

Grouped random parameters Poisson-Lindley model with spatial effects addressing crashes at intersections: Insights from visual environment features and spatiotemporal instability

Chenzhu Wang, Mohamed Abdel-Aty, Lei Han *

Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, FL 32816, United States



ARTICLE INFO

Keywords:

Intersection safety
Crash frequency modeling
Drivers' Visual Environment
Grouped Random Parameters
Spatiotemporal variation

ABSTRACT

This study investigates the unobserved heterogeneity and spatiotemporal variations in the effects of visual environment features on intersection crash frequency. A Grouped Random Parameters Poisson-Lindley model with Spatial Effects is developed to account for spatial variations at both the macro (county) and micro (intersection) levels. The analysis utilizes crash data from 2,044 intersections across 12 Florida counties, collected between 2020 and 2022, along with explanatory variables including traffic flow, geometric design characteristics, and visual environment features (extracted from Google Street View images). Comparing to existing methods (e.g., Fixed, Random Parameters, and Grouped Random Parameters Poisson-Lindley models), the proposed approach, which incorporates both macro- and micro-level spatial effects, demonstrates significantly improved model performance. Additionally, the temporal variations of explanatory variables over the three-year period are clearly identified through out-of-sample predictions and marginal effects analysis. Two visual environment features, Vegetation and Grass, result in the identification of grouped random parameters, highlighting the varying impact of these features on intersection crash frequency across the 12 counties. The findings also reveal a strengthening of micro-level spatial effects, indicating heightened spatial correlations between adjacent intersections following the COVID-19 pandemic. Key factors influencing crash frequency include traffic volume, four-legged intersections, major roads with more than four lanes, wider minor roads, and a higher proportion of vehicles in the drivers' field of vision. These results provide valuable insights into the influence of drivers' visual environment on intersection safety and offer policy recommendations for enhancing traffic safety.

1. Introduction

As critical nodes in transportation network, intersections are characterized by the convergence of multiple streams of traffic including vehicles, pedestrians, cyclists (Cai et al., 2018). The complex nature of intersections, involving various maneuvers including turning, crossing and merging, inherently creates numerous traffic conflicts (Yuan and Abdel-Aty, 2018). Consequently, intersections are frequently exposed to a disproportionately high incidence of traffic crashes. Over half of the combined total of fatal and injury crashes occurred at or near intersections (FHWA, 2023). Understanding the dynamics of intersection crashes is essential for developing

* Corresponding author.

E-mail addresses: chenzhu.wang@ucf.edu (C. Wang), M.aty@ucf.edu (M. Abdel-Aty), le966091@ucf.edu (L. Han).

targeted interventions that enhance safety and reduce frequency of these incidents.

Recent studies have highlighted a diverse range of factors influencing intersection safety, including roadway geometric factors (Yuan and Abdel-Aty, 2018), traffic volume/composition (Huang et al., 2017), signal control devices (Abdel-Aty and Wang, 2006), socioeconomic features (Megat-Johari et al., 2018; Su et al., 2021). While existing research efforts primarily account for static infrastructure and traffic conditions, the crucial influence of drivers' visual perception of the surrounding driving environment is frequently neglected. During road navigation, the visual context plays a fundamental role in shaping driver behavior, which subsequently has a direct impact on safety outcomes (Abdel-Aty et al., 2024; Cai et al., 2020). As such, Yu et al. (2019) examined the relationship between driver's visual environment and speeding behaviors by extracting semantic and distance features. Their findings indicate that the drivers are more likely to speed in open areas devoid of trees and buildings. Cai et al. (2022) introduced several visual measures derived from Google Street View (GSV) images to capture the drivers' visual environment and assessed its impact on speeding crash frequency along urban arterials. Their analysis revealed that an increased proportion of trees in drivers' visual fields could mitigate speeding crashes, whereas a higher level of visual complexity was associated with a rise in such incidents. These findings underscore the critical role of incorporating drivers' visual environment features into crash analysis and the development of traffic safety policies. Given the complex interweaving of multiple roads and the presence of diverse traffic participants, the visual environment at intersections becomes increasingly intricate, exerting a substantial influence on drivers' behavior and safety outcomes. Despite this, no studies to date have specifically examined the impact of drivers' visual environment features on intersection safety, highlighting a critical area that warrants further investigation.

The ongoing methodological advancements in crash occurrence research primarily concentrate on capturing unobserved heterogeneity and incorporating spatial and temporal patterns affecting the occurrence of crashes (Mannering and Bhat, 2014; Mannering et al., 2016). Temporal/spatial instability may be driven by variations in driver behaviors as well as broader trends related to economic and macro-environmental factors (Mannering et al., 2018). Neglecting these fundamental temporal/spatial elements could cause biased interpretation and erroneous conclusions.

By allowing parameters to vary across observations, random parameters models are extensively employed and have demonstrated superior performance in addressing unobserved heterogeneity in crash analysis (Fountas et al., 2018a; Hou et al., 2022; Wang et al., 2022a; Wang et al., 2024). Among these, the random parameters Poisson/Negative Binomial models are among the most widely utilized models for count data (Alnawmasi and Mannering, 2022; Coruh et al., 2015; Wang et al., 2020). In addition to models addressing unobserved heterogeneity, recent statistical approaches such as the Zero-inflated Negative Binomial (Anastasopoulos, 2016), Poisson-Gamma (Yu and Abdel-Aty, 2013), and Poisson-Lognormal models (El-Basyouny et al., 2014) have been employed to tackle issues like over-dispersion and excess zero observations in crash data. Although these alternative frameworks do not explicitly account for unobserved heterogeneity, they provide valuable methodologies for analyzing complex crash datasets with distinctive distributional challenges.

To simultaneously address multiple data characteristics—such as overdispersion, correlations among dependent variables, the prevalence of zeros, and unobserved heterogeneity—recent studies have developed models incorporating mixed distributions within a same framework (Islam et al., 2022a; Lord and Geedipally, 2011). As such, Geedipally et al. (2012) employed a Negative Binomial generalized linear model with Lindley mixed effects (NB-L GLM) to analyze traffic crash data, providing a comparison with traditional negative binomial and zero-inflated models. To address underlying heterogeneity and excess over-dispersion, Shaon et al. (2018) proposed a random parameters Negative Binomial-Lindley model, offering enhanced insights into contributing factors with superior goodness-of-fit. In the research of Khodadadi et al. (2022), an Empirical Bayesian framework was developed for the Negative Binomial Lindley model, demonstrating comparable precision to Full Bayesian (FB) estimates while significantly reducing computational costs. Moreover, Tahir et al. (2024) proposed a Poisson lognormal-Lindley model for crash count by accounting for the overdispersion by the extra Poisson variation parameter of the lognormal distribution.

Grouped random parameters have been developed in recent studies to address unobserved heterogeneity stemming from spatial or temporal shifts across different groups of observations (Heydari et al., 2018). These models exhibit superior capabilities in addressing unobserved heterogeneity and yield more precise estimations of model coefficients. Numerous studies have leveraged grouped random parameters models to conduct a robust analysis of crash frequency, highlighting their efficacy in capturing underlying complexities in the data. For instance, Cai et al. (2018) adopted a grouped random parameters spatial model to examine the zonal level crash count data by introducing traffic data and socio-demographic information. Fountas et al. (2018b) proposed a dynamic correlated grouped random parameters binary logit approach explored the mixed effects of variables with time varying or non-time varying features on crash occurrence. Furthermore, Islam et al. (2023) introduced a grouped random parameters Negative Binomial-Lindley model to estimate crash data, demonstrating a significantly superior fit compared to conventional models.

Mutually connected by road segments, adjacent intersections may share unobserved factors related to traffic characteristics, environmental conditions, and weather patterns, which are likely to induce spatial correlations (Ziakopoulos and Yannis, 2020). Previous research has introduced various spatial weight structures such as Spatial Autoregressive (Bhat et al., 2017), Geographically Weighted Regression models (Tang et al., 2020) for modeling relationships between road segments, including those based on adjacency and distance, to effectively capture the spatial correlations among road entities. The Conditional Autoregressive (CAR) prior structure is widely adopted within the Bayesian framework to address spatial dependency in random effects (Yang et al., 2021; Zeng et al., 2019). By leveraging information from spatially related observations, such as crashes occurring at segments or intersections, the CAR prior enhances the precision of estimates (Zeng et al., 2023). Additionally, by clustering geographically proximate regions, road segments, or intersections with similar characteristics, this structure could effectively account for spatial clustering (Ziakopoulos and Yannis, 2020).

Given this, the current paper is motivated to address the research gaps in four key areas:

- a) Propose a Grouped Random Parameters Poisson-Lindley model with Spatial Effects regarding intersections to account for unobserved heterogeneity and spatial correlations in crash occurred at intersections.
- b) Estimate both macro spatial effects at the county level and micro spatial effects at the intersection level to provide accurate performance for intersection crash modeling.
- c) Examine the spatiotemporal instability of model parameters across counties and three-year intervals, and micro spatial variations across intersections through intersection-level spatial effects, with temporal instability using out-of-sample predictions.
- d) Assess the effects of drivers' visual environment features along with multiple geometric design, traffic, environment and socio-economic features on crash occurrence at intersections and propose effective countermeasures.

The study begins by reviewing previous research on methodological approaches, spatial heterogeneities, and contributing factors to intersection crashes. Following this, a detailed description of the dataset and the methodological approaches employed is provided. The results of the model estimation are then discussed, focusing on addressing spatiotemporal aspects within the crash count model. Finally, conclusions are drawn, offering interpretations of the findings and their implications, along with recommendations for future research directions.

2. Literature review

2.1. Review of contributing factors affecting intersection safety

Recently, an extensive body of literature has explored the contributing factors on occurrence of crashes at intersections, in terms of roadway geometric, traffic volume, signal control, socioeconomic and environmental characteristics (Lord and Manning, 2010). Recent research has uncovered important insights into the factors influencing intersection safety. For instance, Huang et al. (2017) developed a multivariate spatial model to analyze crashes in Hillsborough, Florida, and found that incorporating multivariate heterogeneous effects improved the precision of crash frequency estimates for each crash type. Islam et al. (2022b) proposed hierarchical crash frequency models to examine the contributing factors of right-turn crashes. Their findings indicate that the number of right-turn crashes at signalized intersections is likely to be lower at four-legged intersections compared to three-legged ones. Additionally, variables such as traffic volume (Dong et al., 2014), AADT (Guo et al., 2010), signal control mode (Kabir et al., 2021), and population characteristics (Cai et al., 2018) have also been observed to influence crash frequency at intersections in recent research efforts.

In addition to static geometry, traffic, and environmental conditions, drivers' visual perception of the surrounding environment significantly affects driving behavior, and consequently, driving safety (Abdel-Aty et al., 2024; Anciaes, 2023). Visual features play a critical role in shaping driving behaviors and have a substantial impact on road safety. For instance, drivers' speed is observed to be highly correlated to visual road geometry and visual roadside environment (Yu et al., 2019). Drivers are more likely to engage in speeding in open areas devoid of trees and buildings (Yu et al., 2019). Kwon and Cho (2020) used street view images to extract features of the built environment and explore factors influencing children's perceived crash risk. They found that a higher proportion of visible sky significantly reduced perceived crash risk in urban areas. Cai et al. (2022) developed several visual metrics using Google Street View (GSV) images to capture the visual environment and assess its impact on speeding crash frequency on urban arterials. Their findings indicate that a higher proportion of trees in a driver's view may reduce speeding crashes, potentially due to both the visual effect of calming drivers and the physical presence of trees as roadside obstacles, which may encourage more cautious driving behavior.

These studies underscore the importance of incorporating drivers' visual environment features into crash analysis and the development of traffic safety policies. At intersections, this issue becomes even more pronounced. As vehicle speeds decrease, drivers' perception of the surrounding environment changes significantly (Bella, 2013). The complex visual environment at intersections, characterized by multi-lane configurations and the presence of diverse traffic participants, can profoundly influence driver behavior and safety (Li et al., 2024).

Given the limited research on the effects of drivers' visual environments on intersection safety, addressing this gap through further investigation is essential. Understanding the impact of visual environments at intersections will offer valuable insights into the underlying factors contributing to crashes, facilitating the development of more effective safety measures to mitigate risks and improve overall traffic safety at intersections.

2.2. Review of methodological approaches in crash frequency analysis

A variety of methodological approaches have been developed and applied over the years to model crash frequency and address the complexity of traffic environments (Washington et al., 2020). Poisson models are one of the most common approaches used in crash frequency analysis, while the crash frequency is assumed to allow a Poisson distribution, where the mean and variance are equal. By introducing an additional parameter to account for over-dispersion and skewness, the negative binomial model is developed to address the over dispersion (Mohammadi et al., 2014). To capture the unobserved heterogeneity, random parameters (Xie et al., 2014; Bhowmik et al., 2019), grouped random parameters (Cai et al., 2018), correlated random parameters (Kabir et al., 2021), negative binomial-Lindley distribution (Geedipally et al., 2012; Dzinyela et al., 2024; Tahir et al., 2024), Bayesian estimation methods (Yuan and Abdel-Aty, 2018) have been proposed in recent studies.

Specifically, the Bayesian framework offers a flexible approach for estimating complex models by incorporating prior information and updating it as new data becomes available. Bayesian hierarchical models are especially valuable when modeling random

parameters and temporal/spatial effects (Cui and Xie, 2021), as they facilitate the estimation of multi-level structures within crash data (Cai et al., 2023). As such, Ma et al. (2017) developed a Bayesian multivariate space-time model to estimate crash frequencies by different injury severity levels, which is observed to outperform multivariate random effects and multivariate spatial models. Using hierarchical Bayesian methods, Song et al. (2020) employed a model that allows all parameters to vary randomly across observations or spatiotemporal groups, significantly improving model fit and prediction accuracy in estimating the spatiotemporal features of pedestrian-vehicle crashes. Gurumurthy et al. (2022) developed a Bayesian Hierarchical model to estimate animal-vehicle collisions, demonstrating its scalability for large datasets containing key explanatory factors with segment-specific factors.

Additionally, grouped random parameters models are a special case of random parameters approach addressing the spatial and temporal variations in different groups of observations in data (Heydari et al., 2018). Such model is observed to demonstrate superiority in capturing spatial heterogenous effects across areas compared to traditional models (Islam et al., 2023). For instance, Ali et al. (2022) and Song et al. (2022) both highlighted the superiority of grouped random parameters in capturing unobserved heterogeneity. Cai et al. (2018) explored the observed and unobserved effects of zone based on grouped random parameters multivariate spatial model. Meng et al. (2021) proposed correlated grouped random parameters binary logit models with space-time variations to estimate the injury severity of non-truck-involved crashes, finding that the model effectively addressed space-time-varying heterogeneities. Intini et al. (2020) proposed grouped random parameters multinomial logit approaches at the segment/intersection levels, demonstrating their applicability in identifying roadway sites with anomalous tendencies or high-risk locations. Islam et al. (2023) developed a grouped random parameters Negative Binomial-Lindley model to account for regional variations in the contributing factors affecting lane departure crashes, including deicing use, terrain, pavement condition, and driver behaviors that are difficult to observe.

Thus, a grouped random parameters model within the Bayesian framework offers a robust approach for capturing unobserved heterogeneities and addressing spatiotemporal variations in crash frequency analysis. This combination not only enhances the accuracy of the findings but also provides deeper insights into the underlying factors contributing to crash occurrences, allowing for more nuanced interpretations and explanations.

2.3. Review of spatiotemporal heterogeneities in crash frequency analysis

Nowadays, growing evidence suggests that both spatial and temporal heterogeneities—the variations in crash occurrences across different geographic locations and time periods—play a significant role in shaping crash risks (Mannering, 2018; Wang et al., 2022b). Neglecting these variations can result in biased estimates and ineffective safety interventions (Mannering et al., 2016; Mannering, 2018).

To date, temporal instability has been well-documented in numerous studies, with several contributing factors showing variability across different time periods (Mannering, 2018; Song et al., 2020; Wang et al., 2024). To test the temporal instability, pairwise and global likelihood ratio tests are widely estimated (Alnawmasi and Mannering, 2019; Fountas et al., 2020; Washington et al., 2020). Otherwise, out-of-sample prediction is another adequate method used to estimate the probability differences (Alnawmasi and Mannering, 2022; Hou et al., 2022). All of these methods have been proven effective in testing temporal instability.

Spatial heterogeneities refer to the geographic variability in crash frequency resulting from differences in road characteristics (Mohammadnazir et al., 2021), environmental factors (Liu et al., 2017), and driver behavior across regions (Ghasemzadeh and Ahmed, 2019). Roadway segments, intersections, and neighborhoods often exhibit distinct patterns of crash occurrences due to localized factors, which must be accounted for in crash models. Road geometry (e.g., intersections, curves, lane widths) and road types (e.g., rural vs. urban roads, highways vs. local roads) significantly influence crash risks (Cafiso et al., 2010; Bella and Silvestri, 2017). For example, urban intersections tend to experience higher rates of pedestrian crashes (Murphy et al., 2017), whereas there is significant differences in the crash frequency among urban and rural highways (Khorashadi et al., 2005; Champahom et al., 2020). Additionally, regions with harsh weather conditions (e.g., snow, ice, heavy rainfall) frequently see elevated crash rates during certain seasons (Andrey, 2010; Dey et al., 2014), while coastal areas may exhibit unique crash patterns due to fog or rain (Jackson and Sharif, 2016).

Additionally, spatial heterogeneity also arises from spatial dependency (Ziakopoulos and Yannis, 2020)—the concept that crash occurrences at one location may be influenced by those at neighboring locations (Xu and Huang, 2015). This spatial autocorrelation underscores the fact that crashes are not spatially independent (Saha et al., 2018). For example, crash-prone intersections in one area may elevate the likelihood of crashes at nearby intersections due to spillover effects (Alhomaidat et al., 2020), such as increased traffic congestion or suboptimal road design (Atumo et al., 2023).

To account for spatial heterogeneities, several advanced modeling approaches have been developed, such as grouped random parameters (Islam et al., 2023), conditional autoregressive (Saha et al., 2018; Zeng et al., 2019), geographically weighted regression (Hezaveh et al., 2019; Huang et al., 2018), hierarchical Bayesian models (Haque et al., 2010; Li et al., 2018; Song et al., 2020).

Overall, drivers' visual environment features, along with variables such as geometric design, traffic conditions, environmental factors, and socioeconomic characteristics, significantly influence crash occurrences at intersections. To account for unobserved and spatiotemporal heterogeneity, this paper develops Grouped Random Parameters Poisson-Lindley models with Spatial Effects that address spatial heterogeneities at both macro and micro levels. Grouped random parameters are incorporated at the county level, while spatial effects are captured at the intersection level. Furthermore, the spatiotemporal instability of the Lindley parameters is examined, spatial variations across intersections are estimated through intersection-level spatial effects, and temporal instability is assessed using out-of-sample predictions.

3. Methodology

This section documents the derivations of the Grouped Random Parameters Poisson-Lindley model with Spatial Effects (G-RPP-LS) and its characteristics in modeling intersection crash data. To better explain the formulation process, brief discussions on the typical Random Parameters Poisson-Lindley (RPP-L) and the Grouped Random Parameters Poisson-Lindley (G-RPP-L) model are first introduced.

3.1. Random parameters Poisson-Lindley model

Before introducing Random Parameters Poisson-Lindley model, let us start with the brief formulations of the fixed parameters Poisson model. In the fixed parameters Poisson model, the crash frequency that occurs on the i -th intersection is assumed to follow a Poisson distribution:

$$y_i|\lambda_i \sim \text{Poisson}(\lambda_i) \quad (1-1)$$

$$\ln(\lambda_i|\beta_0, \beta_1, \dots, \beta_M) = \beta_0 + \sum_{j=1}^M \beta_j X_{ij} \quad (1-2)$$

where λ_i is the expected Poisson mean for i -th intersection; X_{ij} denotes the j -th variables at i -th intersection; M is total number of model variables; β_0 is the intercept and β_j is the corresponding j -th model coefficients.

The fixed parameters model, however, fail to adequately account for unobserved heterogeneity (Alnawmasi and Mannerling, 2022; Islam et al., 2023; Mannerling et al., 2016). To address such issue, Random Parameters models has been proposed to account for unobserved heterogeneity by allowing parameters to vary across observations. In existing crash frequency modeling studies, the Random Parameters Poisson (RPP) model has been one of the typical models to address unobserved heterogeneity (Hou et al., 2021; Tahir et al., 2024). This model allows the coefficients of the fixed parameters Poisson model to vary among different observations. Thus, the Random Parameters Poisson model can be represented as the following hierarchical Bayesian model:

$$y_i|\lambda_i \sim \text{Poisson}(\lambda_i) \quad (2-1)$$

$$\ln(\lambda_i|\beta_0, \beta_1, \dots, \beta_M, \xi_{i1}, \dots, \xi_{iM}) = \beta_0 + \sum_{j=1}^M \beta_j X_{ij} + \sum_{j=1}^{M'} \xi_{ij} Z_{ij} \quad (2-2)$$

$$\xi_{ij}|\mu_j, \sigma_j \sim \text{Normal}(\mu_j, \sigma_j^2) \quad (2-3)$$

where X_{ij} is the j -th fixed parameters variable at i -th intersection; β_j is the fixed coefficient for the j -th fixed parameters variable; M is total number of fixed parameters variables; Z_{ij} is the j -th random parameters variable at i -th intersection; ξ_{ij} is the random coefficient for the j -th random parameters variable at i -th intersection; μ_j and σ_j denote the mean and standard deviation of the j -th random parameters variable, respectively. M' is total number of random parameters variables.

Although the Random Parameters Poisson model can effectively account for the unobserved heterogeneity, it cannot deal with the datasets containing a large amount of zero observations, which is a common issue in crash data (Barua et al., 2016; Mannerling et al., 2016). To overcome such limitation, the Lindley error term was introduced into Random Parameters Poisson model, resulting in the Random Parameters Poisson-Lindley (RPP-L) model, which has demonstrated superior performance. The Random Parameters Poisson-Lindley model is written as a mixture model:

$$P(Y = y_i|\lambda_i, \theta) = \int \text{Poisson}(y_i|\lambda_i e_i) \text{Lindley}(e_i|\theta) de_i \quad (3)$$

According to Zamani and Ismail (2010), the Lindley distribution with parameter θ can be written as a mixture of two gamma distributions:

$$e_i|\theta \sim \frac{1}{1+\theta} \text{gamma}(2, \theta) + \frac{\theta}{1+\theta} \text{gamma}(1, \theta) \quad (4)$$

Therefore, the multi-level hierarchical representation of the Random Parameters Poisson-Lindley model can be given as follows (Geedipally et al., 2012):

$$y_i|\lambda_i e_i \sim \text{Poisson}(\lambda_i e_i) \quad (5-1)$$

$$e_i|z_i, \theta \sim \text{gamma}(1 + z_i, \theta) \quad (5-2)$$

$$z_i|\theta \sim \text{Bernoulli}\left(\frac{1}{1+\theta}\right)$$

$$\ln(\lambda_i|\beta_0, \beta_1, \dots, \beta_M, \xi_{i1}, \dots, \xi_{iM}) = \beta_0 + \sum_{j=1}^M \beta_j X_{ij} + \sum_{j=1}^{M'} \xi_{ij} Z_{ij}$$

$$\xi_{ij} | \mu_j, \sigma_j \sim \text{Normal}(\mu_j, \sigma_j^2) \quad (5-5)$$

Where $\theta > 0$ is the Lindley parameter; the other variables retain the same definitions as described above.

3.2. Grouped random parameters Poisson-Lindley model

Researchers have shown that the Random Parameters Poisson-Lindley model provides additional flexibility to capture the unobserved heterogeneity and address the excess number of zero responses problem. However, the coefficients of the covariates may vary from one group of observations to another due to unobserved heterogeneity (Cai et al., 2018; Intini et al., 2020). In addition, the Lindley term may also vary among groups of observations as different regions may include different percentages of zero observations. Hence, different Lindley terms might also be needed to account for the number of zeros in different regions, instead of one unique or universal term (Islam et al., 2023). Likewise, let us assume "K" groups of observations in data. Then, let the symbol $k(i)$ denote the k -th group of observations that the i -th site is associated with (i.e., the i -th intersection is part of the k -th group or region) (Islam et al., 2023). The Grouped Random Parameters Poisson-Lindley model with varying coefficients and Lindley terms across groups can be reformulated from equation (5) following the study of Islam et al. (2023):

$$y_i | \lambda_i e_i \sim \text{Poisson}(\lambda_i e_i) \quad (6-1)$$

$$e_i | z_{k(i)}, \theta_{k(i)} \sim \text{gamma}(1 + z_{k(i)}, \theta_{k(i)}) \quad (6-2)$$

$$z_{k(i)} | \theta_{k(i)} \sim \text{Bernoulli}\left(\frac{1}{1 + \theta_{k(i)}}\right) \quad (6-3)$$

$$\ln(\lambda_i | \beta_0, \beta_1, \dots, \beta_M, \xi_{i1,k(i)}, \dots, \xi_{iM,k(i)}) = \beta_0 + \sum_{j=1}^M \beta_j X_{ij} + \sum_{j=1}^M \xi_{ij,k(i)} Z_{ij} \quad (6-4)$$

$$\xi_{ij,k(i)} | \mu_{ij,k(i)}, \sigma_{ij,k(i)} \sim \text{Normal}(\mu_{ij,k(i)}, \sigma_{ij,k(i)}^2) \quad (6-5)$$

where $\theta_{k(i)} > 0$ is the Lindley parameter for the k -th group of observations; $\xi_{ij,k(i)}$ represents the grouped random coefficient for the k -th group of the j -th variable at i -th intersection, and $\mu_{ij,k(i)}$ and $\sigma_{ij,k(i)}$ are the corresponding mean and standard deviation, respectively; The remaining variables follow the same definitions as described above.

3.3. Grouped random parameters Poisson-Lindley model with spatial effects

Although the Grouped Random Parameters models have demonstrated better model fit compared to non-grouped Random Parameters Negative Binomial Lindley models (Dzinyela et al., 2024), it can only capture the unobserved heterogeneity across different groups, which is specifically at the macro-spatial level (e.g., county level, traffic analysis zonal level). Given the traffic connection between intersections, it is believed there is a certain spatial correlation between crash frequency at the adjacent micro interaction level (Cai et al., 2018; Manner and Bhat, 2014). To integrate such micro-level spatial correlation into the Grouped Random Parameters Poisson-Lindley model, the widespread Besag-York-Mollie (BYM) spatial random effects v_i are added to the link function (6-4) to capture the spatial correlation between neighboring intersections (Besag et al., 1991):

$$\ln(\lambda_i | \beta_0, \beta_1, \dots, \beta_M, \xi_{i1,k(i)}, \dots, \xi_{iM,k(i)}, v_i) = \beta_0 + \sum_{j=1}^M \beta_j X_{ij} + \sum_{j=1}^M \xi_{ij,k(i)} Z_{ij} + v_i \quad (7)$$

Following to existing studies (Cai et al., 2022; Cheng et al., 2018; Dong et al., 2016; Wang et al., 2024), v_i is set to follow the conditional autoregressive (CAR) formulation. Thus, the Grouped Random Parameters Poisson-Lindley model with Spatial Effects can be written as:

$$y_i | \lambda_i e_i \sim \text{Poisson}(\lambda_i e_i) \quad (8-1)$$

$$e_i | z_{k(i)}, \theta_{k(i)} \sim \text{gamma}(1 + z_{k(i)}, \theta_{k(i)}) \quad (8-2)$$

$$z_{k(i)} | \theta_{k(i)} \sim \text{Bernoulli}\left(\frac{1}{1 + \theta_{k(i)}}\right) \quad (8-3)$$

$$\ln(\lambda_i | \beta_0, \beta_1, \dots, \beta_M, \xi_{i1,k(i)}, \dots, \xi_{iM,k(i)}, v_i) = \beta_0 + \sum_{j=1}^M \beta_j X_{ij} + \sum_{j=1}^M \xi_{ij,k(i)} Z_{ij} + v_i \quad (8-4)$$

$$\xi_{ij,k(i)} | \mu_{ij,k(i)}, \sigma_{ij,k(i)} \sim \text{Normal}(\mu_{ij,k(i)}, \sigma_{ij,k(i)}^2) \quad (8-5)$$

$$v_i | \tau_i \sim \text{Normal}(\bar{v}_i, \tau_i^2) \quad (8-6)$$

$$\bar{v}_i = \frac{\sum_{i \neq j} v_j^* w_{ij}}{\sum_{i \neq j} w_{ij}} \quad (8-7)$$

$$\tau_i = \frac{\tau_c}{\sum_{i \neq j} w_{ij}} \quad (8-8)$$

where τ_c is precision parameter in CAR prior; w_{ij} is an element in the spatial weight matrix W , representing the spatial weight between intersections i and j ; The remaining variables in Grouped Random Parameters Poisson-Lindley model with Spatial Effects follow the same definitions as described above. For the spatial weight matrix W , existing research has used various spatial weight structures (e.g., adjacency-based in Cheng et al., 2018; Truong et al., 2016, distance-based in Dong et al., 2015; Xie et al., 2014, and parametric geographic weighting models in Tang et al., 2020) to analyze the spatial correlation between adjust units at different scales. At the micro intersection level, the analysis of crash frequency often adopts an adjacency based spatial weight structure (Cai et al., 2018; Munira et al., 2020). Therefore, this study also uses this structure to study spatial correlation as shown in Fig. 1. In this case, if intersection j is adjacent/connected to the target intersection i through an adjacent road section, their spatial weight w_{ij} is set to 1, otherwise the weight between two non-adjacent intersections w_{ij} would be 0.

In the above formulation, the known input parameters include the X_{ij} , Z_{ij} , w_{ij} and y_i . The unknown parameters need to be inferreded include β_0 , $\beta_{j=1,\dots,M}$, $\mu_{j=1,\dots,M,k(i)=1,\dots,K}$, $\sigma_{j=1,\dots,M,k(i)=1,\dots,K}$, τ_c , and $\theta_{k(i)}$. Let us assume a normal prior $N(0, 0.01)$ on fixed parameters coefficients (β), a uniform prior $U(0.01, 1)$ on parameters $1/(1 + 1/\theta_{k(i)})$ and a gamma prior $Gamma(0.5, 0.005)$ on τ_c . For the random parameters, let us assume a normal prior $N(0, 0.01)$ on $\mu_{j,k(i)}$ and a uniform prior $U(0, 5)$ on $\sigma_{j,k(i)}$. Then, these unknown parameters in hierarchical Bayesian G-RPP-LS model can be estimated using the Monte Carlo Markov Chain (MCMC) approach in WinBUGS (Lunn et al., 2000). Referring to existing studies (Chen et al., 2021; Hasan and Abdel-Aty, 2024), a total of 3 Markov chains were considered while estimating these models during the MCMC samplings. For each Markov chain, 30,000 iterations were performed, and the first 5000 samples were the burn-in samples. To ensure the convergence of the estimated models, the Gelman-Rubin statistics threshold value was selected as less than 1.1 and the Monte Carlo error for each estimated parameter was adopted as less than 3 percent of the posterior standard deviation.

3.4. Model performance evaluation

Referring to existing studies (Barua et al., 2016; Hasan and Abdel-Aty, 2024; Tahir et al., 2024), four measures were utilized to comprehensively compare the model goodness of fit in this study:

Deviance Information Criteria (DIC) was considered as a local goodness of fit measure to compare the estimated models. The DIC is the hierarchical modeling generalization of Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) which penalizes the larger parameter model. Mathematically, DIC can be formulated as follows.

$$DIC = D(\bar{\theta}) + 2pD = \bar{D} + pD \quad (9)$$

where $D(\bar{\theta})$ is the deviance using the posterior mean values of the parameters of interest $\bar{\theta}$. \bar{D} is the posterior mean deviance for assessing the model fit, pD is the effective number of parameters in the model. Generally, a model with lower DIC is considered superior among the candidate models. DIC differences between 5 and 10 are considered substantial, while differences more than 10 indicate significant outperformance by the model with a lower DIC.

To evaluate the prediction performance of candidate models, three global goodness of fit measures include the Mean Absolute Error (MAE), and Mean Squared Error (MSE), and Widely Applicable Information Criterion (WAIC). For, the lower the values of MAE, MSE and WAIC indicate a better the model in terms of prediction performance. These three measures can be calculated as follow:

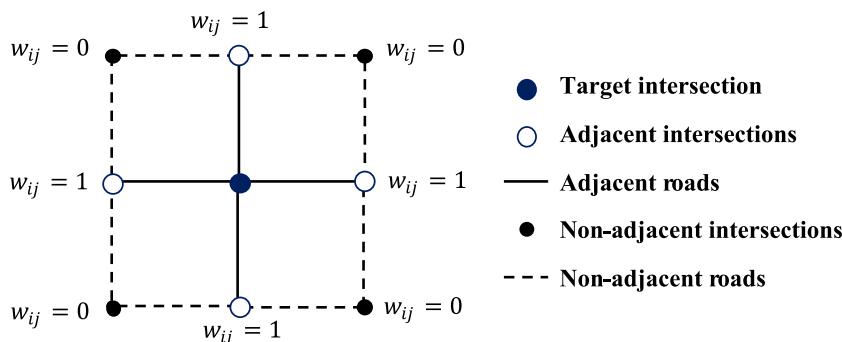


Fig. 1. Definition of the spatial weight of intersections.

$$MAE = \frac{1}{N} \sum |y_i - \hat{y}_i| \quad (10)$$

$$MSE = \frac{1}{N} \sum (y_i - \hat{y}_i)^2 \quad (11)$$

$$WAIC = -2(lppd - p_{waic}) = -2 \left(\sum_{i=1}^N \log \left(\frac{1}{S} \sum_{s=1}^S p(y_i | \theta^{(s)}) \right) - \sum_{i=1}^N Var_{s=1}^S [\log p(y_i | \theta^{(s)})] \right) \quad (12)$$

Here, N represents the number of observations, y_i and \hat{y}_i denote the observed and predicted crash count for intersection i , respectively. \bar{y}_i denote the global mean of crash count for all studied intersections. S denotes the number of posterior samples drawn from the model's prior distribution. $\theta^{(s)}$ means the parameter vector sampled from the posterior distribution at the s -th iteration. $p(y_i | \theta^{(s)})$ represents the likelihood of the i -th intersection y_i given the sampled parameter $\theta^{(s)}$, representing the model's predictive performance for a specific intersection under the sampled parameter. $Var_{s=1}^S$ denotes the variance of the log-likelihood values across the posterior samples, which reflects the uncertainty in the model's predictions for the i -th intersection. Moreover, p_{waic} denotes the effective number of parameters as a penalization of model for its complexity.

4. Data description

In this study, a total of 2044 signalized intersections in 12 counties in Florida (FL) state, USA, were selected as shown in Fig. 2. The counties include Citrus, Hernando, Hillsborough, Lake, Pasco, Pinellas, Polk, Sumter, Brevard, Orange, Seminole, and Osceola.

According to the Florida Traffic Crash Data, although these counties only account for 21.46 % of Florida state's area, they have over 45 % of all intersection and intersection-related crashes in Florida (Table 1). As a result, these counties provide a good real-world example for the application of the proposed model. For each intersection, there is one target variable, i.e., crash frequency and four major categories of explanatory variables: (a) traffic flow features, (b) intersection geometric design, (c) socioeconomic characteristics, and (d) drivers' visual environment features.

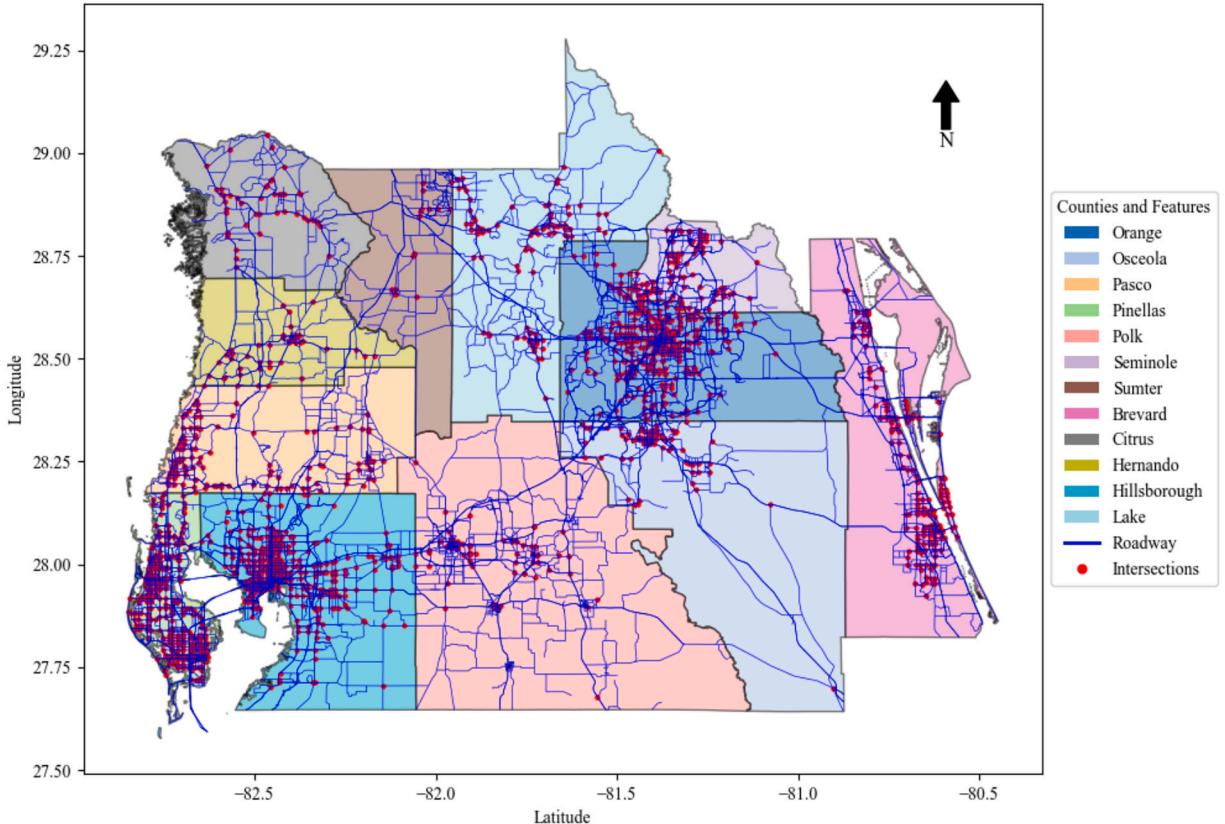


Fig. 2. Studied intersections in 12 FL counties.

Table 1

Counts of intersections and intersection crashes in 12 Florida counties.

County	Area (km ²)	Number of intersections	Number of intersection crashes			
			2020	2021	2022	SUM
Citrus (2)	2122.85	46	465	543	508	1516
Hernando (8)	1652.40	46	579	686	720	1985
Hillsborough (10)	3566.41	473	6807	8136	8728	23,671
Lake (11)	3914.31	94	989	1117	1208	3314
Pasco (14)	2565.23	119	2035	2360	2287	6682
Pinellas (15)	963.01	379	4893	5513	5419	15,825
Polk (16)	6700.29	79	1179	1301	1294	3774
Sumter (18)	1960.95	16	154	165	211	530
Brevard (70)	3536.86	187	2263	2617	2663	7543
Orange (75)	3378.45	420	6648	7160	6715	20,523
Seminole (77)	1165.50	102	1521	1736	1718	4975
Osceola (92)	5029.18	83	1331	1647	1715	4693
Subtotal	36555.44	2044	28,864	32,981	33,186	95,031
Florida	170310	4016	62,270	72,017	72,543	206,830
Proportion (Subtotal/FL)	21.46 %	50.90 %	46.35 %	45.80 %	45.75 %	45.95 %

4.1. Crash data

The crash data, covering the period from 2020 to 2022, were obtained from the Signal Four Analytics (S4A) system of Florida DOT ([S4A, 2014](#)). This online platform includes all crash records in Florida. For each crash, it has the crash time, location, type, severity, number of involved vehicles, and other detailed information. According to the S4A system, crashes that occur within 250 ft of the stop line are defined as “intersection-related crashes”. Thus, within intersection and intersection-related crashes were identified and used in the analysis. [Table 1](#) provides the counts of studied signalized intersections and their corresponding intersection crashes in the 12 Florida counties.

The intersection data and traffic volume information were obtained from the “Intersections” and “Annual Average Daily Traffic” shapefiles in the open-source FDOT [Roadway Characteristics Inventory \(RCI\)](#). This study selected intersections with available traffic volume data for both major and minor roads, resulting in 4016 intersections from whole Florida. These 12 counties account for 21.46 % of Florida’s total area but represent 50.9 % of the intersections in the selected dataset and 45.95 % of intersection-related crashes statewide during 2020–2022. This disproportionality underscores the concentration of crashes and intersections in regions characterized by dense road networks and high traffic volumes.

At the 2,044 intersections in these counties, a total of 95,031 intersection crashes were recorded over the three-year period. Among them, Hillsborough County has the most intersections (473) and, consequently, the highest number of intersection crashes among these counties. Orange County follows with 420 intersections and 20,523 intersection crashes, ranking second. In contrast, Citrus, Hernando, and Sumter counties have the fewest intersections and intersection crashes. Notably, Sumter County has only 16 intersections, resulting in just 530 intersection crashes over the three years.

4.2. Traffic flow and intersection geometric design features

Traffic flow include the traffic volume at each intersection. Existing studies show that they can be represented by the average annual daily traffic (AADT) over three years (2020–2021), including AADT and Truck AADT from major roads and minor roads ([Abdel-Aty and Wang, 2006; Huang et al., 2017; Yuan and Abdel-Aty, 2018](#)). These data were collected from the open-source FDOT [Roadway Characteristics Inventory \(RCI\)](#). The intersection geometric design features can also be obtained from the FDOT RCI system. For example, the posted speed limit is available for each roadway, allowing this information to be matched with the major and minor roadways at the intersections. Finally, a total of 26 geometric design features, which are believed to be related to the crash occurrence ([Gu et al., 2023; Nightingale et al., 2017](#)), were extracted as shown in [Table A1](#).

4.3. Socioeconomic characteristics surrounding intersections

Socioeconomic data reflect the regional economy and population features surrounding the intersections. Multiple socioeconomic features (e.g., population, proportion of poverty, property of unemployed) at census-tract-level were extracted from the dataset of USDOT Census Bureau ([U.S. Census Bureau](#)). Since a single intersection may be near multiple census tracts, a 0.5-mile buffer was created around each intersection, as illustrated in [Fig. 3](#). Socioeconomic features from the census tracts that spatially overlapped with this buffer were aggregated for the intersections. A weighted average was utilized as suggested by existing studies ([Huang et al., 2017](#)). As an example, the average population variable E_i for the intersection buffer i can be calculated:

$$E_i = \sum_j \frac{A_{j,i} * E_j}{\sum_j A_j} \quad (13)$$

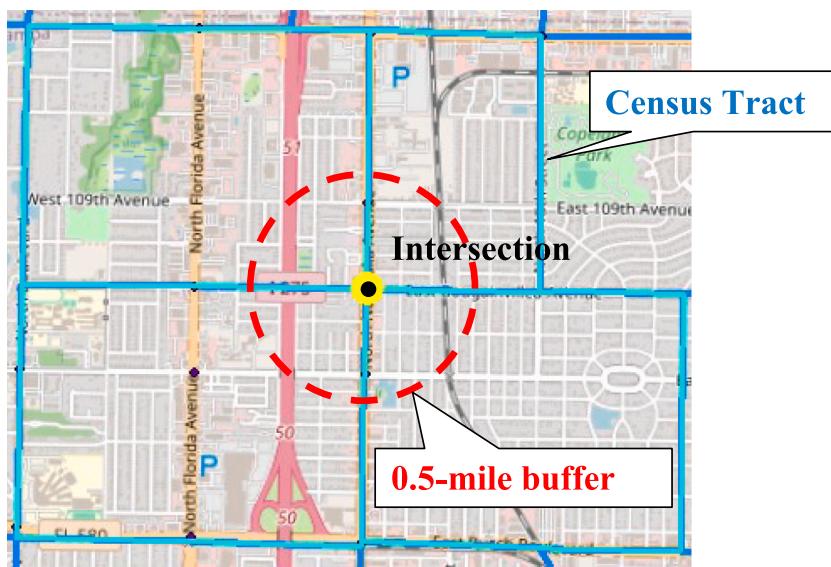


Fig. 3. Spatial overlay of census tracts on a 0.5-mile buffer around an intersection.

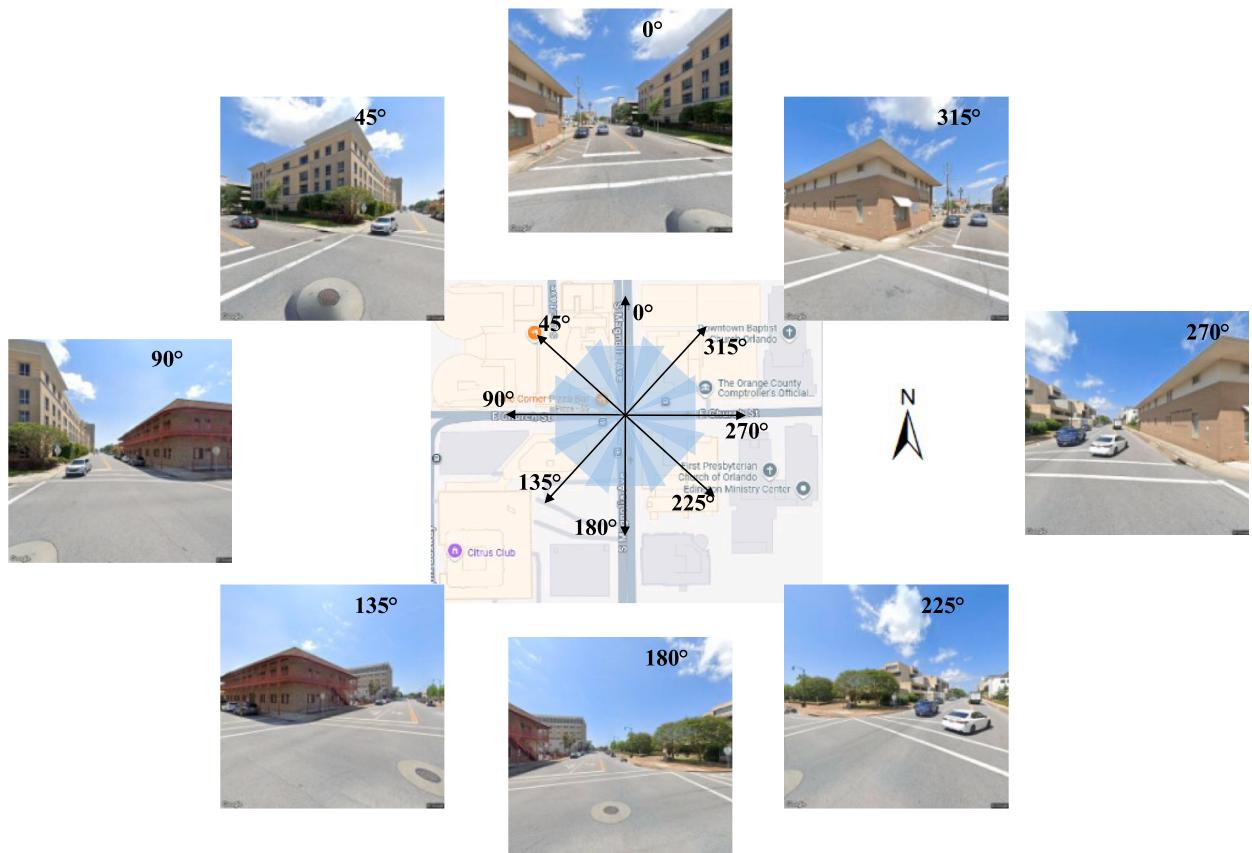


Fig. 4. Illustration of GSV image collection for each intersection.

where E_j is the population of census tract j , $A_{j,i}$ is the area of census tract j within buffer i , and A_j is the area of the census tract j .

4.4. Drivers' visual environment features

Existing studies have shown that the drivers' visual perception of the surrounding driving environment significantly impact the influence drivers' driving behaviors and thus safety (Abdel-Aty et al., 2024; Cai et al., 2020; Fan et al., 2023; Zhao and Khattak, 2018). To extract the drivers' visual environment features, the Google Street View (GSV) images has been widely used to reflect the real-world drivers' visual perception (Cai et al., 2022; Li et al., 2015). Therefore, GSV images were collected to capture the entire intersection environment in this study. Leveraging advanced image segmentation techniques, seven types of objects were identified in the images: sky, road, building, vegetation, grass, vehicle, and walk area. Their pixel proportions were then aggregated as the drivers' visual environment's features.

First, to collect the GSV images of each intersection, a Python script was developed to automatically download these images from the HTTP URL form using the Google API (Google API). For each intersection, its center point was identified as the viewpoint origin, and 8 GSV images form ranges from 0° (north), 45° (northeast), ..., 315° (northwest) were obtained to fully capture the drivers' visual perception, as shown in Fig. 4. It is worth noting that other quantities of GSV images (e.g., 2, 4, 16) were also tested, but 8 images provided the optimal coverage of the intersection without overlaps. To get images similar to the drivers' view, the horizontal field of view and pitch were set as 50° and 0° , respectively (Cai et al., 2022; Li et al., 2015; Yang et al., 2009). Finally, a total of $8 \times 2044 = 16352$ GSV images were collected at studied intersections.

GSV images are sourced from Google's publicly available archive, which may include variations in environmental conditions, such as time of day, weather, or season, at the time of image capture. While this is a limitation inherent to the use of GSV data, previous studies have demonstrated the robustness of GSV images in capturing visual features for transportation and urban studies (Cai et al., 2022; Li et al., 2015). For consistency, all images in this study were taken with a uniform field of view, pitch, and direction settings, and the pixel proportions for each object type were calculated relative to the total number of pixels in the eight images per intersection, minimizing the potential impact of lighting or weather differences.

It is also worth noting that GSV image archives may not fully capture temporal changes in the visual environment, such as newly constructed buildings or vegetation growth. Therefore, we recommend future research using supplementary data sources, such as recent drone imagery or field-collected images, to validate and enhance the analysis. Despite these potential limitations, GSV remains an accessible and widely used tool for researchers to analyze drivers' visual environments across diverse locations. The reliance on

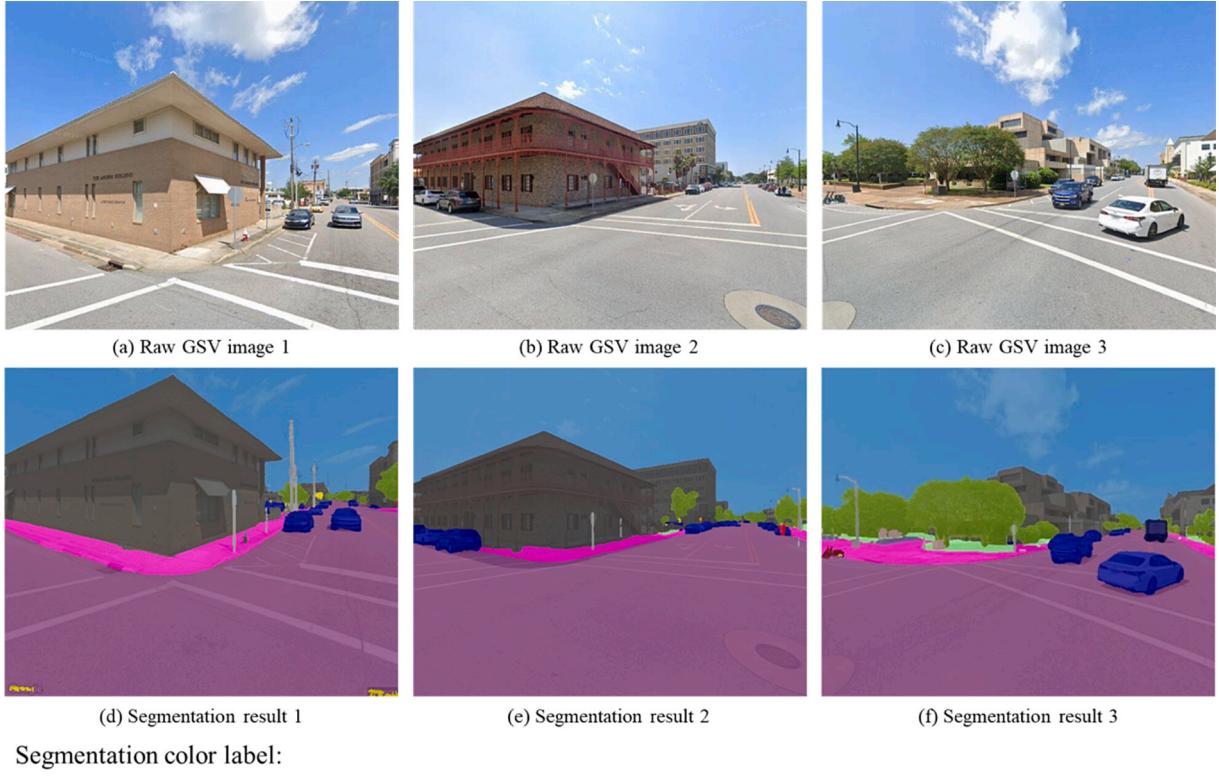


Fig. 5. Illustration of semantic segmentation of intersection GSV images.

publicly available GSV images also makes the methodology replicable for other researchers. This approach allows for broader application in areas where direct field data collection is infeasible, while still providing valuable insights into the visual environments at intersections.

Recently, semantic segmentation of objects from images has been growing in the field of computer vision and multiple deep learning methods have been developed (Dosovitskiy et al., 2021; Long et al., 2015; Strudel et al., 2021). In this study, “Segmenter,” a transformer-based segmentation model, was used to cluster objects from GSV images. Proposed by Strudel et al. (2021), Segmenter can integrate global context information into image segmentation, outperforming convolution-based methods and remaining state-of-the-art. During the image segmentation in the Segmenter model, the raw GSV image is split into a sequence of pixel patches, which are flattened and projected to a series of embeddings. A pre-trained transformer encoder, Vision Transformer (Dosovitskiy et al., 2021), takes the embeddings to extract their contextualized and position features, which are then fed into a transformer decoder to predict their segmentation classes. As shown in Fig. 5, a total of seven types of objects in the drivers' view environment were labeled (with different colors): sky, road, building, vegetation, grass, vehicle, and walk area. Based on the clustering results, the object type by each pixel can be determined in the image. Since each intersection has 8 GSV images, the proportion of object type $Prop_k$ was calculated based on the sum of pixels in the 8 images:

$$Prop_k = \frac{P_k}{\sum_{k=1}^6 P_k} = \frac{\sum_{b=1}^8 P_{k,b}}{\sum_{k=1}^6 \sum_{b=1}^8 P_{k,b}} \quad (14)$$

where $P_{k,b}$ is the number of pixels of object type k in the b -th GSV intersection image. P_k is the total pixel number of object type k in all intersection images. k can be sky, road, buildings, vegetation, vehicle, or walk area. Detailed descriptive statistics of these visual environment features in the modeling data can be seen in Table A1.

5. Results and discussion

5.1. Model evaluation

Before discussing the model results, the models' performances are summarized and presented in Table 2. To assess the goodness-of-fit, Deviance Information Criterion (DIC), Widely Applicable Information Criterion (WAIC), Mean Absolute Error (MAE), Mean Squared Error (MSE), are compared.

Across all three years, the Grouped Random Parameters Poisson-Lindley model with Spatial Effects consistently achieves the lowest WAIC and DIC values, indicating a better overall goodness-of-fit compared to other models. Specifically, in 2022, the DIC for the Grouped Random Parameters Poisson-Lindley model with Spatial Effects is 12,130.9, lower than 12,187.1 for Grouped RPNB-Lindley model. Similarly, its WAIC is 11,780.2, compared to 11,789.3 for Grouped RPNB-Lindley model.

Furthermore, the Grouped Random Parameters Poisson-Lindley model with Spatial Effects demonstrates the lowest MAE and MSE across all years, except for the year 2020,¹ highlighting its superior predictive accuracy and lower residual error. In 2022, it achieves a lower MAE (2.846) and MSE (15.390) compared to the Grouped RPNB-Lindley model, which has a higher MAE (2.854) and MSE (15.710), further confirming its improved predictive performance.

Additionally, the dispersion parameters (α) of Grouped RPNB-Lindley model remain between 0.011 and 0.012 across all three years, which is close to zero, indicating an absence of substantial overdispersion. This suggests that a Poisson-based distribution is a more appropriate choice for modeling the data.

These findings emphasize that incorporating grouped random parameters and spatial effects at both county and intersection levels significantly enhances model performance. Overall, the Grouped Random Parameters Poisson-Lindley model with Spatial Effects demonstrates the best fit and predictive accuracy among all considered models.

In Table 2, the small values of the Lindley parameter (θ) in this study are consistent with findings from previous research. For instance, Khan et al. (2023) employed a Random Parameters Negative Binomial Lindley model and reported a similarly θ value of 0.032 in their study on design consistency and run-off-road crashes. This highlights that small θ values are not unusual in models specifically designed to handle highly dispersed crash data. Furthermore, the Lindley parameter (θ) in the Grouped RPP-Lindley model with Spatial Effects demonstrated a substantial reduction compared to the values observed in the RPP-Lindley and Grouped RPP-Lindley models. For example, the Lindley parameter (θ) for the RPP-Lindley model was 0.794 in 2020, whereas the Grouped RPP-Lindley model with Spatial Effects yielded a considerably lower value of 0.085. The significant reduction indicates that incorporating Grouped Random Parameters and Spatial Effects effectively captures unobserved heterogeneity and spatial dependencies, reducing the need for over-dispersion adjustments and improving model fit.

Despite the θ , our model outperforms alternative approaches, in terms of DIC, WAIC, MAE, and MSE, validating its robustness. Furthermore, concerns regarding potential numerical instability or extreme ϵ values were addressed by rigorous convergence diagnostics. The estimation process successfully converged, as indicated by Gelman-Rubin statistics below the commonly accepted threshold of 1, with no irregularities observed. The trace plots for all key parameters exhibit well-mixed chains with no visible trends or divergence, confirming good mixing and stationarity. These results collectively affirm the reliability of our model in handling the

¹ For the year 2020, the MSE value of the Grouped Random Parameters Poisson-Lindley model with Spatial Effects is 0.02 higher than that of the Grouped Random Parameters Poisson-Lindley model.

Table 2

Comparisons of goodness-of-fit measures for six Poisson models.

Models	Random Parameters Poisson (RPP)	Grouped RPNB ³	RPP-Lindley	Grouped RPP-Lindley	Grouped RPNB-Lindley ³	Grouped RPP-Lindley with Spatial Effects
2020						
DIC	11963.7	11780.5	11822.9	11360.1	11839.8	11355.1
WAIC	11840.7	11636.7	11445.2	11434.8	11455.5	11427.4
MAE	2.825	2.741	2.652	2.648	2.651	2.643
MSE	14.250	14.190	13.620	13.620	13.621	13.640
α^1	—	0.230	—	—	0.012	—
θ^2	—	—	0.794	0.163	0.106	0.085
2021						
DIC	12344.9	12158.3	12208.4	12206.5	12223.0	12148.4
WAIC	12215.7	11973.61	11788.6	11802.0	11814.8	11779.0
MAE	3.043	2.944	2.850	2.848	2.852	2.845
MSE	17.440	16.200	15.540	15.570	15.570	15.510
α^1	—	0.228	—	—	0.011	—
θ^2	—	—	0.327	0.137	0.098	0.073
2022						
DIC	12314.1	12146.3	12188.2	12163.1	12187.1	12130.9
WAIC	12144.9	11964.7	11792.4	11787.1	11789.3	11780.2
MAE	3.036	2.942	2.857	2.852	2.854	2.846
MSE	17.490	16.34	15.810	15.700	15.710	15.390
α^1	—	0.218	—	—	0.011	—
θ^2	—	—	0.232	0.126	0.119	0.086

Note: Bold values indicate a better fit. ¹ denotes the values of dispersion parameters. ² denotes the values of Lindley parameters. ³ The NB (Negative Binomial) models follow the function of Poisson-Gamma distribution (Geedipally et al., 2014).

complexities of the dataset while maintaining numerical stability.

Based on the Grouped Random Parameters Poisson-Lindley model with Spatial Effects, several variables are identified as contributing factors over the three-year period (2020–2022). As shown in Table 3, the variables Log_Major_AADT (The log value of AADT on major road (pcu)), Log_Minor_AADT (The log value of AADT on minor road (pcu)), Legs_4 (4-legged), Major_lanes > 4, Minor_width, Vehicle have positive interactions with intersection crashes, indicating an increase in these variables would cause more crashes. Vegetation (Proportion of vegetation in the driver's vision (%)), Grass (Proportion of grass area in the driver's vision (%)), P_Over65 (Percent of population 65 years or older (%)) have negative effects on crash frequency.

Table 3

Estimates and standard deviations of Grouped Random Parameters Poisson-Lindley model with Spatial Effects.

Variables	2020		2021		2022	
	Mean	S. D. ¹	Mean	S. D. ¹	Mean	S. D. ¹
Intercept	-2.786	0.884	-2.519	0.844	-2.977	0.788
Log_Major_AADT	0.102	0.072	0.085	0.033	0.107	0.031
Log_Minor_AADT	0.121	0.022	0.108	0.027	0.169	0.037
P_Over65	<i>-0.004²</i>	0.003	-0.007	0.003	-0.005	0.002
Legs_4	0.322	0.046	0.259	0.048	0.276	0.050
Major_lanes > 4	0.400	0.088	0.438	0.051	0.402	0.054
Minor_width	0.008	0.002	0.009	0.002	0.005	0.002
Vehicle	0.039	0.011	0.052	0.011	0.043	0.011
Vegetation ³	-0.050	0.008	-0.053	0.009	-0.049	0.009
Std. Dev of Vegetation	0.017	0.004	0.017	0.004	0.015	0.005
Grass ³	-0.077	0.022	-0.066	0.026	-0.083	0.024
Std. Dev of Grass	0.048	0.010	0.039	0.009	0.047	0.012
Lindley Parameter ³ (θ)	0.085	0.017	0.073	0.022	0.086	0.017
Spatial Parameter (τ_c)	0.071	0.019	0.064	0.015	0.053	0.011
Model Performance						
DIC	11355.1		12148.4		12130.9	
WAIC	11447.4		11779.0		11780.2	
MAE	2.643		2.845		2.846	
MSE	13.640		15.510		15.390	

Note: ¹ Standard deviation. ² Italic font indicates that the mean value of the parameter is not statistically significant at the 95% confidence level. ³ Mean of the grouped parameters.

5.2. Heterogeneity and spatiotemporal effects

5.2.1. Macro spatial variations (county level)

Allowing the model parameter to vary across groups of observations enables the Poisson model to better account for spatiotemporal variations and unobserved heterogeneity of different counties. Table 4 shows the values of Lindley parameter (θ) for each group of observations (county), along with the two grouped random parameters including Vegetation and Grass. To establish a better trigger image, Fig. 6 shows the value of the Lindley parameter for each group of observations (county). As shown in this figure, the value of the Lindley parameter varies from one county to another. Specifically, the value of this parameter varies from 0.07 for Brevard to 0.12 for Citrus for year 2020. To provide a clearer representation, Fig. 6 illustrates the value of the Lindley parameter for each group of observations (county). As depicted in this figure, the Lindley parameter varies across counties. Specifically, in 2020, the parameter ranges from 0.07 for Brevard County to 0.12 for Citrus County.

Additionally, a comprehensive evaluation of MCMC sampling configurations confirmed that the statistical insignificance and observed bimodality in the Lindley parameter (θ) for Sumter County are not due to chain convergence issues but rather an inherent characteristic of the dataset. This finding underscores the robustness of our proposed MCMC approach and highlights the importance of accounting for latent data heterogeneity in similar studies.

In addition to spatial variations, temporal shifts are also evident in the Lindley parameters. The parameter values remain relatively stable over the three years in counties such as Sumter, Brevard, and Osceola. In contrast, significant changes are observed in counties like Citrus and Hernando, indicating a certain temporal shift. For example, in Citrus County, the parameter values fluctuate considerably, from 0.124 in 2020 to 0.061 in 2021.

Similarly, the coefficients for Vegetation and Grass exhibit significant variations across the 12 counties and over the 3-year period. For example, the coefficients for Vegetation range from -0.050 to -0.103 across different counties in 2020, while the coefficients for Grass range from -0.033 to -0.120 . Moreover, notable variations are also observed across different years within a single county.

To further compare the spatiotemporal variations of the two visual environment features, their marginal effects are estimated based on the estimated coefficients. Similar to elasticities, which measure the effect of a 1 % change in the independent variable (only appropriate for continuous variables) on the expected crash frequency, marginal effects reflect the effects of a “one unit” change in independent variables on crash frequency, which are easier to interpret compared to elasticities, particularly for models with categorical variables (Washington et al., 2020). The marginal effects of variable x_i can be calculated by $\partial\lambda_i/\partial x_i$. Although the marginal effects can be computed for all the observations, it is common to present the average marginal effects of the overall observation populations. In this study, the marginal effects (ME) were the average effects of a unit increase in an independent variable on the crash frequency on an intersection as shown in Table 5.

Fig. 7 shows marginal effects of the Vegetations (Proportion of vegetation in the driver's vision (%)) across 12 different Florida

Table 4

The estimates and standard deviations of Grouped Random Parameters Poisson-Lindley model with Spatial Effects for visual environment features.¹

Florida counties	Lindley Parameter			Grouped random parameters					
				Vegetation			Grass		
	2020	2021	2022	2020	2021	2022	2020	2021	2022
Citrus	0.124 (0.048)	0.061 (0.026)	0.093 (0.035)	<i>-0.019</i> ² (0.021)	-0.060 (0.021)	-0.064 (0.024)	-0.020 (0.055)	-0.040 (0.023)	0.012 (0.051)
Hernando	0.083 (0.025)	0.075 (0.026)	0.116 (0.037)	<i>-0.008</i> (0.017)	-0.011 (0.012)	-0.026 (0.017)	-0.120 (0.058)	-0.100 (0.053)	0.091 (0.073)
Hillsborough	0.092 (0.018)	0.078 (0.029)	0.080 (0.019)	<i>-0.053</i> (0.008)	-0.047 (0.006)	-0.040 (0.007)	-0.033 (0.033)	<i>-0.017</i> (0.014)	-0.060 (0.032)
Lake	0.075 (0.019)	0.092 (0.029)	0.062 (0.019)	<i>-0.078</i> (0.017)	-0.041 (0.017)	-0.084 (0.012)	0.007 (0.049)	<i>-0.016</i> (0.032)	-0.006 (0.055)
Pasco	0.062 (0.018)	0.055 (0.016)	0.066 (0.013)	<i>-0.103</i> (0.015)	-0.102 (0.015)	-0.069 (0.020)	-0.073 (0.049)	<i>-0.047</i> (0.045)	-0.172 (0.037)
Pinellas	0.096 (0.023)	0.092 (0.023)	0.115 (0.023)	<i>-0.050</i> (0.006)	-0.031 (0.008)	-0.031 (0.005)	-0.111 (0.020)	-0.116 (0.025)	-0.146 (0.027)
Polk	0.094 (0.033)	0.058 (0.030)	0.084 (0.020)	<i>-0.004</i> (0.026)	-0.085 (0.027)	-0.065 (0.020)	-0.113 (0.044)	-0.117 (0.060)	-0.068 (0.023)
Sumter	0.060 (0.050)	0.064 (0.044)	0.064 (0.057)	<i>-0.057</i> (0.048)	-0.103 (0.038)	-0.015 (0.036)	-0.182 (0.112)	0.013 (0.068)	-0.353 (0.095)
Brevard	0.068 (0.014)	0.064 (0.028)	0.066 (0.014)	<i>-0.090</i> (0.012)	-0.055 (0.017)	-0.070 (0.011)	-0.077 (0.029)	-0.093 (0.031)	-0.080 (0.026)
Orange	0.083 (0.015)	0.073 (0.026)	0.089 (0.017)	<i>-0.037</i> (0.009)	-0.043 (0.008)	-0.042 (0.007)	-0.100 (0.047)	-0.104 (0.027)	-0.093 (0.044)
Seminole	0.068 (0.019)	0.060 (0.025)	0.084 (0.022)	<i>-0.054</i> (0.011)	-0.049 (0.010)	-0.046 (0.011)	-0.062 (0.028)	-0.081 (0.039)	-0.062 (0.032)
Osceola	0.115 (0.038)	0.109 (0.042)	0.110 (0.035)	<i>-0.046</i> (0.015)	-0.014 (0.021)	-0.038 (0.016)	-0.044 (0.056)	<i>-0.081</i> (0.054)	-0.066 (0.062)

Note: ¹ Standard deviations are shown in parenthesis. ² Italic font indicates that the mean value of the parameter is not statistically significant at the 95% confidence level.

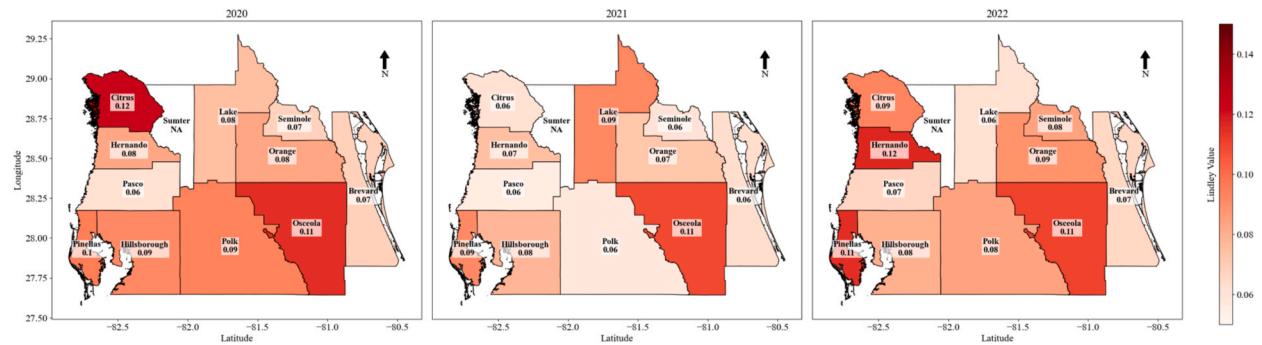


Fig. 6. Variations of the Lindley parameters across 12 different Florida counties.

Table 5
Marginal effects of grouped random parameters across 12 Florida counties.

Florida counties	Vegetation			Grass		
	2020	2021	2022	2020	2021	2022
Citrus	-0.195	-0.705	-0.706	-0.198	-0.468	0.135
Hernando	-0.096	-0.161	-0.413	-1.515	-1.458	1.429
Hillsborough	-0.763	-0.813	-0.741	-0.482	-0.290	-1.098
Lake	-0.826	-0.492	-1.081	0.074	-0.188	-0.075
Pasco	-1.762	-2.021	-1.330	-1.244	-0.932	-3.306
Pinellas	-0.652	-0.455	-0.442	-1.440	-1.695	-2.092
Polk	-0.065	-1.399	-1.063	-1.684	-1.934	-1.109
Sumter	-0.544	-1.057	-0.188	-1.745	0.129	-4.373
Brevard	-1.085	-0.775	-1.001	-0.932	-1.308	-1.144
Orange	-0.587	-0.727	-0.679	-1.581	-1.773	-1.482
Seminole	-0.798	-0.835	-0.775	-0.923	-1.385	-1.039
Osceola	-0.736	-0.278	-0.776	-0.701	-1.600	-1.358

Note: Italic font indicates that the mean value of the parameter is not statistically significant at the 95% confidence level.

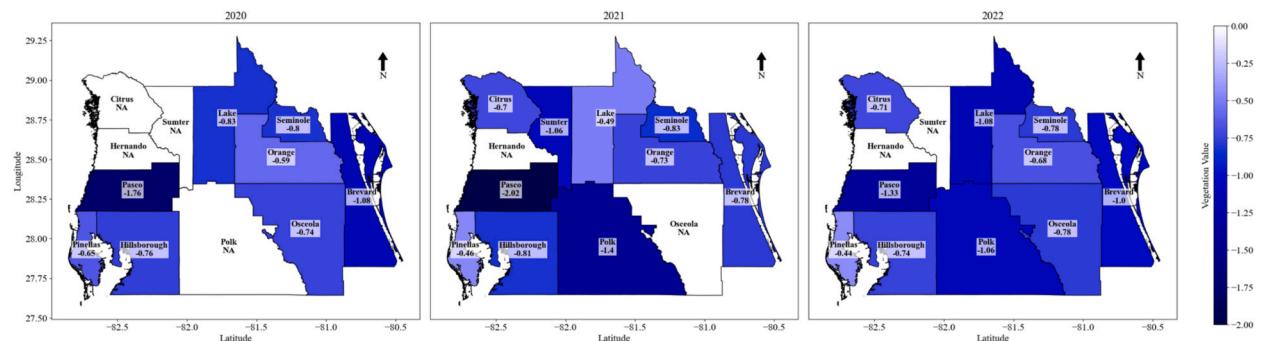


Fig. 7. Marginal effects of the Vegetations (Proportion of vegetation in the driver's vision (%)) across 12 different Florida counties.

counties. The results show the effect of this variables is the greatest for Pasco County, with changes across three years. In Pasco County, vegetation may function as a natural buffer between roadways and potential hazards, such as pedestrians, cyclists, or roadside objects (Khattak et al., 2021). This buffering effect could contribute to the greater reduction in crashes observed in Pasco compared to other counties, where such a buffer may be less prominent or effective.

Fig. 8 illustrates the marginal effects of Grass (the proportion of grass area in the driver's vision) across 12 different Florida counties. The impact of this variable varies across counties and over different time periods. The value for Polk County is the highest in 2020 and 2021, whereas Pasco County shows the highest value in 2022. Thus, the Grass proportion results in a higher reduction in crash frequency within such dimensions. Both Polk and Pasco counties may have implemented effective roadside grass management strategies, especially for intersections. Besides, both counties could have more rural or suburban roads where the visual environment, including grass, plays a larger role in influencing driver behavior compared to more urban areas (Hanson et al., 2013). Roads with grassy medians or shoulders provide greater visibility, more space for vehicles to recover from potential errors, and a calming driving environment, all of which could contribute to a higher reduction in crash frequency (Abdel-Aty et al., 2024).

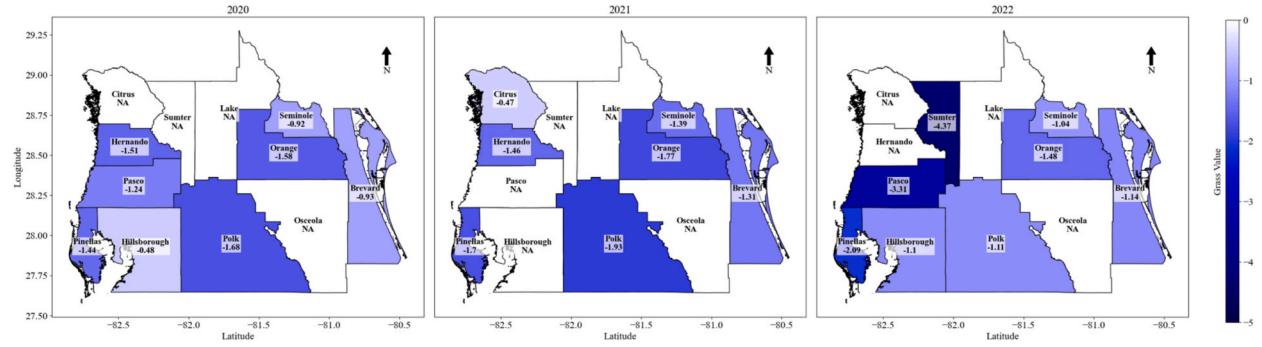


Fig. 8. Marginal effects of the Grass (Proportion of grass area in the driver's vision (%)) across 12 different Florida counties.

Thus, significant macro spatial variations are observed through the identification of grouped random parameters, revealing the varying effects of vegetation and grass on crash frequency across different counties.

5.2.2. Micro spatial effects (intersection level)

Other than macro spatial variations captured by Lindley parameters and grouped random parameters, the micro spatial effects are also addressed by estimating spatial errors of intersections. Fig. 9 presents the spatial errors of the 2,044 intersections across three years in 12 Florida counties. Specifically, the color of each intersection represents the magnitude of its spatial error. Specifically, red indicates positive values, meaning that the adjacent intersections' features contribute to an increase in crashes at the intersection, with darker shades of red reflecting larger values. Conversely, blue indicates negative values, reflecting the impact of neighboring intersections in reducing crash frequency, with darker shades of blue corresponding to smaller values. As shown in Fig. 9, varying spatial effects are observed among the 2,044 intersections, indicating that each intersection is influenced by its surrounding intersections to different degrees. Moreover, most intersections still experience significant fluctuations in spatial errors from 2020 to 2022 (e.g., shifting from positive (red) to negative (blue) and vice versa). This suggests that the spatial effects are not static but change from a temporal perspective.

To further investigate the spatiotemporal variation of spatial effects, Fig. 10 illustrates the spatial errors of intersections over three years in three typical counties: Hillsborough, Orange, and Pinellas. Each sub-figure includes a zoom-in box to highlight the spatial effects of specific intersections within their urban areas. The absolute value of the spatial errors can be used to compare the mutual spatial influence between adjacent intersections, with the depth of color indicating the strength of this influence. Based on Fig. 10, we can find that:

- For the urban intersections in Hillsborough County (zoom-in boxes in Fig. 10 (a)), the spatial errors at northern intersections decrease from over 0.4 to below 0.1, indicating a gradual weakening of the connections between such intersections over the three-year period. On the contrary, the absolute value of spatial errors at southern intersections (closer to the downtown area) increase significantly from 0 in 2020 to over 0.2 in 2022, suggesting that the spatial correlations between adjacent intersections are becoming more pronounced.
- In Orange County (zoom-in boxes in Fig. 10 (b)), the absolute value of spatial errors at the majority of urban intersections increase from 2020 to 2022. Particularly in the high-density northern intersections, the spatial errors rise from 0 to approximately 0.5. This clearly indicates that the spatial impacts and connections between surrounding intersections become more pronounced over the three years.
- Additionally, a similar pattern can be observed in Pinellas County, as shown in the zoom-in boxes in Fig. 10 (c). The spatial errors of all urban intersections increase from 0.2 to 0.4–0.5, reflecting the strengthening connections of these intersections within road networks.

Overall, although some spatial errors decrease, the majority of intersections exhibit increasing absolute values of spatial errors, indicating stronger correlations and enhanced connections between adjacent intersections following the COVID-19 pandemic. A possible explanation for this is that, compared to during the pandemic (2020), there was a significant increase in travel activities (Yang et al., 2023) and higher crash rates (Dong et al., 2022) after the pandemic (2021 and 2022). As a result, under these conditions of higher traffic volumes, intersections are more likely to be influenced by surrounding roads and nearby intersections, leading to stronger spatial connections.

Thus, spatial heterogeneity at the intersection level is addressed by incorporating spatial effects in the Grouped Random Parameters Poisson-Lindley models, while the potential relationship between spatial effects at the macro level (county) and the micro level (intersections) is explored.

5.2.3. Out-of-sample prediction and marginal effects

Out-of-sample prediction is a powerful method for testing temporal instability (Alnawmasi and Mannering, 2022; Hou et al., 2022),

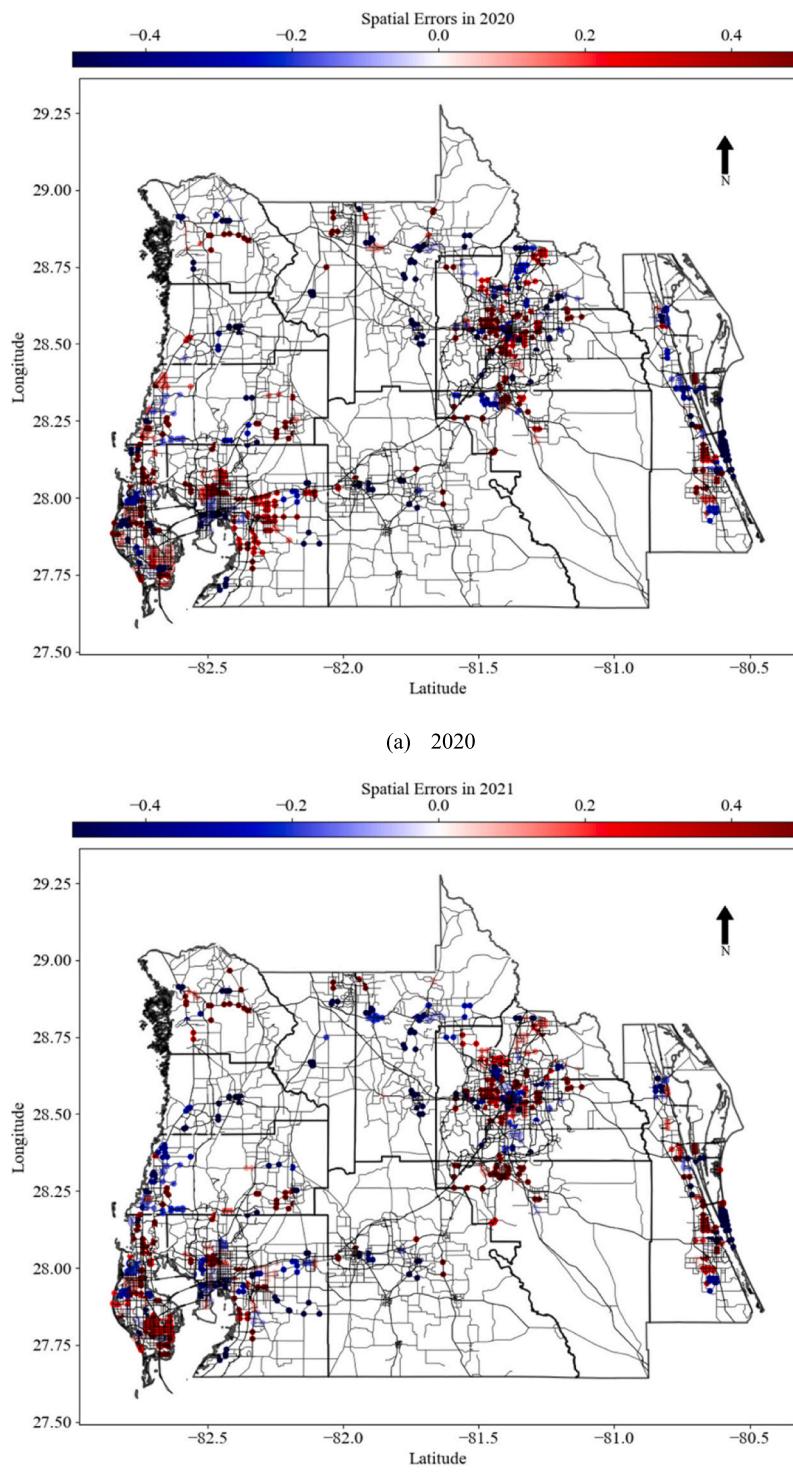


Fig. 9. Spatial errors of intersections across three years in 12 different FL counties.

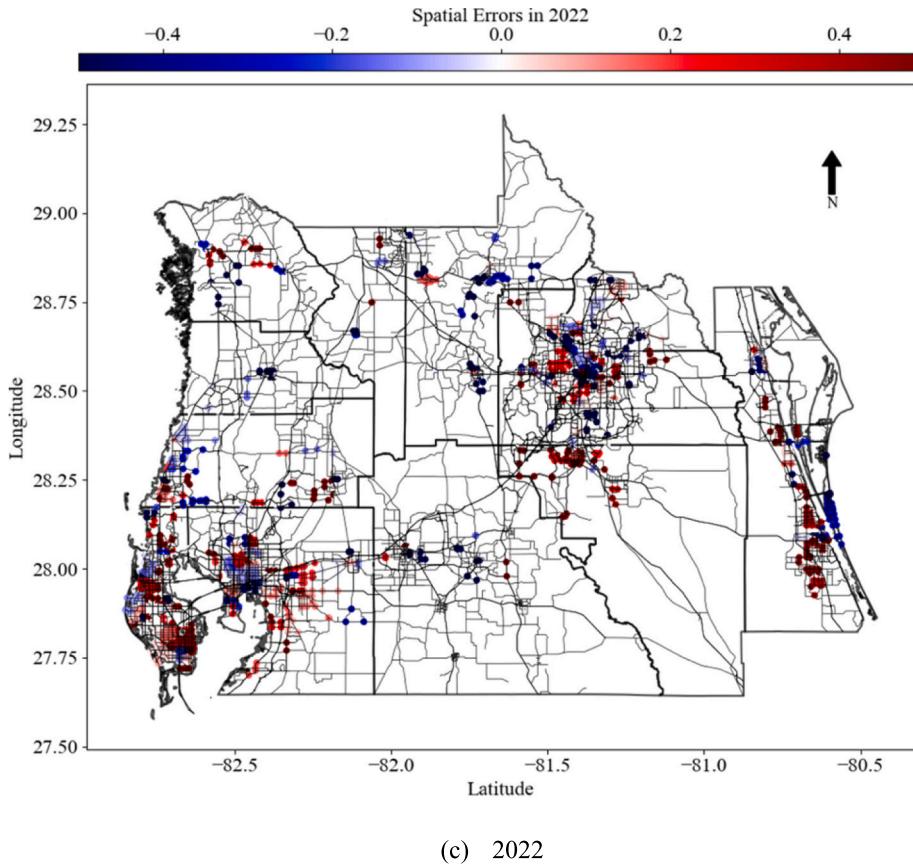


Fig. 9. (continued).

particularly when analyzing how effects of variables change over time (Wang et al., 2024). To test the temporal instability, a model is typically trained on data from one time period (in-sample data) and then used to predict the outcomes for another time period (out-of-sample data) (Alnawasi and Mannering, 2022).

Fig. A1 illustrates the predictions based on the parameters of the Grouped Random Parameters Poisson-Lindley models with Spatial Effects. Fig. A1 (a) shows the predicted crashes using the model parameters from 2020 to predict the 2020 crash data, which is also referred to as Absolute Error. In this figure, values below the 45-degree line represent over-predictions, while values above the line represent under-predictions. Fig. A1 (b) indicates the predicted crashes for 2021 using the 2020 model with the predicted crashes for 2020. Specifically, the mean difference in the predicted values for the 2,044 intersections is 0.069 (2021 values minus 2020 values), with a standard deviation of 0.781 in the comparison between 2020 and 2021, as shown in Table 6. The comparison reveals a significant mean difference, particularly between the during-pandemic period (2020) and the post-pandemic period (2021 and 2022), with differences exceeding 0.06. Besides, the high standard deviations show significant volatility among the predicted values and the original values.

Given this, the findings in Table 6 indicate the temporal variations in predicted crash frequencies of the 2044 intersections among the three years.

Table 7 presents the average marginal effects of significant variables in Grouped Random Parameters Poisson-Lindley model with Spatial effects. However, the effects of some contributing factors vary significantly across the three years. For instance, the Log_Minor_AADT increases the crash frequency by around 1.7 for year 2020 and 2021, while increases that by 2.739 in year 2022. Similarly, the Minor_width and Vehicle also show substantial changes across the three years.

Overall, significant temporal variations are observed through out-of-sample predictions and the presentation of marginal effects over the three-year period.

5.3. Parameter estimation and interpretation

5.3.1. Traffic flow volume

With respect to traffic flow, the average annual daily traffic volumes on both major and minor roads are positively correlated with intersection crash frequency. This finding is consistent with the expectation that higher traffic volumes increase exposure to risky vehicle interactions, thereby leading to more crashes at intersections (Wang and Abdel-Aty, 2008; Cai et al., 2018; Wang et al., 2020).

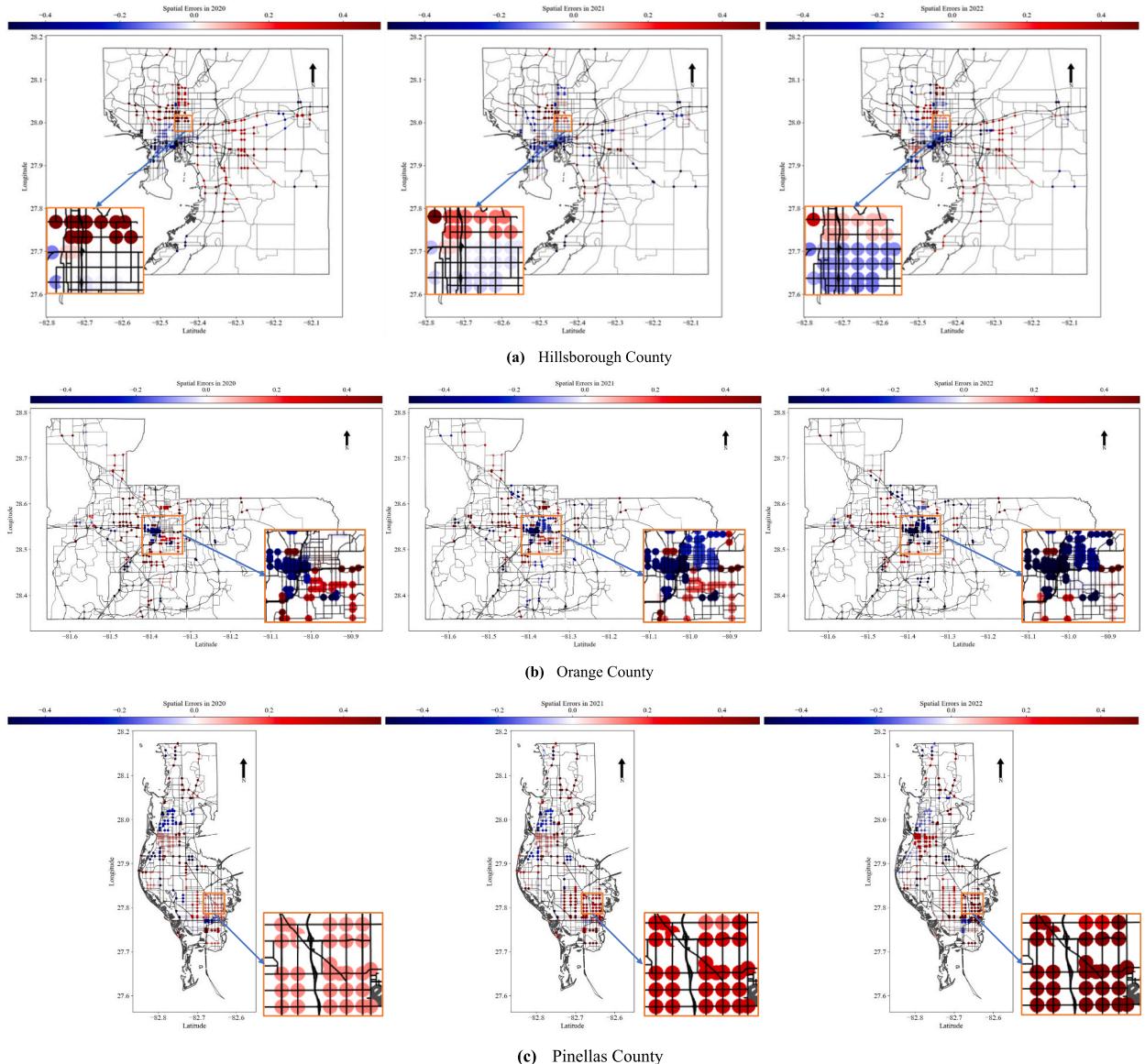


Fig. 10. Spatial errors of intersections in Hillsborough, Orange and Pinellas County across three years. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6

Difference in probabilities by temporal instability using model parameters to estimate another year crash.

Base (crash data) year	Predict (model parameter) year					
	2020		2021		2022	
	Mean	S. D	Mean	S. D	Mean	S. D
2020	—	—	0.069	0.781	0.109	0.994
2021	-0.064	0.646	—	—	0.013	0.489
2022	-0.083	0.932	-0.022	0.525	—	—

Additionally, the traffic volumes of minor roads have greater effects on crash frequency compared to major roads. As such, the marginal effect of Log_Minor_AADT is 2.739, compared to 1.729 for Log_Major_AADT in 2022.

Intersections involving minor roads typically have less sophisticated traffic control measures (e.g., fewer traffic signals, stop signs, or dedicated lanes) compared to those on major roads (Stipancic et al., 2021). This can result in less structured traffic flow, increasing

Table 7

Average marginal effects of significant variables in Grouped Random Parameters Poisson-Lindley model with Spatial effects.

Variables	2020	2021	2022
Log_Major_AADT	1.444	1.379	1.729
Log_Minor_AADT	1.714	1.736	2.739
P_Over65	-0.062	-0.107	-0.089
Legs_4	4.551	4.175	4.487
Major_lanes > 4	5.658	7.068	6.533
Minor_width	0.116	0.136	0.086
Vehicle	0.549	0.835	0.694
Vegetation	-0.705	-0.863	-0.800
Grass	-1.093	-1.073	-1.354

Note: Variables shown in italic font are not statistically insignificant at the 95% confidence level.

the likelihood of conflicts. Additionally, drivers on minor roads may exercise less caution when approaching intersections (Yan et al., 2007), contributing to a higher risk of collisions. At intersections where major roads intersect with minor roads, traffic flow is often prioritized for the major road. However, vehicles from minor roads still enter or cross the intersection, creating potential conflict points (Caliendo and Guida, 2012), especially when traffic volumes on the minor roads are high. Furthermore, minor roads may lack sufficient infrastructure, such as turning lanes, adequate lighting, or traffic signals (Oviedo-Trespalacios et al., 2017), which can exacerbate the risks at intersections, particularly during periods of high traffic volume. This combination of limited control measures and infrastructure inadequacies can significantly increase the danger of crashes at intersections involving minor roads (Papadimitriou et al., 2019).

5.3.2. Socioeconomic variables

P_Over65 (Percent of population 65 years or older (%)) is observed to consistently decrease the intersection crash frequency across three years. Older drivers are less likely to engage in high-risk behaviors like speeding, quick lane changes, or running red lights (Hong et al., 2008; Souders et al., 2020), which are common contributors to intersection crashes. Intersections in communities with a large elderly population might experience lower overall traffic volumes during peak hours, as older adults may rely more on public transportation or alternative mobility pattern like walking and cycling (Zhou et al., 2022).

5.3.3. Geometric design features

Regarding geometric design, Legs_4, Major_lanes > 4, and Minor_width are all observed to positively increase the intersection crash frequency. A 4-legged intersection inherently contains more conflict points (locations where traffic streams intersect or merge) compared to a 3-legged intersection. Furthermore, 4-legged intersections are often situated on busier roads with more complex traffic movements (Kozev et al., 2016), such as turning maneuvers, and present more challenging dynamics for both drivers and vulnerable users.

Major roads with more than four lanes significantly increase intersection crash frequency, with marginal effects ranging from 5.658 to 7.068. Wide roads increase intersection complexity for drivers and pedestrians by impairing navigation, decision-making, and visibility, leading to a higher likelihood of errors and misjudgments (Sheykhfard et al., 2021). They also pose significant crossing challenges for pedestrians and cyclists. Moreover, the higher speed limits and faster driving often associated with wider roads elevate crash frequency and severity at intersections.

The increased width of minor roads is positively associated with a higher frequency of crashes due to the expanded roadway area, which creates additional conflict points and increases the potential for driver misjudgment. Wider roads may also lead to reduced driver caution by fostering a false sense of security or encouraging higher speeds. Furthermore, larger intersections resulting from wider minor roads complicate vehicle maneuvering and pedestrian crossing, further amplifying crash risks (Biswas et al., 2017).

5.3.4. Visual environment features

Vehicle (Proportion of vehicles in the driver's vision (%)), Vegetation (Proportion of vegetation in the driver's vision (%)) and Grass (Proportion of grass area in the driver's vision (%)) are all observed to be significant.

The Vehicle is observed to increase the intersection crash frequency ranging from 0.549 to 0.835 across three years. When a higher proportion of the driver's field of vision is occupied by other vehicles, it can result in visual clutter and distractions (Raddaaoui and Ahmed, 2020). Additionally, the cognitive load on drivers increases, potentially overwhelming their decision-making and judgment (Briggs et al., 2018), which may lead to driving errors and subsequent crashes. As the number of vehicles in the driver's vision rises, the available reaction time and space to respond to sudden changes in traffic flow (e.g., sudden braking, lane changes, or merging) diminishes.

Otherwise, Vegetation and Grass show significantly negative effects on the intersection crash frequency. This is consistent with recent research (Abdel-Aty et al., 2024), as the increase in proportion of vegetation and grass in the drivers' point of view could decrease the crash frequency (Cai et al., 2022). Moreover, the estimated Vegetation and Grass are identified as grouped random parameters showing spatial variations across the 12 Florida counties.

More surrounding vegetation, especially roadside grasses, can make drivers feel a narrow road and exercise more caution when

approaching intersections (Ewing and Dumbaugh, 2009; Treese-II et al., 2017). Meanwhile, roadside vegetation is found to have a positive psychological effect to reduce the stress and anxiety of drivers (Abdel-Aty et al., 2024), thereby to reduce speeding behaviors and crashes.

6. Conclusions

This paper aims to examine the unobserved heterogeneity and spatiotemporal variations in the effects of visual environment features on intersection crash frequency. Grouped random parameters are employed to account for spatial variations at the county level, while spatial effects, capturing correlations among adjacent intersections, are also introduced. Then, a Grouped Random Parameters Poisson-Lindley model with Spatial Effects is developed and compared against other models, including the Fixed Parameters Poisson, Random Parameters Poisson, Random Parameters Poisson-Lindley, and Grouped Random Parameters Poisson models. The analysis focuses on crash data from 2,044 intersections across 12 Florida counties between 2020 and 2022. Explanatory variables are selected based on traffic flow, geometric design characteristics, and visual environment features at the intersections.

The proposed model's performance is evaluated using several measures, including the Deviance Information Criterion (DIC), Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 . The results clearly demonstrate that incorporating both macro- and micro-level spatial effects significantly enhances the model's performance in analyzing intersection crash frequency. Spatial variations are effectively captured by the Lindley parameters, grouped random parameters, and spatial effects among adjacent intersections. Furthermore, temporal variations across the three-year period are clearly identified through out-of-sample predictions and the calculation of marginal effects. A set of variables are found to have significant effects on intersection crash frequency. Among them, two visual environment features—Vegetation and Grass—result in the identification of grouped random parameters across 12 counties. These findings align with and reinforce existing literature on intersection safety. Based on the results, it is recommended that signal control measures be optimized for minor roads and that infrastructure improvements be prioritized to enhance overall safety. Furthermore, strategically adding vegetation (e.g., trees and grass) along intersection approaches may offer traffic calming benefits by encouraging reduced vehicle speeds and improving driver attention, ultimately contributing to lower crash frequency.

The current study underscores the importance of accounting for spatiotemporal heterogeneity and unobserved heterogeneity (spatial variation at both macro and micro levels) in intersection crash analysis. It also highlights the superiority of the Grouped Random Parameters Poisson-Lindley model with Spatial Effects in addressing these complexities. Incorporating these considerations allows for a more accurate representation of the spatial and temporal variations in crash occurrences, improving model performance and providing a deeper understanding of the factors influencing intersection safety.

Nevertheless, the current study has some limitations. For instance, additional variables such as the timing and phasing of signalized intersections should be considered. Furthermore, more extensive analysis is needed to examine the correlation of spatial effects on crashes across both road segments and intersections. Lastly, more advanced statistical or data-driven method could help improve the crash model performance and eliminate the spatial heterogeneity and unobserved heterogeneity.

CRediT authorship contribution statement

Chenzhu Wang: Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Mohamed Abdel-Aty:** Writing – review & editing, Writing – original draft, Funding acquisition, Data curation. **Lei Han:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Data curation, Conceptualization.

Declaration of competing interest

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

Appendix A

Table A1
Descriptive Statistics of the modeling variables.

Variables	Description	Min			Max			Mean			S. D.		
		2020	2021	2022	2020	2021	2022	2020	2021	2022	2020	2021	2022
Crash data													
Intersection crashes	Total crashes per intersection	0	0	0	92	117	123	14.12	16.14	16.23	13.52	15.03	15.47
Traffic flow volume													
Log_Major_AADT	The log value of AADT on major road (pcu)	7.00	7.00	7.00	11.43	11.46	11.46	9.89	9.91	9.92	0.71	0.71	0.71

(continued on next page)

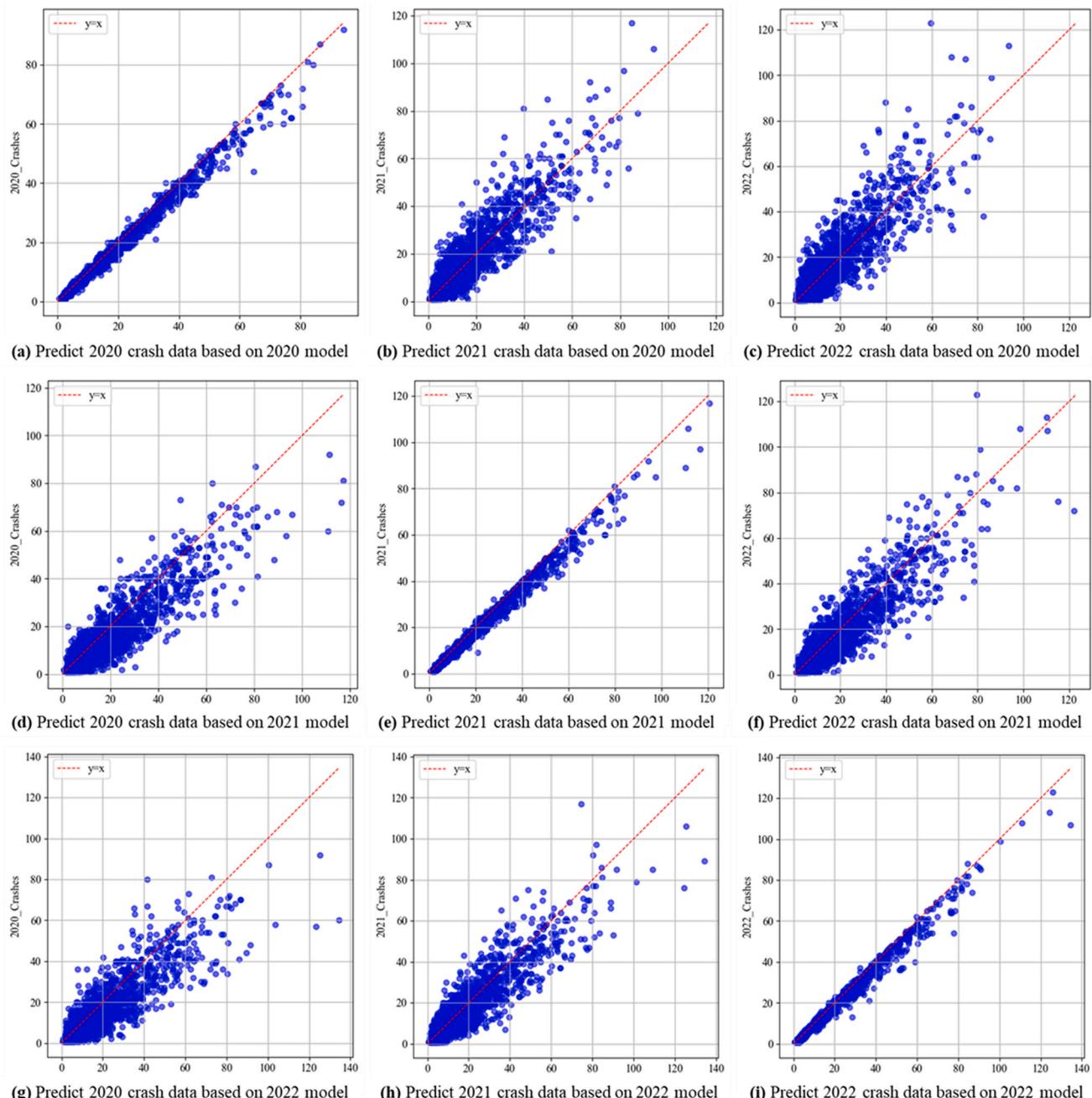
Table A1 (continued)

Variables	Description	Min			Max			Mean			S. D.		
		2020	2021	2022	2020	2021	2022	2020	2021	2022	2020	2021	2022
Log_Minor_AADT	The log value of AADT on minor road (pcu)	5.53	5.53	5.30	11.14	11.13	11.10	8.83	8.82	8.84	0.87	0.88	0.88
Socioeconomic variables													
Log_population	The log value of average population	6.91	6.89	6.93	9.58	9.75	9.87	8.33	8.34	8.36	0.38	0.39	0.39
P_Over65	Percent of population 65 years or older (%)	0.50	0	0	83.80	88.90	85.30	18.80	18.81	19.27	9.34	9.27	9.48
P_Under17	Percent of population 17 years or younger (%)	0.26	0.46	0	37.90	38.00	34.24	19.06	18.91	18.39	6.08	6.11	5.98
P_Unemployed	Percent of people age 16 + unemployed (%)	0	0	0	23.80	23.88	21.60	5.83	5.71	5.37	3.17	3.00	2.81
P_Poverty	Percent of population with income below average poverty level (%)	0.30	1.30	0.70	50.90	50.40	50.80	15.75	15.54	15.16	8.14	8.02	7.73
P_Uneducation	Percent of people age 25 + with less than a high school diploma (%)	0	0	0	73.80	76.60	78.30	14.23	14.17	13.35	10.26	10.75	10.22
P_Traveltime45	Percent of people with travel time to work > 45 min (%)	0	0	0	48.08	46.76	47.50	15.69	15.60	15.71	7.44	7.60	7.49
P_Mobile_Homes	Percent of total housing units that are mobile homes (%)	0	0	0	82.23	83.07	82.47	7.69	7.45	7.34	12.33	12.00	11.88
P_Nocar	Percent of households with no car (%)	0	0	0	41.20	42.90	49.20	8.33	8.03	7.81	6.37	6.08	5.80
P_walk	Percent of people walked to work (%)	0	0	0	24.1	20.4	25.5	2.20	2.03	2.11	3.09	2.74	3.18
P_bicycle	Percent of people riding bicycle to work (%)	0	0	0	15.40	10.10	10.70	0.69	0.63	0.61	1.01	0.94	0.93
Variables	Description							Min	Max		Mean	S. D.	
Geometric design features													
Legs_4	4-legged (yes = 1)							0	1		0.66	0.47	
Major_lanes > 4	Major road lanes > 4 (yes = 1)							0	1		0.25	0.44	
Minor_lanes > 4	Minor road lanes > 4 (yes = 1)							0	1		0.04	0.21	
Major_speed_very_low	Speed limit on major road < 30 mph (yes = 1)							0	1		0.03	0.17	
Major_speed_low	Speed limit on major road in 30–40 mph (yes = 1)							0	1		0.34	0.47	
Major_speed_medium	Speed limit on major road in 40–50 mph (yes = 1)							0	1		0.51	0.50	
Major_speed_high	Speed limit on major road 50–60 mph (yes = 1)							0	1		0.11	0.32	
Major_speed_very_high	Speed limit on major road > 60 mph (yes = 1)							0	1		0.01	0.08	
Minor_speed_very_low	Speed limit on major road < 30 mph (yes = 1)							0	1		0.11	0.31	
Minor_speed_low	Speed limit on minor road in 30–40 mph (yes = 1)							0	1		0.55	0.50	
Minor_speed_medium	Speed limit on minor road in 40–50 mph (yes = 1)							0	1		0.31	0.46	
Minor_speed_high	Speed limit on minor road 50–60 mph (yes = 1)							0	1		0.04	0.19	
Minor_speed_very_high	Speed limit on minor road > 60 mph (yes = 1)							0	1		0.01	0.02	
Major_width	The surface width of major road (feet)							12	98		45.27	17.83	
Minor_width	The surface width of minor road (feet)							12	96		32.67	12.41	
Major_minor_collector	Major road class is minor collector (yes = 1)							0	1		0.03	0.16	
Major_major_collector	Major road class is major collector (yes = 1)							0	1		0.22	0.41	
Major_minor_arterial	Major road class is minor arterial (yes = 1)							0	1		0.34	0.47	
Major_major_arterial	Major road class is major arterial (yes = 1)							0	1		0.42	0.49	
Minor_minor_local	Minor road class is local road (yes = 1)							0	1		0.01	0.10	
Minor_minor_collector	Minor road class is minor collector (yes = 1)							0	1		0.14	0.34	
Geometric design features													
Minor_major_collector	Minor road class is major collector (yes = 1)							0	1		0.53	0.50	
Minor_minor_arterial	Minor road class is minor arterial (yes = 1)							0	1		0.24	0.43	
Minor_major_arterial	Minor road class is major arterial (yes = 1)							0	1		0.07	0.26	
Major_median_marking	Major road median type is traffic marking (yes = 1)							0	1		0.32	0.47	
Major_median_separator	Major road median type is raised traffic separator (yes = 1)							0	1		0.28	0.45	
Major_median_curb	Major road median type is curb and vegetation (yes = 1)							0	1		0.24	0.43	
Minor_median_marking	Minor road median type is traffic marking (yes = 1)							0	1		0.42	0.49	
Minor_median_separator	Minor road median type is raised traffic separator (yes = 1)							0	1		0.18	0.38	
Minor_median_curb	Minor road median type is curb and vegetation (yes = 1)							0	1		0.14	0.34	

(continued on next page)

Table A1 (continued)

Variables	Description	Min	Max	Mean	S. D.
Visual environment features					
Sky	Proportion of sky visible in the driver's vision (%)	9.06	48.43	39.62	6.04
Road	Proportion of road in the driver's vision (%)	16.67	46.27	41.40	3.34
Buildings	Proportion of buildings in the driver's vision (%)	0	37.00	2.56	4.40
Vegetation	Proportion of vegetation in the driver's vision (%)	1.11	43.20	8.53	5.31
Grass	Proportion of grass area in the driver's vision (%)	0	20.36	2.48	2.05
Vehicle	Proportion of vehicles in the driver's vision (%)	0	29.94	1.91	1.89
Walk	Proportion of walk area in the driver's vision (%)	0	14.98	1.17	1.24

**Fig. A1.** Prediction results of crash frequencies in 2044 intersections

Data availability

The authors do not have permission to share data.

References

- Abdel-Aty, M., Ugan, J., Islam, Z., 2024. Exploring the influence of drivers' visual surroundings on speeding behavior. *Accident Analysis and Prevention* 198, 107479.
- Abdel-Aty, M., Wang, X., 2006. Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors. *Transportation Research Record* 1953, 98–111.
- Alhomaidat, F., Kwigizile, V., Oh, J., Van Houten, R., 2020. How does an increased freeway speed limit influence the frequency of crashes on adjacent roads? *Accident Analysis and Prevention* 136, 105433.
- Alnawmasi, N., Mannerling, F., 2019. A statistical assessment of temporal instability in the factors determining motorcyclist injury severities. *Analytic Methods in Accident Research* 22, 100090.
- Ali, Y., Haque, M., Zheng, Z., Afghari, P., 2022. A Bayesian correlated grouped random parameters duration model with heterogeneity in the means for understanding braking behaviour in a connected environment. *Analytic Methods in Accident Research* 35, 100221.
- Alnawmasi, N., Mannerling, F., 2022. The impact of higher speed limits on the frequency and severity of freeway crashes: accounting for temporal shifts and unobserved heterogeneity. *Analytic Methods in Accident Research* 34, 100205.
- Anastasopoulos, P., 2016. Random parameters multivariate tobit and zero-inflated count data models: addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. *Analytic Methods in Accident Research* 11, 17–32.
- Anciaes, P., 2023. Effects of the roadside visual environment on driver wellbeing and behaviour—a systematic review. *Transport Reviews* 43 (4), 571–598.
- Andrey, J., 2010. Long-term trends in weather-related crash risks. *Journal of Transport Geography* 18 (2), 247–258.
- Atumo, A., Li, H., Jiang, X., 2023. Segment-level spatial heterogeneity of arterial crash frequency using locally weighted generalized linear models. *Transportation Research Record* 2677 (3), 1637–1653.
- Barua, S., El-Basyouny, K., Islam, M., 2016. Multivariate random parameters collision count data models with spatial heterogeneity. *Analytic Methods in Accident Research* 9, 1–15.
- Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43, 1–20.
- Bella, F., 2013. Driver perception of roadside configurations on two-lane rural roads: effects on speed and lateral placement. *Accident Analysis and Prevention* 50, 251–262.
- Bella, F., Silvestri, M., 2017. Interaction driver–bicyclist on rural roads: effects of cross-sections and road geometric elements. *Accident Analysis and Prevention* 102, 191–201.
- Bhat, C., Astroza, S., Lavieri, P., 2017. A new spatial and flexible multivariate random-coefficients model for the analysis of pedestrian injury counts by severity level. *Analytic Methods in Accident Research* 16, 1–22.
- Bhowmik, T., Yasmin, S., Eluru, N., 2019. Do we need multivariate modeling approaches to model crash frequency by crash types? A panel mixed approach to modeling crash frequency by crash types. *Analytic Methods in Accident Research* 24, 100107.
- Biswas, S., Chandra, S., Ghosh, I., 2017. Effects of on-street parking in urban context: a critical review. *Transportation in Developing Economies* 3, 1–14.
- Briggs, G., Hole, G., Turner, J., 2018. The impact of attentional set and situation awareness on dual tasking driving performance. *Transportation Research Part F* 57, 36–47.
- Cafiso, S., Graziano, A., Silvestro, G., Cava, G., Persaud, B., 2010. Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables. *Accident Analysis and Prevention* 42 (4), 1072–1079.
- Cai, Q., Abdel-Aty, M., Lee, J., Wang, L., Wang, X., 2018. Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. *Analytic Methods in Accident Research* 19, 1–15.
- Cai, Q., Abdel-Aty, M., Yuan, J., Lee, J., Wu, Y., 2020. Real-time crash prediction on expressways using deep generative models. *Transportation Research Part C* 117, 102697.
- Cai, Q., Abdel-Aty, M., Yuan, J., Lee, J., Wu, Y., 2022. Applying machine learning and google street view to explore effects of drivers' visual environment on traffic safety. *Transportation Research Part C* 135, 1003541.
- Cai, Z., Wei, F., Guo, Y., 2023. A full Bayesian multilevel approach for modeling interaction effects in single-vehicle crashes. *Accident Analysis and Prevention* 193, 107331.
- Caliendo, C., Guida, M., 2012. Microsimulation approach for predicting crashes at unsignalized intersections using traffic conflicts. *Journal of Transportation Engineering* 138 (12), 1453–1467.
- Champahom, T., Jomnonkwo, S., Watthanaklang, D., Karoonsoontawong, A., Chatpattananan, V., Ratanavaraha, V., 2020. Applying hierarchical logistic models to compare urban and rural roadway modeling of severity of rear-end vehicular crashes. *Accident Analysis and Prevention* 141, 105537.
- Cheng, W., Gill, G., Zhang, Y., Cao, Z., 2018. Bayesian spatiotemporal crash frequency models with mixture components for space-time interactions. *Accident Analysis and Prevention* 112, 84–93.
- Chen, T., Wong, Y., Shi, X., Yang, Y., 2021. A data-driven feature learning approach based on Copula-Bayesian Network and its application in comparative investigation on risky lane-changing and car-following maneuvers. *Accident Analysis and Prevention* 154, 106061.
- Coruh, E., Bilgic, A., Tortum, A., 2015. Accident analysis with aggregated data: the random parameters negative binomial panel count data model. *Analytic Methods in Accident Research* 7, 37–49.
- Cui, H., Xie, K., 2021. An accelerated hierarchical Bayesian crash frequency model with accommodation of spatiotemporal interactions. *Accident Analysis and Prevention* 153, 106018.
- Dey, K., Mishra, A., Chowdhury, M., 2014. Potential of intelligent transportation systems in mitigating adverse weather impacts on road mobility: a review. *IEEE Transactions on Intelligent Transportation Systems* 16 (3), 1107–1119.
- Dong, C., Clarke, D., Yan, X., Khattak, A., Huang, B., 2014. Multivariate random-parameters zero-inflated negative binomial regression model: an application to estimate crash frequencies at intersections. *Accident Analysis and Prevention* 70, 320–329.
- Dong, N., Huang, H., Zheng, L., 2015. Support vector machine in crash prediction at the level of traffic analysis zones: assessing the spatial proximity effects. *Accident Analysis and Prevention* 82, 192–198.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv preprint arXiv:2010.11929*.
- Dong, N., Zhang, J., Liu, X., Xu, P., Wu, Y., Wu, H., 2022. Association of human mobility with road crashes for pandemic-ready safer mobility: a New York City case study. *Accident Analysis and Prevention* 165, 106478.
- Dzinyela, R., Shirazi, M., Das, S., Lord, D., 2024. The negative Binomial-Lindley model with time-dependent parameters: accounting for temporal variations and excess zero observations in crash data. *Accident Analysis and Prevention* 207, 107711.
- El-Basyouny, K., Barua, S., Islam, M., 2014. Investigation of time and weather effects on crash types using full Bayesian multivariate Poisson lognormal models. *Accident Analysis and Prevention* 73, 91–99.
- Ewing, R., Dumbaugh, E., 2009. The built environment and traffic safety: a review of empirical evidence. *Journal of Planning Literature* 23 (4), 347–367.
- Fan, Z., Zhang, F., Loo, B., Ratti, C., 2023. Urban visual intelligence: uncovering hidden city profiles with street view images. *Proceedings of the National Academy of Sciences* 120 (27), e2220417120.
- Fountas, G., Anastasopoulos, P., Abdel-Aty, M., 2018a. Analysis of accident injury-severities using a correlated random parameters ordered probit approach with time variant covariates. *Analytic Methods in Accident Research* 18, 57–68.

- Fountas, G., Sarwar, T., Anastasopoulos, P., Blatt, A., Majka, K., 2018b. Analysis of stationary and dynamic factors affecting highway accident occurrence: a dynamic correlated grouped random parameters binary logit approach. *Analytic Methods in Accident Research* 113, 330–340.
- Fountas, G., Fonzone, A., Gharavi, N., Rye, T., 2020. The joint effect of weather and lighting conditions on injury severities of single-vehicle accidents. *Analytic Methods in Accident Research* 27, 100124.
- FHWA. 2023. Intersection Safety. Federal Highway Administration. <https://highways.dot.gov/research/research-programs/safety/intersection-safety>. Accessed SEP 3, 2024.
- Geedipally, S., Lord, D., Dhavala, S., 2012. The negative binomial-Lindley generalized linear model: characteristics and application using crash data. *Accident Analysis and Prevention* 45, 258–265.
- Geedipally, S., Lord, D., Dhavala, S., 2014. A caution about using deviance information criterion while modeling traffic crashes. *Safety Science* 62, 495–498.
- Ghasemzadeh, A., Ahmed, M., 2019. Quantifying regional heterogeneity effect on drivers' speeding behavior using SHRP2 naturalistic driving data: a multilevel modeling approach. *Transportation Research Part C* 106, 29–40.
- Google API. Google Maps Platform. <https://developers.google.com/maps>.
- Gu, Y., Liu, D., Arvin, R., Khattak, A., Han, L., 2023. Predicting intersection crash frequency using connected vehicle data: a framework for geographical random forest. *Accident Analysis and Prevention* 179, 106880.
- Guo, F., Wang, X., Abdel-Aty, M., 2010. Modeling signalized intersection safety with corridor-level spatial correlations. *Accident Analysis and Prevention* 42 (1), 84–92.
- Gurumurthy, K., Bansal, P., Kockelman, K., Li, Z., 2022. Modelling animal-vehicle collision counts across large networks using a Bayesian hierarchical model with time-varying parameters. *Analytic Methods in Accident Research* 36, 100231.
- Hanson, C., Noland, R., Brown, C., 2013. The severity of pedestrian crashes: an analysis using Google Street View imagery. *Journal of Transport Geography* 33, 42–53.
- Haque, M., Chin, H., Huang, H., 2010. Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. *Accident Analysis and Prevention* 42 (1), 203–212.
- Hasan, T., Abdel-Aty, M., 2024. Short-term safety performance functions by random parameters negative Binomial-Lindley model for part-time shoulder use. *Accident Analysis and Prevention* 199, 107498.
- Heydari, S., Fu, L., Thakali, L., Joseph, L., 2018. Benchmarking regions using a heteroskedastic grouped random parameters model with heterogeneity in mean and variance: applications to grade crossing safety analysis. *Analytic Methods in Accident Research* 19, 33–48.
- Hezaveh, A., Arvin, R., Cherry, C., 2019. A geographically weighted regression to estimate the comprehensive cost of traffic crashes at a zonal level. *Accident Analysis and Prevention* 131, 15–24.
- Hong, I., Kurihara, T., Iwasaki, M., 2008. Older drivers' perceptions, responses, and driving behaviours during complex traffic conditions at a signalized intersection. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering* 222 (11), 2063–2076.
- Hou, Q., Huo, X., Tarko, A., Leng, J., 2021. Comparative analysis of alternative random parameters count data models in highway safety. *Analytic Methods in Accident Research* 30, 100158.
- Hou, Q., Huo, X., Leng, J., Mannering, F., 2022. A note on out-of-sample prediction, marginal effects computations, and temporal testing with random parameters crash-injury severity models. *Analytic Methods in Accident Research* 33, 100191.
- Huang, Y., Wang, X., Patton, D., 2018. Examining spatial relationships between crashes and the built environment: a geographically weighted regression approach. *Journal of Transport Geography* 69, 221–233.
- Huang, H., Zhou, H., Wang, J., Chang, F., Ma, M., 2017. A multivariate spatial model of crash frequency by transportation modes for urban intersections. *Analytic Methods in Accident Research* 14, 10–21.
- Intini, P., Berloco, N., Fonzone, A., Fountas, G., Ranieri, V., 2020. The influence of traffic, geometric and context variables on urban crash types: a grouped random parameter multinomial logit approach. *Analytic Methods in Accident Research* 28, 100141.
- Islam, A., Shirazi, M., Lord, D., 2022a. Finite mixture negative Binomial-Lindley for modeling heterogeneous crash data with many zero observations. *Accident Analysis and Prevention* 175, 106765.
- Islam, S., Washington, S., Kim, J., Haque, M., 2022b. A comprehensive analysis on the effects of signal strategies, intersection geometry, and traffic operation factors on right-turn crashes at signalised intersections: an application of hierarchical crash frequency model. *Accident Analysis and Prevention* 171, 106663.
- Islam, M., Shirazi, M., Lord, D., 2023. Grouped random parameters negative Binomial-Lindley for accounting unobserved heterogeneity in crash data with preponderant zero observations. *Analytic Methods in Accident Research* 37, 100255.
- Jackson, T., Sharif, H., 2016. Rainfall impacts on traffic safety: rain-related fatal crashes in Texas. *Geomatics, Natural Hazards and Risk* 7 (2), 843–860.
- Kabir, R., Remias, S., Lavrenz, S., Waddell, J., 2021. Assessing the impact of traffic signal performance on crash frequency for signalized intersections along urban arterials: a random parameter modeling approach. *Accident Analysis and Prevention* 149, 105868.
- Khan, S., Afghari, A., Yasmin, S., Haque, M., 2023. Effects of design consistency on run-off-road crashes: an application of a random parameters negative Binomial Lindley model. *Accident Analysis and Prevention* 186, 107042.
- Khattak, A., Ahmad, N., Wali, B., Dumbaugh, E., 2021. A taxonomy of driving errors and violations: evidence from the naturalistic driving study. *Accident Analysis and Prevention* 151, 105873.
- Khorashadi, A., Niemeier, D., Shankar, V., Mannering, F., 2005. Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis. *Accident Analysis and Prevention* 37 (5), 910–921.
- Khodadadi, A., Tsapakis, I., Shirazi, M., Das, S., Lord, D., 2022. Derivation of the Empirical Bayesian method for the Negative Binomial-Lindley generalized linear model with application in traffic safety. *Accident Analysis and Prevention* 170, 106638.
- Kozey, P., Xuan, Y., Cassidy, M., 2016. A low-cost alternative for higher capacities at four-way signalized intersections. *Transportation Research Part C* 72, 157–167.
- Kwon, J., Cho, G., 2020. An examination of the intersection environment associated with perceived crash risk among school-aged children: using street-level imagery and computer vision. *Accident Analysis and Prevention* 146, 105716.
- Li, L., Zhang, Z., Xu, Z., Yang, W., Lu, Q., 2024. The role of traffic conflicts in roundabout safety evaluation: a review. *Accident Analysis and Prevention* 196, 107430.
- Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., Zhang, W., 2015. Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry and Urban Greening* 14 (3), 675–685.
- Li, Z., Chen, C., Ci, Y., Zhang, G., Wu, Q., Liu, C., Qian, Z., 2018. Examining driver injury severity in intersection-related crashes using cluster analysis and hierarchical Bayesian models. *Accident Analysis and Prevention* 120, 139–151.
- Liu, J., Khattak, A., 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 and Prevention* 109, 132–142.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 3431–3440.
- Lord, D., Geedipally, S., 2011. The negative binomial-Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros. *Accident Analysis and Prevention* 43 (5), 1738–1742.
- Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A* 44 (5), 291–305.
- Lunn, D., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 10, 325–337.
- 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.
- Mannering, F., 2018. Temporal instability and the analysis of highway accident data. *Analytic Methods in Accident Research* 17, 1–13.
- Mannering, F., Bhat, C., 2014. Analytic methods in accident research: methodological frontier and future directions. *Analytic Methods in Accident Research* 1, 1–22.
- Mannering, F., Shankar, V., Bhat, C., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research* 11, 1–16.

- Megat-Johari, M., Bazargani, B., Kirsch, T., Barrette, T., Savolainen, P., 2018. An examination of the safety of signalized intersections in consideration of nearby access points. *Transportation Research Record* 2672 (17), 11–21.
- Meng, F., Sze, N., Song, C., Chen, T., Zeng, Y., 2021. Temporal instability of truck volume composition on non-truck-involved crash severity using uncorrelated and correlated grouped random parameters binary logit models with space-time variations. *Analytic Methods in Accident Research* 31, 100168.
- Mohammadi, M., Samaranayake, V., Bham, G., 2014. Crash frequency modeling using negative binomial models: an application of generalized estimating equation to longitudinal data. *Analytic Methods in Accident Research* 2, 52–69.
- Mohammadnazar, A., Mahdinia, I., Ahmad, N., Khattak, A., Liu, J., 2021. Understanding how relationships between crash frequency and correlates vary for multilane rural highways: estimating geographically and temporally weighted regression models. *Accident Analysis and Prevention* 157, 106146.
- Munira, S., Sener, I., Dai, B., 2020. A Bayesian spatial Poisson-lognormal model to examine pedestrian crash severity at signalized intersections. *Accident Analysis and Prevention* 144, 105679.
- Murphy, B., Levinson, D., Owen, A., 2017. Evaluating the safety in numbers effect for pedestrians at urban intersections. *Accident Analysis and Prevention* 106, 181–190.
- Nightingale, E., Parvin, N., Seiberlich, C., Savolainen, P., Pawlovich, M., 2017. Investigation of skew angle and other factors influencing crash frequency at high-speed rural intersections. *Transportation Research Record* 2636, 9–14.
- Oviedo-Trespalacios, O., Haque, M., King, M., Washington, S., 2017. Effects of road infrastructure and traffic complexity in speed adaptation behaviour of distracted drivers. *Accident Analysis and Prevention* 101, 67–77.
- Papadimitriou, E., Filtness, A., Theofilatos, A., Ziakopoulos, A., Quigley, C., Yannis, G., 2019. Review and ranking of crash risk factors related to the road infrastructure. *Accident Analysis and Prevention* 125, 85–97.
- Raddaoui, O., Ahmed, M., 2020. Evaluating the effects of connected vehicle weather and work zone warnings on truck drivers' workload and distraction using eye glance behavior. *Transportation Research Record* 2674 (3), 293–304.
- Roadway Characteristics Inventory (RCI) database. Geographic Information System (GIS). FDOT. <https://www.fdot.gov/statistics/gis/default.shtml>.
- Saha, D., Alluri, P., Gan, A., Wu, W., 2018. Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models. *Accident Analysis and Prevention* 118, 166–177.
- Shaon, M., Qin, X., Shirazi, M., Lord, D., Geedipally, S., 2018. Developing a random parameters negative binomial-lindley model to analyze highly over-dispersed crash count data. *Analytic Methods in Accident Research* 18, 33–44.
- Sheykhard, A., Haghghi, F., Papadimitriou, E., Gelder, P., 2021. Review and assessment of different perspectives of vehicle-pedestrian conflicts and crashes: Passive and active analysis approaches. *Journal of Traffic and Transportation Engineering (English Edition)* 8 (5), 681–702.
- Signal Four Analytics (S4A.), 2014. FDOT Transportation Data Symposium. GeoPlan Center, University of Florida. Available at: <https://www.fdot.gov/docs/default-source/statistics/symposium/2014/Signal4Analytics.pdf>.
- Stipancic, J., St-Aubin, P., Ledezma-Navarro, B., Labbe, A., Saunier, N., Miranda-Moreno, L., 2021. Evaluating safety-influencing factors at stop-controlled intersections using automated video analysis. *Journal of Safety Research* 77, 311–323.
- Strudel, R., Garcia, R., Laptev, I., Schmid, C., 2021. Segmenter: Transformer for Semantic Segmentation. In: Presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7262–7272.
- Song, L., Li, Y., Fan, W., 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.
- Song, P., Sze, N., Zheng, O., Abdel-Aty, M., 2022. Addressing unobserved heterogeneity at road user level for the analysis of conflict risk at tunnel toll plaza: a correlated grouped random parameters logit approach with heterogeneity in means. *Analytic Methods in Accident Research* 36, 100243.
- Souders, D., Charness, N., Roque, N., Pham, H., 2020. Aging: older adults' driving behavior using longitudinal and lateral warning systems. *Human Factors* 62 (2), 229–248.
- Su, J., Sze, N., Bai, L., 2021. A joint probability model for pedestrian crashes at macroscopic level: roles of environment, traffic, and population characteristics. *Accident Analysis and Prevention* 150, 105898.
- Tahir, H., Yasmin, S., Haque, M., 2024. A Poisson Lognormal-Lindley model for simultaneous estimation of multiple crash-types: application of multivariate and pooled univariate models. *Analytic Methods in Accident Research* 41, 100315.
- Tang, J., Gao, F., Liu, F., Han, C., Lee, J., 2020. Spatial heterogeneity analysis of macro-level crashes using geographically weighted Poisson quantile regression. *Accident Analysis and Prevention* 148, 105833.
- Truong, L., Kieu, L., Vu, T., 2016. Spatiotemporal and random parameter panel data models of traffic crash fatalities in Vietnam. *Accident Analysis and Prevention* 94, 153–161.
- U.S. Census Bureau, n.d. American Community Survey Data. The United States Census Bureau. <https://www.census.gov/data.html>.
- Treese-II, J., Koeser, A., Fitzpatrick, G., Olexa, M., Allen, E., 2017. A review of the impact of roadway vegetation on drivers' health and well-being and the risks associated with single-vehicle crashes. *Arbicultural Journal* 39 (3), 179–193.
- Wang, X., Abdel-Aty, 2008. Modeling left-turn crash occurrence at signalized intersections by conflicting patterns. *Accident Analysis and Prevention* 40 (1), 7–88.
- Wang, C., Chen, F., Cheng, J., Bo, W., Zhang, P., Hou, M., Xiao, F., 2020. Random-parameter multivariate negative binomial regression for modeling impacts of contributing factors on the crash frequency by crash types. *Discrete Dynamics in Nature and Society* 2020 (1), 6621752.
- Wang, C., Chen, F., Zhang, Y., Wang, S., Yu, B., Cheng, J., 2022a. Temporal stability of factors affecting injury severity in rear-end and non-rear-end crashes: a random parameter approach with heterogeneity in means and variances. *Analytic Methods in Accident Research* 35, 100219.
- Wang, C., Chen, F., Zhang, Y., Cheng, J., 2022b. Spatiotemporal instability analysis of injury severities in truck-involved and non-truck-involved crashes. *Analytic Methods in Accident Research* 34, 100214.
- Wang, C., Abdel-Aty, M., Han, L., 2024. Effects of speed difference on injury severity of freeway rear-end crashes: Insights from correlated joint random parameters bivariate probit models and temporal instability. *Analytic Methods in Accident Research* 42, 100320.
- Washington, S., Karlaftis, M., Mannerling, F., Anastasopoulos, P., 2020. Statistical and Econometric Methods for Transportation Data Analysis, 3rd edition. CRC Press, Taylor and Francis Group, New York, NY.
- Xie, K., Wang, X., Ozbay, K., Yang, H., 2014. Crash frequency modeling for signalized intersections in a high-density urban road network. *Analytic Methods in Accident Research* 2, 39–51.
- Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: random parameter versus geographically weighting. *Accident Analysis and Prevention* 75, 16–25.
- Yan, X., Radwan, E., Guo, D., 2007. Effects of major-road vehicle speed and driver age and gender on left-turn gap acceptance. *Accident Analysis and Prevention* 39 (4), 843–852.
- Yang, D., Xie, K., Ozbay, K., Yang, H., 2021. Fusing crash data and surrogate safety measures for safety assessment: development of a structural equation model with conditional autoregressive spatial effect and random parameters. *Accident Analysis and Prevention* 152, 105971.
- Yang, J., Zhao, L., McBride, J., Gong, P., 2009. Can you see green? Assessing the visibility of urban forests in cities. *Landscape and Urban Planning* 91 (2), 97–104.
- Yang, C., Wan, Z., Yuan, Q., Zhou, Y., Sun, M., 2023. Travel before, during and after the COVID-19 pandemic: exploring factors in essential travel using empirical data. *Journal of Transport Geography* 110, 103640.
- Yu, R., Abdel-Aty, M., 2013. Investigating different approaches to develop informative priors in hierarchical Bayesian safety performance functions. *Accident Analysis and Prevention* 56, 51–58.
- Yu, B., Chen, Y., Bao, S., 2019. Quantifying visual road environment to establish a speeding prediction model: an examination using naturalistic driving data. *Accident Analysis and Prevention* 129, 289–298.
- Yuan, J., Abdel-Aty, M., 2018. Approach-level real-time crash risk analysis for signalized intersections. *Accident Analysis and Prevention* 119, 274–289.
- Zamani, H., Ismail, N., 2010. Negative binomial-Lindley distribution and its application. *Journal of Mathematics and Statistics* 6 (1), 4–9.
- Zeng, Q., Gu, W., Zhang, X., Wen, H., Lee, J., Hao, W., 2019. Analyzing freeway crash severity using a Bayesian spatial generalized ordered logit model with conditional autoregressive priors. *Accident Analysis and Prevention* 127, 87–95.

- Zeng, Q., Wang, Q., Zhang, K., Wong, S., Xu, P., 2023. Analysis of the injury severity of motor vehicle–pedestrian crashes at urban intersections using spatiotemporal logistic regression models. *Accident Analysis and Prevention* 189, 107119.
- Zhao, S., Khattak, A., 2018. Injury severity in crashes reported in proximity of rail crossings: the role of driver inattention. *Journal of Transportation Safety and Security* 10 (6), 507–524.
- Zhou, Y., Yuan, Q., Ding, F., Chen, M., Yang, C., Guo, T., 2022. Demand, mobility, and constraints: exploring travel behaviors and mode choices of older adults using a facility-based framework. *Journal of Transport Geography* 102, 103368.
- Ziakopoulos, A., Yannis, G., 2020. A review of spatial approaches in road safety. *Accident Analysis and Prevention* 135, 105323.