Received: 14 July 2023 | Revised: 18 January 2024 | Accepted: 24 January 2024

DOI: 10.1111/tgis.13138

Transactions in GIS WILEY

RESEARCH ARTICLE

Analyzing urban crash incidents: An advanced endogenous approach using spatiotemporal weights matrix

Reza Mohammadi[1] | Mohammad Taleai[1,2] | Philipp Otto[3] | Monika Sester[3]

[1]Spatial Decision Making & Smart Cities Lab, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran

[2]School of Built Environment, Faculty of the Arts, Design & Architecture, University of New South Wales (UNSW), Sydney, New South Wales, Australia

[3]Institute of Cartography and Geoinformatics, Leibniz University Hannover, Hannover, Germany

Correspondence
Mohammad Taleai, Spatial Decision Making & Smart Cities Lab, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, No. 1346, Vali-asr Avenue, Mirdamad Cross, P.O. Box: 15875-4416, Tehran 1996715433, Iran.
Email: taleai@kntu.ac.ir

Abstract

Contemporary spatial statistics studies often underestimate the complexity of road networks, thereby inhibiting the strategic development of effective interventions for car accidents. In response to this limitation, the primary objective of this study is to enhance the spatiotemporal analysis of urban crash data. We introduce an innovative spatial-temporal weight matrix (STWM) for this purpose. The STWM integrates external covariates, including road network topological measurements and economic variables, offering a more comprehensive view of the spatiotemporal dependence of road accidents. To evaluate the functionality of the presented STWM, random effect eigenvector spatial filtering analysis is employed on Boston's traffic accident data from January to March 2016. The STWM improves analysis, surpassing distance-based SWM with a lower residual standard error of 0.209 and a higher adjusted R^2 of 0.417. Furthermore, the study emphasizes the influence of road length on crash incidents, spatially and temporally, with random standard errors of 0.002 for spatial effects and 0.026 for non-spatial effects. This is particularly evident in the north and center of the study area

during specific periods. This information can help decision-makers develop more effective urban development models and reduce future crash risks.

1 | INTRODUCTION

Traffic accidents are a critical global issue, causing approximately 1.3 million deaths and 20 to 50 million injuries annually. Despite various safety measures, the complexity of urban traffic networks and the dynamic nature of accidents pose a significant challenge in reducing accident rates. Recognizing the severity of this issue, the United Nations General Assembly aims to reduce the global death and injury rate from traffic accidents by 2030 (WHO, 2016). In alignment with this global endeavor, scholarly research has increasingly concentrated on the spatial aspect of urban car accidents. Several studies (Guo et al., 2010; Sun et al., 2016; Wang et al., 2016; Wang, Yuan, et al., 2019; Xie et al., 2013) have explored spatial dependence in urban car accidents. They state that spatial statistics could be employed to identify the most effective approach to reducing the crash rate. Current spatial statistical methods enable the identification of geographical patterns and structures in crash data, providing insights into traffic accident dynamics. These methods depend on constructing a spatial weight matrix (SWM) that captures the intensity of spatial relationships among observations in a vicinity. The challenge lies in creating an SWM that accurately reflects the real-world structure of traffic phenomena.

While various SWMs have been proposed, many studies (Getis & Aldstadt, 2004; Mawarni & Machdi, 2016) rely solely on exogenous geographical factors, such as proximity or distance between samples. However, this approach neglects the influence of time-varying economic variables that are often endogenous to the spatial system. To bridge this gap, recent studies suggest combining geographical factors with regional economic or social variables to define time variables and endogenous matrices (Qu et al., 2017, 2021). This approach acknowledges that the degree of spatial dependence may also depend on these dynamic economic variables, which are intertwined with the outcome variables. Zhang et al. (2021) demonstrate that integrating endogenous economic variables in SWM definitions enriches the depiction of diverse influence modes, encompassing geographic, socioeconomic, and cultural effects. Furthermore, Kelejian and Piras (2014) argue that exclusively relying on exogenous SWMs for estimating economic systems is insufficient, as it leads to oversimplification by presuming the weighting matrix to be exogenous.

Studies by Olubusoye and Salisu (2016), Pljakić et al. (2019), Wang, Chen, et al. (2019), and Sandoval-Pineda et al. (2022) have utilized endogenous variables to generate SWM and analyze crash data at a macro-scale traffic analysis zone (TAZ). These SWMs are uniformly weighted for all neighboring TAZ units. This uniform weighting approach might be sensible for regular lattices but is not suitable for mapping difficulties in complex metropolitan road networks, according to Raftery and Banfield (1991) and Corpas-Burgos and Martinez-Beneito (2020). Urban road networks are comprised of various geographical and temporal components and interactions (Shang et al., 2020). To address this complexity, Zhang and Wang (2017) and Corpas-Burgos and Martinez-Beneito (2020) proposed SWMs based on endogenous economic variables to model the unequal spillover effect of spatial dependence. These SWMs are typically designed for regionally scattered data and are unsuitable for evaluating phenomena distributed across a road network. The utilization of the distance function to establish these endogenous SWMs is unsuitable for determining the attributes of neighboring entities in the road network, primarily due to significant disparities in road lengths. Additionally, these SWMs overlook the road network structure's impact on the phenomenon distributed on the road network. The necessity to integrate endogenous variables into the SWM is clear, as this is crucial for accurately estimating spatial dependence in traffic accident data. Our study aims to enhance the SWM by integrating these endogenous factors, thereby providing a more comprehensive model for spatiotemporal analysis of traffic accidents.

The primary objective of this research is to advance spatiotemporal analysis and improve the understanding of urban crash incidents at the micro-level, with a focused case study on the Roxbury neighborhood in Boston. This study focuses on the development and application of a spatial-temporal weight matrix (STWM). This STWM, distinct from traditional distance-based Spatial Weight Matrices (SWMs), integrates endogenous socioeconomic factors and road network topology to provide a nuanced spatiotemporal analysis of crash data. It is designed to reveal the most hazardous streets by capturing the unequal spillover effect in both spatial and temporal dimensions, thereby refining the autocorrelation of crash data. This enhanced modeling approach aims to yield a more precise depiction of the spatiotemporal structure of crash incidents. Employing ESRI's best practices for spatial relationship conceptualization, this study utilizes auxiliary crash data to identify high-risk streets. These streets are then used to weigh the impact of neighboring streets on the target street, depending on the defined topological properties of the road network. Furthermore, economic variables such as road type, land use, speed limit, and weather data are employed to determine an endogenous STWM for modeling the time variation of crash data.

Moreover, as a secondary objective, this research marks the first effort to employ three advanced models: Eigenvector Spatial Filtering (ESF), Random Effect Eigenvector Spatial Filtering (RE-ESF), and spatially and non-spatially varying coefficients (SNVC) in the analysis of crash data. The ESF and RE-ESF models demonstrate superior capability in capturing the spatiotemporal patterns of crash data compared to traditional measures such as Moran's I. Since crash data do not necessarily follow a normal distribution, these models are appropriate for analysis. Additionally, the SNVC is utilized to estimate coefficients that vary both spatially and non-spatially, thereby accounting for the residual spatial dependence structure not adequately captured by the RE-ESF method. The knowledge uncovered by this proposed spatiotemporal analysis can help traffic police agencies conduct comprehensive analyses and research on the traffic situation in specific locations. Moreover, the analysis provides traffic police agencies with a better understanding of geographical and temporal variations in collision hotspots.

State-of-the-art studies are reviewed in the next section. The theory of the proposed spatiotemporal model is described in Section 4. Section 5 explains the presented STWM, followed by a description of Boston's crash dataset in Section 6. Section 7 presents the results of employing STWM for the spatiotemporal analysis of urban crash data. Moreover, to assess the performance of the proposed model, ESF, RE-ESF, and SNVC are used to evaluate the proposed STWM against the traditional distance-based SWM. Finally, the article's conclusion is discussed in Section 7.

2 | LITERATURE REVIEW

In this section, we will provide an overview of the current state of research on various types of endogenous and exogenous SWMs. Additionally, we will conduct a review of the crucial features of car accident analysis.

2.1 | Literature on SWM in car accident analysis

Moran (1950) was one of the first researchers to introduce the term “spatial correlation” (Getis, 2008). This foundation was later expanded by Whittle (1954) in a geographical framework. The concept of “spatial autocorrelation” found its academic roots in the conference presentation of Cliff and Ord (1968). The seminal work of Tobler (1970) was a pivotal moment, underlining the importance of exploring spatial dependence in neighboring regions or entities. This concept is a fundamental principle in spatial analysis and has influenced subsequent research, including the development of Spatial Weight Matrices (SWMs). The 1990s marked significant growth in spatial analysis, notably with the introduction of Spatial Lag Models (SLMs) by Anselin (1995). SLMs evaluate spatial relationships between observations and neighboring entities by defining an SWM. While Anselin's contribution was instrumental in understanding spatial dependencies, it also opened avenues for further methodologies, particularly in economic and environmental contexts. Malczewski (2000) provided an early review of studies on SWMs, emphasizing the growing importance of SWMs. Cohn and Jackman (2011) utilized an SWM and the local Moran's *I* statistic to investigate the impact of the Modifiable Areal Unit Problem (MAUP) on income segregation. Furthermore, Guo and Wang (2011) used SWMs and Empirical Bayes smoothing to identify spatial patterns of cancer prevalence across various regions. Moreover, the scope of SWMs was further broadened by Halleck Vega and Elhorst (2015). They extended the application of SWMs to various economic phenomena, thus significantly enriching the scope and applicability of spatial econometric models.

In addition to more contemporary research, we focus on recent developments in SWM in the context of car accident analysis. A comprehensive literature review conducted by Ziakopoulos and Yannis (2020), prior to 2019, examined the application of different areal unit levels, such as street and zonal levels, in spatial road safety studies. However, the review did not address the incorporation of endogenous parameters in defining SWM and spatial–temporal analysis. Table 1 provides an overview of the current literature on SWMs in car accident analysis. The exploration of SWM in crash data analysis has evolved through various stages. Exogenous SWMs, an early approach, are typically determined based on the distance or contiguity of samples. Distance-based conceptualization considers neighborhood features impacting a target feature based on inverse distance. Halleck Vega and Elhorst (2015) added a theoretical dimension by raising concerns about the lack of a specific rule governing the decrease of spatial dependence with increasing distance, although they acknowledged that distance-based effects are intrinsically linked to spatial interactions. Harizi et al. (2016) advanced the field by concentrating on distance-based conceptualization, which is suitable for point-type features. However, they acknowledged that this approach is unsuitable for more complex geometries, such as polygons or line features. For region-based spatial analysis, a contiguity (or adjacency)-based SWM is preferred, indicating whether polygons are adjacent to one another through an edge or corner, with zero denoting no relationship, as outlined by Alarifi et al. (2018) and Abokifa and Sela (2019). Following this concept, Alkahtani et al. (2019) and Wang, Yuan, et al. (2019) employed rook and queen contiguity SWMs at the TAZ level to assess the spatial dependency of car accidents between TAZs. These studies also utilized a Bayesian spatial Poisson-lognormal model and linear regression to investigate risk factors associated with urban traffic accidents. Recent studies have expanded this concept to street-level crash data analysis. Wen et al. (2019) incorporated a first-order adjacency concept to construct a contiguity-based SWM and employed Conditional Autoregressive (CAR) models to examine spatial autocorrelation and spillover effects. Similarly, Almasi and Behnood (2022) reevaluated the relevance of distance-based SWMs, particularly for point-type features. More recently, Gilardi et al. (2023) developed a contiguity-based SWM and utilized a Bayesian Hierarchical Model for crash analysis. Xiong et al. (2023) developed an SWM based on the distances between census areas and applied a spatial Durbin model to extract the spatial relationship between traffic accidents and low-income and minority communities. Wu et al. (2024) utilized a decay function to define an exogenous SWM within a macro-level Middle-Super-Output-Area (MSOA), which is delineated as a spatial unit averaging 8000 inhabitants and used a spatial Random Forest to predict crash incidents.

Several researchers have utilized exogenous time-varying variables to define SWMs. Starting in 2016, Huang et al. (2016) laid the groundwork at the TAZ level by using time-varying first-order adjacent TAZs to define a contiguity-based SWM. They employed a Bayesian spatial joint model for zonal crash data analysis. In the subsequent year, Liu et al. (2017) focused on the street level and generated an SWM using a distance function that accounts for both spatial and temporal dimensions of crashes. They used Geographically Weighted Negative Binomial Regression in their analysis. Concurrently, Soltani and Askari (2017) utilized an inverse distance function at the TAZ level to construct an SWM and then employed global Moran's *I* and local Moran's Getis-Ord Gi* to identify crash hotspots across spatial and temporal dimensions. Meanwhile, Ma et al. (2017) developed an exogenous time-variant SWM based on the first-order contiguity concept at the street level. They then applied a Bayesian multivariate space–time model for space–time modeling of crash data. Building on this, Blazquez et al. (2018) used a similar distance function approach as Liu et al. (2017) but employed Moran's *I* and the Getis-Ord Gi* index to identify crash hotspots. Advancing the concept further, Song et al. (2020) utilized an inverse distance squared function to create an exogenous time-varying SWM

TABLE 1 A review of SWMs in spatial analysis of car accident data.

Author(s) and publication year	SWM type and category	Scale	Spatial analysis model
Exogenous time-invariant SWM			
Harizi et al. (2016)	Distance between crashes	Street level	Global Moran's *I* and local Moran's Getis-Ord Gi*
Alarifi et al. (2018)	Contiguity and distance-based SWM	Street level	Hierarchical Poisson-lognormal joint model with spatial effects
Alkahtani et al. (2019)	Contiguity SWM	TAZ level	Moran's *I* and a Bayesian spatial Poisson-lognormal model
Wang, Yuan, et al. (2019)	Rook and queen contiguity of TAZs	TAZ level	Linear regression model
Wen et al. (2019)	Contiguity SWM	Street level	Conditional autoregressive (CAR)
Almasi and Behnood (2022)	Distance between TAZs	TAZ level	Geographic weighted Poisson regression
Gilardi et al. (2023)	Contiguity SWM	Street level	Bayesian hierarchical model
Xiong et al. (2023)	Distance between censuses	Census level	Spatial Durbin model
Wu et al. (2024)	Distance decay function between MSOAs	Middle-Super-Output-Area (MSOA)	Spatial random forest
Exogenous time-variant SWM			
Huang et al. (2016)	Contiguity SWM	TAZ level	Bayesian spatial joint model
Liu et al. (2017)	Inverse distance function in spatial and temporal dimensions of crashes	Street level	Geographically weighted negative binomial regression
Soltani and Askari (2017)	Inverse distance function in spatial and temporal dimensions of TAZs	TAZ level	Global Moran's *I* and local Moran's Getis-Ord Gi*
Ma et al. (2017)	Contiguity SWM	Street level	Bayesian multivariate space–time model
Blazquez et al. (2018)	Inverse distance function in spatial and temporal dimensions of crashes	Street level	Global Moran's *I* and local Moran's Getis-Ord Gi*
Song et al. (2020)	Inverse distance function of defined space–time cubes	Street level	A hierarchical Bayesian random-effects model
Feizizadeh et al. (2022)	Distance between crashes	Street level	Kernel density estimation method
Gilardi et al. (2022)	Contiguity SWM	Street level	Bayesian multivariate space–time model
Exogenous SWM and endogenous variables regression model			
Wang, Chen, et al. (2019)	Hybrid of continuity and Euclidean distance of exogenous variables	TAZ level	Used endogenous variable in spatial regression models

TABLE 1 (Continued)

Author(s) and publication year	SWM type and category	Scale	Spatial analysis model
Pljakić et al. (2019)	Euclidean distance of exogenous variables	TAZ level	Used endogenous variable in spatial regression models
Alves et al. (2021)	Euclidean distance of exogenous variables	Street level	Used endogenous variable in difference-in-differences approach
Sandoval-Pineda et al. (2022)	Euclidean distance of exogenous variables	Territorial units	Used endogenous variables in a support vector regression model
Endogenous variables in SWM and spatial–temporal model of crash analysis			
Olubusoye and Salisu (2016)	Euclidean distance of endogenous variables	Local government area	Used endogenous variables in spatial autoregressive (SAR) model
Endogenous variables in time-variant SWM and spatial–temporal model in non-crash applications			
Qu et al. (2017)	Hybrid of contiguity and distance	Not available	Used quasi-maximum-likelihood (QML) to model the spatial dynamic of panel data
Merk and Otto (2020)	Inverse distance	Location of monitoring sites	Used spatial autoregressive to model the effect of daily wind direction and speed on $PM_{2.5}$ data
Billé et al. (2020)	Negative exponential distance function	Country level	Used spatial autoregressive models to model house prices
Zhou et al. (2022)	A combination of a gravity model incorporating endogenous socioeconomic variables and Euclidean geographical distance for exogenous variables	TAZ level	Used spatial autoregressive binary model to investigate the travel flow differences between morning and evening peaks

based on a space–time cube, with each cube representing a specific distance and time interval within the study area. They used a hierarchical Bayesian random-effects model to investigate the factors contributing to crashes. More recently, Feizizadeh et al. (2022) employed the distances between crash records to define an SWM and applied a Kernel Density Estimation method to identify the risk of traffic accident hotspots. Finally, Gilardi et al. (2022) returned to the first-order contiguity concept at the street level, incorporating a Bayesian multivariate space–time model similar to that used by Ma et al. (2017).

Since economic systems are dynamic, exogenous SWMs may produce inaccurate results in static spatial models. Although geographical locations are time-invariant, the strength of spatial dependence may vary depending on time-varying economic variables. When geographical factors are combined with regional economic or social factors, the relationship becomes time variable and, more importantly, endogenous (Qu et al., 2017, 2021). According to Zhang et al. (2021), incorporating endogenous economic variables into the SWM definition helps to describe different modes of influence, including geographical, economic, and cultural effects, as emphasized. Instead of using a low-order weight matrix (exogenous variables), the endogenous SWM approach provides a more comprehensive understanding of complex relationships among variables. Therefore, using an exogenous SWM to estimate economic systems is likely to be invalid.

Several researchers have explored the use of exogenous SWMs while incorporating the influence of endogenous variables into their model estimates. Wang, Chen, et al. (2019) developed an exogenous SWM that considers both contiguity and Euclidean distance. Utilizing this SWM, they employed various spatial regression models, such as Ordinary Least Squares (OLS), Spatial Lag Model (SLM), Spatial Error Model (SEM), and Spatial Durbin Model (SDM), to analyze spatial correlations among TAZs in Tianjin, China. Similarly, Pljakić et al. (2019) designed an exogenous SWM using contiguity and Euclidean distance and integrated endogenous variables into spatial regression models, aiming to examine the relationship between several variables and traffic crashes at the macro-level TAZ. Alves et al. (2021) utilized the Euclidean distances of exogenous variables to construct an SWM at the street level. Armed with endogenous variables such as population, weather data, and temperature, they applied a difference-in-differences approach to explore the impact of highway concessions on road crashes. Sandoval-Pineda et al. (2022) used the mean distance within the TAZ area to analyze the impact of socioeconomic, land use, and mobility factors on traffic accidents using vector support regression models.

To the best of the author's knowledge, there is limited research on the development of an SWM based on endogenous variables, particularly in the context of utilizing these endogenous variables in regression models to analyze urban crashes. Olubusoye and Salisu (2016) constructed an SWM incorporating endogenous variables such as travel density, land area, road length, and population. They subsequently applied these endogenous variables within a spatial autoregressive (SAR) model to identify hotspots and explore the influence of crashes in nearby local government areas on the crash frequency in other areas. The formulated SWM model uniformly distributed weights to adjacent road segments. However, this approach overlooks the potential contribution of hazardous streets in adjacent areas. These hazardous streets may play a significant role in the occurrence of crashes on the target street. This observation aligns with the findings of Raftery and Banfield (1991) and Corpas-Burgos and Martinez-Beneito (2020), which argued that while equal weighting may be suitable for regular lattices, it proves inadequate for complex systems such as urban road networks.

The application of SWMs with varied weights for car accident analysis remains underexplored, despite some research having been conducted in other domains. Addressing this gap, Zhang and Wang (2017) developed an SWM that accounts for both geographical and economic distances, aiming to capture the unequal spillover effects of spatial dependence among housing market units in China. While this method is designed for regionally dispersed data, it is not suitable for analyzing road network phenomena. Building on this concept, Corpas-Burgos and Martinez-Beneito (2020) proposed an SWM tailored for the spatial analysis of disease mapping that assigns random weights to adjacent areas. Although this approach was innovative in the health sector, it was not designed to address the complexities of urban road networks. In addition to these issues, Shang et al. (2020) explicitly noted that the urban road network encompasses numerous spatial and temporal components and interactions, thereby posing several challenges to the development of SWMs for such complex systems.

Furthermore, significant research has been conducted on utilizing endogenous parameters in time-varying SWMs and spatial–temporal analysis models, extending beyond accident applications. Exploring these studies provides a more comprehensive understanding of how exogenous variables are incorporated into generating SWMs and spatial–temporal models. One notable example is the study by Qu et al. (2017), which employed a time-invariant SWM utilizing endogenous variables of contiguity and distance. In a Monte Carlo experiment, they applied the quasi maximum-likelihood (QML) method to analyze panel data spatial dynamics. Additionally, Merk and Otto (2020) developed a time-varying SWM by incorporating endogenous variables such as inverse distance, wind direction, and bearing. Their study utilized spatial autoregressive techniques to model the daily effects of wind direction and speed on particulate matter ($PM_{2.5}$) in panel data. In another study, Billé et al. (2020) constructed a dynamic SWM model using negative exponential functions of endogenous variables. This model was specifically designed to analyze the spatial and temporal structures of house prices in regional areas of the United Kingdom. Recently, Zhou et al. (2022) employed a spatial autoregressive binary model with an endogenous SWM to examine the variances in travel flow between morning and evening peak times on weekdays and weekends. They developed an SWM that quantifies the relative weights among all TAZ pairs based on endogenous variables. This SWM was constructed by integrating a gravity model—which accounts for endogenous socioeconomic factors—with exogenous factors measured by Euclidean geographical distance.

2.2 | Literature on the important features in car accident analysis

Table 2 summarizes four types of characteristics that significantly contribute to urban car accidents at both the street level (micro-scale) and the TAZ level (macro-scale). Traffic characteristics play a significant role in car accidents.

TABLE 2 Contributing factors to car accidents in urban areas.

Feature category	Features, author(s) & publication year	Scale
Traffic characteristics	• Average Annual Daily Traffic (AADT) (Alarifi et al., 2018; Huang et al., 2016; Liu et al., 2017; Mahmud et al., 2019; Xiong et al., 2023) • Vehicle Miles Traveled (VMT) (Xu et al., 2019) • Running red lights (Retting et al., 1999)	• Street level (Alarifi et al., 2018) • Zonal level (Huang et al., 2016; Liu et al., 2017; Mahmud et al., 2019; Retting et al., 1999; Xiong et al., 2023; Xu et al., 2019)
Road characteristics	• One-way streets, bus and bike lanes, road quality (WHO, 2018) • Speed limit (Almasi & Behnood, 2022; Huang et al., 2016; Liu et al., 2017; Ma et al., 2017; Mahmud et al., 2019; Rahman et al., 2023) • Road length [m] (Alarifi et al., 2018; Huang et al., 2016; Liu et al., 2017; Xie & Yan, 2008) • Proximity to intersections (Li et al., 2019) • Number of intersections (Hasan et al., 2022; Shariat-Mohaymany et al., 2015) • Road type (Hasan et al., 2022; Huang et al., 2016; Li et al., 2019; Wu et al., 2024) • Number of road lanes (Alarifi et al., 2018; Huang et al., 2016; Ma et al., 2017) • Presence of a median on roads (Alarifi et al., 2018; Huang et al., 2016) • Vertical grade, curvature of roads (Wen et al., 2019) • Pavement condition (Huang et al., 2016; Ma et al., 2017)	• Street level (Alarifi et al., 2018; Huang et al., 2016; Li et al., 2019; Liu et al., 2017; Ma et al., 2017; Wen et al., 2019) • Zonal level (Almasi & Behnood, 2022; Hasan et al., 2022; Mahmud et al., 2019; Rahman et al., 2023; Shariat-Mohaymany et al., 2015; Wu et al., 2024; Xie & Yan, 2008)
Socioeconomic characteristics	• Population (Almasi & Behnood, 2022; Alves et al., 2021; Feizizadeh et al., 2022; Huang et al., 2016; Mahmud et al., 2019; Wu et al., 2024; Zhou et al., 2022) • Median household income (Huang et al., 2019) • Land use (Almasi & Behnood, 2022; Feizizadeh et al., 2022; Sandoval-Pineda et al., 2022; Wang, Yuan, et al., 2019) • POI (Almasi & Behnood, 2022; Sandoval-Pineda et al., 2022; Wang, Yuan, et al., 2019; Zhou et al., 2022)	• Street level (Alves et al., 2021; Feizizadeh et al., 2022; Huang et al., 2016) • Zonal level (Almasi & Behnood, 2022; Huang et al., 2016, 2019; Mahmud et al., 2019; Sandoval-Pineda et al., 2022; Wang, Yuan, et al., 2019; Wu et al., 2024; Zhou et al., 2022)
Weather data	• Precipitation, snow depth, temperature, wind speed, visibility, cloud cover (Alves et al., 2021; Hasan et al., 2022)	• Street level (Alves et al., 2021) • Zonal level (Hasan et al., 2022)

According to Retting et al. (1999), over one million crashes occur annually at traffic signals in the United States. Alarifi et al. (2018), Huang et al. (2016), Liu et al. (2017), Mahmud et al. (2019), Xu et al. (2019) and Xiong et al. (2023) found that both Average Annual Daily Traffic (AADT) and Vehicle Miles Traveled (VMT) are statistically significant predictors of crashes. AADT represents the average number of vehicles that pass through a specific road segment in a year, while VMT measures the total miles driven by vehicles within a particular area or over a specific period. Furthermore, running red lights, by either drivers or pedestrians (Retting et al., 1999), can increase crash rates.

Furthermore, road characteristics have been identified as significant factors in the analysis of urban car accidents over the years. According to the World Health Organization's report (WHO, 2018), one-way streets, bus and bike lanes, and poor road quality (i.e., potholes) can confuse drivers who are unfamiliar with local street directions. Several studies have established relationships between crashes and various factors, such as speed limits (Almasi & Behnood, 2022; Huang et al., 2016; Liu et al., 2017; Ma et al., 2017; Mahmud et al., 2019; Rahman et al., 2023), road length (Alarifi et al., 2018; Huang et al., 2016; Liu et al., 2017; Xie & Yan, 2008), proximity to intersections (Li et al., 2019), the number of intersections (Hasan et al., 2022; Shariat-Mohaymany et al., 2015), and road type (Hasan et al., 2022; Huang et al., 2016; Li et al., 2019; Wu et al., 2024). Additionally, relationships have been reported between crashes and the number of road lanes (Alarifi et al., 2018; Huang et al., 2016; Ma et al., 2017), the presence of a median on roads (Alarifi et al., 2018; Huang et al., 2016), vertical grade, the curvature of segments (Wen et al., 2019), and pavement condition (Huang et al., 2016; Ma et al., 2017).

Socioeconomic factors are also critical to understanding the causes of car accidents. Prior research has revealed that population, employment rates, and the number of uneducated residents (Almasi & Behnood, 2022; Alves et al., 2021; Feizizadeh et al., 2022; Huang et al., 2016; Mahmud et al., 2019; Wu et al., 2024; Zhou et al., 2022), as well as median household income (Huang et al., 2019), correlate with pedestrian accidents. In addition, specific land use and points of interest, such as hospitals and schools, have been shown to affect traffic accidents in TAZs, according to studies conducted by Almasi and Behnood (2022), Feizizadeh et al. (2022), Sandoval-Pineda et al. (2022), Wang, Yuan, et al. (2019) and Zhou et al. (2022). Hasan et al. (2022) and Alves et al. (2021) have also suggested that weather conditions like rain, snow, frost, and fog can negatively impact accidents due to decreased visibility and difficulty controlling a vehicle.

Some studies mentioned in Table 2 focus on constructing models for macro-scale accident prediction, using TAZ as the unit of analysis. However, our study aims to conduct a micro-level safety analysis at the street level. Therefore, we will utilize the significant factors identified in both TAZ and street-level analyses and subsequently employ feature selection techniques to determine the most crucial factors for street-level microanalysis. This approach will allow us to gain a more comprehensive understanding of the characteristics affecting car accidents at the street level.

3 | THEORY CONCEPT OF SPATIOTEMPORAL ANALYSIS

Spatial autocorrelation refers to the relationship between attributes at a specific location and surrounding locations (Barua et al., 2015; De Knegt et al., 2010; Huang et al., 2019). Positive spatial autocorrelation occurs when adjacent observations have similar values, while negative spatial autocorrelation occurs when neighboring observations have different values. By analyzing spatial autocorrelation, we can identify consistent groups of objects based on their attributes (Mohamed et al., 2013).

Moran's *I* is a statistical measure used to evaluate whether a spatial pattern of STRs is clustered, dispersed, or random (Rodriguez Rangel & Sanchez Rivero, 2020). This is accomplished by defining two hypotheses. The null hypothesis (H0) represents no spatial structure of the values associated with the geographical features in the study area. The alternative hypothesis (H1) refers to the data being more spatially clustered than randomly distributed. Moran's *I* test (Moran, 1950) assesses whether the observed spatial autocorrelation significantly differs from the expected value under the null hypothesis. Moran's *I* test is given by

$$I = \frac{N}{S_0}\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\left(y_i - \overline{y}\right)\left(y_j - \overline{y}\right)}{\sum_{i=1}^{N}\left(y_i - \overline{y}\right)}, \tag{1}$$

where w_{ij} represents the specific element of the SWM corresponding to pair i and j, w_{ii}=0 for all i, $\overline{y}$ is the average value of the variable, N *is* the total number of observations, and S_0 is the number of all positive spatial weights. Since Moran's I is a weighted Pearson correlation, z-score and p-values are used to interpret the spatial autocorrelation result. The z-score is calculated based on the data standard deviations. If the p-value is sufficiently small, and the absolute value of the z-score exceeds the threshold necessary to lie outside the specified confidence interval, it is justifiable to reject the null hypothesis. The estimated z-score provides a basis for the statistical assumption about a standard normal distribution. The expected value used to separate positive and negative spatial autocorrelation is based on the null hypothesis of a normal distribution (Jacquez & Oden, 1994). However, count data, such as crash incidents, are rarely expected to follow a normal distribution (Chun, 2014). Therefore, the spatial filtering method is a better measurement to describe the count data spatiotemporal pattern.

3.1 | Eigenvector spatial filtering (ESF)

The objective of spatial filtering is to characterize spatial data by considering three primary components: mean trends, spatially structured random components, and random noise (Griffith & Chun, 2012). To achieve this, the analysis separates spatially structured random components from the overall trend and random noise. Eigenvector Spatial Filtering (ESF) is one of the most well-known spatial filtering variations that utilizes eigenvectors derived from a SWM (Griffith, 2003). Eigenvectors are obtained from eigenfunctions. These eigenfunctions serve as synthetic covariates and are used to compute non-zero spatial autocorrelation in regression residuals (Griffith, 2021). They are given by

$$\left(I - \frac{11^T}{n}\right)W\left(I - \frac{11^T}{n}\right), \tag{2}$$

where W represents a generic $n \times n$ SWM, where the main diagonal is comprised of zeros. Moreover, I denotes the $n \times n$ identity matrix, 1 is an $n \times 1$ vector that contains 1, and T represents the matrix transpose operator.

ESF utilizes eigenvectors derived from a doubly centered SWM to capture spatial autocorrelation in regression residuals (Griffith & Chun, 2012). ESF decomposes the explained variable y_i, measured at the ith sample site, into three components: a regression component $\sum_{k=1}^{k} x_{i,k}\beta_k$, a spatial process $f_{MC(F)}\left(S_i\right)$, which depends on location S_i, and the noise term ε_i. For all $i = 1, \ldots, n$, the equation is:

$$y_i = \sum_{k=1}^{k} x_{i,k}\beta_k + f_{MC(F)}\left(S_i\right) + \varepsilon_i, \text{ with } \varepsilon_i \sim N\left(0, \sigma^2\right). \tag{3}$$

The spatial process in ESF serves to eliminate remaining spatial dependence in the residuals and to accurately estimate the regression coefficients β_k. Specifically, the ESF calculates $f_{MC(F)}\left(S_i\right)$ using the Moran coefficient spatial process to efficiently reduce residual spatial dependence (Murakami, 2020). More precisely, $f_{MC(F)}\left(S_i\right)$ is defined as follows:

$$f_{MC(F)}\left(S_i\right) = \sum_{l=1}^{L} e_l\left(S_i\right)\gamma_{l(F)}, \tag{4}$$

where L represents the number of positive eigenvalues of SWM, $e_l\left(S_i\right)$ is the ith element of the lth eigenvector e_l, and $\gamma_{l(F)}$ is a fixed coefficient to model the spatial process. By considering spatial dependence via $f_{MC(F)}\left(S_i\right)$, statistical significance can be prevented from being overestimated.

Random Effect Eigenvector Spatial Filtering (RE-ESF) is a recent enhancement of ESF that accounts for spatial autocorrelation in both the response and predictor variables. According to Murakami (2017), RE-ESF results may be more reliable than those of ESF due to its superior performance in accurately estimating the standard errors of regression coefficients, especially when predictors exhibit spatial autocorrelation. RE-ESF utilizes a Moran coefficient-based spatial random process to eliminate any remaining spatial dependence by specifying $f_{MC(F)}(S_i)$.

3.2 | Theory of spatially and non-spatially varying coefficient model (SNVC)

Although RE-ESF accounts for spatial dependence in the dependent variable, residual spatial dependence may still exist in the model residuals. This suggests the presence of additional spatial patterns in the data that the RE-ESF method does not capture. To overcome this limitation, the spatially and non-spatially varying coefficient (SNVC) model (Murakami & Griffith, 2020) has been introduced. This model can estimate spatially varying coefficients that change according to the residual spatial dependence structure. For example, it may identify that the impact of road type on crashes varies depending on the surrounding land use or population density. The SNVC model can capture these complex relationships by estimating different coefficients for different locations in the study area.

Furthermore, RE-ESF only estimates the coefficients of filtered variables and does not directly estimate non-spatially varying coefficients that may vary based on independent variables. The SNVC model can address this by allowing the estimation of non-spatially varying coefficients that vary based on independent variables. In this study, the SNVC can help determine the effect of independent variables (such as road type, land use, population density, and weather data) on crash risk while accounting for the possibility that the effect may vary depending on other variables. For example, we can estimate the impact of road type on crash risk separately for different land use types. This would enable the capture of any variation in road type's effect on crash risk depending on the surrounding land use. In addition, it would estimate a single coefficient for road type across the entire study area.

The SNVC model estimates residual spatial dependence, constant coefficients, spatially varying coefficients, and non-spatially varying coefficients, which are defined as (Murakami & Griffith, 2020), i.e.,

$$\begin{gathered} y_i = \sum_{k=1}^{k} x_{i,k}\beta_{i,k} + f_{MC}(S_i) + \varepsilon_i, \\ \beta_{i,k} = b_k + f_{MC,k}(S_i) + f(x_{i,k}) \text{ with } \varepsilon_i \sim N(0, \sigma^2). \end{gathered} \tag{5}$$

where $\beta_{i,k}$ represents the kth regression coefficient at the ith site, b_k denotes constant the coefficients, $f_{MC,k}(S_i)$ represents spatially varying coefficients, and $f(x_{i,k})$ represents the non-spatially varying coefficient term. Combining both spatial and non-spatial aspects of the coefficients has prevented this method from suffering from any possible spurious correlations among spatially varying coefficients (Murakami, 2017).

3.3 | Spatial-temporal weight matrix (STWM)

Spatiotemporal panel data analysis relies on accurately defining the spatial relationships between features. One crucial aspect of spatiotemporal analysis is determining the most suitable SWM to accurately analyze the data (Mawarni & Machdi, 2016). There are two main methods for generating SWM: distance-band and contiguity-based. The distance-band SWM assigns weights to features based on their inverse proportionality to their distance from the target feature. Closer features have a greater impact on the target feature (Mawarni & Machdi, 2016). The

K-nearest neighbor weight matrix is a well-known distance-band SWM. In this approach, the *k* closest features to the target feature are considered its neighbors.

On the contrary, contiguity-based SWM comes in two types: queen and rook contiguity. Queen contiguity is the simplest form, which defines neighbor relationships based on shared edges. A shared edge is represented by 1, while no relationship is represented by 0 (Abokifa & Sela, 2019). Similarly, in the rook contiguity weight matrix, two features are considered neighbors if they share a corner point. These SWM, known as first-order contiguity, only consider direct neighbors when defining the spatial relationships between units. In contrast, higher order contiguity SWM incorporates additional levels of proximity, including secondary, tertiary, or more distant relationships.

Contiguity-based or distance-band SWMs are often customized in spatiotemporal datasets by incorporating a time window. The relationships between features in STWM are defined based on whether a feature falls within a predefined distance and time window.

4 | INTEGRATING TOPOLOGY AND ECONOMIC VARIABLES IN ENDOGENOUS STWM

We propose an endogenous STWM and employ spatiotemporal analysis to analyze urban road crashes to identify high-risk areas and vulnerable times. Figure 1 illustrates the multiple steps involved in the proposed

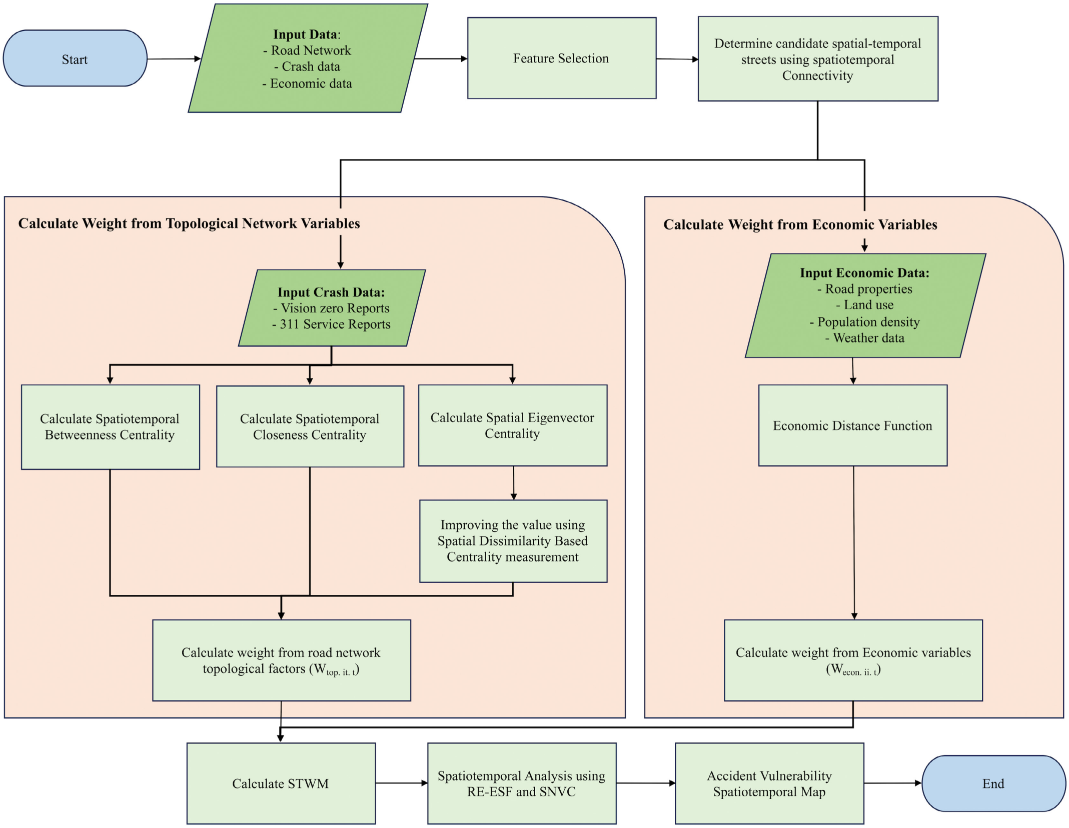

Start
Input Data:
- Road Network
- Crash data
- Economic data
Feature Selection
Determine candidate spatial-temporal streets using spatiotemporal Connectivity
Calculate Weight from Topological Network Variables
Input Crash Data:
- Vision zero Reports
- 311 Service Reports
Calculate Spatiotemporal Betweenness Centrality
Calculate Spatiotemporal Closeness Centrality
Calculate Spatial Eigenvector Centrality
Improving the value using Spatial Dissimilarity Based Centrality measurement
Calculate weight from road network topological factors ($W_{top.\ it.\ t}$)
Calculate Weight from Economic Variables
Input Economic Data:
- Road properties
- Land use
- Population density
- Weather data
Economic Distance Function
Calculate weight from Economic variables ($W_{econ.\ ii.\ t}$)
Calculate STWM
Spatiotemporal Analysis using RE-ESF and SNVC
Accident Vulnerability Spatiotemporal Map
End

FIGURE 1 Workflow for integrating topological and economic factors to generate STWM and perform spatiotemporal analysis of urban road crashes.

STWM method. First, we collect and store input data—comprising road network, crash, and economic data—in a spatiotemporal database. Next, feature selection is applied to identify the most crucial variables for spatiotemporal analysis. Subsequently, we define candidate spatial–temporal streets (*STRs*) based on the spatiotemporal connectivity relationships within the road network. We then estimate the weight of each STR, represented as $str_{i,t}$ ($str_{i,t} \in STRs$), by considering its topological factors (w_{top}) and economic characteristics (w_{econ}).

The topological weight (w_{top}) is determined through three key metrics: spatial eigenvector centrality, spatiotemporal closeness centrality, and betweenness centrality. For this purpose, we calculate spatial eigenvector centrality for the road network and improve the results by spatial dissimilarity-based centrality. The spatial dissimilarity-based centrality factor helps to determine the *STRs* with a higher degree that represents the most critical *STRs* in the road network. Afterward, we calculate the spatiotemporal betweenness and closeness centrality to identify the network's streets (hubs) that contribute to the transfer of accident effects within the road network. To achieve this, we identify crash-prone areas by analyzing Vision Zero data and non-emergency 311 reports. These datasets contain essential information, including date, time, and precise latitude and longitude coordinates of each crash incident and report. Finally, we calculate the weight w_{top} using the determined values of spatial eigenvector centrality, spatiotemporal betweenness centrality, and closeness centrality.

In the second step, economic variables are incorporated into the analysis to model their effect on crashes. Endogenous economic variables, including road type, road width, road length, speed limit, land use, population density, and weather data, are employed to determine the economic weight of *STRs*, denoted as w_{econ}. This is achieved through the application of an inverse distance function.

Finally, the STWM is generated by combining the weight of *STRs* based on its topological factors w_{topo} and its economic characteristics w_{econ}. Afterward, RE-ESF and SNVC techniques are applied to the STWM to generate an accident vulnerability spatiotemporal map. The results can help inform policymakers and transportation professionals in developing targeted interventions to improve road safety and reduce the number of crashes on urban roadways.

4.1 | The structure of STWM

This study utilizes the concept of high-order contiguity weight to develop the STWM that considers the spatial and temporal relationships between streets' neighbors. By considering high-order contiguity, the STWM captures spatial and temporal interactions that extend beyond immediate neighbors. This expanded perspective enables a more comprehensive analysis of spatial and temporal relationships, revealing patterns that may not be evident when only first-order neighbors are considered.

Considering n as the number of streets in the study area, the elements of the SWM, represented by an $n \times n$ matrix, indicate the influence of crashes that occurred on one street on the estimation of crashes on the target street. To model the temporal dimension of accident data in the STWM, we consider the streets on an hourly temporal scale (denoted by STRs). Note that the location of streets does not change over time. More precisely, $str_{i,t} \in STRs$ is a mapping of the individual street i to its spatial and temporal dimensions using function $f(i, t)$, where streets i: $\{1, 2, \ldots, n\} \rightarrow i \subseteq R+ = \{x | x > 0\}$ and time variable t: $\{1, 2, \ldots, T\} \rightarrow t \subseteq R+ = \{x | x > 0\}$. To consider the self-prediction effect of street i at time t at time $t+1$ ($w_{ii,tt+1} \neq 0$), we define the STWM matrix as a $2nT \times 2nT$ non-negative element. However, to prevent self-prediction of street i at the same time t, STWM is a zero-diagonal matrix ($w_{ii,tt} = 0$). Therefore, the element on the ith row and jth column of STWM is denoted as $w_{ij,tt}$, which signifies the magnitude of the link between street i and j at the same time t. On the contrary, $w_{ij,tt+1}$ represents the effect of street i at time t on street j at time $t+1$. The higher value of SWM elements is related to observations which have greater importance in estimating the target individual. Therefore, STWM could be defined as:

$$\mathrm{STWM}=\begin{bmatrix} w_1 & 0 & \dots & 0 & 0 \\ 0 & w_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & w_{T-1} & 0 \\ 0 & 0 & \dots & 0 & w_T \end{bmatrix}, \tag{6}$$

where t: $\{1, 2, \dots, T\}$ and,

$$w_t=\begin{bmatrix} w_{11,tt} & w_{11,tt+1} & w_{12,tt} & w_{12,tt+1} & \dots & w_{ij,tt} & w_{ij,tt+1} \\ w_{11,t+1t} & w_{11,t+1t+1} & w_{12,t+1t} & w_{12,t+1t+1} & \dots & w_{ij,t+1t} & w_{ij,t+1t+1} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ w_{n1,tt} & w_{n1,tt+1} & w_{n2,tt} & w_{n2,tt+1} & \dots & w_{nn,tt} & w_{nn,tt+1} \\ w_{n1,t+1t} & w_{n1,t+1t+1} & w_{n2,t+1t} & w_{n2,t+1t+1} & \dots & w_{nn,t+1t} & w_{nn,t+1t+1} \end{bmatrix}. \tag{7}$$

To gain a better understanding of the generated STWM, consider a scenario where street i is spatially connected to street j, and street k is also connected to street j. Furthermore, assume that street k is the second-order nearest neighbor to street i. It is important to note that accidents occurring on streets n (where $n \neq i, j, k$) at time t and $t+1$ will not impact street i at time t and $t+1$. This is reflected in the elements $w_{in,tt}$, $w_{in,tt+1}$, $w_{ni,tt}$ and $w_{ni,tt+1}$, which are all set to 0.

Additionally, we assume that accident situations in the future are unaffected by accident events in the past. Specifically, for street i, the accident data that occurs on streets j and k at time $t+1$ will not affect the situation at time t. Consequently, the elements $w_{ii,t+1t}$, $w_{ij,t+1t}$, $w_{ji,t+1t}$, $w_{ik,t+1t}$, and $w_{ki,t+1t}$ are all set to 0.

Furthermore, to prevent self-prediction of accidents on street i at the same time t, we utilize a zero-diagonal matrix represented by w_t. This ensures that the elements $w_{ii,tt}$, $w_{ii,t+1t+1}$, $w_{jj,tt}$, $w_{jj,t+1t+1}$, $w_{kk,tt}$ and $w_{kk,t+1t+1}$ are all set to 0. Hence, the weight matrix can be written as:

$$w_t=\begin{bmatrix} 0 & w_{ii,tt+1} & w_{ij,tt} & w_{ij,tt+1} & w_{ik,tt} & w_{ik,tt+1} & 0 & \dots & 0 \\ 0 & 0 & 0 & w_{ij,t+1t+1} & 0 & w_{ik,t+1t+1} & 0 & \dots & 0 \\ w_{ji,tt} & w_{ji,tt+1} & 0 & w_{jj,tt+1} & w_{jk,tt} & w_{jk,tt+1} & 0 & \dots & 0 \\ 0 & w_{ji,t+1t+1} & 0 & 0 & 0 & w_{jk,t+1t+1} & 0 & \dots & 0 \\ w_{ki,tt} & w_{ki,tt+1} & w_{kj,tt} & w_{kj,tt+1} & 0 & w_{kk,tt+1} & 0 & \dots & 0 \\ 0 & w_{ki,t+1t+1} & 0 & w_{kj,t+1t+1} & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

In the following sections, we will provide a detailed explanation of the values assigned to the elements of the STWM.

4.2 | Calculate weights from topological network variables

4.2.1 | Spatiotemporal connectivity relationship

Connectivity is a geometric property that characterizes the linkages between line features such as road networks. It refers to the degree of links between nodes (Bamford & Robinson, 1978), with nodes having higher degrees being more critical in a road network.

Definition 1. To operationalize this concept, we define the spatiotemporal Connectivity Relationship, which identifies the neighbors (NID_i) of a given spatial–temporal road ($str_{i,t}$). The Connectivity index of each $str_{i,t}$ is determined by the average Connectivity of its neighbors, which are nodes that are spatially and temporally connected to $str_{i,t}$. Formally, this can be represented as a graph $G=(V, STRs)$ with vertices V and edges STRs connecting them. Each edge $str_{i,t}$ links vertex v_k to vertex v_z at time t. To calculate the number of neighbors of a node v_k, we use the formula NID $(v_k)=(t\times(n_s+1))-1$, where t is the temporal scale (in this study, we use one hour as the temporal scale) and n_s represents the number of spatially connected nodes to v_k. To avoid considering v_k as its own neighbor at the same time t, we subtract 1 from the result of t times n_s+1. Simplifying the equation, we get NID $(v_k)=(2\times(n_s+1))-1$. Consequently, the spatiotemporal Connectivity of the node v_k is given by

$$K(v_k) = \sum_{z=1}^{\mathrm{NID}(v_k)} r_{kz}, \tag{8}$$

where NID (v_k) represents the total number of spatiotemporal neighbors of the node v_k, and r_{kz} indicates the relation of nodes v_k and v_z. The value of r_{kz} is 1 if nodes v_k and v_z are connected, indicating the presence of a connection between them. Conversely, if v_k and v_z are not connected, the value of r_{kz} is 0, indicating the absence of a connection. Finally, the spatiotemporal Connectivity Index (SCI) of $str_{i,t}$ can be calculated as:

$$\mathrm{SCI}(str_{i,t}) = \frac{\mathrm{SCI}(v_{k,t}) + \mathrm{SCI}(v_{z,t})}{2}, \tag{9}$$

where $\mathrm{SCI}(v_{k,t})$ and $\mathrm{SCI}(v_{z,t})$ are the spatiotemporal Connectivity Index of nodes v_k and v_z at time t, respectively.

4.2.2 | Spatial eigenvector centrality

Eigenvector Centrality is a method for measuring a node's centrality in a network by considering the centrality of its neighbors (Bonacich, 1987; Rings et al., 2022). Specifically, eigenvector Centrality is based on the idea that connections to highly central nodes contribute more to a node's centrality score than connections to fewer central nodes. Therefore, nodes with high Eigenvector Centrality scores are those connected to other highly central nodes in the network. In this study, we use Eigenvector Centrality to identify clusters of highly central nodes in the road network. These clusters represent areas where the network is most vulnerable to disruptions or failures.

Definition 2. The Eigenvector Centrality for node k ($k \in V$) is determined by solving the linear system of equations as shown below:

$$AX = \lambda X, \tag{10}$$

where A represents the adjacency matrix of graph G, and λ is the eigenvalue obtained using the Perron-Frobenius theorem. The solution X to this system is unique, and its entries are positive if λ is the largest eigenvalue of A (Newman, 2008).

4.2.3 | Spatial dissimilarity-based centrality

Dissimilarity-based Centrality allows for the quantification of each node's topological contribution to the centrality of a given node (Alvarez-Socorro et al., 2015). To improve the ranking of nodes in a network, we employ Dissimilarity-based Centrality with the Spatial Eigenvector Centrality measure.

Definition 3. Dissimilarity-based Centrality measures the difference between the neighborhoods of two nodes, v_k and v_z, by defining a distance metric. In this study, we utilize the Jaccard index (Jaccard, 1912) as the dissimilarity measure, which is calculated using Equation (11). In the Jaccard index, two nodes, v_k and v_z, are considered close if they share common neighbors. Therefore, the Dissimilarity-based Centrality of node v_k is given by

$$D_{kz} = 1 - \frac{|v_k \cap v_z|}{|v_k \cup v_z|}. \tag{11}$$

Hence, we can calculate the contribution made by node v_z in the neighborhood of node v_k by assigning a weight, denoted as W_{kz}, which is expressed by the following equation:

$$W_{kz} = A_{kz}\, D_{kz}, \tag{12}$$

where D_{kz} refers to the dissimilarity matrix and A_{kz} is the adjacency matrix. This approach allows us to account for the relative importance of each node in the neighborhood of the node v_k. The spatiotemporal Dissimilarity-based Centrality index for node v_k is calculated based on the centrality of its neighboring nodes, weighted by their contributions. The centrality of a node is proportional to the total cumulative centrality of its neighbors. Therefore, the spatiotemporal Dissimilarity-based Centrality index for node v_k is given by

$$\mathrm{DC}_k = \frac{1}{\lambda_{\max}} \sum_{k=1}^{n} W_{kz}\, \mathrm{EC}_k, \quad k = 1, 2, \ \dots\ n, \tag{13}$$

where n is the number of nodes in the network and EC_k is the spatial Eigenvector Centrality of the node v_k. The largest eigenvalue of the adjacency matrix A (see Equation 10) is denoted as $\lambda_{\max}$. Finally, to calculate the spatial Dissimilarity-based Centrality index of a street $str_{i,t}$, the average of its relevant nodes v_k and v_z is taken into consideration.

4.2.4 | Spatiotemporal betweenness centrality

The Eigenvector Centrality and Dissimilarity-based Centrality measurements vary depending on the geometric structure of the road network and remain constant over time. To address this issue, we incorporated time variables into the weight calculation of STRs' neighborhoods by utilizing the betweenness centrality properties of the road network. The betweenness centrality metric identifies nodes that contribute to the transfer of accident effects within the road network. It can also be used to obtain the edge betweenness, which helps identify the community structure of networks (Cardillo et al., 2006). Essentially, the betweenness centrality metric measures how often a node falls on the shortest path between other nodes (Freeman, 2002). Nodes with high betweenness centrality values are typically crucial and likely to be used by drivers (Shang et al., 2020), making them vulnerable parts of the road network system.

Definition 4. To determine the betweenness centrality of $str_{i,t}$, we calculate the shortest path between vulnerable areas by considering 311 reports and Vision Zero reports as vulnerable accident areas. A higher betweenness centrality value for a given $str_{i,t}$ indicates a greater level of significance. Consequently, these STRs represent the most vulnerable parts of the entire network system. They have the potential to exert a larger influence on neighboring STRs compared to other STRs. Given the vulnerable areas as origins O and destinations D, the betweenness centrality for $str_{i,t}$ is the shortest path that passes through street i at time t and can be calculated as follows:

$$\mathrm{BC}\left(str_{i,t}\right) = \sum_{O \neq D} \frac{n_{OD}\left(str_{i,t}\right)}{n_{OD}}, \tag{14}$$

where n_{OD} refers to the total number of available shortest paths from the origin vulnerable area O to the destination vulnerable area D at time t. Meanwhile, $n_{OD}\left(str_{i,t}\right)$ denotes the number of paths that traverse a street i at time t.

4.2.5 | Spatiotemporal closeness centrality

The Closeness Centrality measure is used to determine how close a node is to all other nodes in a network. This is done by calculating the sum of the distances between a node and all other nodes in the network using the shortest paths between all pairs of nodes. The nodes with high Closeness Centrality scores are those with the shortest distances to all other nodes in the network.

Definition 5. We propose a method for calculating the Closeness Centrality of each node based on the 311 non-emergency reports and Vision Zero reports, which serve as indicators of vulnerable areas. The Closeness Centrality of each node v_k is the average shortest path between the node v_k and the vulnerable areas. In other words, it measures how close a node is to the vulnerable areas at time t. The formula for calculating the Closeness Centrality of a node $v_{k,t}$ is given by

$$\mathrm{CC}\left(v_{k,t}\right) = \frac{n-1}{\sum_{k=1}^{n-1} d\left(v_k, x, t\right)}, \tag{15}$$

where $d(v_k, x, t)$ represents the shortest path distance between node v_k and vulnerable area x at time t, and n is the number of vulnerable areas that can be reached by node v_k at time t. The spatiotemporal Closeness Centrality of $str_{i,t}$ is calculated by the average of Closeness Centrality value of its relevant nodes v_k and v_z at time t.

4.2.6 | Combine topological measurements

The weight of $str_{j,t}$ on $str_{i,t}$, based on the topological variables, is calculated as the Euclidean Distance of each topological factor of streets i and j at the same time t,

$$W_{\text{top } ij,tt} = \sqrt{\left(\mathrm{DC}\left(str_j\right) - \mathrm{DC}\left(str_i\right)\right)^2 + \left(\mathrm{BC}\left(str_{j,t}\right) - \mathrm{BC}\left(str_{i,t}\right)\right)^2 + \left(\mathrm{CC}\left(str_{j,t}\right) - \mathrm{CC}\left(str_{i,t}\right)\right)^2}, \tag{16}$$

where $\mathrm{DC}\left(str_{ij}\right)$, $\mathrm{BC}\left(str_{ij,t}\right)$, and $\mathrm{CC}\left(str_{ij,t}\right)$ are time-invariant spatial eigenvector centrality (improved by dissimilarity-based centrality), spatiotemporal betweenness centrality, and closeness centrality, respectively, and the impact of street i at time t $\left(str_{i,t}\right)$ on street j at time $t+1$ ($str_{j,t+1}$) is defined as:

$$W_{\text{top } ij,tt+1} = \left(\left(\text{DC}(str_j) - \text{DC}(str_i) \right)^2 + \left(\text{BC}(str_{j,t+1}) - \text{BC}(str_{i,t}) \right)^2 + \left(\text{CC}(str_{j,t+1}) - \text{CC}(str_{i,t}) \right)^2 \right)^{-1/2}. \tag{17}$$

4.3 | Calculate weights from economic variables

The weight between two observations from the point of economic variables can be calculated using the inverse distance function introduced by Hines and Rosen (1993). This function incorporates economic variables and assigns decreasing weight values ($W_{\text{econ } ij,tt}$) to pairs of observations with increasing economic distance. Specifically, the weight assigned to a pair of observations decreases in proportion to their economic distance. This function is defined as:

$$W_{\text{econ } ij,tt} = \frac{1}{\left| z_{j,t} - z_{i,t} \right|}, \tag{18}$$

where $W_{\text{econ } ij,tt}$ represents the magnitude of the connection of street i at time t and street j at the same time t for economic purposes. Additionally, the intensity of the effect of street i at time t on street j at time $t+1$ is defined as:

$$W_{\text{econ } ij,tt+1} = \frac{1}{\left| z_{j,t+1} - z_{i,t} \right|}, \tag{19}$$

where $z_{i,t}$ and $z_{j,t}$ represent the economic variable of interest for street i and street j at time t, respectively. The economic distance between streets i and j at time t is calculated as the absolute difference between $z_{j,t}$ and $z_{i,t}$, denoted as $\left| z_{j,t} - z_{i,t} \right|$. Furthermore, $z_{j,t+1}$ represents the economic variable of interest for street j at time $t+1$, and $\left| z_{j,t+1} - z_{i,t} \right|$ represents the economic distance between street i at time t and street j at time $t+1$.

4.4 | Integrating topological network and economic variables to calculate weights

Sun et al. (2016) suggested two methods for combining the weights of endogenous and exogenous variables. One of these methods, initially proposed by Case et al. (1993), employs an additive spatial weight function, denoted as $w(a_{ij}, b_{ij})$, where a_{ij} and b_{ij} represent the geographical and economic distances between street i and street j, respectively. Another approach proposed by Qu and Lee (2015) involves a multiplicative spatial weight function expressed as $w(a_{ij}, b_{ij}) = w_1(a_{ij})\ w_2(b_{ij})$. This function uses the exogenous geographical distance a_{ij} and the endogenous economic variable b_{ij} to calculate the SWM. Notably, an additive spatial weight function may still produce non-zero spatial weights even if one of the spatial covariates is irrelevant. For instance, if $w_2(b_{ij}) = 0$, $w(a_{ij}, b_{ij})$ may not be equal to zero, but $w_1(a_{ij})$ is not equal to zero for certain i and j. In contrast, a multiplicative spatial weight function will yield a zero weight if either $w_1(a_{ij}) = 0$ or $w_2(b_{ij}) = 0$ occurs for certain i and j.

In this study, we utilize the exponential version of the multiplicative weight function. The exponential weight function allows for the combination of weights in a manner that emphasizes larger values while downplaying smaller values. This function is beneficial when dealing with weights that vary extensively, as it helps balance the effects of extreme values. Specifically, in this study, using the exponential function helps to focus on the relationship of streets with their nearest neighbors. Therefore, we calculate the magnitude of the spillover effect of a crash incident on street i, on street j at the same time t, as follows:

$$w_{ij,tt} = e^{W_{\text{top } ij,tt}} \times e^{W_{\text{econ } ij,tt}}. \tag{20}$$

We also calculate the magnitude of the spillover effect of a crash incident on street i at time t, on street j at time $t+1$, as follows:

$$w_{ij,tt+1} = e^{w_{\text{top } ij,tt+1}} \times e^{w_{\text{econ } ij,tt+1}}, \tag{21}$$

where w_{top} represents the weight derived from the topological factors of the road network, including the Spatiotemporal Connectivity relationship, Spatial Dissimilarity-based Centrality, and Spatiotemporal Betweenness Centrality. On the contrary, w_{econ} indicates the weight obtained from the economic variables by utilizing the economic distance.

Finally, the STWM is computed using Equations (20) and (21). For further analysis, a normalized version of the STWM is typically employed, which is achieved by rescaling the STWM using the min-max function.

4.5 | Feature selection

In machine learning, feature selection is crucial since using all features may not be desirable to build an accurate predictor model. Several measures have been proposed to identify the most important features. Kuhn and Johnson (2019) and Zheng and Casari (2018) suggest using Pearson's and Kendall's rank coefficients, respectively, to calculate the correlation between numerical and categorical variables. Specifically, Pearson's correlation coefficient is the ratio of the covariance of two variables to the product of their standard deviations (Lee Rodgers & Nicewander, 1988). Pearson's correlation coefficient is denoted as r_{xy} and is defined as:

$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n \sum y_i^2 - \left(\sum y_i\right)^2}}, \tag{22}$$

where n represents the number of observations, and x_i and y_i are the values of x and y for ith observation. The value of Pearson's correlation coefficient ranges between −1 and 1.

On the contrary, Kendall's rank coefficient, also known as the Tau-b statistic, is a non-parametric measure that does not rely on assumptions about the distributions of two features (Corder & Foreman, 2011). Kendall's rank coefficient can be computed as:

$$\tau_B = \frac{n_c - n_d}{\sqrt{\left(n_0 - n_1\right)\left(n_0 - n_2\right)}}, \tag{23}$$

where $n_0 = \frac{n(n-1)}{2}$, $n_1 = \sum_i \frac{t_i(t_i - 1)}{2}$, $n_2 = \sum_j \frac{u_i(u_i - 1)}{2}$, and n_c represent the number of concordant pairs, and n_d denotes the number of discordant pairs. In addition, t_i indicates the number of tied values in the ith group of ties for the first quantity, and u_j represents the number of tied values in the jth group of ties for the second quantity.

5 | CASE STUDY: CRASH INCIDENT, 311 REPORTS, AND SOCIOECONOMIC DATA IN BOSTON

Boston, with a population of <700,000, serves as Massachusetts' industrial and commercial center. This research focuses on the Roxbury neighborhood of Boston during the winter of 2016 (January, February, and March). The study utilizes a crash dataset collected by the Vision Zero project to analyze traffic safety in the Roxbury neighborhood. Figure 2 illustrates the study area along with Vision Zero crash and Vision Zero-entry data. Vision Zero crash is a multinational road traffic safety initiative that aims to eliminate traffic crashes by allocating city resources to established strategies. The Vision Zero crash dataset consists of records of incidents requiring public safety response due to injuries or fatalities. This dataset contains information, including the date, time, and precise

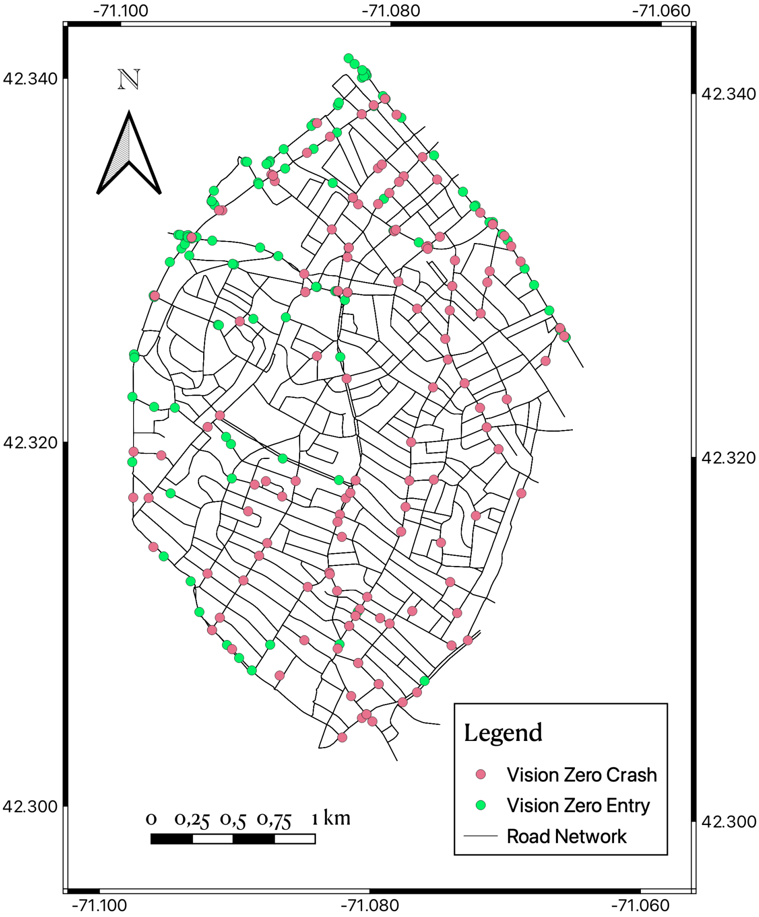


FIGURE 2 Case study and dataset.

latitude and longitude coordinates of each crash incident. During the winter of 2016, a total of 163 crash incidents were reported in Roxbury. Henceforth, we shall utilize the term "crash" in lieu of "Vision Zero crash" for concision. Additionally, the Vision Zero-entry dataset, collected by citizens, reports accident-prone areas. It provides information on the date, time, status, and location of each observation. The "status" attribute indicates whether the reported issue has been resolved or not. For this study, any Vision Zero-entry data that remains unresolved is considered a vulnerable area. A total of 123 Vision Zero-entry data points were gathered for the Roxbury Neighborhood case study.

Table 3 presents descriptive statistics of the explanatory variables used in this study. The 311 non-emergency request service is a North American municipal hotline that enables citizens to provide feedback to their municipal governments regarding the performance of urban services. The 311 data comprise essential details, including the date, time, subject, latitude, and longitude coordinates of each report. Reports are classified into 11 general topics and referred to the relevant departments. This study analyzed 311 reports related to accident-related topics.

In addition, road network properties, including road type, road width, road length, and speed limit, are incorporated into the model. These properties vary in the spatial dimension but are constant in the temporal dimension. Socioeconomic characteristics, such as dominant land use, which varies in the spatial dimension but remains constant in the temporal dimension, are also included. The model also incorporates normalized population density for STRs, which is calculated by dividing the population count by the road length. Population density data vary spatially and remain constant temporally, as they are reported annually.

TABLE 3 Descriptive statistics of data.

Variable	Descriptive statistical value	SD	Description
Vision zero-crash data	[0.00 < 0.00038 < 1]	0.02	The number of crashes on STRs. This variable is temporally and spatially varying. *Source*: https://data.boston.gov/organization/vision-zero-boston-program
Vision zero-entry data	[0 < 0.0002 < 3]	0.018	The number of vision zero-entry on STRs. Accessed from https://data.boston.gov/dataset/vision-zero-entry
311 request service	[0 < 0.0005 < 10]	0.03	The number of non-emergency 311 request services on STRs. Accessed from https://data.boston.gov/dataset/311-service-requests
Road type	Accessed from https://koordinates.com/layer/96131-boston-massachusetts-street-edges/		
Residential road	80.58%		The ratio of residential roads on STRs. Residential roads indicate the proportion of roads specifically intended for accessing residential areas
Secondary road	4.637%		The ratio of secondary roads in streets refers to the proportion of roads that are classified as secondary, which typically have two lanes and a central line that separates traffic from both directions (https://wiki.openstreetmap.org/wiki/Tag:highway%3Dsecondary)
Tertiary road	14.78%		The ratio of tertiary roads in streets represents. Tertiary roads serve as connections between small settlements and link local centers to larger settlements. Additionally, tertiary roads often connect minor streets to more prominent roads, facilitating transportation and access within the street network (https://wiki.openstreetmap.org/wiki/Tag:highway%3Dtertiary)
Road width [m]	[0 < 3.69 < 8.47]	1.2306	Road width varies in the spatial dimension but remains constant in the temporal dimension
Road length [m]	[16.71 < 331.36 < 5106]	491.39	Road length varies in the spatial dimension but remains constant over time
Speed limit [mph/h]	[15 < 21 < 35]	3.1176	The speed limit varies across different spatial locations but remains constant over time
Dominant land use	Dominant land-use variables exhibit spatial variation while remaining constant over time. Accessed from https://data.boston.gov/dataset/parcels-2016-data-full		
Residential	56.52%		The ratio of residential-dominant land use for STRs
Exempt	33.04%		The ratio of exempt dominant land use for STRs
Commercial	8.98%		The ratio of commercial dominant land use for STRs
Mixed residential and commercial	0.87%		The ratio of mixed residential and commercial dominant land use for STRs
Industrial	0.58%		The ratio of industrial dominant land use for STRs
Population density	[0 < 0.68 < 11.23]	0.84	The normalized population density for STRs is calculated by dividing the population count by the road length. Population density data exhibit spatial variation while remaining constant over time

TABLE 3 (Continued)

Variable	Descriptive statistical value	SD	Description
Traffic lights	[0 < 0.258 < 1]	0.4375	The count of traffic lights in STRs varies spatially and remains constant over time. The data for traffic lights can be accessed from the following source: https://data.boston.gov/dataset/traffic-signals
Weather data	The weather data vary in the temporal dimension and remain constant in the spatial dimension. The weather data can be accessed from the following source: Accessed from https://visual-crossing-weather.p.rapidapi.com/history		
Clear	16.66%		The ratio of clear weather conditions in STRs refers to the proportion of instances when the sky is free from clouds. Clear weather is characterized by the absence of clouds in the sky
Overcast	58.33%		The ratio of overcast weather conditions in STRs indicates the proportion of instances when the sky is predominantly covered by clouds, typically accounting for over 95% of the sky
Clear overcast	8.33%		The ratio of clear overcast weather conditions in STRs refers to the proportion of instances when the sky is predominantly covered by clouds, with no rain or other precipitation
Cloudy	16.66%		The ratio of cloudy weather conditions in STRs refers to the proportion of instances when the sky is predominantly covered by clouds, resulting in less sunshine
Precipitation	0	0	The amount of recorded precipitation in STRs is measured in millimeters (mm)
Temperature	[−4 < 34.76 < 63]	10.702	The recorded temperature for STRs is measured in degrees celsius (°C)
Wind speed	[0.1 < 11.38 < 23.8]	5.145	The wind speed for STRs is measured in kilometers per hour (km/h)
Visibility	[1.9 < 9.29 < 9.9]	1.8251	The visibility for STRs is measured in meters
Snow depth	[0 < 0.72 < 5.68]	1.428	The snow depth in STRs is measured in millimeters (mm)
Cloud cover	[0 < 59.6 < 100]	42.164	The cloud cover percentage represents the extent of the sky covered by clouds. It is measured on a scale from 0 to 100, where higher values indicate a greater amount of cloud cover

Note: [Min < Mean < Max].

Finally, weather data, including precipitation, snow depth, temperature, wind speed, visibility, and percentage of cloud cover, are used in the spatiotemporal analysis of crash data. The weather data is collected from one weather monitoring station and thus varies in the temporal dimension but is constant over the spatial dimension of the case study.

The spatiotemporal database utilized in this study consists of street-level data known as Streets for an Hour of a Day (STRs) from January to March 2016. The dataset is stored in a PostgreSQL database. The spatiotemporal analysis of the data is conducted using R software, employing the ESF, RE-ESF, and SNVC analysis methods, which are implemented through the *spdep* and *spmoran* packages. Before spatiotemporal analysis, the data are pre-processed using PyCharm in a Python 3.6 environment.

6 | RESULTS AND DISCUSSION

6.1 | Temporal window tuning

According to Antczak (2018) and Kooijman (1976), the optimal spatiotemporal window size for STWM can be achieved by maximizing the spatiotemporal structure value. In this study, we determined the most suitable temporal window for the spatiotemporal database by evaluating the STWM matrix, which was calculated based on crash count data with a one-hour temporal scale. Since there is no autocorrelation in the crash count data at a one-hour temporal scale, we looked for the temporal window that exhibits the highest spatiotemporal structure in the data. To accomplish this, we first mapped STRs to a time series while holding the spatial dimension constant. The values in the time series represent crash counts from January to March 2016. Afterward, we used the Augmented Dickey-Fuller test (ADF) (Cheung & Lai, 1995) to check the stationarity of the time series. The test results indicate an ADF statistic of −5.3256 with a *p*-value of 0.000007 and a critical value of −3.51 at 1%. The more negative the ADF statistic, the more confidence we have in rejecting the null hypothesis. In this case, the test statistic is lower than the critical value, and the *p*-value is below the significance level of 0.05. As a result, we reject the null hypothesis and conclude that the time series database is stationary.

As depicted in Figure 3, the parameter tuning of the Fuzzy Time Series (FTS) model reveals that the minimum root mean square error (RMSE) values of 2.05 and 2.72 are achieved when the number of fuzzy sets and the time window parameters are set to 10 and 7 days, respectively. This indicates that a seven-day temporal window is optimal for conducting ESF, RE-ESF, and SNVC analyses.

Figure 4 illustrates the daily time series of crash data for streets (STRs) from January 2016 to March 2016. The plot includes a red line representing the seasonal decomposition based on a four-day moving average. Figure 4 shows that crash values increased in the first week of February, decreased on February 3, and then increased again in the first week of March 2016. Additionally, the seasonal decomposition indicated in Figure 4 reveals that there is no repeating short-term cycle in the series.

To understand the temporal dependencies within the time series, we employed the autocorrelation function (ACF) to identify lags with significant correlations and discern patterns. The ACF plot illustrates the similarity between crash data at a specific time lag and the crash data point at zero lag. The autocorrelation coefficient can vary between −1 and 1, where −1 represents a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 indicates no correlation.

Figure 5a represents the ACF plot for the time series, with bars representing the size and direction of correlations. The plot indicates that autocorrelations for all lags significantly differ from zero for at least one lag, implying that the time series data do not exhibit white noise (i.e., crash accident samples are not purely random). Consequently, a time series analysis is necessary to appropriately model the pattern in the data.

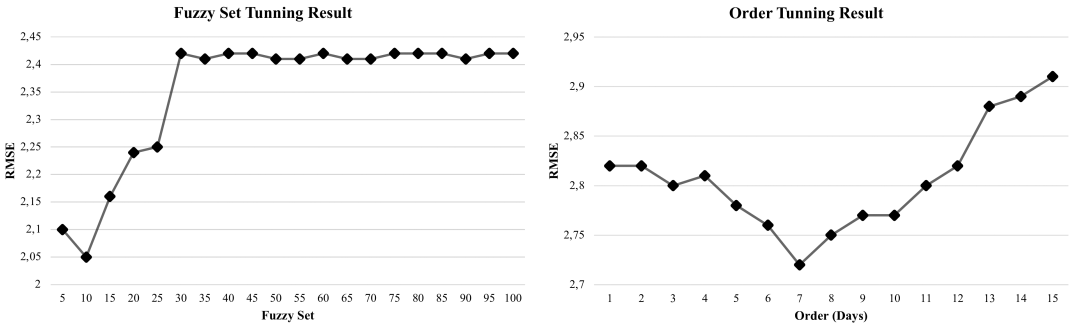


FIGURE 3 Results of parameters tuning for fuzzy time series.

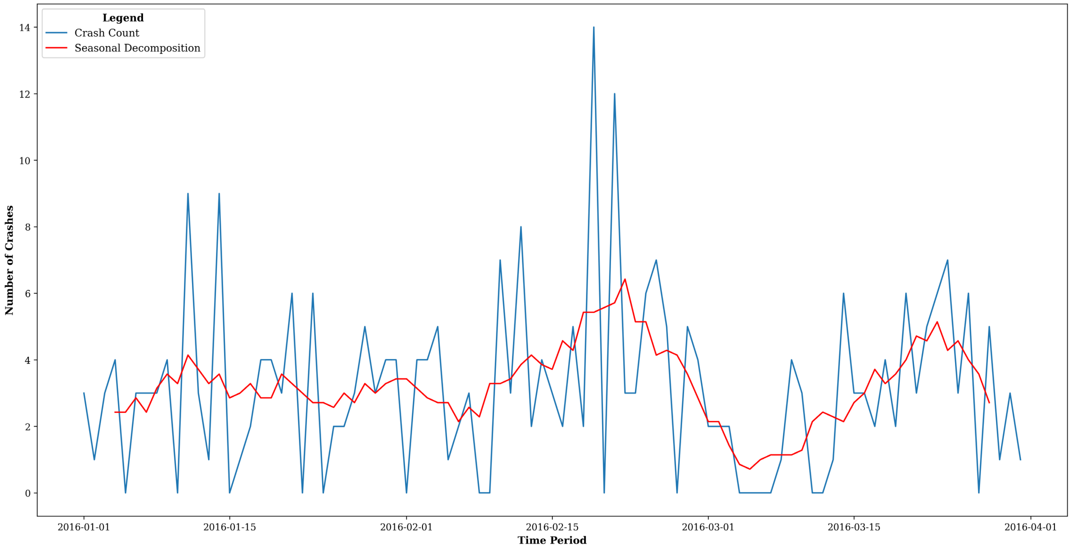


FIGURE 4 Time series of the number of crashes from January to March 2016.

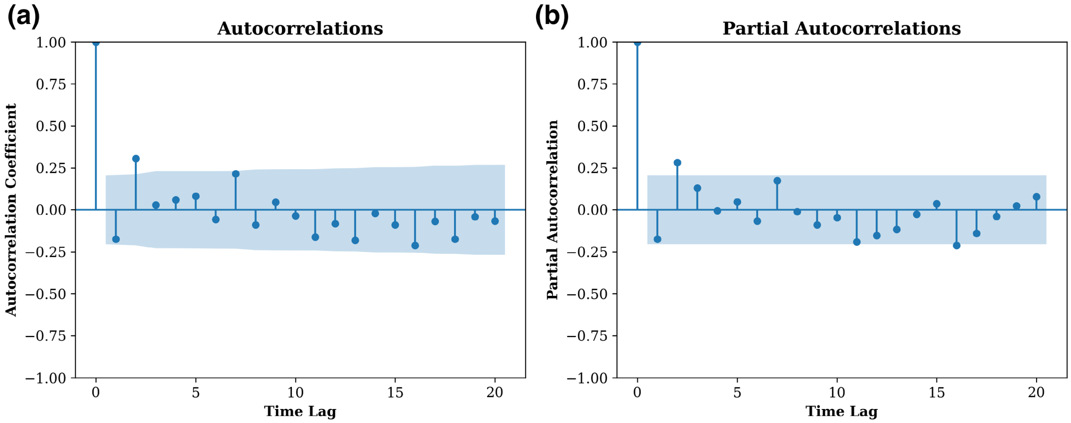


FIGURE 5 (a) The autocorrelation and (b) partial autocorrelation plot for the time series data.

Figure 5b examines the partial autocorrelations of the time series. The partial autocorrelation at a specific lag measures the degree of correlation between data points at that lag while controlling for the influence of shorter lags. In other words, it quantifies the unique correlation between observations at a given lag after accounting for correlations at shorter lags. The results reveal that the partial autocorrelation for lag 2 is statistically significant. This suggests that the value at time t is influenced by the value at time $t-2$. However, the statistical significance of the influence decreases with subsequent lags. Therefore, these findings suggest applying a second-order autoregressive model.

6.2 | Correlation between non-emergency 311 reports and crash data

The collected data from non-emergency 311 reports are classified into seven categories. Utilizing Pearson correlation statistics, we identified reports with a high correlation to crash data and utilized them as economic variables. Table 4 presents the Pearson correlation between crash data, 311 reports, and Vision Zero-entry data. The results

TABLE 4 Pearson correlation between crashes and 311 reports and vision zero-entry data.

Dependent variables	Vision zero-entry	Animal control	Public works department	Mayor's 24-h hotline	Property management	Parks recreation department	Inspectional services	Transportation traffic division
Crash	0.71	0.04	0.54	0.28	0.22	0.25	0.36	0.50

demonstrate a strong positive linear correlation between Vision Zero-entry and crash data with a value of 0.71. Additionally, the categories “Public Works Department” and “Transportation Traffic Division” in the 311 reports exhibit positive linear correlations with crash data, with values of 0.54 and 0.50, respectively. Consequently, the “Public Works Department” and “Transportation Traffic Division” reports, along with Vision Zero-entry data, were employed to investigate high-risk areas on the road network.

6.3 | The results of feature selection

Before conducting spatiotemporal analysis, it is crucial to apply feature selection to eliminate irrelevant features from the regression model. Table 5 presents the results of the feature selection based on Pearson's and Kendall's rank coefficients. To assess the correlation between numerical features such as crash number, topological variables, road width, road length, and speed limit, a p-value <0.05 indicates a significant correlation. A Pearson correlation coefficient value close to +1 suggests a strong positive relationship, while a value close to −1 suggests a strong negative relationship. A value of zero indicates no association between the features. The results show that Dissimilarity-based Centrality is positively correlated with crash data, with a Pearson value of 0.564. Furthermore, the results reveal that betweenness centrality has a moderate correlation with crash data, while Closeness Centrality has the weakest correlation (Pearson = 0.460). Additionally, concerning economic variables, road length exhibits the strongest positive correlation (Pearson = 0.401) with crash numbers.

Regarding the numerical-categorical feature correlation (crash number, road type, dominant land use, and weather condition), Kendall's correlation coefficient is examined if the p-value <0.05. Kendall's correlation coefficient results can be interpreted similarly to the Pearson coefficient. The findings of the feature selection show that the Residential Road type feature has a slightly negative correlation (−0.180) with crash numbers, suggesting that fewer crashes occurred on roads with the Residential Road type. As for precipitation data, no rainfall was recorded for more than 90% of the days; hence, it is not statistically possible to calculate the relationship between precipitation and crash data.

To prevent singularity in spatiotemporal analysis, it is crucial to investigate the correlations between independent feature variables before applying the analysis. Highly correlated features can be more dependent and have a similar effect on the dependent variable. In such cases, it is necessary to eliminate one of the features. Figure 6 presents the correlation results between independent variables. The analysis reveals a significant correlation (0.78) between the properties “Road Width” and “Lanes.” However, based on the correlation values presented in Table 5, “Road Width” has a higher correlation with crash data (0.173) than “Lanes” (0.059). Therefore, it is necessary to eliminate the “Lanes” attribute before further analysis. This step is crucial to ensuring the accuracy and validity of the analysis results.

6.4 | Spatial regression random effect eigenvector spatial filtering (RE-ESF)

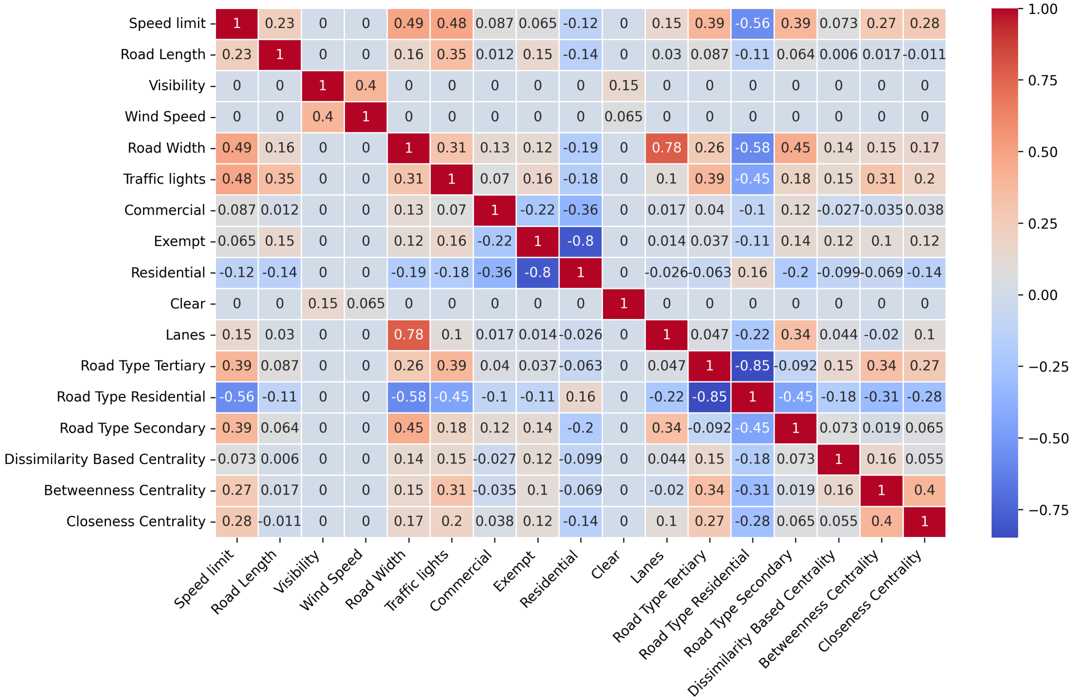
Understanding the spatial and temporal structure of crash data is crucial, and this requires an investigation of Moran's eigenvectors (MEs). MEs play a vital role in filtering spatial autocorrelation within the data, enabling

TABLE 5 Feature selection results (dependent variable: crash number numerical variable).

		Pearson's rank coefficient		Kendall's rank coefficient		
Independent features	Type	Pearson	*p*-value	Kendall	*p*-value	Relation
Topological network variables						
Dissimilarity-based centrality	Numerical	0.564	0.000			Correlated
Betweenness centrality	Numerical	0.512	0.000			Correlated
Closeness centrality	Numerical	0.460	0.000			Correlated
Economic variables						
Road type	Categorical					
Tertiary				0.114	0.000	Correlated
Residential				−0.180	0.000	Correlated
Secondary				0.146	0.000	Correlated
Road width [m]	Numerical	0.173	0.000			Correlated
Lanes	Numerical	0.059	0.000			Correlated
Road length [m]	Numerical	0.401	0.000			Correlated
Speed limit [mph]	Numerical	0.208	0.000			Correlated
Dominant land use	Categorical					
Commercial				0.076	0.000	Correlated
Exempt				0.075	0.000	Correlated
Industrial				−0.018	0.245	
Mixed residential commercial				0.001	0.955	
Residential				−0.112	0.000	Correlated
Traffic lights	Categorical	0.209	0.000			Correlated
Population density	Numerical	−0.006	0.704			
Weather condition	Categorical					
Clear				−0.039	0.011	Correlated
Clear overcast				0.026	0.091	
Overcast				0.012	0.421	
Partially cloudy				0.003	0.826	
Precipitation [mm]	Numerical					
Temperature [°C]	Numerical	−0.003	0.859			
Wind speed [km/h]	Numerical	0.034	0.030			Correlated
Visibility [m]	Numerical	−0.017	0.048			Correlated
Snow depth [m]	Numerical	0.008	0.629			
Cloud cover [0–10]	Numerical	0.013	0.390			

RE-ESF to estimate the true relationship between the dependent variable (the number of crashes) and independent variables such as road type, road length, land use, population, and weather.

Each eigenvector represents a distinct spatial scale or pattern found within the data. After analyzing the MEs using the provided STWM, we identified the first ME (e_1), which corresponds to the largest eigenvalue (λ_1) that captures the most dominant spatial autocorrelation pattern within the data. This eigenvector describes a large-scale map pattern exhibiting the highest positive Moran coefficient (Cliff & Ord, 1972). Additionally, we extracted



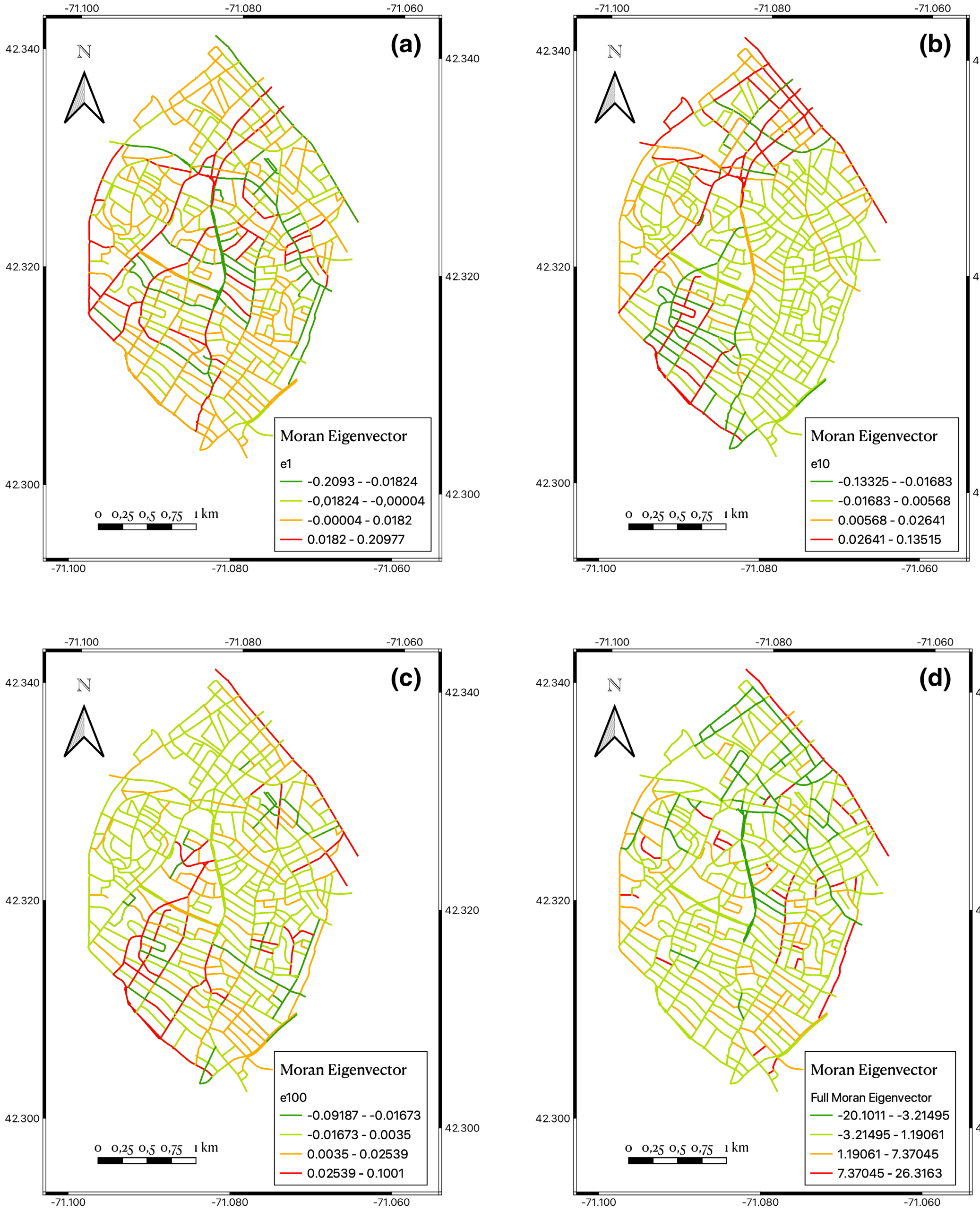
	Speed limit	Road Length	Visibility	Wind Speed	Road Width	Traffic lights	Commercial	Exempt	Residential	Clear	Lanes	Road Type Tertiary	Road Type Residential	Road Type Secondary	Dissimilarity Based Centrality	Betweenness Centrality	Closeness Centrality
Speed limit	1	0.23	0	0	0.49	0.48	0.087	0.065	-0.12	0	0.15	0.39	-0.56	0.39	0.073	0.27	0.28
Road Length	0.23	1	0	0	0.16	0.35	0.012	0.15	-0.14	0	0.03	0.087	-0.11	0.064	0.006	0.017	-0.011
Visibility	0	0	1	0.4	0	0	0	0	0	0.15	0	0	0	0	0	0	0
Wind Speed	0	0	0.4	1	0	0	0	0	0	0.065	0	0	0	0	0	0	0
Road Width	0.49	0.16	0	0	1	0.31	0.13	0.12	-0.19	0	0.78	0.26	-0.58	0.45	0.14	0.15	0.17
Traffic lights	0.48	0.35	0	0	0.31	1	0.07	0.16	-0.18	0	0.1	0.39	-0.45	0.18	0.15	0.31	0.2
Commercial	0.087	0.012	0	0	0.13	0.07	1	-0.22	-0.36	0	0.017	0.04	-0.1	0.12	-0.027	-0.035	0.038
Exempt	0.065	0.15	0	0	0.12	0.16	-0.22	1	-0.8	0	0.014	0.037	-0.11	0.14	0.12	0.1	0.12
Residential	-0.12	-0.14	0	0	-0.19	-0.18	-0.36	-0.8	1	0	-0.026	-0.063	0.16	-0.2	-0.099	-0.069	-0.14
Clear	0	0	0.15	0.065	0	0	0	0	0	1	0	0	0	0	0	0	0
Lanes	0.15	0.03	0	0	0.78	0.1	0.017	0.014	-0.026	0	1	0.047	-0.22	0.34	0.044	-0.02	0.1
Road Type Tertiary	0.39	0.087	0	0	0.26	0.39	0.04	0.037	-0.063	0	0.047	1	-0.85	-0.092	0.15	0.34	0.27
Road Type Residential	-0.56	-0.11	0	0	-0.58	-0.45	-0.1	-0.11	0.16	0	-0.22	-0.85	1	-0.45	-0.18	-0.31	-0.28
Road Type Secondary	0.39	0.064	0	0	0.45	0.18	0.12	0.14	-0.2	0	0.34	-0.092	-0.45	1	0.073	0.019	0.065
Dissimilarity Based Centrality	0.073	0.006	0	0	0.14	0.15	-0.027	0.12	-0.099	0	0.044	0.15	-0.18	0.073	1	0.16	0.055
Betweenness Centrality	0.27	0.017	0	0	0.15	0.31	-0.035	0.1	-0.069	0	-0.02	0.34	-0.31	0.019	0.16	1	0.4
Closeness Centrality	0.28	-0.011	0	0	0.17	0.2	0.038	0.12	-0.14	0	0.1	0.27	-0.28	0.065	0.055	0.4	1

FIGURE 6 The results of correlation between independent variables.

e_{10} and e_{100} eigenvectors that are orthogonal and uncorrelated with the previously identified eigenvectors, maximizing the Moran coefficient. While the initial eigenvectors captured the most dominant spatial patterns, the subsequent eigenvectors revealed more subtle spatial patterns and maximized the negative Moran coefficient. Therefore, it is crucial to investigate a multi-scale analysis of crash data in spatial and temporal structures.

Figure 7a–c illustrate the spatial distribution of the 1st, 10th, and 100th largest MEs, respectively. The spatial pattern depicted by e_1 (Figure 7a) suggests significant large-scale autocorrelation, with a strong concentration in the central and western regions of the study area. This finding indicates a higher likelihood of accidents on highways in these regions. Several factors, such as major traffic routes and major commercial and residential hubs, may be responsible for this finding. Identifying these zones is crucial for implementing large-scale strategies, such as significant infrastructure developments and wide-ranging traffic management plans. Figure 7b represents intermediate-scale spatial autocorrelation in the northern and southwestern areas of the study area. Notably, this intermediate-scale autocorrelation could not be captured by e_1 alone. According to these findings, while the broader region has its patterns, certain areas within the study area have unique characteristics that influence crash data. Possible explanations for these nuances might derive from regional factors like road designs and specific intersections. Furthermore, Figure 7c displays the spatial pattern of e_{100}, which reveals low-scale autocorrelation primarily in the southwest. The occurrence of low-scale spatial autocorrelation patterns might arise from localized phenomena such as specific intersections, particular road conditions, or potentially transient factors like construction zones. By comparing the spatial patterns of e_1, e_{10}, and e_{100}, we gain insight into the dynamic spatial structure inherent in crash data at different scales. This analysis underscores the importance of examining crash data at various scales to identify potential high-risk zones, thereby enhancing road safety initiatives.

In Figure 7d, the cumulative sum of all MEs is presented. This cumulative sum serves as a model encapsulating spatial autocorrelation across various scales. The RE-ESF and SNVC utilize the cumulative sum of MEs for further analysis. This cumulative sum enhances the accuracy and reliability of spatial analysis by capturing both large and small-scale spatial patterns. Practically, interventions can be tailored to specific scales. For example, while



FIGURES 7 Spatial distribution of the (a) 1st, (b) 10th, and (c) 100th largest MEs, and (d) all MEs.

large-scale patterns might require policy changes, micro-level patterns could be addressed with local engineering solutions. As a result of this integrated approach, effective crash prevention strategies can be developed and implemented, which is the primary objective of this study.

Figure 8 illustrates the temporal distribution of the 1st, 10th, and 100th largest MEs, as well as the cumulative sum of all MEs from January to March 2016. Notably, during the period from January 1 to February 19, 2016, e_1, e_{10}, e_{100}, and the cumulative sum of all MEs exhibited negative values. These negative eigenvectors indicate that the crash distribution deviated from the expected clustering pattern during the timeframe. Conversely, the period

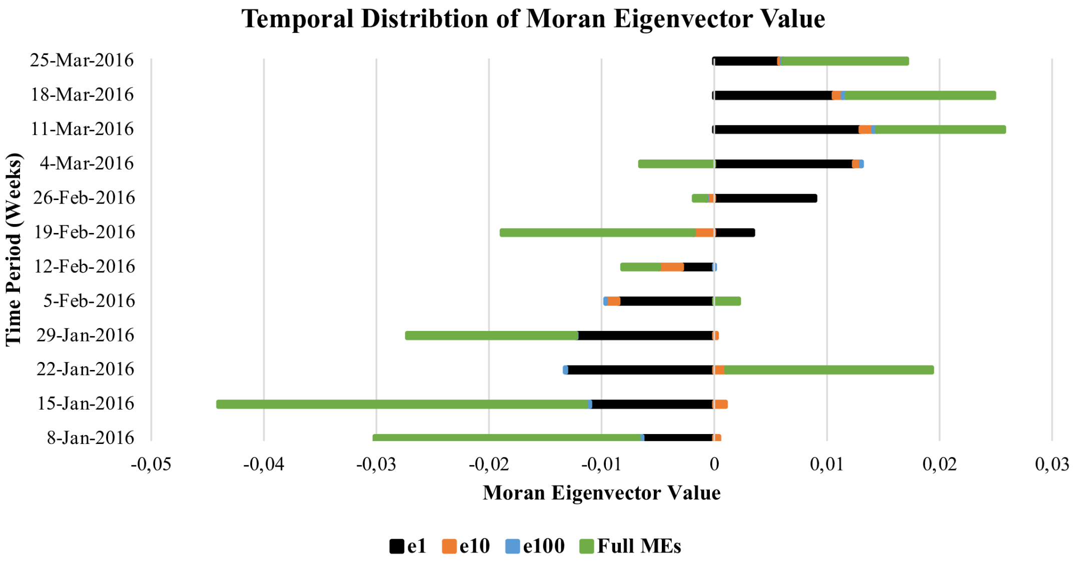


FIGURE 8 The temporal distribution of the 1st, 10th, and 100th largest and all MEs.

following February 19th saw these eigenvectors predominantly shift to positive values. The transition from negative to positive suggests that crash patterns are stabilizing. This might infer that the crash temporal distribution adopted a more predictable pattern after an initial unstable period. Recognizing periods of stability versus volatility in crash patterns provides valuable insights for road safety authorities. A stable pattern might support broader safety interventions, whereas volatile periods might require targeted interventions based on the specific reasons for the observed changes.

The negative eigenvectors observed during the initial period can aid in identifying potential spatial and temporal clustering of crashes that may have occurred during that time. In contrast, the positive eigenvectors in the subsequent period indicate a more predictable and stable pattern, offering valuable insights for future crash prevention efforts. These findings hold relevance for RE-ESF and SNVC analyses, as they consider the spatial and temporal structure of the data.

By presenting the temporal distribution of MEs and analyzing the signs of eigenvectors, this study unveils crucial information about changing patterns in crash occurrences. It provides valuable guidance for understanding and mitigating crash risks.

Table 6 presents the estimated coefficients and statistical significance for the different independent variables in the RE-ESF model. The results reveal interesting findings about the relationships between these variables and crash numbers. There is a positive correlation between crashes and the topological variables, specifically betweenness centrality (0.5406) and closeness centrality (0.1605), at the 0.5% significance level. Regarding betweenness centrality, the finding suggests that areas serving as major transition or connection points (higher betweenness centrality) in the traffic network are prone to a higher number of crashes. This could be due to the convergence of different traffic flows, creating more potential conflict points. Regarding closeness centrality, the positive correlation at the 0.5% significance level suggests that more central areas within the network, which are accessible and often frequented, experience a higher number of crashes. Regarding economic variables, the results demonstrate a positive relationship between Commercial land use (+0.063) and the number of crashes, as well as a negative correlation between Residential Road type (−0.0431) and the number of crashes, both at the 0.1% significance level. Areas with commercial land use tend to see more crashes. This could be due to increased vehicular and pedestrian activity combined with diverse traffic movements (turns, parking, and stops) in commercial areas. Conversely, residential roads show a lower number of crashes, possibly due to their lower speeds.

TABLE 6 The estimated coefficients on independent variables, standard errors, *t*-values, and *p*-values for RE-ESF (dependent variable: crash number).

	Estimated coefficients	Standard errors	*t*-value	*p*-value
(Intercept)	−0.3710	0.0887	−4.1841	0.0000
Speed limit	0.0059	0.0016	3.6191	0.0003
Road length	0.0002	0.0000	22.2755	0.0000
Visibility	0.0282	0.0083	3.4077	0.0007
Wind speed	−0.0024	0.0028	−0.8616	0.0890
Road width	0.0002	0.0012	0.1941	0.0461
Traffic signal	−0.0016	0.0108	−0.1462	0.1837
Commercial land use	0.0630	0.0345	1.8239	0.0083
Exempt land use	0.0264	0.0314	0.8410	0.0404
Residential land use	0.0240	0.0314	0.7644	0.0447
Clear weather	−0.0147	0.0114	−1.2827	0.1997
Residential road	−0.0431	0.0139	−3.1072	0.0564
Secondary road	0.0045	0.0240	0.1888	0.0019
Dissimilarity-based centrality	0.0529	0.1791	0.2955	0.0803
Betweenness centrality	0.5406	0.1118	0.7781	0.0476
Closeness centrality	0.1605	0.1346	0.5437	0.0365

Furthermore, among all the independent variables, road length exhibits the highest *t*-statistic value of 22.2755 at the 0.1% significance level. Since the null hypothesis is rejected, this indicates a relationship between the road length variable and the crash incidents. However, the small coefficient value of 0.0002 suggests that this correlation is not particularly strong. Therefore, longer roads might not necessarily lead to more crashes.

Furthermore, variables such as speed limit, visibility, wind speed, and road width also correlate with crashes, but their significance varies. Lower *p*-values for speed limit and visibility suggest stronger evidence for their association with crash numbers relative to the null hypothesis. Conversely, higher *p*-values indicate a weaker correlation between traffic signals, clear weather, and dissimilarity-based centrality variables, and crash incidents.

The intercept in the model represents the average expected value of the crash variable when all the independent variables are equal to zero. However, it is important to note that interpreting the negative intercept value of −0.3710 is not meaningful in this context, as land use and road type variables cannot realistically be zero in the real world.

The estimated standard error and scaled Moran coefficient (Moran. I/max (Moran. I)) of the RE-ESF spatiotemporal process $f_{MC(F)}(S_i)$ are 0.1078 and 0.3871, respectively. These results suggest the existence of residual spatial and temporal dependence in crash data analysis. Moreover, the value of the scaled Moran coefficient (0.3871) indicates that there is moderate-scale residual spatiotemporal dependence present in the estimation, implying the potential for clustering of accidents in certain regions.

In Figure 9a, we can observe the spatial distribution of the estimated spatially dependent component. The results indicate that the northeast and central parts of the study area (region 25) exhibit the strongest spatial dependence and are more prone to accidents. This spatial concentration underscores the necessity of understanding regional specifics, as certain areas might present unique risk factors that cause higher accident rates. Specifically, approximately 41% of the roads with high spatial dependence are tertiary roads, while residential roads account for about 36%. Tertiary roads serve as a connection between major and minor roads, accommodating a combination of traffic of both types. The nature of combined traffic may contribute to spatial dependence. In addition, given the domestic nature of residential roads, a high accident rate could indicate speeding or a lack of proper

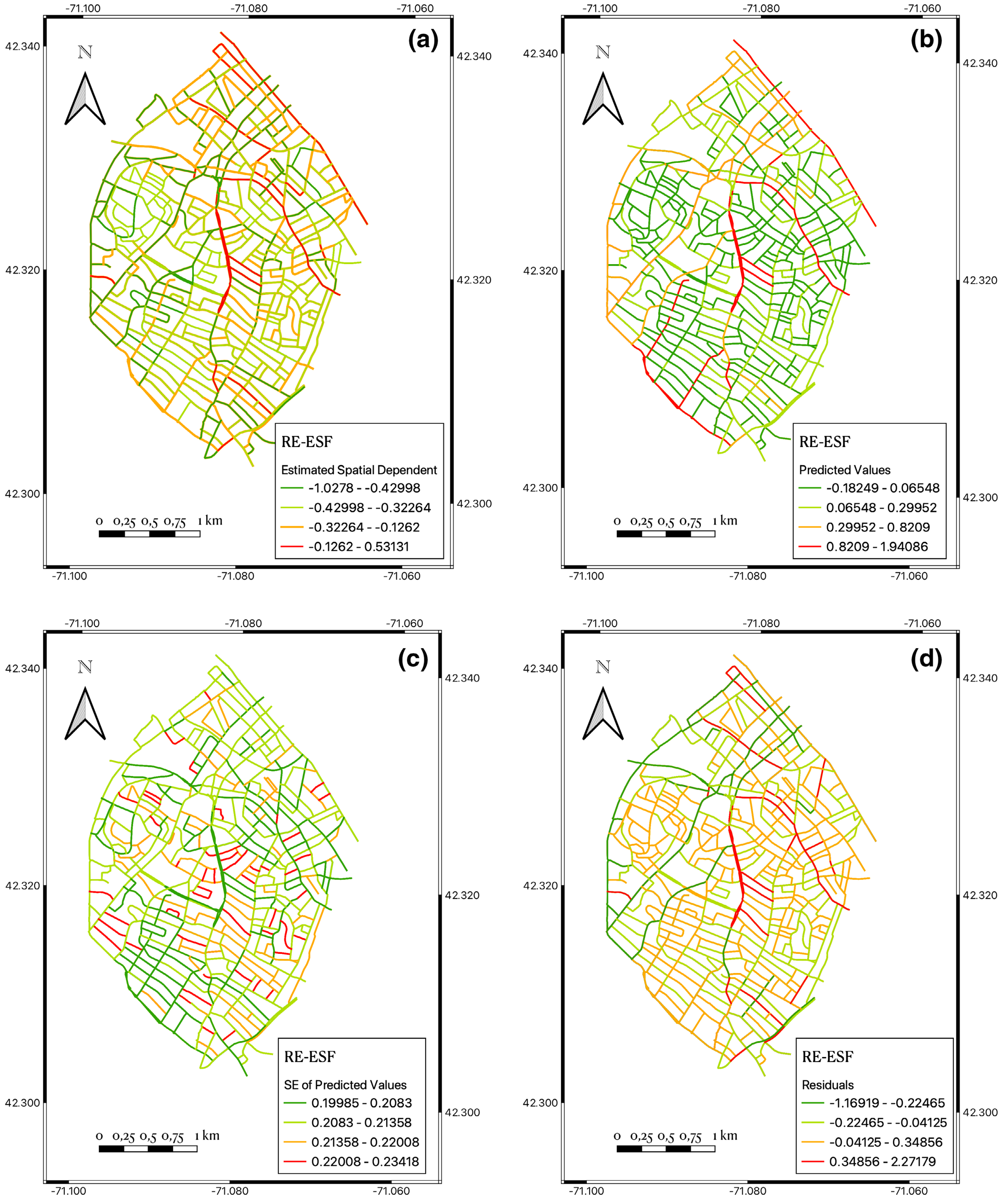


FIGURE 9 (a) Spatial distribution of estimated spatial dependent component, (b) the predicted crash incident values, (c) the standard error of the predicted values, and (d) residual.

signage. These findings suggest that accidents are more likely to happen on tertiary and residential roads in the northern and central parts of the study area. Moreover, most roads with a higher degree of spatial dependence are in areas dominated by exempt land use (57%), or by commercial dominance (25%). This may suggest a correlation between land use and accident susceptibility.

Figure 9b,c present the predicted crash incident values and the standard error of the prediction, respectively. Figure 9b suggests that highways in the center, northeast, and southwest of the study area are more susceptible to accidents. The highlighted variability in accident susceptibility across different regions underscores the heterogeneity of factors affecting accidents. Moreover, Figure 9c shows that residential roads in the central area have high standard error predictions, indicating lower accuracy in predicting accidents on these roads. Combining these results indicates that while some areas, like highways in the northeast, show high susceptibility, residential roads in the central region are difficult to predict accurately. Notably, more than 90% of streets with high standard prediction errors are associated with residential-dominant land use. Consequently, it is crucial to implement appropriate strategies to enhance safety in these areas. Additionally, the high standard error predictions for residential roads in the central area underscore a potential limitation of the RE-ESF model in these areas. Besides, most high prediction errors are related to streets with residential land use, suggesting the need for model refinement in these contexts.

Figure 9d depicts the spatial distribution of residuals, representing the disparities between the actual observed crash values and the predicted values. Notably, higher residual values are observed in the center of region 25, indicating the model's limitations. This shows a significant discrepancy between the predicted and actual numbers of accidents. In other words, the RE-ESF model fails to capture the real accident count accurately in these areas. This discrepancy suggests the presence of additional factors or variables not considered in the model. Such discrepancies are not merely statistical but have real-world implications. High residuals indicate areas potentially more hazardous than the model predicts, underscoring the need for increased safety measures and interventions. This limitation highlights the importance of dynamic model validation and updating. Consequently, this information is crucial for identifying areas that require additional attention and resources to enhance road safety.

Figure 10 presents the average temporal (spatially constant) statistical results of the RE-ESF from January to March 2016. In Figure 10a, a negative temporal structure of the estimated spatiotemporal dependent variable is observed. These negative values were predominantly observed during the first week of January and the second and third weeks of March. Notably, there was a 72% cloud cover during these weeks, which indicated overcast weather conditions. Additionally, the highest negative values were associated with an average maximum wind

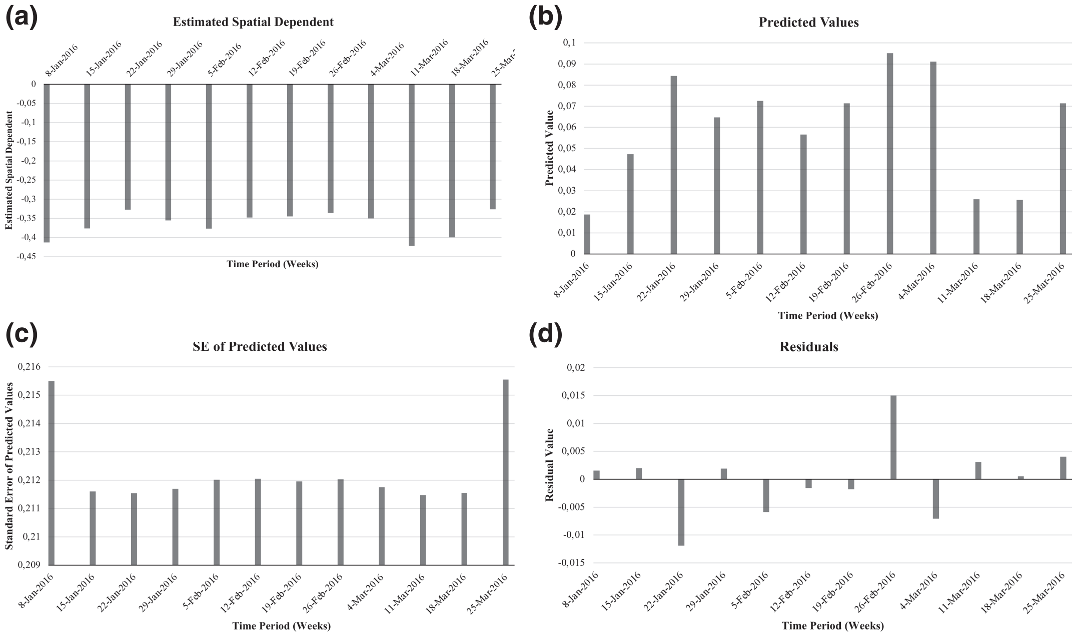


FIGURE 10 The temporal average (spatially constant) statistical results of RE-ESF. (a) estimated spatial dependent, (b) predicted value, (c) SE of predicted value, and (d) residual.

speed of 11.1 km/h. These findings suggest that adverse weather conditions, such as overcast skies and high wind speeds, may act as deterrents to accidents, thus contributing to the observed negative correlation.

The negative temporal correlation has significant implications for decision-making regarding traffic safety policies and resource allocation. Understanding periods when lower accident rates are observed, such as the first week of January and the second and third weeks of March, can help prioritize resources and interventions during other high-risk periods or locations. By focusing efforts on times or areas with higher accident rates, authorities can allocate resources effectively and implement targeted interventions to reduce accidents and enhance road safety.

Furthermore, Figure 10b illustrates the predicted values, highlighting the last week of February and the first week of March as particularly vulnerable periods with average predicted values of 0.095 and 0.091, respectively. Conversely, the first week of January and the second and third weeks of March appear to be safer periods. The presence of larger standard errors in the regression during these weeks, which denote lower accuracy in the model's predictions, supports this observation (see Figure 10c). The larger standard errors observed during specific weeks suggest increased uncertainty in the model predictions for those periods. Such large standard errors could result from various factors, including unaccounted variables or outliers, potentially influencing accident rates during that time.

Additionally, Figure 10d presents the average of residuals, reflecting the disparity between the predicted and actual number of accidents. The third week of March exhibits the lowest average of residuals, suggesting a better alignment between the model's predictions and the actual data for that week. In contrast, the third week of January and the last week of February show the highest average of residuals, indicating a large discrepancy between predicted and observed accident counts. There are several reasons for such high residuals. It is possible that some factors not included in the model (e.g., traffic flow) had a significant impact on accident counts during these periods. Furthermore, unforeseen temporary events may have affected accident rates during these weeks, which were not factored into the model.

In summary, as demonstrated in the last week of February and the first week of March, periods of heightened accident risk are critical for decision-making. Recognizing these vulnerable periods can guide traffic safety policies, and public advisories can warn of increased traffic.

In conclusion, the RE-ESF model provides significant insights into the spatial and temporal dependence of crash incidents. The spatial results of the RE-ESF underscore the significance of Land use in analyzing accident probability. Additionally, the negative temporal correlation suggests safer time zones and helps prioritize resources for other higher-risk periods. However, despite these advantages, the RE-ESF model reveals limitations in predictive accuracy for specific locations within the study area and during certain periods, as evidenced by large standard errors and residuals. These limitations suggest the need to incorporate additional variables into the model to enhance its predictive capability. The findings underscore the necessity for developing dynamic and adaptive strategies to enhance road safety.

6.5 | Spatially and non-spatially varying coefficients (SNVC)

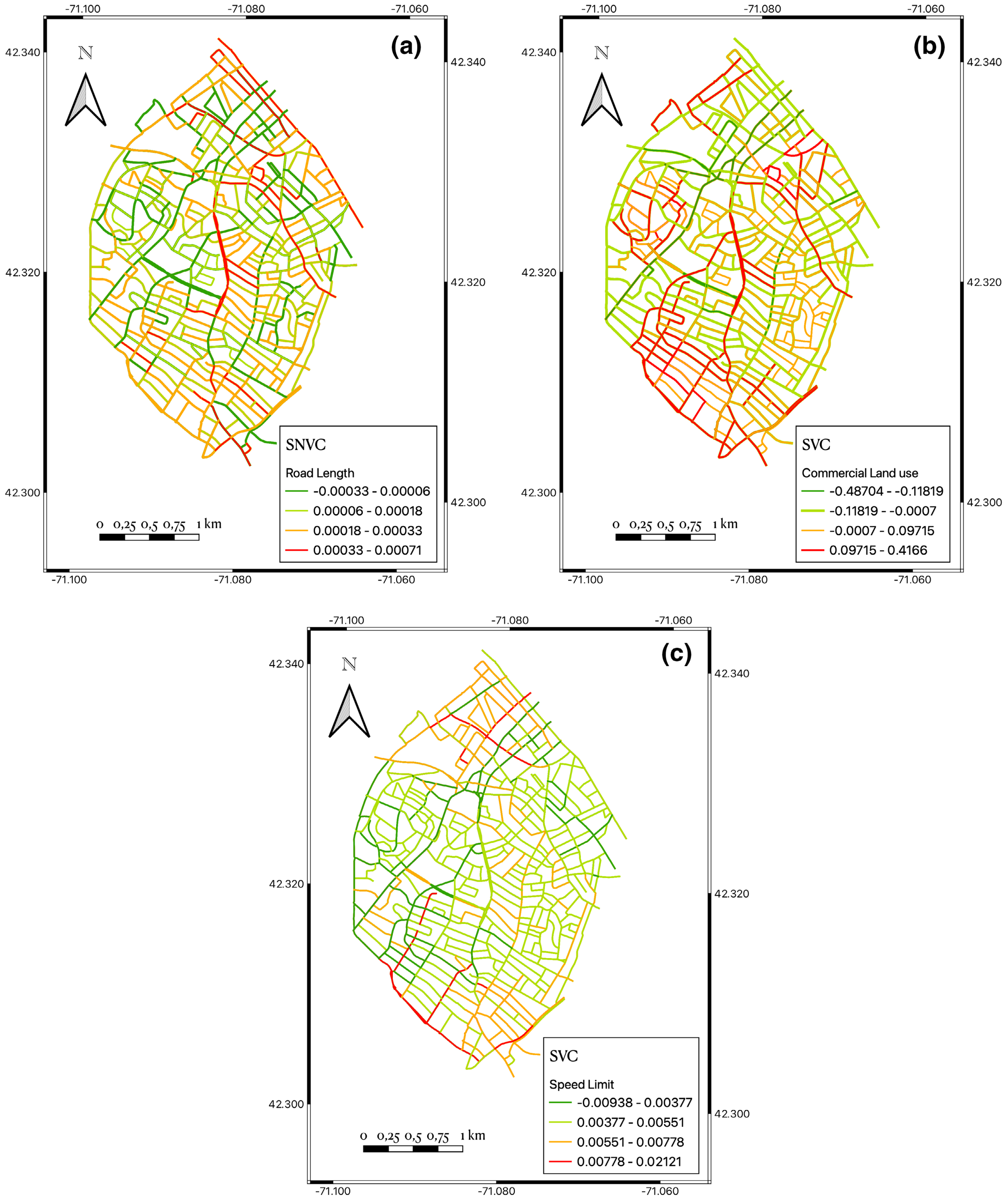
In contrast to RE-ESF, SNVC offers the ability to estimate spatially varying coefficients that change based on the residual spatial dependence structure. This means different coefficients can be estimated for different locations within the study area. The adoption of SNVC in this research highlighted spatial heterogeneity in the influence of independent variables on crash risk. Moreover, SNVC is valuable for determining the impact of independent variables such as road type, land use, population density, and weather data on crash risk in different spatial and non-spatial variations. For example, a busy intersection in an urban area might present a high risk due to factors such as traffic volume and complexity. However, a similar intersection in a suburban area could have a different risk profile due to varying factors such as speed limits or pedestrian activity. Additionally, SNVC acknowledges that the impact of these variables might change based on other factors. In this study, we employed SNVC while assuming spatially and non-spatially varying coefficients for the independent variables. The coefficients of constant variables were assumed to be fixed by default. As weather-related data was collected from a single station, we treated variables like Visibility, Wind Speed, and Clear weather as constants in spatial dimensions. This is crucial because, while certain factors might vary across locations, some remain constant and uniformly influence the entire study area. Recognizing these factors strengthens model reliability. With these assumptions, we aim to identify factors that contribute to the likelihood of accidents, considering both spatially and non-spatially varying factors.

Table 7 highlights the influence of both spatially and non-spatially varying factors on the estimated coefficients between the dependent variable (crash data) and the independent variables (Traffic Signal, land use, Road Type, Dissimilarity-based Centrality, Betweenness Centrality, and Closeness Centrality). Notably, the coefficient of road length exhibits both spatially and non-spatially varying characteristics, with random standard errors (SEs) of 0.002 and 0.026 for the spatial and non-spatial effects, respectively. These results suggest that the impact of road length on the dependent variable can change across different spatial and non-spatial contexts. This could be attributed to various factors, such as different traffic densities, the presence of intersections, or other geographical constraints that may influence its relationship with the dependent variable. Furthermore, the coefficients of speed limit and Commercial land use are identified as spatially varying coefficients with random SEs of 0.004 and 0.311, respectively. This indicates that their effect on the dependent variable varies depending on the spatial location but is constant across different non-spatial contexts. A wide range of factors can affect these spatial variations. For example, the effect of the speed limit on crash data might be influenced by land uses, road characteristics, or regional traffic policies in the vicinity. Similarly, commercial land use in a densely populated urban area might have a different risk profile than in a suburban area. This is due to variations in traffic volume, pedestrian activities, and accessibility. Overall, the results indicate that while the coefficients of the road length variable can be estimated as spatial and non-spatial coefficients, the remaining independent variables exert influence on the dependent variable, which remains consistent irrespective of the spatial and non-spatial context. This indicates that their influence on crash data is relatively stable and does not exhibit heterogeneity based on location or other non-spatial factors.

TABLE 7 The results of estimated spatially and non-spatially varying coefficients.

	Spatial effects (coefficients on *x*)		Non-spatial effects
	Random SE	Moran *I*/max (Moran *I*)	Random SE
(Intercept)	0.000	0.648	0
Speed limit	0.004	0.762	0
Road length	0.002	0.659	0.026
Road width	0.000	NA	0
Traffic signal	0.000	NA	0
Commercial land use	0.311	0.662	0
Exempt land use	0.000	NA	0
Residential land use	0.000	NA	0
Residential road type	0.000	NA	0
Secondary road type	0.000	NA	0
Dissimilarity-based centrality	0.000	NA	0
Betweenness centrality	0.000	NA	0
Closeness centrality	0.000	NA	0

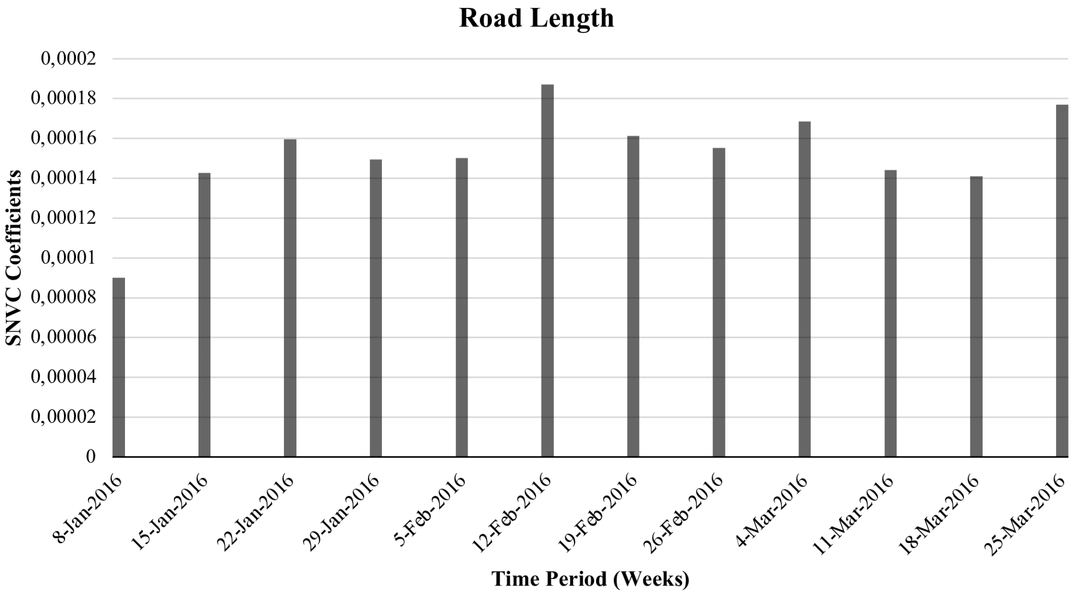
Figure 11 displays the spatial patterns of the SNVC coefficients for road length, Commercial land use, and speed limit. Figure 11a shows that road length exerts a strong influence on crash incidents, particularly in the northern and central parts of the study area. This suggests that these regions might have longer roads, leading to increased crash occurrences. Contributing factors could be higher speeds, longer distances between intersections, or even road characteristics unique to these areas. Similarly, Figure 11b highlights the significant impact of Commercial Land use on crash incidents in the southern and central parts of the study area. Commercial areas



FIGURES 11 (a) The spatial distribution of SNVC on road length and (b) spatially varying coefficients on commercial land use, and (c) speed limit.

typically have higher traffic volumes, frequent stop-and-go traffic, increased pedestrian movement, and parking activities, which contribute to increased accident risk. In addition, commercial areas might exhibit large intersections, multiple entry and exit points, and varied land use in proximity, leading to complex traffic patterns. Furthermore, Figure 11c illustrates the positive effect of speed limit in the southern region of the study area. These results suggest a potential correlation between higher speed limits or more frequent changes in speed limits and an increased number of crashes in the southern region. Higher speeds can reduce the reaction time of drivers, resulting in more severe crashes. Alternatively, it might suggest that roads in the southern area are designed for higher speeds, possibly resulting in riskier driving behavior. In conclusion, these findings emphasize the varying impacts of road length, commercial land-use, and speed limit on crashes across different areas of the study. By leveraging this information, policymakers and urban planners can develop targeted interventions to enhance road safety in specific locations where these factors have the greatest influence. For example, in regions influenced by commercial Land use, urban planners might consider redesigning intersections, enhancing pedestrian safety measures, or diverting heavy traffic away from commercial hubs.

Figure 12 depicts the temporal distribution of SNVC coefficients on road length from January to March 2016. The findings indicate that road length has a significant impact on crash incidents during the second week of February and the last week of March 2016. The degree of road length's coefficient is generally assumed to be constant over time, but the SNVC model indicates spatial and temporal variation. Specifically, the positive effect of road length on increasing crash incidents was most pronounced during the second week of February and the last week of March 2016. This could be attributed to several factors, including seasonal variations, road maintenance or construction activities, large events causing increased traffic, or other temporal phenomena. These findings imply the importance of policymakers and urban planners implementing time-adaptive strategies to enhance urban traffic safety and effectively plan construction activities. While temporal insights are crucial, it is imperative to interpret them in conjunction with spatial distributions, as shown in Figure 11a. The significant influence of road length on accidents was previously emphasized in region 25, particularly in its northern and central sections. When combined with the temporal insights from Figure 12, it becomes evident that during those two peak periods, the risk associated with longer roads in these areas might significantly increase.



FIGURES 12 The temporal distribution of SNVC on road length.

The SNVC analysis underscores the importance of considering spatial and temporal variability in factors influencing crash incidents. By understanding these dynamics, policymakers and urban planners can develop spatio-temporal adaptive safety interventions that address specific conditions and risk factors. However, the SNVC also acknowledges its limitations, including the presumption of static coefficients for certain variables and potential oversight of local variations in weather conditions. These assumptions may overlook subtle, yet impactful, micro-climatic differences or localized weather phenomena, which can significantly affect crash incidences. The findings from the SNVC analysis emphasize the need for localized, data-driven road safety interventions and highlight the critical role of accounting for spatial variability in traffic safety analysis.

Furthermore, while the STWM offers considerable advantages in refining spatial–temporal analysis, it exhibits certain limitations. The current version of STWM is designed to facilitate micro-level spatial analysis, utilizing hourly temporal scales and street-level spatial resolution. Consequently, applying the STWM to broader temporal scopes and larger urban extents may increase its complexity. Therefore, the model's intricacies may limit its applicability to expansive datasets, particularly those encompassing extensive urban areas over long-term durations, such as periods exceeding 1 year. Additionally, the scope of the analysis results is limited by the dataset from January to March 2016, which may not capture long-term trends or seasonal variations.

6.6 | Model evaluation and comparison

To demonstrate the effectiveness of the proposed STWM, we utilized the ESF, RE-ESF, and SNVC models for a comprehensive evaluation against the conventional distance-based SWM. The models' performance was quantitatively assessed using several statistical metrics: residual standard error (RSE), adjusted R^2 (adj R^2), log-likelihood (logLik), akaike information criterion (AIC), and Bayesian information criterion (BIC). These metrics were selected for their capacity to provide a comprehensive evaluation of model performance from various statistical perspectives. Specifically:

- RSE is an indicator of model precision, measuring the standard deviation of prediction errors. Lower RSE values for the STWM suggest a more precise fit to the data, underscoring its predictive strength.
- adj R^2 provides insight into the explained variability of the dependent variable, accounting for the number of predictors. Higher adj R^2 values indicate a robust goodness of fit for the STWM, highlighting the model's explanatory power.
- logLik measures the likelihood of the model having produced the observed dataset. Higher logLik scores for the STWM imply a greater probability that the model accurately represents the underlying processes, thus confirming its statistical validity.
- AIC and BIC serve as tools for model selection by balancing goodness of fit with model complexity. Lower AIC and BIC values for the STWM indicate an optimal balance, suggesting a model that effectively captures the essence of the data.

In conclusion, the model exhibiting lower RSE, AIC, and BIC alongside higher adj R^2 and logLik values is considered superior.

Table 8 presents a comparison between the results obtained using the proposed STWM and the traditional distance-based SWM across the ESF, RE-ESF, and SNVC analysis models. The superiority of STWM is quantitatively demonstrated through several critical statistical measures. Notably, the STWM's performance exhibits significantly lower RSE values of 0.209, 0.1943, and 0.1998 for the ESF, RE-ESF, and SNVC models, respectively. This reduction in RSE suggests a substantial increase in the prediction accuracy of the STWM, as a lower RSE represents a closer fit of the model to the observed data. The adj R^2 values also show a marked improvement, with the STWM achieving 0.417, 0.4034, and 0.4689 for the respective models. These values indicate a stronger explanatory power of the STWM in accounting for the variance in the data after adjusting for the number of

TABLE 8 The statistics comparison results of the STWM performance.

Weight matrix	Analysis model	RSE	Adj R^2	logLik	AIC	BIC
Distance-based SWM	ESF	0.241	0.223	69.189	111.622	9802.678
	RE-ESF	0.245	0.203	−12,980	25,994	26,102
	SNVC	0.242	0.217	−177.707	393.416	513.656
STWM	ESF	0.209	0.417	1509.204	−74.409	9241.070
	RE-ESF	0.1943	0.4034	87.7341	−141.228	−33.5314
	SNVC	0.1998	0.4689	305.3551	−568.709	−435.819

predictors. This is a crucial factor in spatial and temporal analyses, where the complexity of the models can often obscure their interpretability. Furthermore, the logLik values are substantially higher for the STWM (1509.204, 87.7341, and 305.3551), implying that the probability of the STWM producing the observed data is significantly greater than that of the conventional distance-based SWM. This aligns with a better model fit and provides significant evidence for the STWM's methodological robustness. In terms of information criteria, ESF, RE-ESF, and SNVC models utilizing the STWM reflect lower AIC and BIC values, which are −74.409, −141.228, and −568.709 for AIC, and 9241.070, −33.5314, and −435.819 for BIC, respectively. These lower AIC and BIC values show that using the STWM in ESF, RE-ESF and SNVC models achieves a preferable balance between model accuracy and complexity. This balance is essential for the effective modeling of spatial–temporal phenomena.

Interestingly, the comparative analysis of the ESF, RE-ESF, and SNVC models emphasizes the importance of the selected weight matrix—be it distance-based SWM or STWM—in the efficacy of the analysis model. The results indicate that the ESF model yields a lower RSE and AIC and higher adj R^2 and logLik in contrast to the RE-ESF model when utilizing the distance-based SWM, suggesting a tighter fit to the observed data. However, when utilizing the STWM, there is a notably high BIC value (9241.070) for the STWM's ESF model, compared to the substantial negative value found with the RE-ESF model. This warrants further investigation, as it may indicate an area where the ESF model's complexity is not sufficiently offset by its explanatory power when using the STWM. The RE-ESF model emerges as more effective, with lower RSE, AIC, and BIC values. This finding is confirmed by Murakami (2017), who states that RE-ESF generally outperforms the ESF in estimating regression coefficients. These results underscore the significance of selecting an appropriate weight matrix for spatial and temporal economic analysis to enhance the accuracy and reliability of research findings.

7 | CONCLUSIONS AND FUTURE WORKS

In this study, we utilized topological and economic variables of the urban road network to develop STWM for analyzing crash data in Boston. To address the limitations of previous studies, we incorporated road network topological measurements and economic variables instead of relying solely on a distance function to define spatial weight. By considering these variables, we aimed to model the impact of high-risk crash areas on streets.

The results of the feature selection indicate that Dissimilarity-based Centrality exhibits a strong positive correlation with crash numbers (Pearson = 0.564), while Closeness Centrality demonstrates the weakest correlation among the topological measurements (Pearson = 0.460). Additionally, when considering economic factors, road length shows the strongest positive correlation (Pearson = 0.401) with crash numbers, whereas residential road type demonstrates a slightly negative correlation (Kendall = −0.180).

RE-ESF analysis offers a quantitative indicator for identifying periods and locations more prone to crash accidents. It provides valuable insights for decision-makers to implement appropriate strategies to enhance urban road safety in high-risk areas and during critical times. For example, the results of the RE-ESF model indicate a significant positive correlation between crashes and the topological variables of betweenness centrality

(0.5406) and closeness centrality (0.1605) at a significance level of 0.5%. Regarding economic variables, the findings demonstrate a positive relationship between commercial land use (+0.063) and the number of crashes.

Moreover, the SNVC analysis reveals spatial variability in the coefficients of road length and commercial land use, indicating that their impact on crash data varies across different locations. Specifically, road length strongly influences crash incidents in the north and center of the study area. Commercial land use exhibits a significant effect in the south and center of the region. A subsequent analysis of the SNVC data indicates that the coefficients related to road length vary over time and indicate a strong impact on road length during the second week of February and the last week of March 2016. Therefore, considering the spatiotemporal variation of these factors, decision-makers can propose more suitable urban development models to reduce crash risk in the future.

Finally, to assess the functionality of the presented STWM compared to the traditional distance-based SWM, we conducted ESF, RE-ESF, and SNVC analyses. Measurements including Residual Standard Error, adjusted R^2, log-likelihood, AIC, and BIC suggest that employing STWM for estimating regression coefficients yields more reliable results than relying on distance-based SWM.

However, despite these advantages, the RE-ESF and SNVC models exhibit limitations in their predictive accuracy for specific locations and during certain periods. These limitations are due to their assumptions of fixed coefficients for certain factors and the potential oversight of local weather variations. These limitations necessitate the incorporation of additional variables to improve crash prediction functionality. Moreover, while the STWM is developed for micro-level analysis, it faces challenges when scaling up to broader temporal scopes and larger urban areas. Additionally, the scope of the data utilized in this study encompasses January to March 2016, which may overlook long-term trends or seasonal variations.

For future research, we recommend expanding and refining the analysis performed in this study. This will enhance our comprehension of spatiotemporal crash data patterns in urban areas. Firstly, incorporating additional variables like Average Annual Daily Traffic (AADT) and Vehicle Miles Traveled (VMT) can offer a more comprehensive understanding of factors affecting crash occurrences. This augmented dataset will enable a more robust analysis and the identification of additional significant variables contributing to crash incidents. Secondly, integrating advanced machine learning techniques, such as ensemble methods, can improve the accuracy and predictive capability of the developed STWM. These techniques capture complex nonlinear relationships and interactions between variables that may exist in the crash data. By leveraging these advanced methods, we can enhance the model's ability to predict crash incidents and assess associated risks accurately. Furthermore, exploring the adaptability of the developed STWM to other urban areas or different transportation modes, like pedestrian or bicycle networks, can expand its applicability. By conducting these proposed investigations, we can advance the field of spatiotemporal crash data analysis and provide valuable insights for guiding effective strategies to enhance urban road safety.

ACKNOWLEDGMENTS

The results presented here were (partially) achieved by computations carried out on the cluster system at the Leibniz University of Hannover, Germany. Open access publishing facilitated by University of New South Wales, as part of the Wiley &hyphen; University of New South Wales agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

There are no relevant financial or non-financial conflicts of interest to report.

DATA AVAILABILITY STATEMENT

Most of the data that support the findings of this study are openly available in Analyze Boston at https://data.boston.gov/. Besides, Road Network Information and Weather Data are openly available at https://koordinates.com/layer/96131-boston-massachusetts-street-edges/ and https://visual-crossing-weather.p.rapidapi.com/history, respectively.

ORCID

Mohammad Taleai https://orcid.org/0000-0002-8419-4425

Monika Sester https://orcid.org/0000-0002-6656-8809

REFERENCES

Abokifa, A. A., & Sela, L. (2019). Identification of spatial patterns in water distribution pipe failure data using spatial autocorrelation analysis. *Journal of Water Resources Planning and Management*, *145*(12), 04019057. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001135

Alarifi, S. A., Abdel-Aty, M. A., Lee, J., & Wang, X. (2018). Exploring the effect of different neighboring structures on spatial hierarchical joint crash frequency models. *Transportation Research Record*, *2672*(38), 210–222. https://doi.org/10.1177/0361198118776759

Alkahtani, K. F., Abdel-Aty, M., & Lee, J. (2019). A zonal level safety investigation of pedestrian crashes in Riyadh, Saudi Arabia. *International Journal of Sustainable Transportation*, *13*(4), 255–267. https://doi.org/10.1080/15568318.2018.1463417

Almasi, S. A., & Behnood, H. R. (2022). Exposure based geographic analysis mode for estimating the expected pedestrian crash frequency in urban traffic zones; case study of Tehran. *Accident Analysis & Prevention*, *168*, 106576. https://doi.org/10.1016/j.aap.2022.106576

Alvarez-Socorro, A., Herrera-Almarza, G., & González-Díaz, L. (2015). Eigencentrality based on dissimilarity measures reveals central nodes in complex networks. *Scientific Reports*, *5*(1), 1–10. https://doi.org/10.1038/srep17095

Alves, P. J., Emanuel, L., & Pereira, R. H. (2021). Highway concessions and road safety: Evidence from Brazil. *Research in Transportation Economics*, *90*, 101118. https://doi.org/10.1016/j.retrec.2021.101118

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Antczak, E. (2018). Building W matrices using selected geostatistical tools: Empirical examination and application. *Stats*, *1*(1), 112–133. https://doi.org/10.3390/stats1010009

Bamford, C. G., & Robinson, H. (1978). *Geography of transport*. Macdonald and Evans.

Barua, S., El-Basyouny, K., & Islam, M. T. (2015). Effects of spatial correlation in random parameters collision count-data models. *Analytic Methods in Accident Research.*, *5*, 28–42. https://doi.org/10.1016/j.amar.2015.02.001

Billé, A. G., Blasques, F., & Catania, L. (2020). *Dynamic spatial autoregressive models with time-varying spatial weighting matrices*. SSRN 3241470. https://doi.org/10.2139/ssrn.3241470

Blazquez, C. A., Picarte, B., Calderón, J. F., & Losada, F. (2018). Spatial autocorrelation analysis of cargo trucks on highway crashes in Chile. *Accident Analysis & Prevention*, *120*, 195–210. https://doi.org/10.1016/j.aap.2018.08.022

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, *92*(5), 1170–1182. https://doi.org/10.1086/228631

Cardillo, A., Scellato, S., Latora, V., & Porta, S. (2006). Structural properties of planar graphs of urban street patterns. *Physical Review E*, *73*(6), 066107. https://doi.org/10.1103/PhysRevE.73.066107

Case, A. C., Rosen, H. S., & Hines, J. R., Jr. (1993). Budget spillovers and fiscal policy interdependence: Evidence from the states. *Journal of Public Economics*, *52*(3), 285–307. https://doi.org/10.1016/0047-2727(93)90036-S

Cheung, Y.-W., & Lai, K. S. (1995). Lag order and critical values of the augmented Dickey–Fuller test. *Journal of Business & Economic Statistics*, *13*(3), 277–280. https://doi.org/10.2307/1392187

Chun, Y. (2014). Analyzing space–time crime incidents using eigenvector spatial filtering: An application to vehicle burglary. *Geographical Analysis*, *46*(2), 165–184. https://doi.org/10.1111/gean.12034

Cliff, A., & Ord, K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis*, *4*(3), 267–284. https://doi.org/10.1111/j.1538-4632.1972.tb00475.x

Cliff, A. D., & Ord, J. K. (1968). *The problem of spatial autocorrelation*. University.

Cohn, M. J., & Jackman, S. P. (2011). A comparison of aspatial and spatial measures of segregation. *Transactions in GIS*, *15*, 47–66. https://doi.org/10.1111/j.1467-9671.2011.01271.x

Corder, G. W., & Foreman, D. I. (2011). *Nonparametric statistics for non-statisticians*. John Wiley & Sons, Inc. https://doi.org/10.2307/27919868

Corpas-Burgos, F., & Martinez-Beneito, M. A. (2020). On the use of adaptive spatial weight matrices from disease mapping multivariate analyses. *Stochastic Environmental Research and Risk Assessment*, *34*(3), 531–544. https://doi.org/10.1007/s00477-020-01781-5

De Knegt, H., van Langevelde, F. V., Coughenour, M., Skidmore, A., De Boer, W., Heitkönig, I., Knox, N., Slotow, R., Van der Waal, C., & Prins, H. (2010). Spatial autocorrelation and the scaling of species–environment relationships. *Ecology*, *91*(8), 2455–2465. https://doi.org/10.1890/09-1359.1

Feizizadeh, B., Omarzadeh, D., Sharifi, A., Rahmani, A., Lakes, T., & Blaschke, T. (2022). A GIS-based spatiotemporal modelling of urban traffic accidents in Tabriz City during the COVID-19 pandemic. *Sustainability*, *14*(12), 7468. https://doi.org/10.3390/su14127468

Freeman, L. C. (2002). Centrality in social networks: Conceptual clarification. *Social Network: Critical Concepts in Sociology. Londres: Routledge*, *1*, 238–263. https://doi.org/10.1016/0378-8733(78)90021-7

Getis, A. (2008). A history of the concept of spatial autocorrelation: A geographer's perspective. *Geographical Analysis*, *40*(3), 297–309. https://doi.org/10.1111/j.1538-4632.2008.00727.x

Getis, A., & Aldstadt, J. (2004). Constructing the spatial weights matrix using a local statistic. *Geographical Analysis*, *36*(2), 90–104. https://doi.org/10.1111/j.1538-4632.2004.tb01127.x

Gilardi, A., Borgoni, R., Presicce, L., & Mateu, J. (2023). Measurement error models for spatial network lattice data: Analysis of car crashes in Leeds. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *186*(3), 313–334. https://doi.org/10.1093/jrsssa/qnad057

Gilardi, A., Mateu, J., Borgoni, R., & Lovelace, R. (2022). Multivariate hierarchical analysis of car crashes data considering a spatial network lattice. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(3), 1150–1177. https://doi.org/10.1111/rssa.12823

Griffith, D. A. (2003). *Spatial autocorrelation and spatial filtering. Gaining understanding through theory and visualization*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-24806-4

Griffith, D. A. (2021). Interpreting Moran eigenvector maps with the Getis-Ord Gi* statistic. *The Professional Geographer*, *73*(3), 447–463. https://doi.org/10.1080/00330124.2021.1878908

Griffith, D. A., & Chun, Y. (2012). Spatial autocorrelation and eigenvector spatial filtering. In *Handbook of regional science*. Springer. https://doi.org/10.1007/978-3-642-23430-9_72

Guo, D., & Wang, H. (2011). Automatic region building for spatial analysis. *Transactions in GIS*, *15*, 29–45. https://doi.org/10.1111/j.1467-9671.2011.01269.x

Guo, F., Wang, X., & Abdel-Aty, M. A. (2010). Modeling signalized intersection safety with corridor-level spatial correlations. *Accident Analysis & Prevention*, *42*(1), 84–92. https://doi.org/10.1016/j.aap.2009.07.005

Halleck Vega, S., & Elhorst, J. P. (2015). The SLX model. *Journal of Regional Science*, *55*(3), 339–363. https://doi.org/10.1111/jors.12188

Harizi, R., Ouni, F., & M'raihi, R. (2016). Detection and classification of road accident black zones using exploratory spatial data techniques. *International Journal of Trend in Research and Development*, *3*(1), 173–180.

Hasan, A. S., Orvin, M. M., Jalayer, M., Heitmann, E., & Weiss, J. (2022). Analysis of distracted driving crashes in New Jersey using mixed logit model. *Journal of Safety Research*, *81*, 166–174. https://doi.org/10.1016/j.jsr.2022.02.008

Hines, J., & Rosen, H. S. (1993). Budget spillovers and fiscal policy interdependence. *Journal of Public Economics*, *52*, 285–307. https://doi.org/10.1016/0047-2727(93)90036-S

Huang, H., Chang, F., Zhou, H., & Lee, J. (2019). Modeling unobserved heterogeneity for zonal crash frequencies: A Bayesian multivariate random-parameters model with mixture components for spatially correlated data. *Analytic Methods in Accident Research.*, *24*, 100105. https://doi.org/10.1016/j.amar.2019.100105

Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., & Abdel-Aty, M. (2016). Macro and micro models for zonal crash prediction with application in hot zones identification. *Journal of Transport Geography*, *54*, 248–256. https://doi.org/10.1016/j.jtrangeo.2016.06.012

Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New Phytologist*, *11*(2), 37–50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x

Jacquez, G., & Oden, N. (1994). *User manual for STAT: Statistical software for the clustering of health events*. Biomedware. https://doi.org/10.1002/(sici)1097-0258(19960415)15:7/9<951::aid-sim265>3.0.co;2-0

Kelejian, H. H., & Piras, G. (2014). Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. *Regional Science and Urban Economics*, *46*, 140–149. https://doi.org/10.1016/j.regsciurbeco.2014.03.001

Kooijman, S. (1976). Some remarks on the statistical analysis of grids especially with respect to ecology. In *Annals of systems research* (pp. 113–132). Springer. https://doi.org/10.1007/978-1-4613-4243-4

Kuhn, M., & Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*. CRC Press. https://doi.org/10.1080/00031305.2020.1790217

Lee Rodgers, J., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, *42*(1), 59–66. https://doi.org/10.1080/00031305.1988.10475524

Li, X., Goldberg, D. W., Chu, T., & Ma, A. (2019). Enhancing driving safety: Discovering individualized hazardous driving scenes using GIS and mobile sensing. *Transactions in GIS*, *23*(3), 538–557. https://doi.org/10.1111/tgis.12540

Liu, J., Khattak, A. J., & Wali, B. (2017). Do safety performance functions used for predicting crash frequency vary across space? Applying geographically weighted regressions to account for spatial heterogeneity. *Accident Analysis & Prevention*, *109*, 132–142. https://doi.org/10.1016/j.aap.2017.10.012

Ma, X., Chen, S., & Chen, F. (2017). Multivariate space-time modeling of crash frequencies by injury severity levels. *Analytic Methods in Accident Research.*, *15*, 29–40. https://doi.org/10.1016/j.amar.2017.06.001

Mahmud, S. S., Ferreira, L., Hoque, M. S., & Tavassoli, A. (2019). Micro-level safety risk assessment model for a two-lane heterogeneous traffic environment in a developing country: A comparative crash probability modeling approach. *Journal of Safety Research*, *69*, 125–134. https://doi.org/10.1016/j.jsr.2019.03.008

Malczewski, J. (2000). On the use of weighted linear combination method in GIS: Common and best practice approaches. *Transactions in GIS*, *4*(1), 5–22. https://doi.org/10.1111/1467-9671.00035

Mawarni, M., & Machdi, I. (2016). *Dynamic nearest neighbours for generating spatial weight matrix*. International Conference on Advanced Computer Science and Information Systems (ICACSIS), Malang, Indonesia, 2016.

Merk, M. S., & Otto, P. (2020). Estimation of anisotropic, time-varying spatial spillovers of fine particulate matter due to wind direction. *Geographical Analysis*, *52*(2), 254–277. https://doi.org/10.1111/gean.12205

Mohamed, M. G., Saunier, N., Miranda-Moreno, L. F., & Ukkusuri, S. V. (2013). A clustering regression approach: A comprehensive injury severity analysis of pedestrian–vehicle crashes in New York, US and Montreal, Canada. *Safety Science*, *54*, 27–37. https://doi.org/10.1016/j.ssci.2012.11.001

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1/2), 17–23. https://doi.org/10.2307/2332142

Murakami, D. (2017). spmoran: An R package for Moran's eigenvector-based spatial regression analysis. *arXiv preprint arXiv:1703.04467.*

Murakami, D. (2020). Spatial regression using the spmoran package: Boston housing price data examples. *arXivLabs*. https://doi.org/10.48550/arXiv.1703.04467

Murakami, D., & Griffith, D. A. (2020). Balancing spatial and non-spatial variation in varying coefficient modeling: A remedy for spurious correlation. *arXiv preprint arXiv:2005.09981* https://doi.org/10.48550/arXiv.2005.09981

Newman, M. E. (2008). The mathematics of networks. *The New Palgrave Encyclopedia of Economics*, *2*(2008), 1–12. https://doi.org/10.1057/978-1-349-95121-5_2565-1

Olubusoye, O. E., & Salisu, A. A. (2016). Modelling road traffic crashes using spatial autoregressive model with additional endogenous variable. *Statistics*, *17*, 659–670. https://doi.org/10.21307/stattrans-2016-045

Pljakić, M., Jovanović, D., Matović, B., & Mićić, S. (2019). Macro-level accident modeling in Novi Sad: A spatial regression approach. *Accident Analysis & Prevention*, *132*, 105259. https://doi.org/10.1016/j.aap.2019.105259

Qu, X., & Lee, L.-F. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, *184*(2), 209–232. https://doi.org/10.1016/j.jeconom.2014.08.008

Qu, X., Lee, L.-F., & Yang, C. (2021). Estimation of a SAR model with endogenous spatial weights constructed by bilateral variables. *Journal of Econometrics*, *221*(1), 180–197. https://doi.org/10.1016/j.jeconom.2020.05.011

Qu, X., Lee, L.-F., & Yu, J. (2017). QML estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices. *Journal of Econometrics*, *197*(2), 173–201. https://doi.org/10.1016/j.jeconom.2016.11.004

Raftery, A. E., & Banfield, J. D. (1991). Stopping the Gibbs sampler, the use of morphology, and other issues in spatial statistics (Bayesian image restoration, with two applications in spatial statistics)—(Discussion). *Annals of the Institute of Statistical Mathematics*, *43*(1), 32–43. https://doi.org/10.1007/BF00116466

Rahman, M. A., Das, S., & Sun, X. (2023). Understanding the drowsy driving crash patterns from correspondence regression analysis. *Journal of Safety Research*, *84*, 167–181. https://doi.org/10.1016/j.jsr.2022.10.017

Retting, R. A., Williams, A., Farmer, C. M., & Feldman, A. F. (1999). Evaluation of red light camera enforcement in Fairfax, VA, USA. *ITE Journal*, *69*, 30–35. https://doi.org/10.1016/S0001-4575(98)00059-1

Rings, T., Bröhl, T., & Lehnertz, K. (2022). Network structure from a characterization of interactions in complex systems. *Scientific Reports*, *12*(1), 11742. https://doi.org/10.1038/s41598-022-14397-2

Rodriguez Rangel, M. C., & Sanchez Rivero, M. (2020). Spatial imbalance between tourist supply and demand: The identification of spatial clusters in Extremadura, Spain. *Sustainability*, *12*(4), 1651. https://doi.org/10.3390/su12041651

Sandoval-Pineda, A., Pedraza, C., & Darghan, A. E. (2022). Macroscopic spatial analysis of the impact of socioeconomic, land use and mobility factors on the frequency of traffic accidents in Bogotá. *Computers*, *11*(12), 180. https://doi.org/10.3390/computers11120180

Shang, W.-L., Chen, Y., Bi, H., Zhang, H., Ma, C., & Ochieng, W. Y. (2020). Statistical characteristics and community analysis of urban road networks. *Complexity*, *2020*, 1–21. https://doi.org/10.1155/2020/6025821

Shariat-Mohaymany, A., Shahri, M., Mirbagheri, B., & Matkan, A. A. (2015). Exploring spatial non-stationarity and varying relationships between crash data and related factors using geographically weighted Poisson regression. *Transactions in GIS*, *19*(2), 321–337. https://doi.org/10.1111/tgis.12107

Soltani, A., & Askari, S. (2017). Exploring spatial autocorrelation of traffic crashes based on severity. *Injury*, *48*(3), 637–647. https://doi.org/10.1016/j.injury.2017.01.032

Song, L., Li, Y., Fan, W. D., & Wu, P. (2020). Modeling pedestrian-injury severities in pedestrian-vehicle crashes considering spatiotemporal patterns: Insights from different hierarchical Bayesian random-effects models. *Analytic Methods in Accident Research.*, *28*, 100137. https://doi.org/10.1016/j.amar.2020.100137

Sun, J., Li, T., Li, F., & Chen, F. (2016). Analysis of safety factors for urban expressways considering the effect of congestion in Shanghai, China. *Accident Analysis & Prevention*, *95*, 503–511. https://doi.org/10.1016/j.aap.2015.12.011

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*(Suppl. 1), 234–240. https://doi.org/10.2307/143141

Wang, S., Chen, Y., Huang, J., Chen, N., & Lu, Y. (2019). Macrolevel traffic crash analysis: A spatial econometric model approach. *Mathematical Problems in Engineering*, *2019*, 1–10. https://doi.org/10.1155/2019/5306247

Wang, W., Yuan, Z., Yang, Y., Yang, X., & Liu, Y. (2019). Factors influencing traffic accident frequencies on urban roads: A spatial panel time-fixed effects error model. *PLoS ONE*, *14*(4), e0214539. https://doi.org/10.1371/journal.pone.0214539

Wang, X., Yang, J., Lee, C., Ji, Z., & You, S. (2016). Macro-level safety analysis of pedestrian crashes in Shanghai, China. *Accident Analysis & Prevention*, *96*, 12–21. https://doi.org/10.1016/j.aap.2016.07.028

Wen, H., Zhang, X., Zeng, Q., Lee, J., & Yuan, Q. (2019). Investigating spatial autocorrelation and spillover effects in freeway crash-frequency data. *International Journal of Environmental Research and Public Health*, *16*(2), 219. https://doi.org/10.3390/ijerph16020219

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, *41*(3/4), 434–449. https://doi.org/10.2307/2332724

WHO. (2016). *Global status report on road safety*. Retrieved January 3, from https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

WHO. (2018). *10 facts about road safety*. Retrieved January 2, from https://www.who.int/news-room/facts-in-pictures/detail/road-safety

Wu, D., Zhang, Y., & Xiang, Q. (2024). Geographically weighted random forests for macro-level crash frequency prediction. *Accident Analysis & Prevention*, *194*, 107370. https://doi.org/10.1016/j.aap.2023.107370

Xie, K., Wang, X., Huang, H., & Chen, X. (2013). Corridor-level signalized intersection safety analysis in Shanghai, China using Bayesian hierarchical models. *Accident Analysis & Prevention*, *50*, 25–33. https://doi.org/10.1016/j.aap.2012.10.003

Xie, Z., & Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems*, *32*(5), 396–406. https://doi.org/10.1016/j.compenvurbsys.2008.05.001

Xiong, Z., Zhang, R., & Wu, W. (2023). Mechanism analysis of traffic accident prone points based on the spatial Durbin model. *Highlights in Science, Engineering and Technology*, *44*, 103–112. https://doi.org/10.54097/hset.v44i.7272

Xu, C., Ding, Z., Wang, C., & Li, Z. (2019). Statistical analysis of the patterns and characteristics of connected and autonomous vehicle involved crashes. *Journal of Safety Research*, *71*, 41–47. https://doi.org/10.1016/j.jsr.2019.09.001

Zhang, H., & Wang, X. (2017). Combined asymmetric spatial weights matrix with application to housing prices. *Journal of Applied Statistics*, *44*(13), 2337–2353. https://doi.org/10.1080/02664763.2016.1254163

Zhang, J., Qu, X., & Yu, J. (2021). Spatial dynamic panel data models with high order time varying endogenous weights matrices. SSRN 3923264 https://doi.org/10.2139/ssrn.4422160

Zheng, A., & Casari, A. (2018). *Feature engineering for machine learning: Principles and techniques for data scientists*. O'Reilly Media, Inc.

Zhou, Y., He, Z., Chen, J.-Y., Ni, L., & Dong, J. (2022). Investigating travel flow differences between peak hours with spatial model with endogenous weight matrix using automatic vehicle identification data. *Journal of Advanced Transportation*, *2022*, 1–26. https://doi.org/10.1155/2022/7729068

Ziakopoulos, A., & Yannis, G. (2020). A review of spatial approaches in road safety. *Accident Analysis & Prevention*, *135*, 105323. https://doi.org/10.1016/j.aap.2019.105323

How to cite this article: Mohammadi, R., Taleai, M., Otto, P., & Sester, M. (2024). Analyzing urban crash incidents: An advanced endogenous approach using spatiotemporal weights matrix. *Transactions in GIS*, *28*, 368–410. https://doi.org/10.1111/tgis.13138