



Spatiotemporal path analysis of collision frequency using restricted spatial regression with speed difference as a mediator

Changhe Liu^{a,b}, Bo Wang^a, Chi Zhang^{a,*}, Karim Ismail^b, Dibin Wei^a, Rulla Al-Haideri^c

^a School of Highway, Chang'an University, Xi'an 710000, China

^b Department of Civil and Environmental Engineering, Carleton University, Ottawa K1S 5B6, Canada

^c Laboratory of Innovations in Transportation (LiTrans), Toronto Metropolitan University, Toronto M5B 2K3, Canada

ARTICLE INFO

Keywords:

Freeway safety
Collision frequency
Path analysis
Speed difference
Integrated nested laplace approximation

ABSTRACT

Developing reliable collision analysis models is a fundamental task in road safety research. Understanding the spatiotemporal effects of highway collisions is essential for effective prevention and control strategies. This study examines multiple spatiotemporal random effect models using three key approaches: [i] incorporating path analysis, [ii] introducing speed difference as a mediating variable, and [iii] accounting for both structured and unstructured unobserved heterogeneity. Spatial confounding between fixed and random effects was addressed by constructing a restriction matrix to correct spatial random terms. Data were collected from freeways in the Guangxi Zhuang Autonomous Region, China. The results showed that model combining structured spatial and unstructured temporal interaction terms achieved the best performance across all goodness-of-fit measures. The spatial terms explained most of the random variation. Among fixed effects, slope, speed difference, rainfall, interchange, route length, and average daily traffic volume were all positively and significantly associated with collision frequency. Further, the slope factor exhibited indirect effect on collision frequency through its influence on the mediating variable, speed difference. However, given that the freeways are located in relatively flat terrain with large radii (exceeding 750 m), the curvature radius does not exhibit a statistically significant effect on either collision frequency or speed difference. The findings provide deeper insight into the influence of spatiotemporal factors on highway collisions and offer guidance for future highway design and traffic management.

1. Introduction

Road collisions remain a persistent global challenge. Each year, roughly 1.19 million people die in road traffic collisions worldwide, causing major losses of life and property (CDC, 2025). These events are rarely driven by a single cause but rather emerge from interactions among multiple traffic system elements operating in a dynamic environment. This complexity makes it difficult to identify causal relationships, especially when key variables are unobserved or data coverage is limited (Mannering et al., 2016). This can lead to biased or incomplete interpretations of collision mechanisms. Developing a modeling framework that explicitly represents correlations among factors and accounts for unobserved heterogeneity is essential. Such a framework is critical for revealing and predicting the mechanisms driving collision occurrence.

In practice, the situation is even more complex. A wide range of

factors influence collision likelihood, and isolating every contributing variable and its interactions is rarely feasible. Unobserved influences may introduce heterogeneity into model results (Mannering et al., 2016). In addition, roadway attributes may also affect collision risk both directly and indirectly through their influence on driver behaviour (Yang and Han, 2025). In other words, a change in one factor may trigger changes in others, which in turn indirectly affect collision outcomes. For instance, drivers often adjust their speed according to highway geometry, yet both geometry and speed behavior can independently contribute to collision risk. Therefore, highway geometry can exert both direct and indirect effects on collision occurrence (Ekmekeci et al., 2024).

Existing studies have advanced the use of path analysis in traffic safety (Ahmad et al., 2023a; Gargoum and El-Basyouny, 2016; Park et al., 2021; Yu et al., 2019). However, most have overlooked structured spatiotemporal effects and the potential confounding between spatial

* Corresponding author.

E-mail address: zhangchi@chd.edu.cn (C. Zhang).

<https://doi.org/10.1016/j.aap.2026.108718>

Received 3 December 2025; Received in revised form 8 August 2026; Accepted 11 August 2026

0001-4575/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

random effects and fixed covariates. These limitations reduce the reliability of causal interpretation and the precision of estimated effects (Ziakopoulos and Yannis, 2020). In light of these observations, the main objective of this study is to develop a spatiotemporal path analysis framework that integrates restricted spatial regression and employs operating speed difference as a mediating variable to quantify both direct and indirect effects of roadway factors on collision frequency. Specifically, the study addresses the following research questions:

1. How can path analysis be extended to incorporate structured spatiotemporal effects for improved causal interpretation of collision mechanisms?
2. To what extent does operating speed difference mediate the influence of geometric characteristics on collision frequency?
3. How can restricted spatial regression mitigate spatial confounding between fixed and random effects in collision models?
4. Which combination of spatial, temporal, and interaction components best captures unobserved heterogeneity and enhances model predictive performance?

The contribution of this study lies primarily at the methodological level, while also offering empirical findings derived from a unique dataset. Methodologically, this study develops a comprehensive path analysis framework for highway collision risk modeling. The proposed framework incorporates multiple sources of heterogeneity and explicitly considers potential confounding effects, particularly the overlap between observed roadway variables and latent spatial structures, which may otherwise bias the estimated relationships among roadway characteristics, operating speed difference, and collision frequency. To the best of our knowledge, this is among the first studies to introduce a constrained regression approach into highway safety analysis to mitigate spatial confounding. By separating fixed-effect signals from latent spatial structures, this approach helps reduce bias in parameter estimation and enables a more reliable assessment of the effects of geometric, traffic, and environmental factors. At the application level, the proposed method is applied to investigate key spatiotemporal factors associated with highway collisions in flat-terrain regions. The empirical findings provide useful evidence for identifying risk-contributing roadway characteristics and can support highway agencies in developing more targeted safety management strategies.

This remainder of the paper is organized as follows: Section 2 reviews the application of path analysis and spatiotemporal effects in highway traffic safety. Section 3 introduces the methodological framework established in this study. Section 4 introduces the data source and road segment division of this study. Section 5 provides the parameter estimation results and discussion of this study. Section 6 is a summary of the entire article and possible future research directions.

2. Literature review

2.1. Effects of speed and collision frequency

Speed has long been recognized as a key factor affecting both collision frequency and collision severity. Different representations of speed, such as mean speed, speed variation, and operating speed, have been widely examined. However, the research findings remain inconsistent. Table 1 summarizes the relationships between different speed indicators and collision frequency reported in previous research. Quddus (2013) reported a negative relationship between mean speed and collision frequency, while Das et al. (2023) found that annual average speed positively affects collision occurrence. Imprialou et al. (2016) and Yu et al. (2018) pointed out that the level of data aggregation between speed and collision records can significantly influence the results. Hutton et al. (2020) further demonstrated that the specific form of the speed metric matters; they found that speed variation was significantly associated with collision frequency, while mean speed and operating speeds were not.

These inconsistencies suggest that different speed indicators capture distinct aspects of driving behavior. Higher absolute speeds may increase collision severity, while greater speed instability may elevate the probability of collision occurrence. Moreover, collision causation is inherently complex, as speed interacts with multiple factors. In addition to its direct influence, speed may also transmit the effects of road and traffic conditions. Gargoum and El-Basyouny (2016) and Park et al. (2021) applied path analysis models and treated speed as a mediating variable. Their results indicated that road characteristics can indirectly influence collision frequency through speed indicators.

The way speed and collision data are aggregated, the representation of speed used, and the role speed plays within a modeling framework can all affect the observed relationship between speed and collisions. To accurately quantify the contribution of speed to collision occurrence, these aspects must be addressed carefully. Recent studies have shown that models based on pre-collision traffic data tend to yield more reliable results (Imprialou et al., 2016; Yu et al., 2018). However, such detailed datasets are scarce, and most studies still rely on aggregated traffic indicators.

To gain a more comprehensive understanding of collision mechanisms and improve risk prevention, it is necessary to integrate a path analysis framework with explicit consideration of unobserved heterogeneity.

2.2. Path analysis in collision modelling

Path analysis is a statistical modelling method that uses mediating variables to analyze both the direct and indirect effects of influencing factors on an outcome, while accounting for potential confounding relationships among variables. It has been widely used in studies of collision frequency and collision severity within highway safety

Table 1
Effects of speed on collision frequency.

Speed Indicator	Positive Effect	Negative Effect	No Effect or Mixed Effect
Speed value:	Average speed: Imprialou et al. (2016)*; Gargoum & El-Basyouny (2016); Azati et al. (2025) Travel speed: Doecke et al. (2021) Operating speed: Das et al. (2023)*	Average speed: Imprialou et al. (2016)*; Hutton et al. (2020)	Average speed: Quddus (2013) Operating speed: Yu et al. (2018); Hutton et al. (2020); Das et al. (2023)*
Speed difference	Operating speed difference: Camacho-Torregrosa et al. (2013); Khan et al. (2023) Speed limit minus driving speed: Park et al. (2021)	/	Speed limit minus driving speed: Hutton et al. (2020)
Speed dispersion	Speed standard deviation: Quddus (2013); Wei et al. (2024) Speed variance: Hamzeie et al. (2017); Xu et al. (2019) Speed coefficient of variation: Ahmad et al. (2023a)	/	Speed variation: Pei et al. (2012)

*The author used different aggregation methods for the same type of indicators.

research.

In terms of collision frequency, Gargoum and El-Basyouny (2016) studied the direct and indirect impacts of factors such as traffic volume, segment length, medians and horizontal curves on collision frequency using average speed as a mediating variable. Park et al. (2021) further analyzed how traffic and geometric factors influence collision frequency through operating speed and its derivative indicators. Yu et al. (2019) combined the latent class logit model with path analysis to investigate collision causation while accounting for unobserved heterogeneity, using pre-collision traffic indicators as mediating variables. Similarly, Afghari et al. (2023) combined the random parameter model with path analysis to study the collision mechanism with driving behavior as the mediating variable. In another study, Abohassan et al. (2021) analyzed the impact of weather and maintenance conditions on collision frequency, using road friction as the mediating variable.

In the context of collision severity, Das et al. (2023) developed a path analysis model with speed difference as the mediating variable at both annual and daily levels. Lu et al. (2022) examined the effects of driver characteristics and environmental factors on collision severity through pre-collision behavior. Okafor et al. (2023) considered unobserved heterogeneity and established a random parameter path analysis model with pedestrian behavior as the mediating variable. Similarly, Yu et al. (2025) combined the correlated random parameter logit and mean heterogeneity models to analyse the severity of motorcyclist injuries.

In the field of highway traffic safety, whether focusing on collision severity or collision frequency, researchers have gradually moved from using general models to employing more advanced modeling approaches that account for unobserved heterogeneity (Mannering et al., 2016). However, most random parameter path analysis models consider only unstructured heterogeneity by introducing random effects. Among heterogeneity-related studies, growing attention has also been given to structured heterogeneity caused by spatiotemporal factors (Jonathan et al., 2016; Ziakopoulos and Yannis, 2020). In particular, for collision frequency, if its spatiotemporal characteristics are not considered, it is sometimes difficult to distinguish the impact of different unobserved factors (Cui and Xie, 2021; Zeng et al., 2023). Therefore, for driving behavior indicators that often serve as mediating variables, their possible spatiotemporal heterogeneity should be considered.

2.3. Spatial and temporal effects

When addressing spatial, temporal effects, or joint spatiotemporal variations, a common approach is to include random terms that capture underlying heterogeneity. Depending on their structural components, these can be categorized as spatial models, spatiotemporal models or spatiotemporal interaction models.

The spatial effects model includes structures as conditional autoregressive (CAR) or Besag-York-Mollié (BYM) terms. The CAR model captures spatial autocorrelation by assuming that the value at each location depends conditionally on its neighboring locations. The BYM model extends this by incorporating both structured (spatially correlated) and unstructured (spatially independent) random effects to better account for local heterogeneity. For example, Singh et al. (2022) used multivariate CAR (MCAR) terms to capture the spatial effect of traffic guidance design on bicycle and pedestrian collisions. Wang et al. (2025) applied the BYM spatial structure to examine spatial correlation and heterogeneity in intersection collision frequency.

The spatiotemporal extends the spatial model by adding temporal random components such as first-order random walk (RW1), first-order autoregression (AR1), or linear time terms. For example, Cheng et al. (2020) analyzed six years of collision data in California using CAR spatial terms and linear temporal random effects. Similarly, Multiple model configurations based on collision data from four counties in Pennsylvania were compared, and the CAR and RW1 structures were found to better capture spatiotemporal characteristics (Wang et al., 2024).

The spatiotemporal interaction model further integrates interaction terms that represent the combined influence of space and time. For instance, Liu and Sharma (2018) analyzed the spatiotemporal effects in Iowa collision severity data based on the spatiotemporal interaction terms of Gaussian distribution. Cui and Xie (2021) analyzed the spatiotemporal effects of collision frequency in Manhattan, New York based on four spatiotemporal interaction models. These studies demonstrate how such models can capture both localized variations and broader spatiotemporal evolution patterns.

In highway safety research, spatiotemporal modelling approaches have become more complex. Although advanced models often improve predictive performance, incorporating complex spatial random effects may distort the estimation of fixed effects due to overlapping explanatory power. This issue, known as spatial confounding, occurs when fixed covariates (e.g., traffic volume or geometric characteristics) explain part of the spatial variation that random effects simultaneously attempt to model (Hui and Bondell, 2022; Reich et al., 2006). Consequently, fixed effect coefficients may become statistically insignificant or even producing the opposite effect (Zimmerman and Ver Hoef, 2022), potentially leading to misguided safety inferences and policy decisions.

In other fields, several methods have been proposed to mitigate the spatial confounding problem, such as the restricted spatial regression method (Reich et al., 2006). However, spatial confounding has received limited attention in traffic safety research. Given its potential impact on model reliability, addressing spatial confounding is critical for collision causal analysis with a path analysis framework.

Therefore, to address the challenges of [i] insufficient consideration of indirect effects, [ii] latent unobserved heterogeneity, and [iii] spatial confounding, we develop a path analysis framework that explicitly incorporates spatiotemporal random effects, with additional constraints imposed to mitigate the potential collinearity induced by the inclusion of spatial random components.

3. Data preparation

3.1. Data source

To model and predict highway collisions, this study integrates multiple data sources. Those include: freight floating vehicle trajectories, roadway geometric and structural parameters, traffic flow data, and meteorological conditions. All data use in this study was approved by Carleton University Research Ethics Board (CUREB-B; project No. 123194). The study area covers two freeways in Southwest China: the GX freeway and the GY freeway, as shown in Fig. 1.

The floating vehicle trajectory dataset records the latitude and longitude coordinates of freight vehicles at one-second from June 2023 to September 2023. After filtering trajectories based on operating speed, weather and traffic volume conditions, a total of 738 valid trajectories sequences met the requirements.

Roadway alignment and structural data were obtained from highway design documents, and the key alignment parameters are shown in Table 2. Both highways were designed with a design speed of 120 km/h. Traffic data were collected from survey stations and electronic toll collection (ETC) gantries along the two highways. This provided hourly records of traffic volume, headway time, and other flow information for the period of February 2020 to January 2024. Collision data were obtained for the same time span, and after excluding collisions caused by mechanical failures, a total of 1172 valid collisions were included in the analysis. In addition, meteorological conditions were obtained from the same database, covering the period from February 2020 to January 2024. The only meteorological variable retained for modelling at the segment level is rain. Other meteorological data such as fog were not available and addressed as limitations and future work.

Based on the literature review in Section 2.1, which examined the relationship between different forms of speed and collision frequency, this study selects speed difference as the mediating variable.

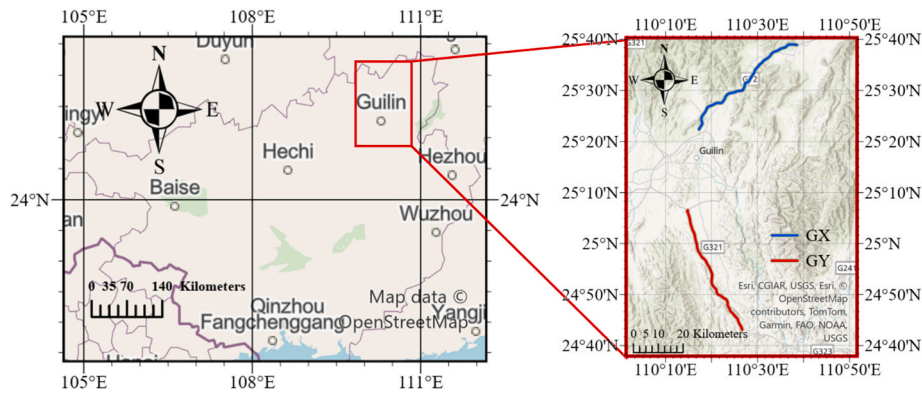


Fig. 1. Floating car trajectory route.

Table 2

Basic route indicators for GX and GY freeways.

Statistical indicator	GX Freeway				GY Freeway			
	R/m	W ^a /m	W _s /m	i	R/m	W ^a /m	W _s /m	i
Average	2137	24.50	2.50	0.05 %	3738	28.00	3.50	−0.07 %
Maximum	4000	24.50	2.50	4.00 %	8615	28.00	3.50	3.6 %
Minimum	1000	24.50	2.50	−2.93 %	752	28.00	3.50	−3.37 %
SD	872	0	0	1.59 %	2314	0	0	1.15 %

Table legend: W represents the cross-section width, R represents the radius, W_s represents the hard shoulder width, i represents the slope.

^a Both highways are four-lane divided highways, with two lanes in each direction, a lane width of 3.75 m, and an earth shoulder width of 0.75 m. However, the median width is 3.0 m on the GX Highway and 4.5 m on the GY Highway.

Considering the characteristics of the trajectory data, the speed difference indicator is constructed from operating speed. Following Khan et al. (2023), it is defined as the mean difference between the operating speed under free-flow conditions on a sunny day and the overall average operating speed of the entire route, as shown in Equation (1). The trajectory data were used to derive the segment level operating speed difference rather than to estimate long term traffic exposure. Since the roadway geometry, speed limits, and traffic organization of the studied sections remained generally stable during the study period, the operating speed difference extracted from the June–September 2023 trajectory data was assumed to represent the typical relative speed-difference characteristics among road sections.

$$\Delta v_i = \frac{1}{K_i} \sum_{k=1}^{K_i} |v_{ik} - \bar{v}| \quad (1)$$

where Δv_i is the speed difference for section i , v_{ik} is the operating speed at the k -th 10 m section i , $\bar{v}$ is the average operating speed of the entire route, and K_i is the total number of 10 m stations in section i . Subscript i refers to the index of road segment along the route, and subscript k refers to the index of a station point within that section.

3.2. Road section division

A total of 229 distinct road segments were delineated in this study using a homogeneity-based segmentation approach. Each segment was defined to maintain a uniform geometric characteristics, and interchange areas were treated as individual segments. According to Hou et al. (2018), the extent of an interchange was defined as the area extending from 450 m upstream of the beginning of the taper section to 450 m downstream of the end of the taper. However, section length has an important impact on modeling accuracy. Segments that are too short often result in many sections with zero collisions. Excessively long segments substantially reduce the number of modeling samples. According to Hou (2019), subdivided segments of equal length within a homogeneous road section are assumed to share the same risk rate. This

implies a proportional relationship between distance and collision frequency. For the derivation from Equation (2) to Equation (3), the local collision-frequency y_i was further assumed to be approximately linearly related to the distance-related variable x_i . Under this assumption, the discrete integral of collision frequency $F(x)$ exhibits a quadratic relationship with distance. Taking the logarithmic transformation gives rise to Equation (3), which corresponds to a line with a slope of 2 in the log-log scale. The maximum section length was determined as 779 m based on Equation (2). And the minimum section length was obtained as 292 m according to Equation (3). As a result, a total of 214 sections were generated, as shown in Fig. 2. The variables used for analysis in this study are shown in Table 3.

$$F(x_i) = \sum_{i=1}^n [y_i \times (x_{i-1} + x_{i+1})/2] \quad (2)$$

$$\ln[F(x)] = 2\ln x + \ln(0.5a) \quad (3)$$

where $F(x_i)$ is the discrete integral value on the i -th road section, representing the cumulative number of collisions up to that section. y_i is the accident frequency (number of collisions per unit length) on the i -th road section. x_{i-1} and x_{i+1} are the lengths of the adjacent road sections, where $x_0=0$ and the last $x=0$, and a is the intercept of the fitted quadratic relationship.

The collision frequency distribution of the GX and GY freeways after segmentation is shown in Fig. 3. A Moran's I test was conducted to evaluate spatial correlation. Results indicate that collision frequency does not exhibit significant spatial autocorrelation at the overall level. However, when further stratified by season (Spring: February–April, Summer: May–June, Fall: August–October, Winter: November–January), a distinct pattern emerges. In spring, collision frequency exhibits a significant positive spatial autocorrelation, characterized by spatial clustering of segments with similar collision frequencies. In contrast, winter demonstrates a negative spatial autocorrelation, reflected by a high–low combination pattern of collision frequencies across adjacent segments, as summarized in Table 4.

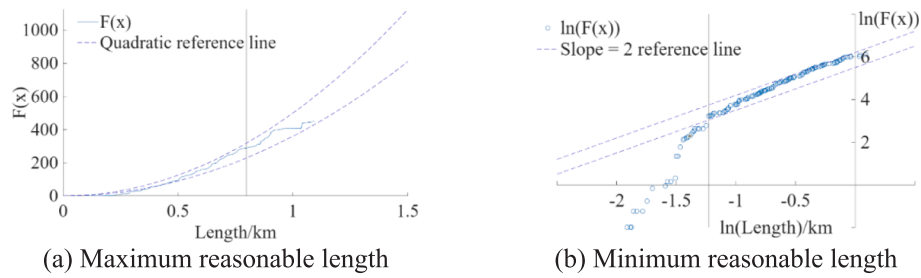


Fig. 2. Reasonable length of road section.

Table 3

Descriptive statistics of variables used for analysis.

Indicators	Unit	Minimum	Maximum	Average
Collision frequency	/	0	16	1.37
Segment length	m	292	779	564
Average daily (total 1460 days) volume	pcu	5587	15,415	10,628
Slope*	/	0.30 %	4.00 %	1.03 %
Curvature	km ⁻¹	0	1.33	0.31
Speed Difference	km/h	0.07	10.35	1.97
Average daily rainfall	mm	2.13	9.47	5.23

* The slope is represented in terms of its magnitude, regardless of direction.

3.3. Preliminary analysis

A correlation analysis was conducted among influencing factors, and a two-dimensional heat maps were plotted to illustrate the relationship between collision frequency and each factor. The results are shown in

Table 4

Moran' I test results.

Collision frequency group	Moran' I	p
Total	0.035	0.282
Spring	0.173	0.004
Summer	0.013	0.399
Fall	0.084	0.100
Winter	-0.190	0.004

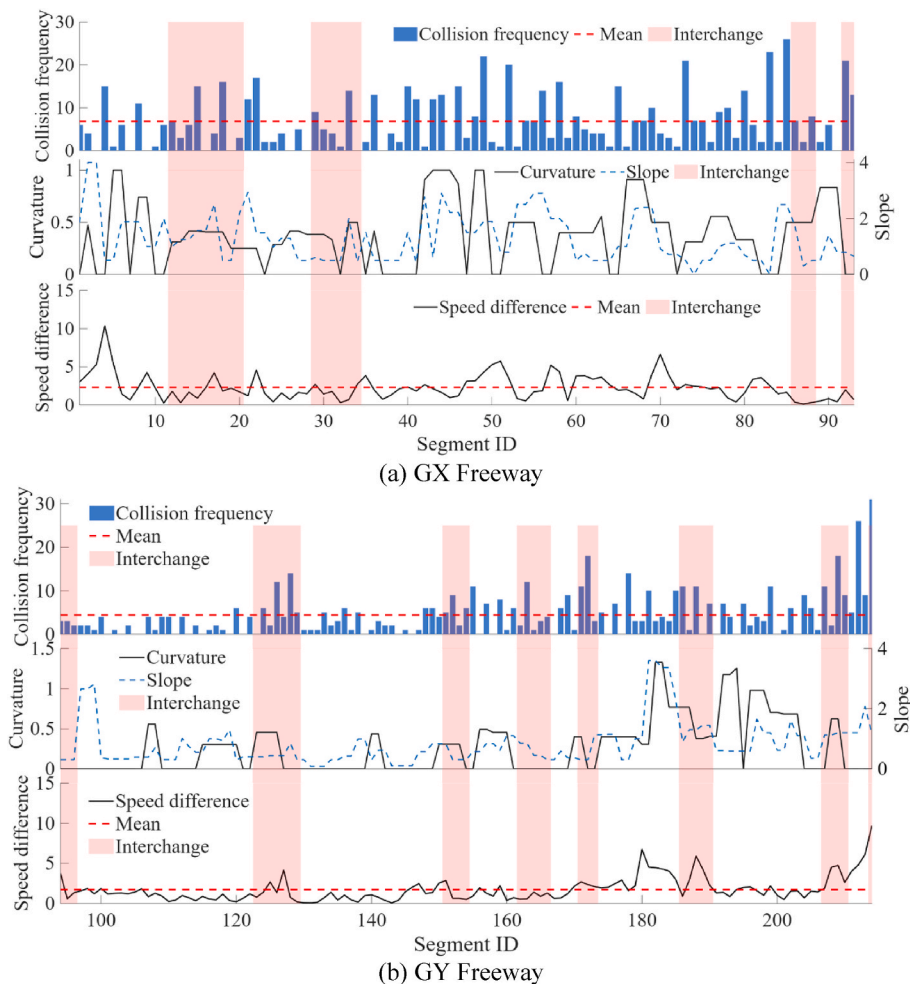


Fig. 3. Collision frequency distribution.

Fig. 4. The correlations among the factors are generally low. The highest correlation coefficient of 0.31 observed between slope and curvature. This indicates that the variables listed in Table 3 are suitable for modeling, as multicollinearity is minimal. Moreover, no clear patterns can be visually identified between collision frequency and the influencing factors in the heat maps, suggesting that a comprehensive multivariate modeling approach is necessary to capture the combined effects of multiple variables.

4. Methodology

The purpose of the methodological framework is to examine how roadway characteristics affect collision frequency both directly and indirectly through operating speed difference. To improve the robustness of the analysis, the model also accounts for unobserved heterogeneity. Collision frequency may vary across road segments and time periods due to factors that are not fully observed in the dataset, such as local traffic signs and markings, pavement condition, drainage facilities, visibility, enforcement intensity, or short-term traffic disturbances. To partially capture these unobserved influences, spatial, temporal, and spatiotemporal interaction random effects are incorporated into the collision-frequency model. For the operating speed difference model, only spatial random effects are considered, because the speed-difference variable is derived at the segment level.

The overall model framework is presented in Fig. 5. The framework consists of five main steps. First, a path analysis model is constructed to evaluate the direct and indirect effects of roadway characteristics on collision frequency through operating speed difference. In this framework, operating speed difference is treated as a mediating variable, linking roadway characteristics to collision frequency. Second, spatial, temporal, and spatiotemporal random effects are sequentially introduced to capture residual heterogeneity in collision frequency that cannot be fully explained by the observed variables. Third, spatial confounding is examined, because spatial random effects may overlap with fixed effects and distort the estimation of roadway-related

coefficients. When spatial confounding is detected, a restriction matrix and projection adjustment are applied to reduce the potential interference of spatial random effects with fixed-effect estimates.

Fourth, all model parameters are estimated under a Bayesian framework using the Integrated Nested Laplace Approximation (INLA) method, which is a deterministic Bayesian inference approach for efficiently approximating posterior distributions in latent Gaussian models. Finally, model performance is evaluated using both goodness-of-fit and prediction-error indicators. The Deviance Information Criterion (DIC) is used to compare model fit while penalizing model complexity, whereas the Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), and Root Mean Square Error (RMSE) are used to assess prediction accuracy. Specifically, MAPE measures the average relative prediction error in percentage terms, MAD measures the average absolute difference between observed and predicted values, and RMSE measures the square root of the average squared prediction error, giving greater weight to larger errors.

4.1. Path analysis

Path analysis quantifies both the direct impact of explanatory variables on the dependent variable and their indirect impact through mediating variables. This method has been widely used in transportation safety research. Xu et al. (2025) integrated path analysis with an interpretable machine learning framework to examine both the direct impact of vehicle speed and weather on collision occurrence, and their indirect effects mediated through pre-driving behaviors. Ahmad et al. (2023b) analyzed the direct effects of distraction duration on safety-critical events, as well as its indirect effects on such events through driving instability.

Path analysis is a special case of the structural equation model, which is used to analyse relationships among manifest variables, which are variables directly observed rather than latent (Gargoum and El-Basyouny, 2016). In this study, the speed difference index constructed based on the operating speed is used as a mediating variable to capture

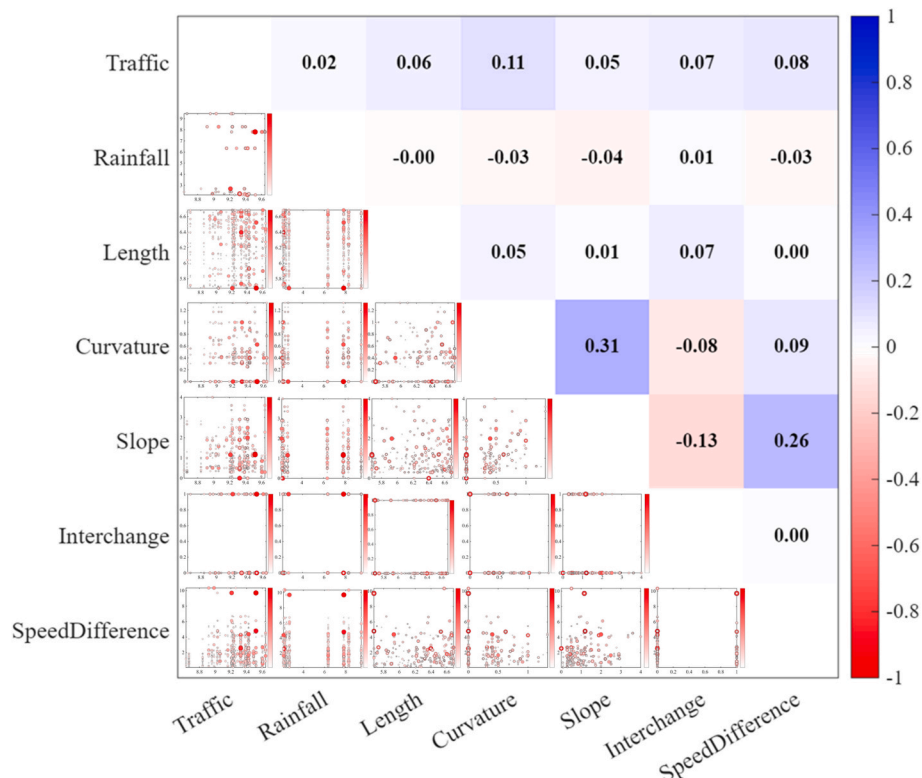


Fig. 4. Correlation matrix and two-dimensional heat map visualization of collision frequency and influencing factors.

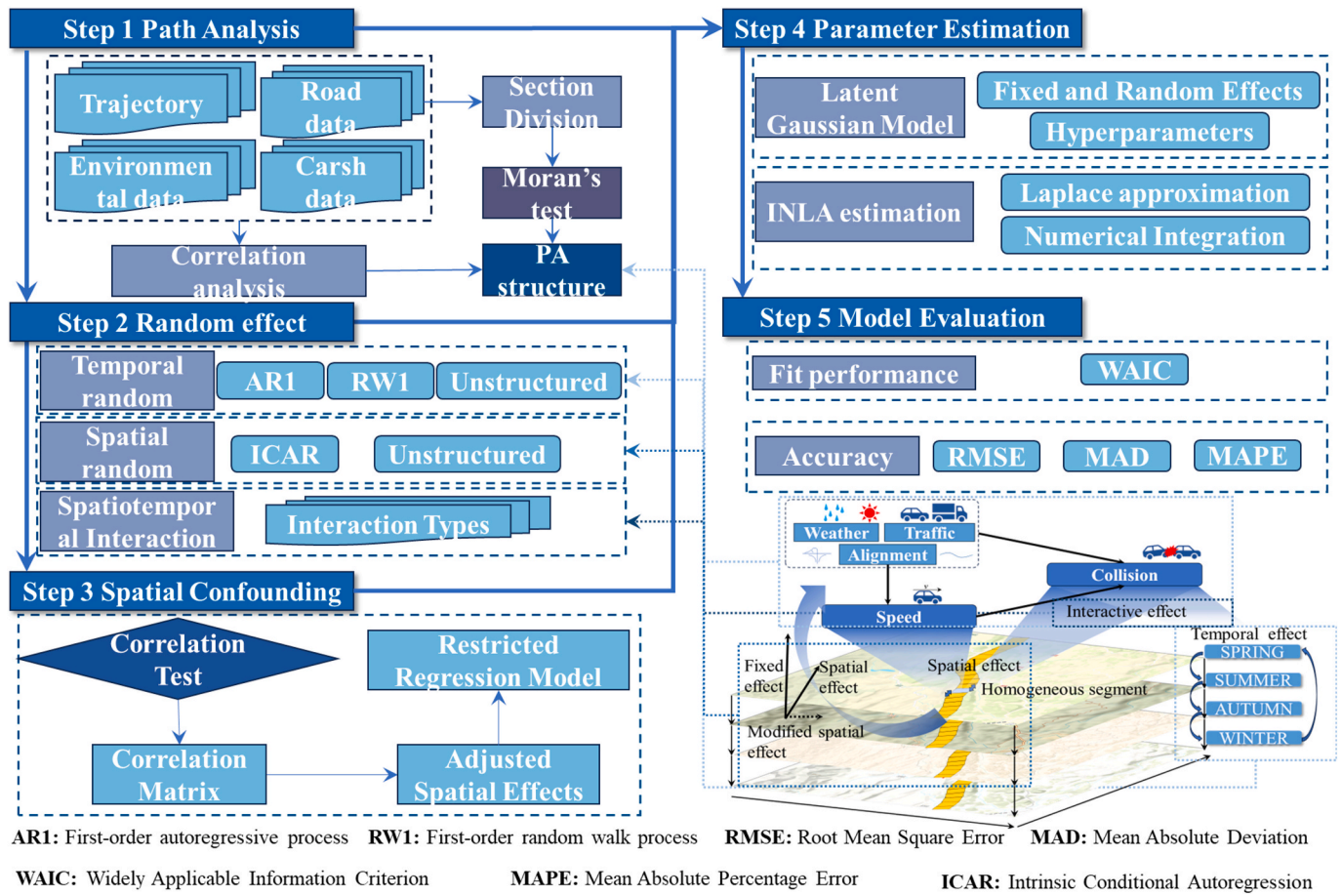


Fig. 5. Conceptual framework of the spatiotemporal path analysis model.

both direct and indirect effects of geometric and environmental factors on collision frequency. At the same time, direct impacts of structural, traffic, and metrological variables on collisions are considered, as shown in Fig. 6. Due to data limitations, local roadway factors such as traffic signs and markings, pavement condition, lighting, drainage facilities, and sight distance could not be explicitly included in the model. To partially account for the potential unobserved heterogeneity caused by these omitted factors, spatial and temporal random effects were incorporated into the modeling framework. The two corridors studied have shared attributes such as number of lane, lane width etc. Therefore, these attributes could not be identified as fixed effects.

The study considers two causal paths: (i) collision - spatiotemporal factors, and (ii) collision - speed difference - geometric indicators.

Among them, the collision frequency is treated as a discrete variable, which can be regarded as the result of the Poisson process (Wang et al., 2025). However, the actual traffic collision frequency has the phenomenon of overdispersion, which is difficult to meet the assumption that the mean of the Poisson distribution is equal to the variance. To address this, the study adopts a negative binomial distribution model specification by introducing a dispersion parameter (r) to effectively capture the heterogeneity and discreteness of the collision frequency data.

In the collision-speed difference-geometric indicators path, since speed difference is a continuous variable, this study uses the linear index radius and slope as influencing factor variables and establishes a model through linear regression. As shown in the following Eqs. (4) and (5):

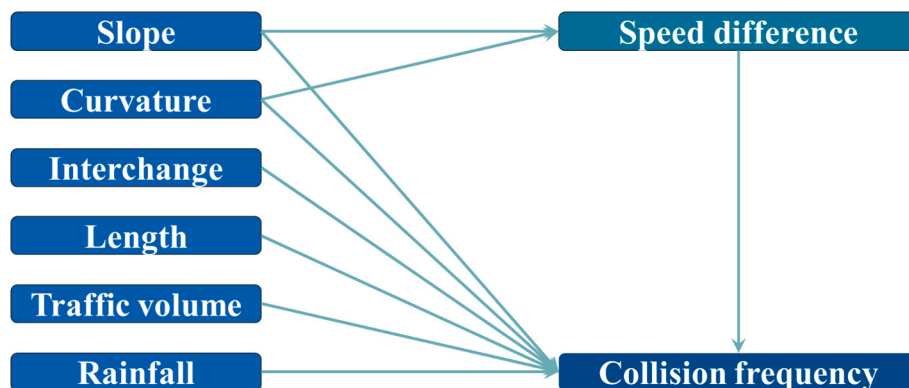


Fig. 6. General structural equation model.

$$P(Y_{it} = y_{it}) = \frac{\Gamma(r + y_{it})}{\Gamma(r) \Gamma(y_{it} + 1)} \left(\frac{r}{r + \mu_{it}} \right)^r \left(\frac{\mu_{it}}{r + \mu_{it}} \right)^{y_{it}} \quad (4)$$

$$\ln(\mu_{it}) = \alpha_0 + \alpha_1 \ln(ADT_{it}) + \alpha_2 \ln(L_i) + \beta^T \mathbf{X}_{it} + \zeta \Delta v_i, \Delta v_i = \delta_0 + \theta^T \mathbf{Z}_i \quad (5)$$

where:

- y_{it} is the collision frequency of the road section i in season t
- r is the dispersion parameter
- μ_{it} is the expected collision frequency of the road section i in season t
- $\Gamma(\cdot)$ is the gamma function
- ADT_{it} is the average daily traffic volume of the road section i in season t
- L_i is the length of the road section i
- $\mathbf{X}_{it}$ is vector of covariates influencing collision frequency (include rainfall, slope and curvature three variables)
- $\mathbf{Z}_i$ is vector of covariates influencing speed difference (include slope and curvature two variables)
- m_i is the operating speed difference
- $\alpha_0, \delta_0, \alpha_1, \alpha_2, \zeta$, are intercepts and coefficients
- β^T and θ^T are coefficient vectors
- $t = \{1, 2, 3, 4\}$ represents spring, summer, fall, and winter

Since the occurrence of traffic collisions is often the result of the combined effects of multiple factors, such as driver behavior, road infrastructure, traffic environment and natural environment, in addition to the influencing factors collected in Section 3.1, there may also be some potential causes that cannot be directly observed (Manning et al., 2016). Ignoring these unobserved factors may lead to bias in parameter estimation. To address this, a random intercept term τ_{it} is introduced to fully represent unobserved heterogeneity of each road section and time period combination. This allows the model to adapt to local variations in baseline collision risk. The final model system is expressed as in Eqs. (6).

$$\ln(\mu_{it}) = \alpha'_0 + \alpha_1 \ln(ADT_{it}) + \alpha_2 \ln(L_i) + \beta^T \mathbf{X}_{it} + \zeta \Delta v_i, \alpha'_0 = \alpha_0 + \tau_{it} \quad (6)$$

where τ_{it} is the random intercept capturing unobserved heterogeneity across road sections, and is assumed to follow a normal distribution.

4.2. Path analysis spatiotemporal model

Even after the fixed effects in the path analysis model are accounted for, a portion of the variation in collision frequency still remains unexplained since not every contributing factor is observed at the segment level. The role of the random components introduced below is to absorb that partial variation. This makes the fixed effect coefficients of interest not biased by the unmodelled spatial or temporal patterns. To capture unobserved spatial and temporal heterogeneity, different random components were sequentially added to the path analysis framework. Accordingly, four hierarchical models were developed: the spatial path analysis model, the temporal path analysis model, the spatiotemporal path analysis model, and the spatiotemporal interaction path analysis model. The stepwise structure allows the contribution of each random component to model performance to be evaluated.

4.2.1. Spatial path analysis model

As discussed in the literature review, both collision frequency and speed may exhibit spatial correlation, meaning that observations from neighbouring road segments are not necessarily independent. To account for this spatial structure, a BYM spatial random term is added to both the collision frequency and speed difference equations. The BYM specification decompose the unobserved spatial variability into two spatial components: (i) a structured spatial effect, represented through an Intrinsic Conditional Autoregression (ICAR), which captures correlations among adjacent road segments; and (ii) an unstructured spatial

effect, which accounts for segment-specific variability that is not spatially correlated. This formulation allows the model to capture systematic spatial differences in safety and speed outcomes beyond the ones explained by geometric, traffic, and environmental variables. The full model is expressed as in Eqs. (7)–(8):

$$\ln(\mu_{it}) = \alpha'_0 + \alpha_1 \ln(ADT_{it}) + \alpha_2 \ln(L_i) + \beta^T \mathbf{X}_{it} + \zeta \Delta v_i + \phi_i + \varepsilon_i, \Delta v_i = \delta_0 + \theta^T \mathbf{Z}_i + \theta_i + \varepsilon'_i \quad (7)$$

$$\phi_i \sim N\left(\frac{\sum_{j \in N_i} \omega_{ij} \phi_j}{\sum_{j \in N_i} \omega_{ij}}, \frac{\delta_s^2}{\sum_{j \in N_i} \omega_{ij}}\right), \theta_i \sim N\left(\frac{\sum_{j \in N_i} \omega_{ij} \theta_j}{\sum_{j \in N_i} \omega_{ij}}, \frac{\delta_a^2}{\sum_{j \in N_i} \omega_{ij}}\right) \quad (8)$$

where ϕ_j and θ_j are the structured spatial effects for collision frequency and speed difference, respectively, each following an ICAR process. ω_{ij} is the adjacent weight between road segments i and j , it is equal to 1 if the two are connected and 0 otherwise. δ_s^2 and δ_a^2 are variances of the structured spatial terms. The random errors ε_i and ε'_i are unstructured spatial effects, independently and identically distributed as $N(0, \sigma^2)$. Each structured spatial effect ϕ_i depends on the weighted mean of its adjacent segments. This reflects the assumption that neighboring segments tend to exhibit similar spatial patterns in collision frequency. Similarly, the structured spatial effect θ_i for operating speed difference. It assumes that adjacent road segments experience comparable speed-related conditions due to shared geometric and environmental characteristics.

4.2.2. Temporal path analysis model

To capture temporal correlation in collision frequency, the dataset was divided into four seasonal periods: spring, summer, fall, and winter. Based on the random-effect path analysis model, the temporal path analysis model incorporates both structured and unstructured temporal random effects. The structured temporal effect accounts for correlations across time, while the unstructured component captures independent temporal variability. The general model is expressed as:

$$\ln(\mu_{it}) = \alpha'_0 + \alpha_1 \ln(ADT_{it}) + \alpha_2 \ln(L_i) + \beta^T \mathbf{X}_{it} + \zeta \Delta v_i + \eta_t + \nu_t \quad (9)$$

where η_t represents the structured temporal effect, ν_t is the unstructured temporal random term. The structured temporal effect η_t can follow two alternative stochastic processes:

- First-order autoregressive process (AR1)

The temporal effect at time t depends on its value at the previous time step, as expressed in Equation (10):

$$\eta_t | \eta_{t-1} \sim N(\rho \eta_{t-1}, \sigma_\eta^2) \quad (10)$$

where ρ is the autoregressive coefficient ($|\rho| < 1$) and σ_η^2 is the variance of the temporal term.

- First-order random walk process (RW1)

The temporal effect evolves gradually as a random walk, as shown in Equation (11):

$$\eta_t | \eta_{t-1} \sim N(\eta_{t-1}, \sigma_\eta^2) \quad (11)$$

where σ_η^2 is the variance representing the magnitude of temporal fluctuation.

4.2.3. Spatiotemporal path analysis model

To jointly capture both spatial and temporal effects in collision frequency, a spatiotemporal path analysis model is established by integrating the structured spatial and temporal components introduced

earlier. This approach allows the model to account for correlations among neighboring road segments and sequential dependence across seasons. The model is expressed in Equation (12):

$$\ln(\mu_{it}) = \alpha'_0 + \alpha_1 \ln(ADT_{it}) + \alpha_2 \ln(L_i) + \beta^T \mathbf{X}_{it} + \zeta \Delta v_i + \phi_i + \varepsilon_i + \eta_t + \nu_t \quad (12)$$

where, ϕ_i represents the structured spatial effect, modeled using ICAR prior to capture spatial autocorrelation between adjacent road segments, ε_i is the unstructured spatial effect, modeled as an independent and identically distributed (i.i.d.) Gaussian term to capture local random variation within individual segments. The temporal effects consist of structured component η_t (modeled using AR1 or RW1 process), and an unstructured component ν_t . This specification allows the model to jointly capture spatial autocorrelation, local spatial heterogeneity, and temporal dependence across seasons.

4.2.4. Spatiotemporal interaction path analysis model

Although the spatial and temporal random effects account for structured and unstructured heterogeneity across road segments and time periods separately, collision frequency may also be influenced by interactions between the spatial and temporal dimensions. This means that spatial differences in collision frequency may not remain constant over time, and temporal variation may not affect all road segments in the same way. To represent this additional source of variation, a spatiotemporal interaction term, $\xi_{i,t}$ was introduced into the model to capture effects that are specific to each road segment and time-period combination. The interaction term is expressed as:

$$\ln(\mu_{it}) = \alpha'_0 + \alpha_1 \ln(ADT_{it}) + \alpha_2 \ln(L_i) + \beta^T \mathbf{X}_{it} + \zeta \Delta v_i + \phi_i + \varepsilon_i + \eta_t + \nu_t + \xi_{i,t} \quad (13)$$

where, $\xi_{i,t}$ represents the spatiotemporal interaction effect, which captures variations specific to each spatial/temporal combination. This ef-

fect is modeled as a multidimensional Gaussian process that can be expressed as the Kronecker product of the spatial and temporal random effect structures. We consider four interaction types depending on whether the spatial and temporal components are structured or unstructured (Fig. 7a-d):

- Type I: Unstructured Time $\times$ Unstructured Space (Fig. 7a)

This interaction represents independent random variations over both space and time, without any structured dependence in either dimension:

$$\xi_{it} \sim N(0, 1/\varphi_\xi) \quad (14)$$

where φ_ξ is the precision parameter controlling the variability of the unstructured space-time interaction term.

- Type II: Structured Time $\times$ Unstructured Space (Fig. 7b)

This interaction allows each road segment to have a distinct temporal trend that evolves smoothly over time but without spatial structure:

$$\xi_{it} \sim \text{AR1}(\rho, 1/\varphi_t) \quad (15)$$

where ρ is the temporal autocorrelation coefficient and φ_t is the precision parameter governing the variability of the temporal component.

- Type III: Unstructured Time $\times$ Structured Space (Fig. 7c)

This interaction captures situations where each season exhibits its own spatial structure, modeled using CAR prior:

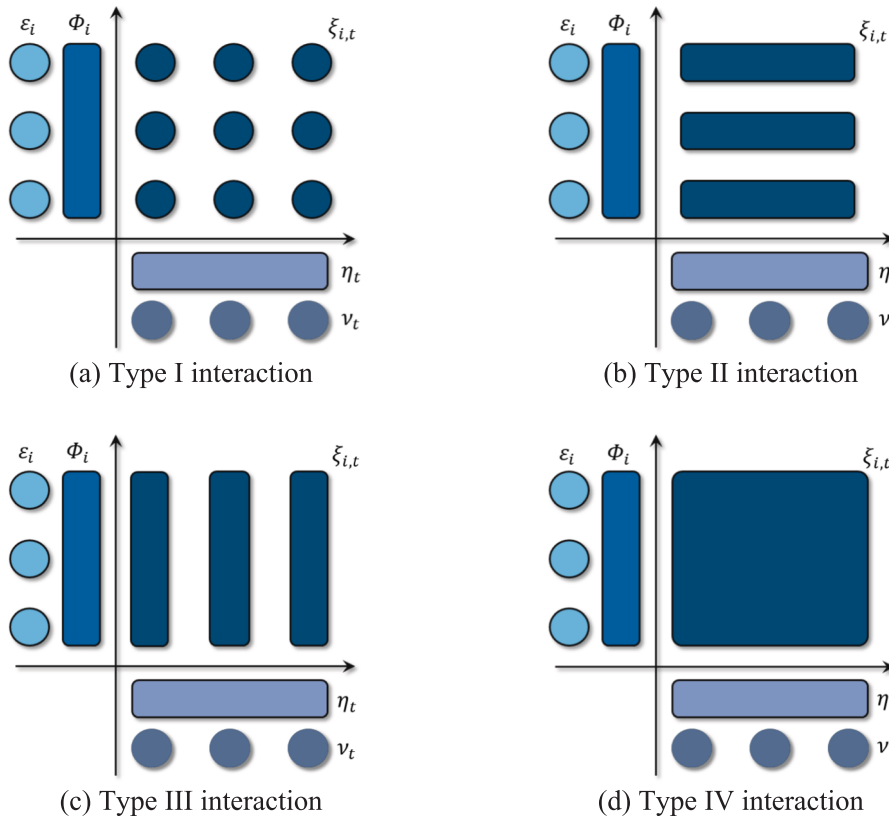


Fig. 7. Spatiotemporal Interaction Item.

$$\xi_{it} \sim \text{CAR}(\psi_i, 1/\varphi_\xi) \quad (16)$$

where ψ_i is the weighted mean of the neighboring spatial units, and φ_ξ is the precision parameter controlling spatial variability.

• Type IV: Structured Time $\times$ Structured Space (Fig. 7d)

This interaction represents the most comprehensive form; it allows both spatial and temporal dependencies to evolve together. It is modeled as the Kronecker product of the CAR (space) and AR1 (time) processes:

$$\xi_{it} \sim \text{CAR}(\psi_i, 1/\varphi_\xi) \otimes \text{AR1}(\rho, 1/\varphi_t) \quad (17)$$

These four formulations allow the model to flexibly capture different patterns of space–time interaction, ranging from purely random fluctuations (Type I) to highly structured spatiotemporal evolution (Type IV).

4.3. Spatial confounding analysis

Spatial confounding refers to the overlap between the explanatory information contained in observed fixed-effect variables and that captured by structured spatial random effects. In the context of this study, roadway geometric, traffic, or environmental variables may exhibit spatial patterns similar to those represented by spatial random effects. When such overlap exists, the spatial random effects may absorb part of the variation that should be attributed to the fixed effects. This can lead to biased or unstable coefficient estimates, enlarged uncertainty, and even changes in the sign of fixed-effect parameters. Consequently, the estimated effects of roadway characteristics on operating speed difference and collision frequency may be difficult to interpret reliably.

To reduce this problem, the potential collinearity between fixed effects and structured spatial random effects was first examined. Specifically, the spatial path analysis model established in Section 4.2 was used as the baseline model, and correlation analysis was conducted between the fixed-effect variables and the estimated structured spatial random effects. The fixed effects that are collinear with spatial random effects and spatial random effects are identified in the mediation model and the outcome model, as shown in Equation (18):

$$\text{corr}(x_m, S) = \frac{\sum_{i=1}^n (x_{im} - \bar{x}_m)(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (x_{im} - \bar{x}_m)^2 \sum_{i=1}^n (S_i - \bar{S})^2}} \quad (18)$$

where x_{im} is the value of the m -th fixed-effect variable for the i -th road section, and S is the structured spatial random effect.

To mitigate this spatial confounding, this study adopts the Restricted Spatial Regression (RSR) framework proposed by Reich et al. (2006). This approach imposes linear constraints on the spatial random effects to ensure orthogonality between the spatial and fixed components. This can effectively remove their overlapping explanatory influence. The process can be expressed as shown in Eqs. (19)–(21).

$$X^T \tilde{S} = 0 \quad (19)$$

$$M_X = I - P_X, \quad P_X = X(X^T X)^{-1} X^T \quad (20)$$

$$\tilde{S} = M_X S \quad (21)$$

where:

- X is the design matrix for fixed effects,
- S is the original structured spatial random-effect vector or spatial basis matrix,
- P_X is the projection matrix onto column space of X ,
- M_X is the projection matrix onto the orthogonal complement of X ,
- I is the identity matrix, and
- $\tilde{S}$ is the adjusted spatial basis that eliminates collinearity between the fixed and spatial random effects.

4.4. INLA Parameter Estimation

The models developed in this study contain multiple random-effect components, including spatial, temporal, and spatiotemporal interaction effects. These components increase the complexity of Bayesian estimation and make sampling-based methods such as Markov Chain Monte Carlo computationally demanding. Therefore, the Integrated Nested Laplace Approximation method was adopted because it provides an efficient deterministic approach for estimating latent Gaussian models with complex random-effect structures.

The INLA parameter estimation method estimates model parameters by using the Laplace approximation technique to fit a Gaussian distribution to a complex posterior distribution (Rue et al., 2009). The method assumes a latent Gaussian model in which both the latent random variables and the coefficients to be estimated follow a joint Gaussian prior. All fixed effects and random effects together constitute a Gaussian Markov random field to ensure that its precision matrix has a sparse structure and is computationally efficient. In addition, the observation layer is assumed to follow an exponential family distribution and be conditionally independent of each other after the latent variable is given. Statistical analyses were conducted in R (version 4.4.2) using the INLA package (INLA 25.04.16) for Bayesian inference.

The INLA algorithm approximates the posterior distribution in three main steps. First, the algorithm is expanded as a second-order Taylor series around its mode, as shown in Equation (22):

$$h(\gamma) \approx h(\hat{\gamma}) + \frac{1}{2}(\gamma - \hat{\gamma})^T \nabla^2 h(\hat{\gamma})(\gamma - \hat{\gamma}) \quad (22)$$

Second, the quadratic term is expressed similar to a Gaussian distribution, as shown in Equation (23):

$$\pi(\gamma|D, \lambda) \sim N(\hat{\gamma}(\lambda), \nabla^2 h(\hat{\gamma})^{-1}) \quad (23)$$

Finally, the marginal posterior distribution is obtained using the classical Gaussian integral, as shown in Equation (24):

$$\tilde{\pi}(\lambda|D) = \pi(\lambda)(2\pi)^{\frac{d}{2}} |H(\hat{\gamma}(\lambda))|^{-\frac{1}{2}} \exp(-h(\hat{\gamma}(\lambda))) \quad (24)$$

This procedure provides an accurate and computationally efficient approximation of the posterior distribution. Compared with traditional Markov Chain Monte Carlo (MCMC) methods, INLA achieves similar accuracy while substantially reducing computational time. The terms used in Eqs. (22)–(24) are defined as follows:

- γ is the latent variable to be inferred,
- D is observed data,
- $h(\gamma)$ is the negative log joint density,
- $\hat{\gamma}$ is the conditional posterior mode,
- $\nabla^2 h(\hat{\gamma})$ is the Hessian matrix evaluated at the mode,
- d denotes the dimension of γ , and
- $\pi(\gamma|D, \lambda)$ is the conditional posterior distribution of γ given D and λ .

4.5. Model evaluation

To assess the performance of the proposed models, four evaluation metrics were used: the Widely Applicable Information Criterion (WAIC), mean absolute deviation (MAD), mean absolute percentage error (MAPE), and root mean square error (RMSE). These metrics evaluate

both model fit and predictive accuracy, as shown in Eqs. (25)–(28):

$$WAIC = -2(lppd - p_{WAIC_2}) \quad (25)$$

$$MAD = \frac{\sum_{t=1}^4 \sum_{i=1}^{214} |Y_{it} - \mu_{it}|}{214 \times 4} \quad (26)$$

$$MAPE = \frac{\sum_{t=1}^4 \sum_{i=1}^{214} \frac{|\mu_{it} - Y_{it}|}{Y_{it}}}{214 \times 4} \quad (27)$$

$$RMSE = \frac{\sum_{t=1}^4 \sum_{i=1}^{214} (Y_{it} - \mu_{it})^2}{214 \times 4} \quad (28)$$

where

- $lppd$ is log point-wise predictive density, which measures model fit by summing the log predictive probabilities across all data points,
- p_{WAIC_2} is the effective number of parameters estimated by summing the variances of individual log predictive densities,
- μ_{it} is the predicted collision frequency for segment i in season t , and
- Y_{it} is the corresponding observed collision frequency.

5. Results and discussion

This section presents and discusses the empirical results of the proposed modeling framework. First, potential factors related to spatial confounding are identified and analyzed. Based on the spatial confounding analysis, restricted regression models are constructed for both spatial and spatiotemporal path analysis frameworks. The candidate models are then compared using goodness-of-fit and prediction-performance indicators. After selecting the optimal model, its fixed effects estimates and random effects components are interpreted in detail and compared with findings reported in previous studies.

5.1. Spatial confounding analysis

After establishing the spatial path analysis model, a correlation analysis was conducted between the spatial effect of the outcome variable (collision frequency) and the mediating variable (speed difference), as shown in Table 5. In the mediating path, both curvature and slope were significantly correlated with the spatial effect. In the collision frequency equation, rainfall, curvature, slope and interchange showed significant correlations with the structured spatial effect. To mitigate these dependencies, restriction matrices were constructed for speed difference and collision frequency. The orthogonally projected spatial effects were then used to establish the restricted spatial path analysis model.

The fixed effect estimation results were compared across the general path analysis model, the spatial path analysis model, and the restricted

spatial path analysis model, as shown in Fig. 8. “PA” denotes the path analysis model, “S” indicates inclusion of spatial random effects, “RE” represents models with random-effect intercepts, and “RS” corresponds to restricted spatial regression models. In the speed-difference equation, curvature was a positive and significant predictor in the general model. In the spatial model, slope was a significant negative factor. This pattern is different from previous findings, as vehicle speed on steep segments is jointly influenced by slope and vehicle performance, leading to higher variability.

In the collision-frequency equation, although the signs of most fixed effects did not change or was not significant, the standard deviations of slope and interchange increased. This led to an increase in the range of 97.5% BCI and an increase in the instability of fixed effects and reducing coefficient stability. This instability is likely due to spatial random terms competing with fixed effects to explain variability. This could be an indicator of spatial confounding. Notably, curvature also become statistically insignificant, which contradicts the existing large number of horizontal curve to higher collision rates (Afghari et al., 2023; Awasthi et al., 2024; Kar et al., 2024). This finding will be re-visited in the fixed effect analysis.

The appearance of spatial confounding in this study is likely to be related to the road segment definition and collision clustering characteristics (Wei et al., 2026). Because each section was homogeneously defined, the slope and curvature within each road segment were the same. In locations with clustered collisions, the spatial random effect forms a structural collinearity with the fixed effect, thereby destabilizing the fixed effect. Since this study uses path analysis to focus on analyzing the confounding effects of collision factors, a restricted space modeling method should be adopted to prevent spatial confounding from distorting the interpretation of fixed effects.

5.2. Model comparison

Based on the models described in Subsection 4.2, this study developed and compared several extended specifications to capture both spatial and temporal dependencies. These include a random effect path analysis model, two temporal path analysis models, two restricted spatiotemporal path analysis models, and four restricted space-time interaction path analysis models (types I–IV). Their corresponding evaluation indicators are shown in Table 6.

According to the WAIC results, the inclusion of random components substantially improved model performance. The WAIC decreased consistently from the general PA model to the REPA and further to the spatiotemporal path analysis models. This could indicate that accounting for structured heterogeneity and latent space–time correlation enhances explanatory power. Although the model with the RW(1) temporal term (ST2-RSREPA) had the lowest WAIC and error values, the differences among the spatiotemporal and spatiotemporal-interaction models were relatively small. This suggests that the marginal benefit of adding additional interaction layers beyond the core spatiotemporal formulation is limited. In addition, for the error-based indicators (RMSE, MAE, MAPE), a similar pattern was observed. Each successive model including random or spatiotemporal components showed reduced prediction error. The ST2-RSREPA model had the lowest overall errors, demonstrating its superior balance between model complexity and predictive accuracy. To further assess model adequacy, Moran’ I spatial autocorrelation tests were applied to the residuals of each model and compared with the observed collision frequencies. The results are shown in Table 7.

The observed collision frequencies exhibited significant spatial clustering in spring, summer, and fall. Winter showed a distinct spatial dispersion pattern. After adding random effects, residuals in the PA model shifted from insignificant clustering to significant negative autocorrelation. This could indicate that the structured spatial component captured most of the large-scale positive dependence, which left small scale contrasting effects (checkerboard-type patterns). Spatial

Table 5
Correlation analysis between fixed factors and structured spatial effects.

Factors	Correlation	P
Speed Difference Equation		
Curvature	0.20	0.00
Slope	0.48	0.00
Collision Frequency Equation		
ln (Traffic)	0.06	0.35
ln (Length)	0.00	0.99
Rainfall	−0.73	0.00
Curvature	0.19	0.00
Slope	0.24	0.00
Interchange	−0.15	0.00

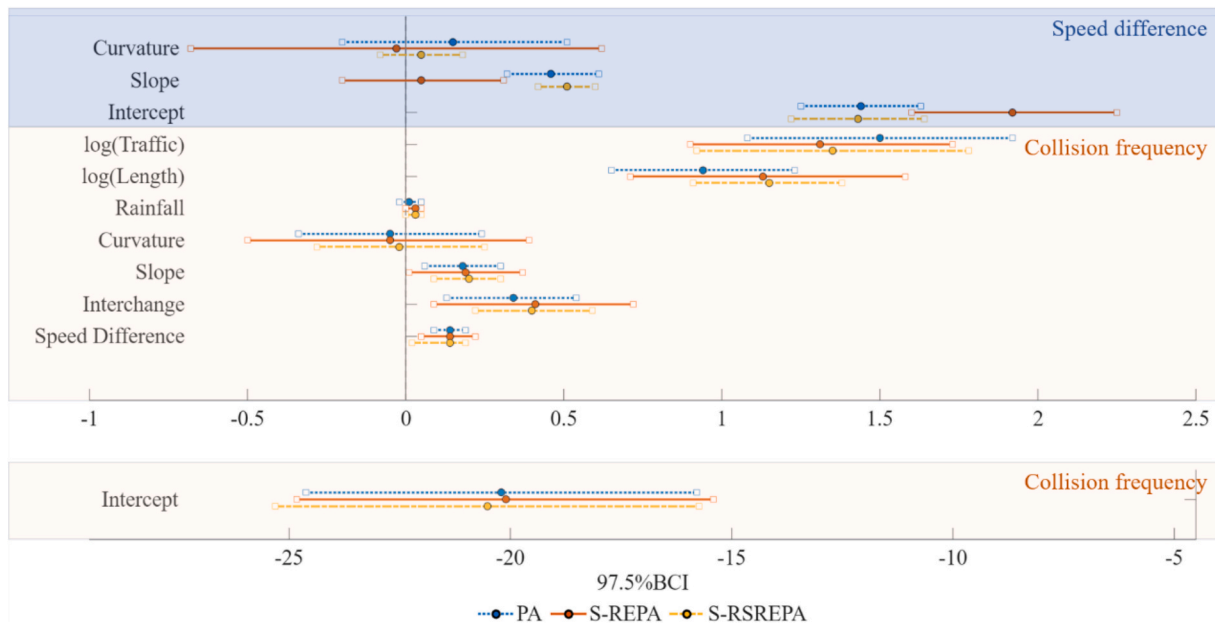


Fig. 8. Comparison between restricted regression model and unrestricted model.

Table 6

Model performance comparison across path analysis models with spatial, temporal, and spatiotemporal components.

Model	WAIC	MAE	MAPE	RMSE
PA	5634.67	1.326	54.44 %	1.903
REPA	-7958.25	1.268	51.55 %	1.825
S-RSREPA	-9174.18	0.908	46.99 %	1.320
T-REPA1	5824.63	1.285	53.88 %	1.851
T-REPA2	5824.74	1.285	53.86 %	1.851
ST1-RSREPA	-9206.44	0.875	47.13 %	1.267
ST2-RSREPA	-9221.44	0.738	39.34 %	1.006
STI1-RSREPA	-9205.59	0.874	47.07 %	1.265
STI2-RSREPA	-9205.71	0.865	46.57 %	1.241
STI3-RSREPA	-9212.48	0.873	47.19 %	1.265
STI4-RSREPA	-9209.25	0.861	46.29 %	1.238

Table legend: T represents the model with time random terms. 1 represents the use of AR1 random terms in the time model and space-time model. 2 represents the use of RW1 random terms. ST represents the model with space-time random terms. STI represents the model with space-time interaction random terms.

clustering or dispersion in residuals occurs when unmodeled spatiotemporal variation remains. Among all spatiotemporal and interaction models, the ST2-RSREPA model produced residuals with no significant spatial clustering or dispersion characteristics of all four seasons. This model included a BYM specification to model both spatially structured

and unstructured heterogeneity, and employed a RW1 process to account for temporal autocorrelation. Its residual Moran's I values were the lowest. This confirms its robust spatial adjustment. When combining the results of Tables 6 and 7, the addition of the spatiotemporal interaction term did not meaningfully improve model performance. This is consistent with previous research results (Wang et al., 2024b). This may indicate that over-parameterized interaction structures can reduce efficiency without enhancing explanatory power. Therefore, the ST2-RSREPA model was selected as the optimal model specification for subsequent analysis.

5.3. Fixed effects

The fixed effect parameters of the ST2-RSREPA model are shown in Table 8. This model includes two causal paths: (1) the collision frequency-spatiotemporal path, and (2) the collision frequency-speed difference-linear geometric condition path. Compared with the general path analysis model, the magnitude and direction of most fixed effects remained consistent. This implies the model's robustness after accounting for spatial and temporal structure. In the Bayesian estimation framework, statistical significance was evaluated according to whether the posterior credible interval excluded zero. Specifically, a parameter was regarded as statistically significant if its 97.5% credible interval was entirely positive or entirely negative.

Table 7

Moran's I analysis of collision frequency and model residuals.

Model	Spring		Summer		Fall		Winter	
	Moran' I	p	Moran' I	p	Moran' I	p	Moran' I	p
<i>Collision frequency</i>	0.17	0.00	0.01	0.40	0.08	0.10	-0.19	0.00
PA	0.07	0.13	-0.13	0.39	0.04	0.27	-0.18	0.06
REPA	-0.16	0.01	-0.11	0.68	-0.01	0.54	-0.18	0.00
S-RSREPA	-0.13	0.03	0.01	0.39	0.09	0.09	0.01	0.20
T-REPA1	-0.15	0.01	-0.11	0.64	-0.15	0.56	-0.21	0.00
T-REPA2	-0.15	0.01	-0.11	0.64	-0.15	0.56	-0.21	0.00
ST1-RSREPA	-0.11	0.68	0.01	0.41	0.11	0.05	-0.05	0.25
ST2-RSREPA	-0.11	0.06	0.00	0.45	0.10	0.07	-0.04	0.28
STI1-RSREPA	-0.12	0.04	0.01	0.41	0.11	0.01	-0.05	0.25
STI2-RSREPA	-0.15	0.02	-0.00	0.48	0.10	0.06	-0.06	0.20
STI3-RSREPA	-0.11	0.07	0.01	0.41	0.11	0.05	-0.05	0.25
STI4-RSREPA	-0.13	0.04	0.01	0.42	0.10	0.05	-0.06	0.22

Table 8
ST-posterior parameter estimates of the RSREPA2 model.

Indicator	Mean	SD	Incidence Rate Ratio	97.5 % BCI
Speed Difference Equation				
Intercept	1.43	0.11	/	(1.22, 1.65)
Curvature	0.05	0.07	1.00*	(−0.08, 0.18)
Slope	0.51	0.05	1.07*	(0.41, 0.60)
Collision Frequency Equation				
Intercept	−20.24	2.89	/	(−25.92, −14.58)
ln (Traffic)	1.35	0.28	1.01	(0.79, 1.90)
ln (Length)	1.14	0.22	1.01	(0.72, 1.59)
Rainfall	0.03	0.04	1.03	(−0.05, 0.10)
Curvature2	−0.01	0.13	0.99	(−0.27, 0.25)
Slope2	0.20	0.05	1.22	(0.09, 0.31)
Interchange	0.42	0.09	1.52	(0.24, 0.60)
Speed Difference	0.14	0.02	1.15	(0.10, 0.19)

*IRR values for Curvature and Slope represent the indirect effects on collision frequency.

Under the specific data and modeling conditions of this study, no statistically significant effects of curvature radius and rainfall on speed difference or collision frequency were observed. It should be noted that the non-significant effects of curvature radius and rainfall in this study should not be interpreted as evidence that horizontal curves or rainfall have generally no influence on driving behavior or collision risk. The present finding is likely to be context-specific and may be related to the characteristics of the study area and dataset. Also, the limited variability of the curvature and climate likely attenuate the estimated coefficients in our dataset, even though the underlying physical effects exist. The incidence rate ratios (IRR) for the direct and indirect effects were derived using Eq. (34).

$$IRR = \begin{cases} e^{\beta}, & \text{direct effects} \\ e^{\beta_{\text{indirect}}}, & \text{indirect effects} \end{cases} \quad (34)$$

where, β represents the posterior mean parameter estimates. Specifically, the IRR for the exposure variables traffic volume and road segment length is 1.01^{α} .

In the collision-spatiotemporal causal path, the slope exhibited a strong and statistically significant effect, with IRR of 1.22. This indicates that for every 1% increase in the absolute value of the slope, the collision frequency will increase by 22%. This result is in line with existing studies (Afghari et al., 2023; Zeng et al., 2024). Moreover, through the indirect path via speed difference, slope also influenced collision frequency, with an additional indirect IRR of 1.07 (computed as $\exp(0.14 \times 0.51)$). Combining both effects, a 1 % increase in slope corresponds to approximately 29 % higher collision risk. This dual impact reflects that steeper grades not only increase speed variability across vehicles but also affect safety margins through longer stopping distances, reduced driver reaction capability, and elevated brake-failure risk on downgrades.

The length of the section and the presence of interchange also significantly affected collision frequency. The IRR for road section length (≈ 1), suggests a linear relationship, consistent with the section length definition criteria used in this study (Hou, 2019). In addition, the IRR of the interchange section was 1.52, reflecting the complexity of traffic merging and diverging flows. However, because the interchange indicator in this study represents a generalized categorical feature, it may include multiple unobserved sub-factors (Hasan and Abdel-Aty, 2024). The IRR of the daily average traffic volume was 1.01, consistent with Hou et al. (2018) and Llopis-Castelló et al. (2019).

The IRR of daily average rainfall was 1.03, but the effect was not statistically significant in the present model. This finding does not necessarily indicate that rainy conditions have no effect on collision risk.

The safety impact of rainfall is typically associated with reduced pavement friction and impaired visibility, but these mechanisms may be moderated by roadway drainage conditions, rainfall intensity, and drivers' adaptive behavior. In the present study, the relatively low rainfall intensity, possible risk compensatory behavior (Jung et al., 2014), and the drainage conditions of highway facilities may have obscured the statistical association between rainfall and collision frequency (Omranian et al., 2018).

The effect of curvature showed no significant impact on either speed difference or collision frequency. This may be attributed to the geometric design characteristics of the study area. The highways analyzed have design speeds of 120 km/h with large average radii (2137 m and 3738 m), resulting in minimal lateral acceleration (≈ 0.05 at 6 % superelevation). One possible explanation is that, for large-radius curves, drivers may have largely completed their speed adjustment before entering the curve or before reaching the curve midpoint, thereby reducing the observable variation in speed behavior. To further examine this possibility, an exploratory correlation analysis was conducted between curvature and the speed difference between the curve entry and midpoint. The Pearson correlation coefficient was -0.014 ($p = 0.91$), indicating no statistically significant linear or monotonic association. This result suggests that curvature-related speed variation was undetectable in the current dataset. However, it does not imply that curvature has no effect on speed behavior in general. Rather, it indicates that the curvature effect could not be clearly identified under the investigated roadway, traffic, and modeling conditions. Previous research focusing on mountainous or low-design-speed highways (Awasthi et al., 2024; Kar et al., 2024) found curvature-safety relationships, suggesting that curvature sensitivity depends on geometric context. Considering that grade and curvature may jointly affect vehicle operation and collision occurrence, an additional model including the interaction term between grade and curvature was tested. The results showed that the interaction term was not statistically significant, as its 97.5% posterior credible interval included zero. Therefore, the interaction effect between grade and curvature was not retained in the final model.

A detailed local analysis was conducted for the GX freeway section from K4 + 900 to K7 + 500, as shown in Fig. 9, which compares the spatial distributions of collision frequency and operating speed with the corresponding elevation and curvature profiles. The speed decreased from approximately 80 km/h to a minimum of about 73 km/h near K6 + 250 and then gradually recovered. The speed trough did not coincide with the location of maximum curvature, suggesting that drivers may respond to the combined upstream and downstream alignment rather than to instantaneous curvature alone, consistent with Wang et al. (2020) and Yan et al. (2017). Segment 4 recorded the highest collision frequency despite its mild downgrade (-0.51%), whereas Segment 3, which contains the steepest uphill grade (4%), recorded no collisions. The speed reduction near the downstream end of Segment 3 and its subsequent recovery in Segment 4 suggest a possible carry-over effect of upstream geometric transitions on downstream traffic instability. Therefore, collision occurrence may be influenced not only by the geometry of the current segment but also by the geometric continuity between adjacent segments.

5.4. Random effects

The random effect parameters of the ST2-RSREPA model are shown in Table 9. The mean values of all random variables were not significant and close to 0, indicating no systematic spatial or temporal bias in either collision frequency or speed difference. Unlike Cui and Xie (2021) who found that there was an overall increasing tendency in the spatial-temporal interaction effects, the present analysis found no such consistent pattern. However, their estimated variances were significantly greater than zero, which implies that there is substantial heterogeneity between different road segment sections and season combinations.

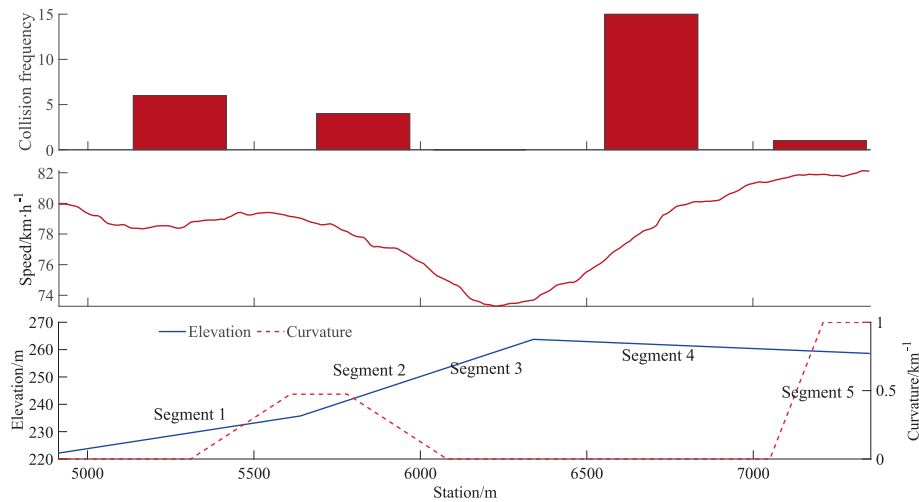


Fig. 9. Analysis of geometry, speed, and collision of K4 + 900 ~ K7 + 500 in GX Expressway.

Table 9
ST2-RSREPA random effects.

Random item	μ	97.5 %BCI	σ^2	97.5 %BCI
Speed Difference				
Spatial structured	-3.61E - 7	(-1.40, 2.52)	0.88	(0.71, 1.11)
Spatial unstructured	4.93E - 6	(-1.03, 1.68)	0.82	(0.68, 1.04)
Collision Frequency				
Spatial structured	-9.23E - 8	(-4.41E - 2, 7.89E - 2)	8.28E - 4	(6.3E - 4, 1.07E - 3)
Spatial unstructured	6.34E - 3	(-0.97, 1.39)	0.68	(0.54, 0.88)
Temporal structured	1.33E - 9	(-1.31E - 3, -6.48E - 9)	2.29E - 5	(1.78E - 5, 3.05E - 5)
Temporal unstructured	-5.76E - 5	(-286E - 1, 2.40E - 1)	4.33E - 2	(3.34E - 2, 5.87E - 2)
Radom intercept	1.60E - 4	(-1.50E - 1, 2.45E - 1)	0.11	(8.68E - 2, 0.15)

For speed variation, the structured spatial effect accounted for 51.8% of the total spatial variance ($0.88 / (0.88 + 0.82)$), slightly exceeding the 48.2% explained by the unstructured component. This indicates that much of the variation in speed differences emerges from smoothly varying spatial patterns across adjacent segments, rather than being entirely attributable to local random fluctuations. In terms of collision frequency, the variances of the structured spatial and temporal effects were smaller than those of their unstructured counterparts, suggesting that most variation is driven by localized, segment-specific randomness. Nevertheless, spatial effects together explained approximately 81.3% of the total variance, with the remainder captured by the random intercept. This implies that unobserved factors underlying collision occurrences exhibit both spatial continuity and considerable local heterogeneity.

Fig. 10 visualizes the exponentiated posterior means of the spatial random effects for both the speed-difference and collision-frequency equations along the GX and GY freeways. From top to bottom, the panels display: (1) structured and unstructured spatial random effects for speed difference, and (2) structured and unstructured spatial random effects for collision frequency. In the speed-difference model, a small number of segments showed relatively large exponentiated posterior means. For visualization, the upper limit of the color scale was capped at 10. These high local residuals likely reflect the limited set of geometric covariates (slope and curvature) included in the model.

The structured component displays a smoothly varying pattern across adjacent segments, consistent with spatially continuous

unobserved heterogeneity. Whereas the unstructured component captures irregular, site-specific deviations. Segments with high values of the structured effect warrant further investigation for spatially continuous latent risk sources, such as consistent clear-zone widths, roadside vegetation (Fitzpatrick et al., 2014; Pathivada et al., 2025), or micro-meteorological conditions not explicitly modeled here. Conversely, segments with large unstructured residuals likely reflect localized disturbances, such as temporary work zones (Wang et al., 2024a) or isolated traffic-control devices, which produce discontinuous spatial patterns.

6. Conclusions

This study developed a spatiotemporal path analysis framework to quantify the direct and indirect effects of road and environmental factors on collision frequency and addressing unobserved heterogeneity. Random effects and structured spatiotemporal components were integrated into the model to capture correlated spatial and temporal effects. Orthogonally projected spatial random terms were implemented using the RSR approach. This was done to mitigate spatial confounding between fixed and random effects. Model parameters were estimated using the INLA method.

Model performance was compared across multiple specifications. This included WAIC, error-based indicators, and Moran's I spatial tests. The ST2-RSREPA model, which combines BYM spatial terms with an RW1 temporal process, achieved the best overall fit. The framework was applied to multi-source data. These included collision records, floating car data, and road attributes from the GX and GY freeways. This was done to assess both direct and mediated effects, with speed difference serving as the mediating variable. The main conclusions are as follows:

1. Spatial confounding was confirmed between fixed and random components. In the speed-difference equation, curvature and slope were significantly correlated with the spatial random terms. In the collision-frequency equation, rainfall, curvature, slope, and interchange also showed significant correlation with the spatial random components. This collinearity caused shifts in coefficient signs and inflated variances. It confirmed that ignoring spatial dependence can bias parameter estimation and distort inference in spatial models.
2. Speed difference mediated the indirect effect of slope on collisions. Other than rainfall and curvature, slope and interchange were found to have positive effects on collision frequency. The interchange attribute had the largest influence. The studied freeways were high-standard facilities with gentle curvature and large curve radii. Under

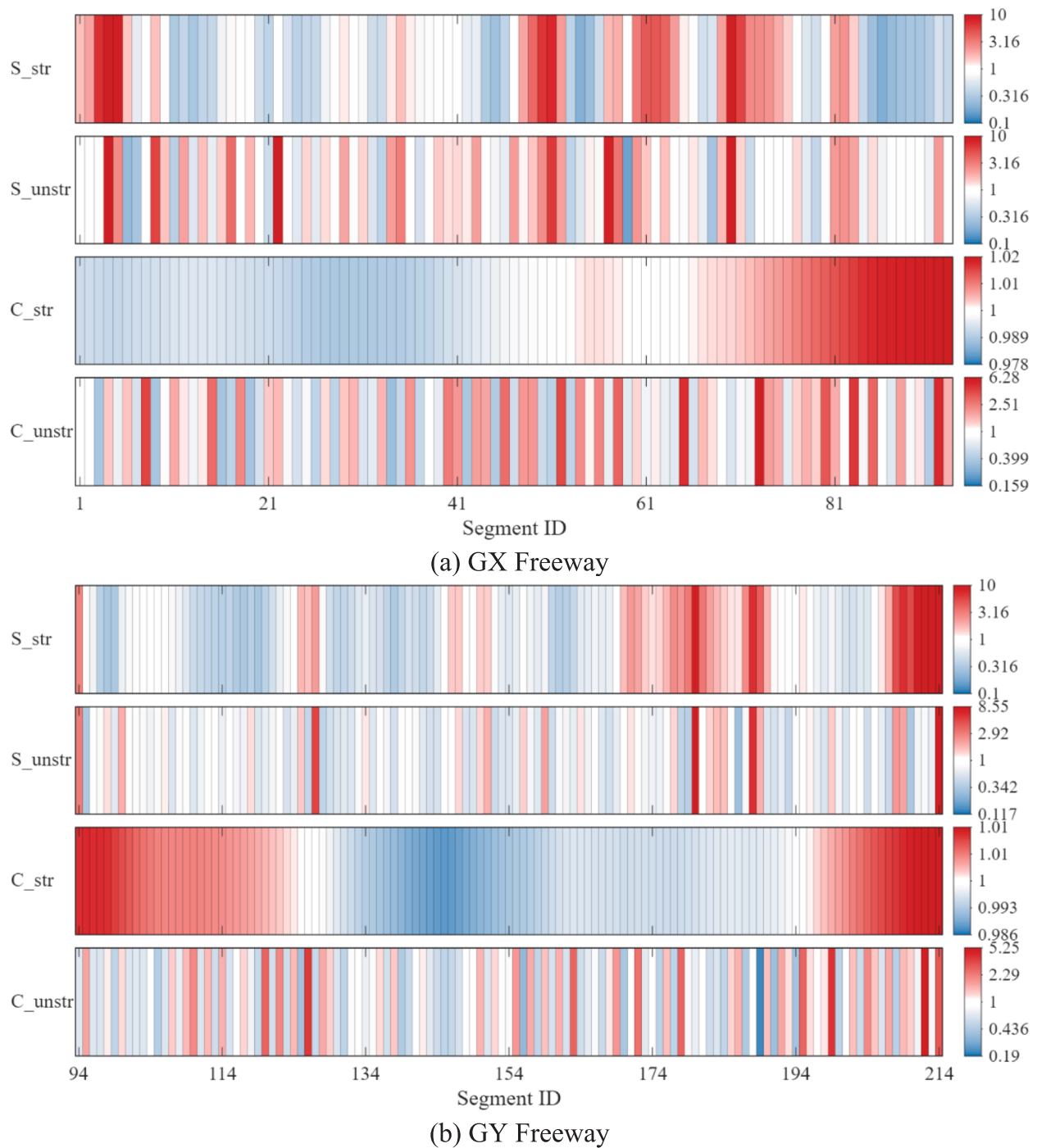


Fig. 10. Spatial-temporal interaction effect diagram.

these conditions, curvature showed no significant relationship with either collision frequency or speed variation.

3. Spatial heterogeneity dominated the random effects. Most variation in collision frequency was captured by spatial random components, with the unstructured term being the highest. The speed-difference heterogeneity was equally attributed to structured and unstructured spatial effects. Segments with elevated mean random effects may indicate unmodeled, spatially continuous risk sources such as roadside conditions or environmental factors that warrant targeted field investigation.

In future work, the current framework can be extended by incorporating a richer set of variables from more comprehensive databases,

thereby enabling the analysis of multiple influencing pathways and the development of a more comprehensive modeling structure. In addition, greater attention should be devoted to the potential correlations among unobserved factors across different model specifications (Ahmad et al., 2023b). Beyond collision frequency, the indirect effects of mediating variables on collision severity also warrant in-depth investigation, which would provide a more holistic understanding of traffic safety outcomes.

CRedit authorship contribution statement

Changhe Liu: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Conceptualization. **Bo Wang:**

Writing – review & editing, Software. **Chi Zhang:** Supervision, Funding acquisition, Data curation. **Karim Ismail:** Writing – review & editing, Supervision. **Dibin Wei:** Visualization. **Rulla Al-Haideri:** Writing – review & editing.

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.

Acknowledgments

A fair amount of the research data analysis works were conducted at Carleton University to which the first author was a visiting student researcher sponsored by the China Scholarship Council. (Grant No. 202406560076). The third author is supported by the Fundamental Research Funds for the Central Universities, CHD (300102215206), National Natural Science Foundation of China (72371036), Xi'an City Social Science Planning Fund Project (25GL123), and Shaanxi Provincial Social Science Fund Project (2025R018).

Data availability

The authors do not have permission to share data.

References

- Abohassan, A., El-Basyouny, K., Kwon, T.J., 2021. Exploring the associations between winter maintenance operations, weather variables, surface condition, and road safety: a path analysis approach. *Accid. Anal. Prev.* 163, 106448. <https://doi.org/10.1016/j.aap.2021.106448>.
- Afghari, A. P., Vos, J., Farah, H., & Papadimitriou, E. (2023). "I did not see that coming": A latent variable structural equation model for understanding the effect of road predictability on crashes along horizontal curves. *Accident Analysis & Prevention*, 187, 107075. <https://doi.org/10.1016/j.aap.2023.107075>.
- Ahmad, N., Arvin, R., Khattak, A.J., 2023a. Exploring pathways from driving errors and violations to crashes: The role of instability in driving. *Accid. Anal. Prev.* 179, 106876. <https://doi.org/10.1016/j.aap.2022.106876>.
- Ahmad, N., Arvin, R., Khattak, A.J., 2023b. How is the duration of distraction related to safety-critical events? Harnessing naturalistic driving data to explore the role of driving instability. *J. Saf. Res.* 85, 15–30. <https://doi.org/10.1016/j.jsr.2023.01.003>.
- Awasthi, D., Parti, R., Mahajan, K., 2024. Effect of spatial relationship between curves on crash severity at horizontal curves in a mountainous terrain. *Accid. Anal. Prev.* 206, 107714. <https://doi.org/10.1016/j.aap.2024.107714>.
- Azati, Y., Wang, X., Ye, X., Zhang, K., 2025. Refining the classification of combined alignment sections on mountainous freeways and analyzing the spatio-temporal effects on crash frequency. *Accid. Anal. Prev.* 221, 108222. <https://doi.org/10.1016/j.aap.2025.108222>.
- Camacho-Torregrosa, F.J., Pérez-Zuriaga, A.M., Campoy-Ungria, J.M., García-García, A., 2013. New geometric design consistency model based on operating speed profiles for road safety evaluation. *Accid. Anal. Prev.* 61, 33–42. <https://doi.org/10.1016/j.aap.2012.10.001>.
- CDC. (2025, August 6). *Global Road Safety*. Transportation Safety. <https://www.cdc.gov/transportation-safety/global/index.html>.
- Cheng, W., Gill, G.S., Zhou, J., Ensich, J.L., Kwong, J., Jia, X., 2020. Alternative multivariate multimodal crash frequency models. *J. Transport. Saf. Sec.* 12 (5), 628–652. <https://doi.org/10.1080/19439962.2018.1525631>.
- Cui, H., Xie, K., 2021. An accelerated hierarchical Bayesian crash frequency model with accommodation of spatiotemporal interactions. *Accid. Anal. Prev.* 153, 106018. <https://doi.org/10.1016/j.aap.2021.106018>.
- Das, S., Park, E.S., Sarkar, S., 2023. Impact of operating speed measures on traffic crashes: annual and daily level models for rural two-lane and rural multilane roadways. *J. Trans. Saf. Sec.* 15 (6), 584–603. <https://doi.org/10.1080/19439962.2022.2098441>.
- Doecke, S.D., Dutschke, J.K., Baldock, M.R., Kloeden, C.N., 2021. Travel speed and the risk of serious injury in vehicle crashes. *Accid. Anal. Prev.* 161, 106359. <https://doi.org/10.1016/j.aap.2021.106359>.
- Ekmekeci, M., Woods, L., Dadashzadeh, N., 2024. Effects of road width, radii and speeds on collisions at three-arm priority intersections. *Accid. Anal. Prev.* 199, 107522. <https://doi.org/10.1016/j.aap.2024.107522>.
- Fitzpatrick, C.D., Harrington, C.P., Knodler Jr, M.A., Romoser, M.R., 2014. The influence of clear zone size and roadside vegetation on driver behavior. *J. Saf. Res.* 49, 97–e1. <https://doi.org/10.1016/j.jsr.2014.03.006>.
- Gargoum, S.A., El-Basyouny, K., 2016. Exploring the association between speed and safety: a path analysis approach. *Accid. Anal. Prev.* 93, 32–40. <https://doi.org/10.1016/j.aap.2016.04.029>.
- Hamzeie, R., Savolainen, P.T., Gates, T.J., 2017. Driver speed selection and crash risk: insights from the naturalistic driving study. *J. Saf. Res.* 63, 187–194. <https://doi.org/10.1016/j.jsr.2017.10.007>.
- Hasan, T., Abdel-Aty, M., 2024. Short-term safety performance functions by random parameters negative binomial-Lindley model for part-time shoulder use. *Accid. Anal. Prev.* 199, 107498. <https://doi.org/10.1016/j.aap.2024.107498>.
- Hou, Q., 2019. Random parameters model for freeway traffic crash considering heterogeneity and endogeneity [Doctoral dissertation]. J. Harbin Inst. Tech.
- Hou, Q., Tarko, A.P., Meng, X., 2018. Investigating factors of crash frequency with random effects and random parameters models: New insights from Chinese freeway study. *Accid. Anal. Prev.* 120, 1–12. <https://doi.org/10.1016/j.aap.2018.07.010>.
- Hui, F.K.C., Bondell, H.D., 2022. Spatial confounding in generalized estimating equations. *Am. Stat.* 76 (3), 238–247. <https://doi.org/10.1080/00031305.2021.2009372>.
- Hutton, J. M., Cook, D. J., Grotheer, J., & Conn, M. (2020). *Research utilizing SHRP2 data to improve highway safety: Development of speed—Safety relationships* (No. FHWA-HRT-20-035). United States. Federal Highway Administration. Office of Safety Research and Development, McLean, Virginia. <https://rosap.nhtl.bts.gov/view/dot/44448>.
- Imprialou, M.-I.-M., Quddus, M., Pittfield, D.E., Lord, D., 2016. Re-visiting crash-speed relationships: a new perspective in crash modelling. *Accid. Anal. Prev.* 86, 173–185. <https://doi.org/10.1016/j.aap.2015.10.001>.
- Jonathan, A.-V., Wu, K.-F.-K., Donnell, E.T., 2016. A multivariate spatial crash frequency model for identifying sites with promise based on crash types. *Accid. Anal. Prev.* 87, 8–16. <https://doi.org/10.1080/00031305.2021.2009372>.
- Jung, S., Jang, K., Yoon, Y., Kang, S., 2014. Contributing factors to vehicle to vehicle crash frequency and severity under rainfall. *J. Saf. Res.* 50, 1–10. <https://doi.org/10.1016/j.jsr.2014.01.001>.
- Kar, P., Venkathurthy, S.P., Chunchu, M., 2024. Crash risk estimation of Heavy Commercial vehicles on horizontal curves in mountainous terrain using proactive safety method. *Accid. Anal. Prev.* 199, 107521. <https://doi.org/10.1016/j.aap.2024.107521>.
- Khan, S.A., Afghari, A.P., Yasmin, S., Haque, M.M., 2023. Effects of design consistency on run-off-road crashes: an application of a random parameters negative binomial lindley model. *Accid. Anal. Prev.* 186, 107042. <https://doi.org/10.1016/j.aap.2023.107042>.
- Liu, C., Sharma, A., 2018. Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity. *Anal. Methods Accid. Res.* 17, 14–31. <https://doi.org/10.1016/j.amar.2018.02.001>.
- Llopis-Castelló, D., Findley, D.J., Camacho-Torregrosa, F.J., García, A., 2019. Calibration of inertial consistency models on North Carolina two-lane rural roads. *Accid. Anal. Prev.* 127, 236–245. <https://doi.org/10.1016/j.aap.2019.03.013>.
- Lu, W., Liu, J., Fu, X., Yang, J., Jones, S., 2022. Integrating machine learning into path analysis for quantifying behavioral pathways in bicycle-motor vehicle crashes. *Accid. Anal. Prev.* 168, 106622. <https://doi.org/10.1016/j.aap.2022.106622>.
- Manning, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. *Anal. Methods Accid. Res.* 11, 1–16. <https://doi.org/10.1016/j.amar.2016.04.001>.
- Okafor, S., Liu, J., Adanu, E.K., Jones, S., 2023. Behavioral pathway analysis of pedestrian injury severity in pedestrian-motor vehicle crashes. *Transp. Res. Interdiscip. Perspect.* 18, 100777. <https://doi.org/10.1016/j.trip.2023.100777>.
- Omrani, E., Sharif, H., Dessouky, S., Weissmann, J., 2018. Exploring rainfall impacts on the crash risk on Texas roadways: A crash-based matched-pairs analysis approach. *Accid. Anal. Prev.* 117, 10–20. <https://doi.org/10.1016/j.aap.2018.03.030>.
- Park, E.S., Fitzpatrick, K., Das, S., Avelar, R., 2021. Exploration of the relationship among roadway characteristics, operating speed, and crashes for city streets using path analysis. *Accid. Anal. Prev.* 150, 105896. <https://doi.org/10.1016/j.aap.2020.105896>.
- Pathivada, B.K., Banerjee, A., Haleem, K., Khan, T., 2025. Safety performance functions for motorcycle crashes at interchange ramp segments. *J. Saf. Res.* 93, 44–54. <https://doi.org/10.1016/j.jsr.2025.02.004>.
- Pei, X., Wong, S.C., Sze, N.-N., 2012. The roles of exposure and speed in road safety analysis. *Accid. Anal. Prev.* 48, 464–471. <https://doi.org/10.1016/j.aap.2012.03.005>.
- Quddus, M., 2013. Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and GIS. *Journal of Transportation Safety & Security* 5 (1), 27–45. <https://doi.org/10.1080/19439962.2012.705232>.
- Reich, B.J., Hodges, J.S., Zadnik, V., 2006. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics* 62 (4), 1197–1206. <https://doi.org/10.1111/j.1541-0420.2006.00617.x>.
- Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B Stat Methodol.* 71 (2), 319–392. <https://doi.org/10.1111/j.1467-9868.2008.00700.x>.
- Singh, M., Zhang, Y., Cheng, W., Li, Y., Clay, E., 2022. Effect of transit-oriented design on pedestrian and cyclist safety using bivariate spatial models. *J. Saf. Res.* 83, 152–162. <https://doi.org/10.1016/j.jsr.2022.08.012>.
- Wang, X., Guo, Q., Tarko, A.P., 2020. Modeling speed profiles on mountainous freeways using high resolution data. *Transp. Res. Part C Emerg. Technol.* 117, 102679. <https://doi.org/10.1016/j.trc.2020.102679>.
- Wang, W., Yang, Y., Yang, X., Gayah, V.V., Wang, Y., Tang, J., Yuan, Z., 2024. A negative binomial Lindley approach considering spatiotemporal effects for modeling traffic

- crash frequency with excess zeros. *Accident Analysis & Prevention* 207, 107741. <https://doi.org/10.1016/j.aap.2024.107741>.
- Wang, B., Chen, T., Zhang, C., Wong, Y.D., Zhang, H., Zhou, Y., 2024a. Toward safer highway work zones: an empirical analysis of crash risks using improved safety potential field and machine learning techniques. *Accid. Anal. Prev.* 194, 107361. <https://doi.org/10.1016/j.aap.2023.107361>.
- Wang, W., Yang, Y., Yang, X., Gayah, V.V., Wang, Y., Tang, J., Yuan, Z., 2024b. A negative binomial Lindley approach considering spatiotemporal effects for modeling traffic crash frequency with excess zeros. *Accid. Anal. Prev.* 207, 107741. <https://doi.org/10.1016/j.aap.2024.107741>.
- Wang, C., Abdel-Aty, M., Han, L., 2025. Grouped random parameters Poisson-Lindley model with spatial effects addressing crashes at intersections: insights from visual environment features and spatiotemporal instability. *Anal. Methods Accid. Res.* 47, 100387. <https://doi.org/10.1016/j.amar.2025.100387>.
- Wei, Z., Das, S., Wu, Y., Li, Z., Zhang, Y., 2024. Modeling the lagged impacts of hourly weather and speed variation factors on the segment crash risk of rural interstate freeways: applying a space-time-stratified case-crossover design. *Accid. Anal. Prev.* 195, 107411. <https://doi.org/10.1016/j.aap.2023.107411>.
- Wei, D., Zhang, C., Luo, R., Wang, B., Zhang, M., Nie, Y., 2026. A multi-scale road segmentation approach to assess MAUP effects on crash frequency and hotspot analysis. *Accid. Anal. Prev.* 236, 108705. <https://doi.org/10.1016/j.aap.2026.108705>.
- Xu, C., Wang, X., Yang, H., Xie, K., Chen, X., 2019. Exploring the impacts of speed variances on safety performance of urban elevated expressways using GPS data. *Accid. Anal. Prev.* 123, 29–38. <https://doi.org/10.1016/j.aap.2018.11.012>.
- Xu, N., Liu, J., Zhang, Z., Jones, S., 2025. How do motorists' pre-crash behaviors contribute to the injury severity of police officers? Using interpretable machine learning to untangle the behavioral pathways in crashes involving police vehicles. *J. Saf. Res.* 94, 193–215. <https://doi.org/10.1016/j.jsr.2025.06.023>.
- Yan, Y., Li, G., Tang, J., Guo, Z., 2017. A novel approach for operating speed continuous predication based on alignment space comprehensive index. *J. Adv. Transp.* 2017 (1), 1–14. <https://doi.org/10.1155/2017/9862949>.
- Yang, W., Han, S., 2025. Direct and indirect effects of road attributes on traffic safety. *J. Saf. Res.* 93, 156–169. <https://doi.org/10.1016/j.jsr.2025.02.017>.
- Yu, R., Quddus, M., Wang, X., Yang, K., 2018. Impact of data aggregation approaches on the relationships between operating speed and traffic safety. *Accid. Anal. Prev.* 120, 304–310. <https://doi.org/10.1016/j.aap.2018.06.007>.
- Yu, R., Zheng, Y., Abdel-Aty, M., Gao, Z., 2019. Exploring crash mechanisms with microscopic traffic flow variables: a hybrid approach with latent class logit and path analysis models. *Accid. Anal. Prev.* 125, 70–78. <https://doi.org/10.1016/j.aap.2019.01.022>.
- Yu, Q., Niu, J., Ou, J., Bai, W., Wang, N., Jiang, X., 2025. Exploring the inter-relationship among hazardous actions, causal factors, and motorcyclist injury severity using path analysis. *Accid. Anal. Prev.* 216, 108033. <https://doi.org/10.1016/j.aap.2025.108033>.
- Zeng, Q., Wang, Q., Zhang, K., Wong, S.C., Xu, P., 2023. Analysis of the injury severity of motor vehicle–pedestrian crashes at urban intersections using spatiotemporal logistic regression models. *Accid. Anal. Prev.* 189, 107119. <https://doi.org/10.1016/j.aap.2023.107119>.
- Zeng, Q., Xu, Z., Liang, Y., Lee, J., Wang, X., 2024. Incorporating spatio-temporal correlation and interaction into a generalized ordered logit model for examining injury severities of traffic accidents. *J. Trans. Saf. Sec.* 16 (11), 1237–1258. <https://doi.org/10.1080/19439962.2024.2311406>.
- Ziakopoulos, A., Yannis, G., 2020. A review of spatial approaches in road safety. *Accid. Anal. Prev.* 135, 105323. <https://doi.org/10.1016/j.aap.2019.105323>.
- Zimmerman, D.L., Ver Hoef, J.M., 2022. On deconfounding spatial confounding in linear models. *Am. Stat.* 76 (2), 159–167. <https://doi.org/10.1080/00031305.2021.1946149>.