrelation in these regions. These findings contribute to our understanding of the spatial distribution of crashes and reinforce the presence of spatial dependencies within the data.

Therefore, there is a need to adopt a modeling approach that induces spatial correlation. GWR and MGWR are the two most popular local spatial regression modeling approaches; they usually perform better than a global regression because these approaches can explore the local relationships and account for spatial non-stationarity characteristics.

5.2. GWR and MGWR model comparison

Given the presence of spatial autocorrelation in the residuals of the OLS model, it became evident that employing the GWR and MGWR models was essential. These models account for spatial relationships during the modeling process. These two models can be used when there is spatial autocorrelation in the residuals from the aspatial regression or the regression coefficients are not stationary. In this study, the performance of these two models was compared. The analysis software used in this study is MGWR 2.2 (Oshan, et al., 2019). Table 7 shows the summary results of the GWR and MGWR model. The following indexes were used to evaluate the model's fitness: Akaike's Information Criterion (AICc), R-squared (R^2), and log-likelihood. For RSS and AIC, smaller values indicate better fitness. For R^2 and log-likelihood, the higher the value, the better the fitness. Collectively, for both FI and PDO levels of fitting, MGWR outperforms the OLS and GWR.

5.3. MGWR modeling results

Utilizing the MGWR model as a local spatial regression model allows for the generation of local coefficient estimates that effectively capture the spatial heterogeneity involved in influencing crash frequencies. The findings obtained through MGWR demonstrate that the relationship between the crash rates and independent variables within the study area undergoes changes in both direction and intensity across different spatial contexts. It is evident that the impact of the same variables on traffic crashes can vary significantly across distinct geographical locations. Table 8 presents a comprehensive summary of the key statistical

Table 8
Summary statistics for the MGWR coefficient estimates of FI crash.

Variable	Coefficient					Significant CBGs (% of Features)[a]	Bandwidth (% of Features)[b]
	Mean	STD	Min	Median	Max		
Intercept	−0.049	0.161	−0.420	−0.077	1.628	8 (0.05)	145 (0.97)
WorkAgeRate	0.005	0.296	−7.258	−0.005	6.179	220 (1.47)	33 (0.22)
ZeroCarRate	−0.024	0.001	−0.029	−0.023	−0.022	14,945 (100.00)	14,945 (100.00)
MultiCarRate	0.033	0.204	−2.298	0.024	9.174	194 (1.30)	48 (0.32)
LowWageRate	0.065	0.001	0.061	0.065	0.066	14,945 (100.00)	14,945 (100.00)
PopDens	−0.070	0.020	−0.137	−0.061	−0.052	14,945 (100.00)	7072 (47.32)
EmpEnt	0.003	0.130	−4.063	0.010	0.569	179 (1.20)	108 (0.72)
EmpHHEnt	0.075	0.033	0.049	0.065	0.196	14,945 (100.00)	4896 (32.76)
AONetDens	0.156	0.249	−2.532	0.135	2.384	1293 (8.65)	72 (0.48)
MMNetDens	0.025	0.235	−6.736	0.036	0.566	200 (1.34)	97 (0.65)
PONetDens	−0.214	0.210	−4.630	−0.198	0.873	2359 (15.78)	89 (0.60)
IntDens	0.076	0.001	0.074	0.076	0.079	14,945 (100.00)	14,945 (100.00)

a: percentage of CBGs that have coefficients with $p \leq 0.05$.
b: local, regional, global scale regression

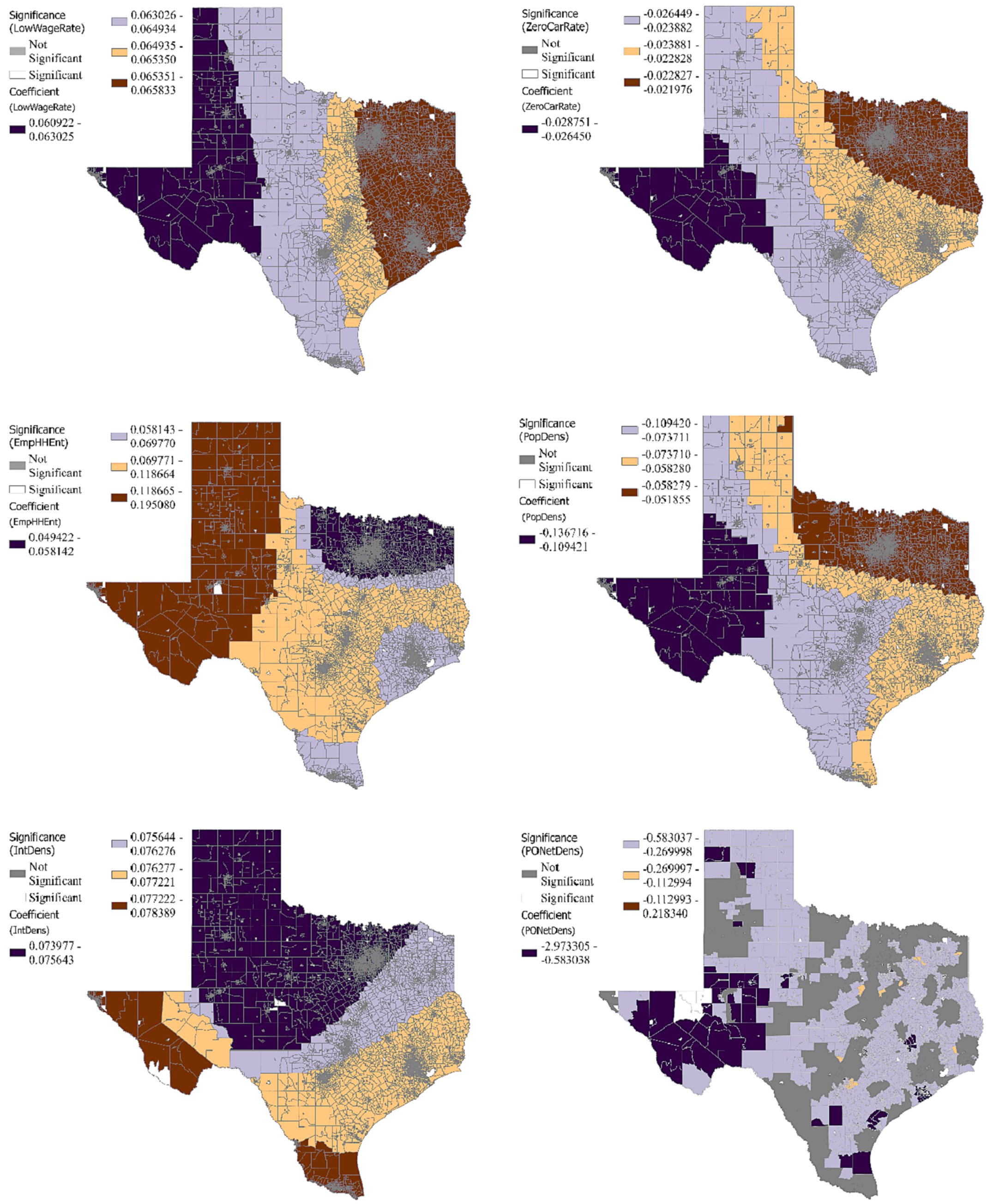


Fig. 5. Local coefficient estimates map of FI level crash from MGWR model.

measures pertaining to the coefficient-based geographical analysis, encompassing mean, maximum, and minimum values, as well as the proportion of CBGs exhibiting significant estimates and the associated bandwidth of the regression coefficients for each predictor variable within the fixed-interval MGWR model. Based on the percentage of CBGs with $p \leq 0.05$ (%), the variables ZeroCarRate, LowWageRate, PopDens, EmpHHEnt, IntDens consistently exhibit values of 100 %, implying that the local coefficient estimates for these variables consistently exhibit significance across all CBGs. This significant consistency underscores the substantial influence of these variables on crash frequency within the study area. Conversely, other variables, such as WorkAgeRate do not demonstrate significance across all CBGs, indicating the association between crash rate and WorkAgeRate within the study area is not consistently significant. Furthermore, the optimal bandwidth for the regression coefficients ranges from 33 to 14,945 CBGs, suggesting that the independent variables operate on distinct scales. Among the examined variables, ZeroCarRate, LowWageRate, and IntDens each have a bandwidth encompassing 100 % of CBGs, indicating a global relationship between these variables and crash rates. This implies that all data points are utilized with equal weight to determine the regression coefficients. For the other variables, a specific point will utilize the nearest bandwidth of data points with non-zero weights for regression estimation. Furthermore, among these variables, PONetDens exhibits the highest mean coefficient, followed by AONetDens, and then IntDens.

The local coefficient maps for the six variables with the largest proportion of CBGs with significant coefficients are presented in Fig. 5. These variables are ZeroCarRate, LowWageRate, PopDens, EmpHHEnt, PONetDens, and IntDens. These factors are broadly categorized into two principal groups: demographic and socio-economic attributes; and roadway infrastructure and traffic characteristics. A detailed analysis of each factor, along with their potential recommendations, is provided.

Table 9
Summary statistics for the MGWR coefficient estimates of PDO crash.

Variable	Coefficient					Percentage of CBG with $p \leq 0.05$ (%)	Bandwidth
	Mean	STD	Min	Median	Max		
Intercept	0.010	0.237	−0.462	−0.033	2.560	27 (0.18)	128 (0.86)
WorkAgeRate	−0.001	0.237	−5.130	−0.009	2.949	269 (1.80)	41 (0.27)
ZeroCarRate	−0.025	0.004	−0.042	−0.024	−0.023	14,945 (100.00)	14,945 (100.00)
MultiCarRate	0.043	0.282	−6.588	0.039	9.031	254 (1.70)	33 (0.22)
LowWageRate	0.107	0.000	0.107	0.107	0.108	14,945 (100.00)	14,945 (100.00)
PopDens	−0.083	0.018	−0.126	−0.090	−0.060	14,945 (100.00)	8099 (54.19)
EmpEnt	0.029	0.123	−3.335	0.026	0.528	320 (2.14)	116 (0.78)
EmpHHEnt	0.106	0.022	0.081	0.098	0.167	14,945 (100.00)	5727 (38.32)
AONetDens	0.253	0.310	−3.219	0.213	2.817	1620 (10.84)	44 (0.29)
MMNetDens	0.049	0.001	0.046	0.049	0.049	14,945 (100.00)	14,945 (100.00)
PONetDens	−0.153	0.023	−0.236	−0.144	−0.122	14,945 (100.00)	5727 (38.32)
IntDens	0.078	0.002	0.074	0.078	0.082	14,945 (100.00)	14,945 (100.00)

5.3.1. Demographic and socio-economic characteristics

For the ratio of low wage workers in a CBG (LowWageRate), every CBG exhibits a significant coefficient, notably, all these coefficients are positive. This indicates that areas with a higher proportion of low-wage workers consistently see higher crash rates across all CBGs. The strength of this positive association is less intense in the western part of Texas but intensifies as one moves eastward, as depicted on the map. The consistent positive correlation between the ratio of low-wage workers and crash rates suggests that areas with a higher concentration of low-wage workers experience more overall FI crashes. Consistent results are found in (Noland and Laham, 2018)'s study. This situation could be associated with various socio-economic factors, such as vehicle maintenance, commuting patterns, or access to safer transportation options. Possible ways to improve this situation include prioritizing infrastructure improvements in areas with a high proportion of low-wage workers, ensuring roads, pedestrian pathways, and public transit facilities are safe and well-maintained. Ensure public transportation is affordable and accessible in areas with high concentrations of low-wage workers.

For the variable representing the ratio of zero-car households in CBGs (ZeroCarRate), every CBG exhibits a significant coefficient, and all these coefficients are negative. This suggests that a higher ratio of zero-car households is consistently associated with lower FI crash rates across all CBGs. This negative association is less pronounced in the eastern part of Texas and intensifies as one moves westward, as visualized on the map. The consistent negative association between zero-car household ratios and crash rates suggests that areas with more households without cars might have fewer FI crashes. This could be due to a variety of factors, such as increased use of public transportation or more walkable neighborhoods. Promoting the safety and walkability of public transportation in areas with higher FI crash rates may potentially reducing the number of vehicles on the road and subsequently, the chances of traffic crashes. Similar results were found on (Merlin, 2020)'s study, where zero car per household were found to has negative association to crashes per capita.

For the variable representing employment and household entropy (EmpHHEnt), every CBG shows a significant positive coefficient. This indicates that areas with higher employment and household diversity consistently experience higher FI crash rates across all CBGs. Notably, the strength of this positive association is particularly pronounced in North Texas, such as Amarillo and Lubbock areas, as reflected in the map. The consistent positive relationship between EmpHHEnt and FI crash rates suggests that areas with diverse employment and household patterns may have varied commuting routines and travel demands, potentially leading to more vehicular interactions and subsequent crashes. The pronounced effect in Amarillo and Lubbock could be due to specific regional factors that further amplify this trend. Given the varied employment and household patterns, providing diverse transit options catering to different commuting needs might help reduce on-road interactions and potential crashes. Given the pronounced effect in Amarillo and Lubbock, targeted interventions in these areas, such as traffic calming measures, improved signage, or public awareness campaigns, could be beneficial.

Gross population density (PopDens) has a range of significant negative coefficient values across Texas. The negative correlation suggests that areas with higher population densities tend to experience fewer FI vehicular crashes. This trend is most noticeable in the West but becomes less distinct toward the East. Higher population densities might correlate with factors such as more public transit use, walkable neighborhoods, or slower driving speeds due to congestion, all of which can reduce crash rates.

5.3.2. Roadway infrastructure and traffic characteristics

In terms of the street intersection density per square mile (IntDens), the results indicate that areas with higher intersection densities consistently witness higher crash frequencies across all CBGs. Comparable findings can be observed in the study conducted by (Siddiqui et al., 2012). It makes sense because intersections often involve complex vehicular and pedestrian interactions, which can increase the potential for collisions. Possible solutions can be to enhance the design of intersections, especially in areas with high densities, to streamline traffic flow and reduce potential conflict points.

For the pedestrian-oriented network density (PONetDens), not all CBGs exhibit a significant coefficient. Specifically, only 15.78 % of CBGs show significant coefficients for PONetDens. Areas with non-significant coefficients are not concentrated but instead are dispersed throughout Texas. However, where PONetDens does have a significant impact, it is all negative, indicating a reduction in FI vehicle crash rates. It indicates that while pedestrian-oriented network density doesn't impact FI crash rates everywhere, where it does have an effect, it tends to reduce crashes. This could be due to design elements that prioritize pedestrian safety, potentially leading to fewer vehicle interactions and collisions. This suggests that considering designing or modifying urban areas to be more pedestrian-friendly could probably reduce the reliance on cars and decrease overall FI crash rates, especially in areas with significant coefficients. (Mahmoudi, et al., 2023) also recommended the construction of new pedestrian- and bicyclist-oriented facilities to reduce the frequency of pedestrian and bicyclist crashes.

Overall, across the entire Texas area, all six variables exhibit coefficients with consistent signs, indicating a uniform direction of associations. Specifically, ZeroCarRate, PopDens, and PONetDens have negative associations, while LowWageRate, EmpHHEnt, and IntDens possess positive associations.

Table 9 presents comprehensive statistics regarding the Mean, Maximum, and Minimum values, as well as the proportion of CBGs exhibiting noteworthy estimates and the bandwidth of the regression coefficients for each predictor variable in the MGWR model, specifically focusing on PDO level crash rates. Notably, the variables percentage of CBG with $p \leq 0.05$ (%), ZeroCarRate, LowWageRate, PopDens,

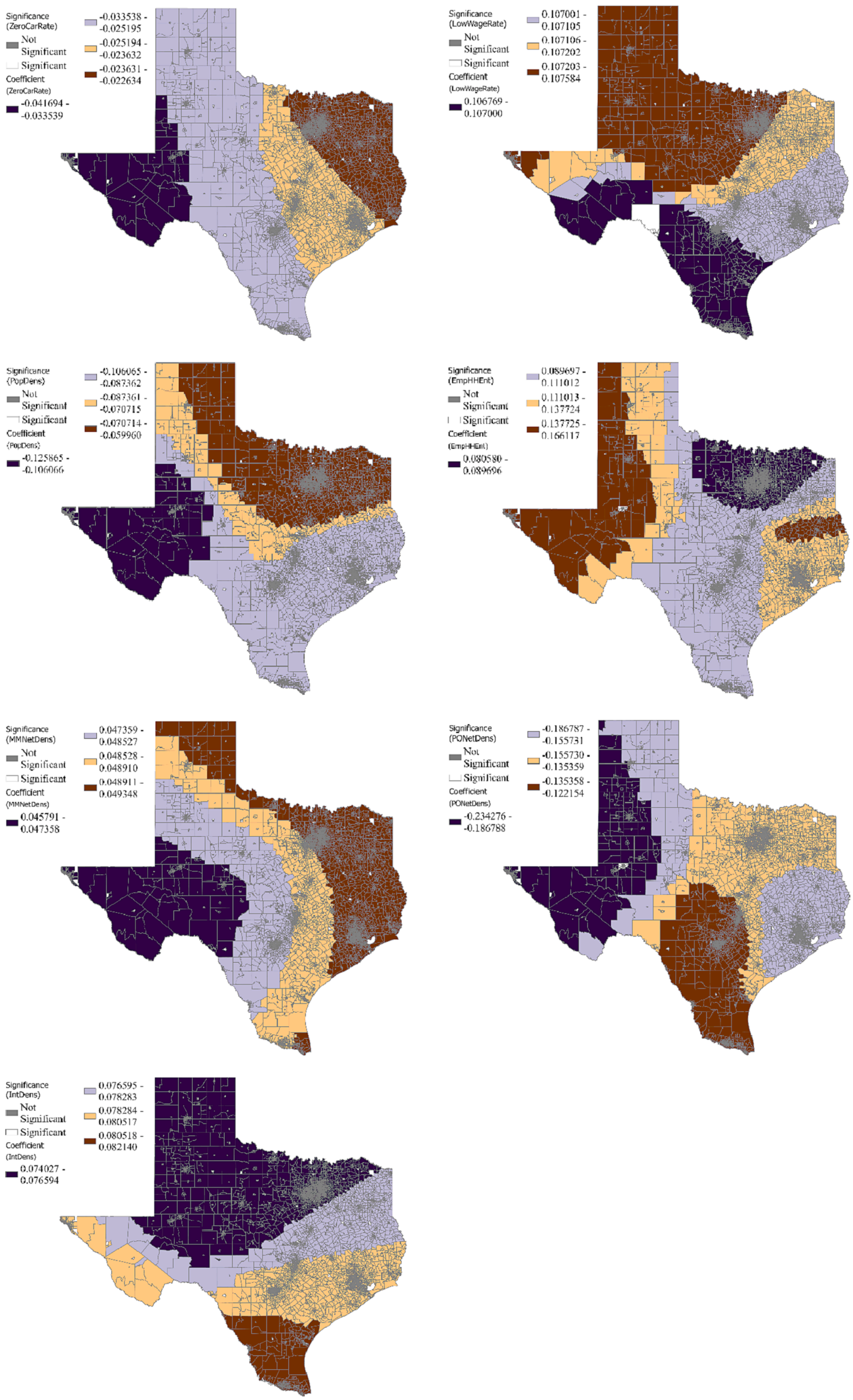


Fig. 6. Local coefficient estimates of PDO level crash map from MGWR models.

EmpHHEnt, MMNetDens, PONetDens and IntDens, all exhibit a value of 100 %. This suggests that the local coefficient has a significant association with crash rates. Among these variables, ZeroCarRate, LowWageRate, MMNetDens, and IntDens have a bandwidth of 100 %, while PopDens, EmpHHEnt, PONetDens have a bandwidth of less than 100 %. This suggests the variation of the association at different scales, 100 % indicating global association while the other indicating a more local association. While for the other variables, they have a very low percentage of significant CBGs; specifically, AONetDens only has 10.84 %, MultiCarRate least, with only 1.70 %.

Conversely, no variable demonstrates insignificance across all CBGs for PDO level crashes. Moreover, the optimal bandwidth ranges from 43 to 14,948 CBGs, aligning with the range observed for fatal injury level crashes. In addition, among these variables, AONetDens has the highest mean, STD absolute coefficient value.

Fig. 6 presents the local coefficient maps of the top seven variables with the highest percentage of CBGs having significant coefficients. These variables include ZeroCarRate, LowWageRate, PopDens, EmpHHEnt, MMNetDens, PONetDens, and IntDens. When compared to FI level crashes, MMNetDens emerges as an additional variable, while the rest maintain their directional associations. Specifically, ZeroCarRate, PopDens, and PONetDens are negatively associated, whereas LowWageRate, EmpHHEnt, and IntDens display positive associations.

The multi-modal oriented network density (MMNetDens) consistently displays a significant positive association throughout Texas. This suggests that areas with denser multi-modal networks experience higher PDO crash rates. This could be due to increased interactions between different modes of transportation, such as cars, bicycles, railways, and trucks, leading to more complex traffic scenarios. Possible recommendations could be enhancing infrastructure to accommodate different modes of transportation more safely, or perhaps simplify the traffic flow.

While other variables maintain consistent directional associations, the intensity of these associations varies spatially. Compared to FI level crash rates, the positive association of LowWageRate is notably less pronounced in the northern half of Texas. Besides, the overall value of coefficient of LowWageRate is higher than FI level crash rates, indicating a stronger negative association than FI level crash rates. Furthermore, PONetDens is significant across all CBGs, in contrast to the mere 15.78 % of CBGs showing significance for FI level crash rates. Notably, all CBGs demonstrate negative associations with PDO crash rates.

6. Conclusion

This study provides a framework to spatially examine the relationships between crash frequency and demographic and socio-economic characteristics, as well as roadway infrastructure and traffic characteristics at the CBG level. Our findings indicate that MGWR outperforms both OLS and GWR in uncovering the spatial relationships between contributing factors and both fatal and injury (FI) crashes as well as property damage only (PDO) crashes. This study also identified key contributing factors associated with both levels of crash rates.

Specifically, the ratio of zero-car households, the gross population density, and the pedestrian-oriented network density are negatively associated, whereas the ratio of low-wage workers, the employment and household entropy, and the street intersection density show positive associations. When comparing PDO to FI level crashes, multi-modal oriented network density emerges as an additional variable, demonstrating a positive association with PDO crash rates. Although these variables maintain similar directional associations across different crash levels, the distribution of the coefficients varies, indicating that different areas may require a distinct focus. Other variables, such as the proportion of the population aged 18 to 64 years, the auto-oriented network density, and the percentage of households with two or more cars, demonstrate significant associations in only a limited number of CBGs.

This study contribute to understand on the connection between social disparities and the frequency of crashes at the marco level. It reveals that areas with a higher concentration of low-wage workers and higher levels of employment and household entropy consistently experience elevated crash rates across all CBGs. The findings also advocate for the improvement of pedestrian-oriented infrastructure, such as sidewalks and crosswalks, as a measure to potentially decrease crash rates. However, concentrating solely on enhancing pedestrian infrastructure may not suffice to mitigate crashes universally. Augmenting the design of intersections, particularly in densely populated regions, to optimize traffic flow and diminish potential conflict points for multimodel traffic, is also integral to crash reduction.

The current study is not without limitation. This study primarily focused on analyzing crash data at the CBG level, which might not capture more detailed regional patterns. It's important to note that the characteristics of CBGs in the studied area may not directly apply to other regions or countries with different demographic or socio-economic compositions. Furthermore, the study has examined 15 key variables to assess their impact on crash rates, but it's essential to recognize that there could be other variables not considered in this analysis that might also play a significant role in influencing crash rates. Additionally, the study did not explicitly factor in temporal elements, such as seasonal fluctuations or annual variations, which can have a substantial influence on crash rates.

Author contribution statement

The authors confirm their contribution to the paper as follows: study conception and design: Jinli Liu, Subasish Das, Md Nasim Khan; data collection: Jinli Liu, Subasish Das; analysis and interpretation of results: Jinli Liu, Md Nasim Khan, Subasish Das; draft manuscript preparation: Jinli Liu. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

References

Ahmad, N., Gayah, V.V., Donnell, E.T., 2023. Copula-based bivariate count data regression models for simultaneous estimation of crash counts based on severity and number of vehicles. Accid. Anal. Prev. 181 https://doi.org/10.1016/j.aap.2022.106928.

Amoh-Gyimah, R., Saberi, M., Sarvi, M., 2016. Macroscopic modeling of pedestrian and bicycle crashes: a cross-comparison of estimation methods. Accid. Anal. Prev. 93, 147–159. https://doi.org/10.1016/j.aap.2016.05.001.

Anselin, L., 1995. Local indicators of spatial association—LISA. Geogr. Anal. 27 (2), 93–115.

Apardian, R.E. and Smirnov, O. (2020) 'An analysis of pedestrian crashes using a spatial count data model', Regional Science [Preprint]. Available at: https://rsaiconnect.onlinelibrary.wiley.com/doi/full/10.1111/pirs.12523?casa_token=usJfUy5F9HIAAAAA%3A7iI2VAxNkqHuMo6FUq41u-uzXDkTrk76s7jwoUvV6yaJ4w890YvywfnxME8q9bnmZ6Nb3IAdhiKlTArD (Accessed: 2 February 2023).

Bauer, K.M., Harwood, D.W., 2013. Safety effects of horizontal curve and grade combinations on rural two-lane highways. Transp. Res. Rec. 2398 (1), 37–49. https://doi.org/10.3141/2398-05.

Cai, Q., et al., 2017. Comparative analysis of zonal systems for macro-level crash modeling. J. Saf. Res. 61, 157–166. https://doi.org/10.1016/j.jsr.2017.02.018.

Chimba, D., Musinguzi, A., Kidando, E., 2018. Associating pedestrian crashes with demographic and socioeconomic factors. Case Studies on Transport Policy 6 (1), 11–16. https://doi.org/10.1016/j.cstp.2018.01.006.

González, S.R., Loukaitou-Sideris, A., Chapple, K., 2019. Transit neighborhoods, commercial gentrification, and traffic crashes: Exploring the linkages in Los Angeles and the Bay Area. J. Transp. Geogr. 77, 79–89. https://doi.org/10.1016/j.jtrangeo.2019.04.010.

Guerra, E., Dong, X. and Kondo, M. (2019) 'Do Denser Neighborhoods Have Safer Streets? Population Density and Traffic Safety in the Philadelphia Region', Journal of Planning Education and Research, p. 0739456X19845043. Available at: https://doi.org/10.1177/0739456X19845043.

Hall, T., Tarko, A.P., 2019. Adequacy of negative binomial models for managing safety on rural local roads. Accid. Anal. Prev. 128, 148–158. https://doi.org/10.1016/j.aap.2019.03.001.

Huang, Y., Wang, X., Patton, D., 2018. Examining spatial relationships between crashes and the built environment: a geographically weighted regression approach. J. Transp. Geogr. 69, 221–233. https://doi.org/10.1016/j.jtrangeo.2018.04.027.

Kim, K., Pant, P., Yamashita, E., 2010. Accidents and accessibility: measuring influences of demographic and land use variables in Honolulu, Hawaii. Transp. Res. Rec. 2147 (1), 9–17. https://doi.org/10.3141/2147-02.

Lee, J., et al., 2019. Transportation safety planning approach for pedestrians: an integrated framework of modeling walking duration and pedestrian fatalities. Transp. Res. Rec. 2673 (4), 898–906. https://doi.org/10.1177/0361198119837962.

Li, X., et al., 2022. 'Do underserved and socially vulnerable communities observe more crashes? A spatial examination of social vulnerability and crash risks in Texas'. Accid. Anal. Prev. 173 https://doi.org/10.1016/j.aap.2022.106721.

Mahmoudi, J., et al., 2023. Modeling the frequency of pedestrian and bicyclist crashes at intersections: big data-driven evidence from maryland. Transp. Res. Rec. 2677 (3), 1245–1260. https://doi.org/10.1177/03611981221122776.

Merlin, L.A., et al., 2020. Residential accessibility's relationships with crash rates per capita. J. Transp. Land Use 13 (1), 113–128.

Noland, R.B., Laham, M.L., 2018. Are low income and minority households more likely to die from traffic-related crashes? Accid. Anal. Prev. 120, 233–238. https://doi.org/10.1016/j.aap.2018.07.033.

O'brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quant. 41, 673–690.

Osama, A., Sayed, T., 2017. Macro-spatial approach for evaluating the impact of socio-economics, land use, built environment, and road facility on pedestrian safety. Can. J. Civ. Eng. 44 (12), 1036–1044. https://doi.org/10.1139/cjce-2017-0145.

Oshan, T.M., et al., 2019. mgwr: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale. ISPRS Int. J. Geo Inf. 8 (6), 269. https://doi.org/10.3390/ijgi8060269.

Pulugurtha, S.S., Duddu, V.R., Kotagiri, Y., 2013a. Traffic analysis zone level crash estimation models based on land use characteristics. Accid. Anal. Prev. 50, 678–687. https://doi.org/10.1016/j.aap.2012.06.016.

Pulugurtha, S.S., Duddu, V.R., Kotagiri, Y., 2013b. Traffic analysis zone level crash estimation models based on land use characteristics. Accid. Anal. Prev. 50, 678–687. https://doi.org/10.1016/j.aap.2012.06.016.

Qu, X., et al., 2021. Exploring the influences of point-of-interest on traffic crashes during weekdays and weekends via multi-scale geographically weighted regression. ISPRS Int. J. Geo Inf. 10 (11), 791.

Shabrina, Z., Buyuklieva, B., Ng, M.K.M., 2021. Short-term rental platform in the urban tourism context: a geographically weighted regression (GWR) and a multiscale GWR (MGWR) approaches. Geogr. Anal. 53 (4), 686–707. https://doi.org/10.1111/gean.12259.

Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accid. Anal. Prev. 45, 382–391. https://doi.org/10.1016/j.aap.2011.08.003.

Texas Department of Transportation (no date) Crash Record Information System. Available at: https://cris.dot.state.tx.us/public/Purchase/app/home (Accessed: 24 September 2023).

United States Census Bureau, 1994. The manual geographic areas reference. https://www2.census.gov/geo/pdfs/reference/GARM/GARMcont.pdf.

United States Census Bureau (2012) 'Geographic Terms and Concepts: Block Groups'.

Wang, X., et al., 2016. Macro-level safety analysis of pedestrian crashes in Shanghai China. Accid. Anal. Prev. 96, 12–21. https://doi.org/10.1016/j.aap.2016.07.028.

Xu, C., et al., 2020. Modeling the spatial effects of land-use patterns on traffic safety using geographically weighted poisson regression. Netw. Spat. Econ. 20 (4), 1015–1028. https://doi.org/10.1007/s11067-020-09509-2.

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.