Project 1920 | February 2021

SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Enhancement of Multimodal Traffic Safety in High-Quality Transit Areas

Yongping Zhang, PhD, P.E.
Wen Cheng, PhD, P.E.
Xudong Jia, PhD, P.E.

CSUTC California State University Transportation Consortium

CalPoly Pomona

CSU TRANSPORTATION CONSORTIUM

transweb.sjsu.edu/csutc

MINETA TRANSPORTATION INSTITUTE

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the Mineta Consortium for Transportation Mobility (MCTM) funded by the U.S. Department of Transportation and the California State University Transportation Consortium (CSUTC) funded by the State of California through Senate Bill 1. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce Development

To ensure the efficient movement of people and products, we must prepare a new cohort of transportation professionals who are ready to lead a more diverse, inclusive, and equitable transportation industry. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Report 21-04

Enhancement of Multimodal Traffic Safety in High-Quality Transit Areas

Yongping Zhang, PhD, P.E.
Wen Cheng, PhD, P.E.
Xudong Jia, PhD, P.E.

February 2021

A publication of the
Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-02

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 21-04	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Enhancement of Multimodal Traffic Safety in High-Quality Transit Areas		**5. Report Date** February 2021 **6. Performing Organization Code**
7. Authors Yongping Zhang https://orcid.org/0000-0002-5935-3834 Wen Cheng https://orcid.org/0000-0001-7225-6169 Xudong Jia https://orcid.org/0000-0001-7911-8869		**8. Performing Organization Report** CA-MTI-1920
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219		**10. Work Unit No.** **11. Contract or Grant No.** ZSB12017-SJAUX
12. Sponsoring Agency Name and Address State of California SB1 2017/2018 Trustees of the California State University Sponsored Programs Administration 401 Golden Shore, 5th Floor Long Beach, CA 90802		**13. Type of Report and Period Covered** Final Report **14. Sponsoring Agency Code**
15. Supplemental Notes DOI: 10.31979/mti.2021.1920		

16. Abstract

Numerous extant studies are dedicated to enhancing the safety of active transportation modes, but very few studies are devoted to safety analysis surrounding transit stations, which serve as an important modal interface for pedestrians and bicyclists. This study bridges the gap by developing joint models based on the multivariate conditionally autoregressive (MCAR) priors with a distance-oriented neighboring weight matrix. For this purpose, transit-station-centered data in Los Angeles County were used for model development. Feature selection relying on both random forest and correlation analyses was employed, which leads to different covariate inputs to each of the two jointed models, resulting in increased model flexibility. Utilizing an Integrated Nested Laplace Approximation (INLA) algorithm and various evaluation criteria, the results demonstrate that models with a correlation effect between pedestrians and bicyclists perform much better than the models without such an effect. The joint models also aid in identifying significant covariates contributing to the safety of each of the two active transportation modes. The research results can furnish transportation professionals with additional insights to create safer access to transit and thus promote active transportation.

17. Key Words Active transportation, transit station, traffic safety, built environment, correlation analysis	**18. Distribution Statement** No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 60	**22. Price**

Form DOT F 1700.7 (8-72)

DOI: 10.31979/mti.2021.1920

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219

Tel: (408) 924-7560
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/research/1920

ACKNOWLEDGMENTS

This study is funded by the California State University Transportation Consortium (CSUTC). The authors are grateful to the Southern California Association of Governments (SCAG) for furnishing the travel demand model data and built environment data, which are critical for the study. The authors would also like to thank Dr. Hilary Nixon for her valuable comments and continuous support. The efforts and comments from the anonymous reviewers and Editing Press are significantly appreciated as well.

Contents

List of Figures

List of Tables

Executive Summary

Active transportation modes such as walking and cycling provide immense benefits associated with health, environmental, and social outcomes, among other benefits. However, active mode users are the most vulnerable segment of road users from motorized traffic due to the lack of a protective structure and differences in vehicular mass. Numerous extant studies are dedicated to enhancing the active transportation modes by considering infrastructure facilities such as intersections, sidewalks, and bike lanes, but there is a lack of sufficient research devoted to safety analysis regarding traffic accidents surrounding transit stations. The current study bridges the gap by developing joint models based on the multivariate conditionally autoregressive (MCAR) priors with a distance-oriented neighboring weight matrix. For this purpose, data centered on high-quality transit stations were used for statistical analysis. Data included built environment characteristics, socioeconomic and demographic information, and crash data aggregated at the level of the 0.5-mile-radius zone surrounding the stations.

To obtain the different covariates to each of the two active transportation modes and to increase the model's flexibility, feature selection was conducted using both random forest and correlation analyses. The research adopted INLAMSM, an Integrated Nested Laplace Approximation (INLA) package with Multivariate Spatial Models. It is a fast Bayesian inference approach within a multivariate spatial framework using the R language. For a comprehensive comparison of the predictive accuracy of the models, different evaluation criteria were utilized, including DIC (deviance information criterion), WAIC (widely applicable information criterion), Dbar (posterior mean deviance), and Pd (effective number of parameters). The results demonstrate that models with a correlation of pedestrians and bicyclists perform much better than the models without such a correlation. The joint models can aid in highlighting the covariates significant to each of the two active transportation modes.

The research results can furnish transportation professionals with additional insights to create safer non-motorized access to transit and promote active transportation across California. The model developed in this study can also be used to identify crash hot spots by ranking the frequency of the collisions related to pedestrians and bicyclists at each transit station. Using the ranking, transit operators and transportation professionals can prioritize improvements near the stations where they are most needed.

I. Introduction

1.1 Research Background

During the year 2014, 32,675 fatalities and more than 2,338,000 injuries occurred on U.S. roads. Road accidents were the leading cause of death in 2014 among people aged 16 through 24 in the U.S. In the SCAG (Southern California Association of Governments) region, each year, about 1,500 fatalities and 120,000 injuries occur due to traffic accidents. These fatalities and injuries reflect a significant number of lives that could have been saved by the application of appropriate safety treatments.

The FAST Act calls for the establishment of performance measures and standards on traffic safety. The Federal Highway Administration (FHWA) is now requiring each state's Department of Transportation (DOT) to work with Metropolitan Planning Organizations (MPOs) to assess fatalities and severe injuries on all public roads and to set annual performance measures. The SCAG 2016–2040 Regional Transportation Plan/Sustainable Communities Strategies (RTP/SCS) was adopted in April 2016. In this regional planning document, the traffic safety issue is considered an important subject. The region's rates of fatal and injury collisions are briefly reviewed and presented as heat maps in the RTP/SCS.

The challenge of meeting the mobility requirements of the 21st century requires a shift in mindset from designing an automobile-focused highway system to operating a transportation network that accommodates all users and modes safely and conveniently. Therefore, transportation agencies of various levels and their partners have been striving to provide more complete streets to all travelers by taking advantage of the many opportunities to go beyond traditional approaches. Typical elements that make up a complete street include sidewalks, bicycle lanes (or wide, paved shoulders), shared-use paths, designated bus lanes, safe and accessible transit stops, and frequent and safe crossings for pedestrians. Amongst the different roadway facilities, the transit stop plays a key role in the successful implementation of complete streets programs due to its unique position as an intermodal interface. It is anticipated that the provision of more, safer, and more easily accessible transit stops would significantly increase the number of active transportation mode users who might heavily rely on public transportation as an economical alternative for relatively long trips.

It is worth noting "active transportation" may refer to any human-powered transportation and low-speed electronic assist devices. In addition to walking and biking, active transportation modes and devices may include but are not limited to: electric bicycle (e-bike), tricycle, wheelchair, scooter, electric scooter (e-scooter), skates, skateboard, push scooter, trailer, and hand cart (SCAG 2020–2045 RTP/SCS). However, for simplicity, this study will use the terms "walking" and "biking," or "pedestrians" and "bicyclists," to represent all modes of active transportation.

However, in contrast to the extensive research centered on the typical mode of motor vehicles and other spatial units, there is considerably less research dedicated to safety analysis along the transit corridors. It still remains unclear which factors constitute the main contributing elements to the multimodal safety conditions in the areas adjacent to transit stops, given the complexity of the influential factors and their interactions. To respond to the urgent need, this study aims to rank the importance and quantify the impact of different variables pertinent to multimodal traffic safety near the transit stops; these factors are imperative for the effective design of complete streets and the development of integrated land use and transportation planning policies.

High-quality transit areas (HQTAs) are defined as areas within a half-mile of a fixed guideway transit stop or a bus transit corridor where buses pick up passengers at a frequency of every 15 minutes or less during peak commuting hours (SCAG, 2016). This project aims to enhance the multimodal traffic safety conditions in HQTAs, which is imperative for successive implementations of complete streets design and integrated land use and transportation planning policies.

While HQTAs account for only 3% of total land area in the SCAG region, they are planned and projected to accommodate 46% of the region's future household growth and 55% of its future employment growth (SCAG, 2016).

According to the National Complete Streets Coalition, complete streets are designed and operated to enable safe access and travel for all users, including pedestrians, bicyclists, motorists, transit users, and travelers of all ages and abilities (Smith et al., 2010). In comparison with motorists, non-motorists are a vulnerable segment of the traveling public due to their lack of a protective structure and a difference in mass between them and motor vehicles, which renders them prone to heightened injury susceptibility in case of a collision (William, 2013). Active-transportation-related crashes represent more than 20% of all road crashes in California. On the other hand, active transportation provides enormous benefits for addressing issues of congestion, health, and the environment (Berrigan et al., 2006). Therefore, encouraging individuals to indulge in active transportation involving walking and bicycling brings with it a societal obligation to protect commuters as they engage in these modes of travel. Ensuring that roads provide safe mobility for all travelers, not just motor vehicles, is at the heart of the complete streets design approach. Unfortunately, in contrast to the wide range of studies dedicated to motorists, less research has focused on investigating the factors impacting non-motorists' safety on roadways (e.g., Lee and Abdel-Aty, 2005; Beck et al., 2007; Moudon et al., 2011; Cai et al., 2016).

1.2 Research Problems

A literature review of the pertinent research reveals the following limitations:

- Most studies modeled the modal crashes separately. Few attempts have been made to combine them into a multimodal approach; combining them allows researchers the flexibility to simultaneously determine the injury risk associated with different travel modes. Ignorance of such correlations among the multiple modes has been shown to reduce the models' efficiency due to less precise parameters (Park and Lord, 2007; Cheng et al., 2017).

- Even though an array of spatial levels have been investigated in past safety research—ranging from microscopic locations such as intersections (Wang and Abdel-Aty, 2006) and road segments (Cheng et al., 2018a) to macroscopic areas such as census tracts (Noland and Quddus, 2004), traffic analysis zones (TAZs) (Abdel-Aty et al., 2011), or counties (Cheng et al., 2018b)—there is little research focusing on the level of transit corridors. Safety analysis focused on such corridors is essential for enhancing the use of active transportation, given that transit provides an economical and viable option that allows people of all abilities and socioeconomic levels to conduct relatively long-distance travel usually available only to motorists.
- Spatial correlations were rarely considered in the context of multimodal safety analyses even though accommodation of spatial dependency has been found by past studies to enhance the model fitness and precision (Gill et al., 2017). It is essential to consider such correlations; nearby locations (such as transit stops) would have more characteristics to share in common. Additionally, most of the spatial dependency was accounted for through the random effects approach, which may be considered as a special restrictive case of a more flexible approach of full random parameters, where each site would be assigned its own coefficient for explanatory variables (Mannering et al., 2016). To the best knowledge of the researchers, there is no literature showing the incorporation of spatial correlations into the random parameter approaches, which is much-needed given the benefits of both approaches.
- Finally, a substantial proportion of safety studies have relied on statistical modeling for feature selection or detection of the significant influential factors for crash occurrence or severities. However, a biased subset of variables is often identified since statistical models are usually subject to omitted variable or multicollinearity issues due to the requirement of a strong assumption of statistical distribution or lack of consideration for high-order variable interactions. Emerging machine learning techniques can provide more accurate variable selection because of their consideration of multiple desirable features to avoid omitted variables or multicollinearity issues.

To summarize, safe and accessible transit stops play an instrumental role in enhancing multimodal use, as they significantly leverage mobility among active transportation users or economically disadvantaged residents who heavily rely on public transportation for long-distance travel or other purposes. Unfortunately, little research is dedicated to the safety analysis of transit corridors while addressing the limitations mentioned above.

1.3 Research Objectives

In response to the limitations mentioned above, this study is designed to satisfy the following main objectives:

- Collect required data in the HQTAs from multiple resources that cover multimodal crash data, socioeconomic data, and built environment data in Los Angeles County.
- Perform a ranking among the influential factors for multimodal safety in the HQTAs using multiple machine learning techniques.
- Develop multivariate multimodal crash statistical models with random parameters and spatiotemporal dependency to provide accurate estimates of the impact of various factors on different modal crashes and fill the current safety research gap.
- Interpret the model results and translate them to the development of good transportation planning policies to enhance safety for active transportation users near the high-quality transit stations.

II. Literature Review

Ensuring a convenient and safe traffic environment for all transportation modes requires a shift from the vehicle-oriented transportation system towards one that accommodates all road users. The goal may be achieved by implementing a complete streets design, which has the flexibility to enhance both traffic safety and strategic urban mobility planning (Smith et al., 2010). Typical elements of complete streets include sidewalks, bicycle lanes (or wide, paved shoulders), shared-use paths, designated bus lanes, safe and accessible transit stops, and frequent and safe crossings for pedestrians and bicyclists. Among the different roadway facilities, the transit stop plays a crucial role in successfully implementing complete streets programs due to its unique intermodal interface position. Compared with other modes, non-motorized transportation modes provide enormous health, environmental, and social benefits, plus many others. However, non-motorists are a vulnerable segment of the traveling public due to the lack of a protective structure and the mass discrepancy between them and motor vehicles, which renders them prone to heightened injury susceptibility in case of a collision (Cai et al., 2017; Mander and Zick, 2014). Therefore, incorporating transportation network attributes into traffic safety would facilitate the development of safety programs and strategies to engender a safe environment for all roadway users.

Previous research studies conducted over the years have striven to obtain valuable insights by considering various empirical and methodological aspects of safety modeling to address this urgent need. Crash frequency models are often used to identify the factors influencing the propensity of active-mode-related crashes. As the crash frequency data are non-negative integers, the most widely used crash frequency models assume the Poisson distribution of crash counts. The initial Poisson regression models are subject to the limitation of equality between the mean and variance of crash counts (Kim et al., 2002; Miranda-Moreno, 2006), which means Poisson models are not capable of handling over-dispersion or under-dispersion and can be adversely influenced by low-sample means. Such issues in data analysis could result in biased and incorrectly estimated parameters, which may lead to invalid inferences and predictions in crash frequency models (Lord and Mannering, 2010; Oh et al., 2006; Cameron and Trivedi, 1998). To overcome possible over-dispersion in the data and to obtain reliable and unbiased results, some researchers have proposed employing a negative binomial or Poisson gamma model as an alternative to the Poisson model (Hauer, 2001; Daniels et al., 2010; El-Basyouny, and Sayed, 2006; Lord and Bonneson, 2007). The negative binomial/Poisson gamma model follows a gamma probability distribution, which allows the researcher to manipulate the relationship between mean and variance by adding a gamma-distributed error term. Even though it is likely the most widely used model in crash frequency modeling, the negative binomial/Poisson gamma model does have its limitations. For example, it cannot handle under-dispersed data and can be adversely affected by a low sample mean and small sample size (Lord, 2006; Lord and Mannering, 2010).

One subsequent enhancement of the negative binomial/Poisson gamma model is the Poisson-lognormal model for modeling crash data (Park and Lord, 2007; Lord and Miranda-Moreno, 2008; Aguero-Valverde and Jovanis, 2008). The Poisson-lognormal model is developed by assuming Poisson-lognormal distribution, which can handle a small sample size and over-dispersion better than the negative binomial/Poisson gamma model due to the heavier tail associated with the lognormal distribution. Nonetheless, the heavier tail associated with the lognormal distribution increases the model complexity, and the model remains affected by small sample sizes and low mean values (Lord and Mannering, 2010; Miaou et al., 2003). Another model extensively adopted by safety researchers is the zero-inflated model (Aguero-Valverde, 2013), as it can handle data characterized by a substantial number of zeros. As data with excess zero density cannot be accommodated by traditional Poisson or negative binomial/Poisson-gamma models, the zero-inflated model can be applied to count data that have an excess of zero counts by splitting the model based on crash-free versus crash-prone propensity, which means excess zeros are yielded independently from a separate process associated with the count values. Besides its wide applicability in diverse situations where the observed data are characterized by large zero densities, some studies have pointed out its limitations, especially in highway safety. For example, Lord and Bonnet and colleagues (2005, 2007) illustrate that the zero-inflated model (both for Poisson and negative binomial models) cannot provide reliable results for crash data in a zero or safe state because the tendency of a long-term mean equal to zero.

The most extensively used format within the above models is the univariate model, which contains only one dependent variable for data interpretation (Anarkooli et al., 2019; Wang et al., 2013). However, the univariate setting cannot address the unobserved heterogeneities which might be common to various crash types or severities (Mannering and Bhat, 2014; Ma et al., 2008). In response, researchers have employed multivariate models extensively to explicitly account for the possible correlations among the distinct responding variables. Some papers have relied on the bivariate framework for various applications in which two crash types are involved such as angled injury security (Russo et al., 2014) and bicycle conflict location (Conway et al., 2013), while others have taken advantage of multivariate models to address a response variable with multiple discrete outcomes, that is, different crash types (Serhiyenko et al., 2016) and crashes involving distinct modes (Huang et al., 2017). Another benefit of using bivariate/multivariate approaches is the explicit consideration of correlation among different outcomes (Song et al., 2006; Bijleveld, 2005; Park and Lord, 2007; Aguero-Valverde and Jovanis, 2009). Even with well-documented benefits over the univariate alternatives (Motherfer et al., 2016), multivariate models are still the focus of studies dedicated to further enhancing them. For example, Wang et al. (2013) developed a Poisson-lognormal conditional autoregressive model for their bivariate spatial analysis of pedestrian crash counts across census tracts in Austin, Texas. The results indicate that a bivariate cross-correlation of serious (fatal and major injury) and non-serious crash rates shows covariates' impacts across severity level are more local in nature (e.g., lighting conditions, or local sight obstructions along with spatially lagged cross-correlation). Ma et al. (2008) used a multivariate Poisson-lognormal specification to investigate different crash counts at different severity levels. Their findings show that the multivariate Poisson-lognormal model aids in showing the statistically significant correlations between crash counts at a different level of injury severity.

Explicit consideration of spatial autocorrelation in the multivariate settings is one popular strategy and includes both random effect (Aguero-Valverde, 2013) and random parameter models (Barua et al., 2016; Imprialou et al., 2016). However, bivariate/multivariate models are complex to estimate as the correlation matrix formulation is required (Lord and Mannering, 2010).

In addition to the models mentioned above, consideration of spatial effects is another aspect of model classification and has been extensively adopted in recent studies. Incorporating geographic information collected by sophisticated computer software such as Geographical Information Systems (GIS) allows the researchers to include influential factors relating to the spatial perspective. However, the data from same geographic area may share unobserved effects and lead to a spatial correlation problem. To analyze the data in a way that addresses the spatial correlation, various studies develop Bayesian spatial models such as conditionally autoregressive (CAR) (Song et al., 2006; Zeng and Huang, 2004; Abdel-Aty and Wang, 2006), simultaneous autoregressive (SAR) (Quddus, 2008; Chiou and Chih-Wei, 2014; Hosseinpour et al., 2018), or multivariate conditionally autoregressive models (MCAR) (Aguero-Valverde, 2013; Jonathan et al., 2016; Cai et al., 2018). CAR is generally appropriate for situations with first-order dependency or relatively local spatial autocorrelation (e.g., city, ZIP code, census ward). In contrast, the spatial regression models are appropriate when second-order dependency or a more global spatial autocorrelation (e.g., state/province, county) is involved. The CAR model is assumed to be auto-normal CAR distributed to account for spatial random effects. The relatively simple model structure leads to its extensive usage in the safety field (Wang et al., 2012). On the other hand, the multivariate conditionally autoregressive model (MCAR), also known as Gaussian Markov random fields as proposed by Jin et al. (2005), allows area-specific heterogeneity by directly analyzing the multivariate spatial count data, such as transit-level pedestrian and bicyclist crash count data. (Boulieri et al., 2016).

Employing another method to account for spatial correlation, many studies consider the random effects models where the common unobserved effects are assumed to be uncorrelated with independent variables and distributed over spatial units. A study conducted by Hausman et al. (1984) examined random effects and fixed effects in negative binomial models for panel data in their research. The findings suggest that random effects help to account for unobserved factors shared by distinct discrete outcomes. In the context of crash frequencies, random effects have been used by a large number of previous studies. For instance, Ma et al. (2017) proposed a series of multivariate models under the Full Bayesian framework with different random effects to predict the crash frequencies of different injury severity levels over one year in Colorado. Cheng et al. (2017) developed the multivariate Poisson-lognormal models with random effects to predict the motor-cycle injury severity crashes using weather data during the years 2008–2013 in San Francisco. Besharati et al. (2020) utilized the bivariate spatial negative binomial model with random effects to examine the association between fuel consumption in the transportation sector and the annual fatal and non-fatal injury counts during the period from 2005–2015 in Iran. However, the main restriction of the random effects approach is that it only influences the intercept of the model. An extension of the random effects model is the random parameter models that allow each estimated parameter of the model to vary across each individual observation in the dataset. Because each observation has its own parameters, the random parameter models help account for unobserved heterogeneity from one observation to another (Anastasopoulos and Mannering, 2009; Milton et al., 2008). However, random parameter models do have their limitations, as they are very complex to estimate; sometimes, they may not improve models' predictive capability, and since the results obtained are observation-specific, the model results are non-transferable to other datasets (Washington et al., 2010; Shugan, 2006; Lord and Mannering, 2010).

To carry out the Bayesian inference, the Markov Chain Monte Carlo (MCMC) simulation method is the most popular approach used in the safety field. Park and Lord (2007) adopted the MCMC simulation method in multivariate Poisson-lognormal models to evaluate covariates' impact on crash counts. Similarly, El-Basyouny and Sayed (2009) used multivariate Poisson-lognormal models with the MCMC simulation approach through the WinBUGS platform to jointly analyze a dataset of crash counts categorized by two injury severity levels. However, the Bayesian framework using the MCMC simulation method may be computationally challenging under a complex model scenario and also time-consuming, especially for large datasets (Narayanamoorthy et al., 2013). To address this issue, the present research adopts an alternative Bayesian approach, Integrated Nested Laplace Approximation or INLA (Rue et al., 2009), to carry out approximate Bayesian inference. The INLA method aids in reducing the computational time and efforts involved in the estimation of complex and large datasets (Serhiyenko et al., 2016; Blangiardo et al., 2013) seven-fold compared to the MCMC simulation method (Serhiyenko, 2015).

All in all, active transportation has gained ever-increasing popularity due to its multiple benefits over the typical motorized modes. However, improving the safety of non-motorists plays a pivotal role in promoting such healthy, economically and environmentally friendly modes, especially at various transit stations where different transportation modes interact with one another. Compared with other infrastructure facilities such as intersections, sidewalks, and bike lanes, there is considerably less research dedicated to safety analysis regarding traffic accidents centered on active transportation modes surrounding transit stations. It remains unclear which factors constitute the main contributors to the walking and biking safety conditions in the areas adjacent to transit stops given the complexity of the influential factors and their interactions. The present study aims to rank the importance and quantify the impact of pedestrians' and bicyclists' traffic-safety-pertinent variables near the transit stops to bridge this gap. This research effort is imperative for the effective design of complete streets and transportation planning policies. For this purpose, multivariate spatial models were chosen owing to the frequent advantages reported in previous research associated with multivariate settings and spatial heterogeneity considerations. Specifically, joint models based on the multivariate conditionally autoregressive (MCAR) priors with a distance-oriented neighboring 14-weight matrix were used.

To take advantage of a substantial reduction in computational time for estimation under a complex model scenario (Serhiyenko et al., 2016; Blangiardo et al., 2013), the current study employs Integrated Nested Laplace Approximation Multivariate Spatial Model (INLAMSM) (Palmí-Perales et al., 2019), an INLA-oriented package, to carry out approximate Bayesian inference within a multivariate spatial framework. In addition, feature selection using both random forest and correlation analyses is employed, yielding different covariates to each of the two active transportation modes and leading to increased model flexibility. Moreover, for the statistical analysis, the researchers used data centered on high-quality transit stations, including built environments and socioeconomic, demographic, and crash data aggregated at the level of the 0.5-mile-radius zone surrounding the stations. Finally, for a comprehensive comparison of the predictive accuracy of models, different evaluation criteria were utilized, which include deviance information criterion (DIC), widely applicable information criterion (WAIC), posterior mean deviance (Dbar), and the effective number of parameters (Pd).

III. Methodology

This section is dedicated to presenting methodological details, including the modeling specification, variable importance by random forest, and evaluation criteria.

3.1 Modeling Specification

In the current study, under the INLA framework (Lindgren and Rue, 2015), the INLAMSM package (Palmí-Perales et al., 2020) is used to draw the inferences due to its ability to address multivariate latent effects. A Bayesian hierarchical approach is used to estimate the Poisson process:

$$y \sim Poisson\ (\boldsymbol{\lambda}) \tag{1}$$

where y is the observed crash count, and λ is the Bayesian mean expected crashes, which can be modeled as a function of the covariates following a lognormal distribution as shown below:

$$\log(\lambda) = \beta_0 + \boldsymbol{\beta X} + \phi + \boldsymbol{\varepsilon} \tag{2}$$

where

β_0 is the global intercept,

$\boldsymbol{\beta}$ is a fixed coefficient vector,

$\boldsymbol{X}$ is the covariate matrix, and

ϕ is the spatially structured error term.

It is fit by the multivariate conditionally autoregressive (MCAR) model, and **ε** represents the white noise matrix.

For the multivariate model, correlated priors in the random effects vector are estimated using multivariate normal priors (Ma and Kockelman, 2006; Park and Lord, 2007):

$$\boldsymbol{\varepsilon} \sim MN(\mu, \Sigma) \tag{3}$$

$$\text{where } \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \tau_1 & \sqrt{\tau_1\tau_2}/\rho_{12} \\ \sqrt{\tau_1\tau_2}/\rho_{12} & \tau_2 \end{pmatrix} \tag{4}$$

In the above equations, $\boldsymbol{\varepsilon}$ is the independent random effect matrix which captures the extra-Poisson heterogeneity among locations, μ is the vector consisting of the mean value for each transportation mode, and Σ is called the precision matrix where the diagonal τ elements represent the marginal precision of each of the transportation modes, while the off-diagonal elements represent the inverse of covariance, calculated as the ratio of $\sqrt{\tau_1\tau_2}$ and ρ_{12} (or, the correlation coefficient between the two response variables). If no correlation between the transportation modes is assumed, the off-diagonal elements can be specified to be zero. In this research, both correlated and non-correlated modes are considered for model performance purposes. This inverse of the precision matrix is defined by:

$$\sum^{-1} \sim Wishart(I, n) \tag{5}$$

where the $\sum^{-1}$ is a symmetric positive definite matrix, I is the scale matrix (Congdon, 2007), and n (n=2) is the degree of freedom, resulting in a non-informative specification (Heydari et al., 2017). The covariates coefficient was specified with normally distributed vague priors N (0,100). Such a diffused normal distribution with mean values of zero and a large variance is commonly employed as a vague prior of posterior estimates in the absence of sufficient knowledge of prior distribution (Osama and Syed, 2017; Cheng et al., 2018).

To tackle the spatial dependency, the authors employed the MCAR algorithm which was initially derived by Mardia (1988) from the results in Besag (1974). Let $\phi^T = (\phi_1^T \ldots\ldots \phi_m^T)$. Then, Φ is $nm \times 1$, with each ϕ being an n-dimensional vector. In the present study, n=2 represents the two transportation modes, and m=655 represents the 655 transit stations selected from Los Angeles County. Considering a multivariate Gaussian distribution for Φ yields:

$$(\Phi) = (2\pi)^{-\frac{nm}{2}} |\boldsymbol{S}|^{\frac{1}{2}} exp\left(-\frac{1}{2}\Phi^{\boldsymbol{T}} \boldsymbol{S}\Phi\right) \tag{6}$$

where **S** is an $nm \times nm$ precision matrix. Following Mardia (1988), the zero-entered MCAR, which has a conditional normal density, is shown as follows:

$$\phi_i \,|\phi_j, \textstyle\sum_i \sim N_k\left(\sum_{j\sim 1} C_{ij}\,, \phi_j, \sum_i\right) \tag{7}$$

where subscripts i and j respectively refer to a transit station and its neighbors, each $\sum_i$ is a positive definite matrix with dimensions of $n \times n$ representing the conditional precision matrix, and C_{ij} is a distance matrix of the same dimensions as $\sum_i$ (Jonathan et al., 2016). The precision matrix $\sum^{-1}$ follows the Wishart distribution, as shown in Equation 5.

3.2 Variable Importance by Random Forest (RF)

Random forest (RF) has become a widely adopted technique for variable importance ranking and is based on the ensemble classifier that consists of many decision trees and outputs the class by individual trees. It is computationally efficient with large datasets and relatively compatible with several software packages (e.g., Python, R Statistics) (Behnamian et al., 2017). The method combines bagging and the random selection of features to construct a collection of decision trees with controlled variation. Using ensembles of predictors for classification or regression has been shown to give more accurate results than a single predictor. A large number of trees ensures that the RF is robust against over-fitting. This technique has an advantage over the traditional decision trees: it obtains unbiased error estimates without separating the cross-validation test dataset. When a particular tree is grown from a bootstrap sample, usually one-third of the training cases are called out-of-bag (OOB) data and are left out and not used in the tree's growing.

The efficient implementation of the RF algorithm relies on two essential components: the number of trees to grow and the number of predictors selected to split each node to produce stable results and a minimum OOB error rate. Generally, the number of predictors (m) at each split is approximately equal to the square root of the total number of predictors (p) from the dataset (James et al., 2017).

$$m \approx \sqrt{p} \tag{8}$$

Once the proper values of the tree number and predictor size were determined, the variables' importance ranking was reported based on the mean decrease in accuracy (MDA) in predictions on the OOB samples when a given variable is excluded from the model using bootstrap aggregation (Liaw and Wiener, 2002).

A similar practice is seen in some previous research in the traffic safety field (Abdel-Aty and Haleem, 2011; Siddiqui et al., 2012; Jiang et al., 2016). Compared with other typical feature selection techniques such as forward or backward selection based on statistical model metrics, the RF is not specific to a particular data distribution and features an enhanced capability to handle data complexity, especially for those sets with a high order of interactions. The RF technique for variable selection was constructed in the R software using the 'randomForest' package.

3.3 Evaluation Criteria

For Bayesian hierarchical model evaluation, Deviance Information Criterion (DIC) is a popular criterion used to assess the models' complexity and goodness of fit (Spiegelhalter et al., 2003). As a hierarchical modeling generalization of the Akaike Information Criterion (AIC), DIC can be expressed as using the following formulation:

$$\mathrm{DIC} = \mathrm{D}(\bar{\theta}) + 2P_D = \bar{D} + P_D \tag{9}$$

where $\mathrm{D}(\bar{\theta})$ is the deviance evaluated at the posterior means of estimated unknowns ($\bar{\theta}$), and posterior mean deviance $\bar{D}$ can be taken as a Bayesian measure of data fitting. P_D denotes the effective number of parameters in a model, the difference between $\mathrm{D}(\bar{\theta})$ and $\bar{D}$. In general, the difference between observed and model-predicted data decreases as the number of parameters in a model increases. Therefore, the P_D term is mainly used to compensate for this effect by favoring models with a smaller effective number of parameters. The larger the DIC value, the worse the model tends to perform. As a general rule of thumb suggested by Lunn et al. (2012), the models with a DIC value of less than five are considered to have the same performance; the models with DIC values greater by 5 and 10 points are slightly worse, and the models with a DIC larger by more than 10 points are significantly worse. Overall, DIC may be regarded as the measure of an indirect assessment of the out-of-sample errors, as it is based on in-sample errors ($\bar{D}$) while also accounting for the bias due to overfitting usually resulting from more model parameters (James et al., 2013).

Like DIC, the widely applicable information criterion (WAIC) is another generalized version of the Akaike Information Criterion (AIC). For Bayesian models, WAIC (Watanabe, 2010) can be viewed as an improvement on the DIC, where the latter is not fully Bayesian since it is based on a point estimate (van der Linde, 2005; Plummer, 2008). By contrast, WAIC is fully Bayesian, invariant to parametrization, and closely approximates Bayesian cross-validation using leave-one-out techniques. Like DIC, the model with a smaller WAIC is preferred (Gelman et al., 2013). WAIC was also used in the present research as an additional criterion to assess model performance from a different perspective.

IV. Data Analysis

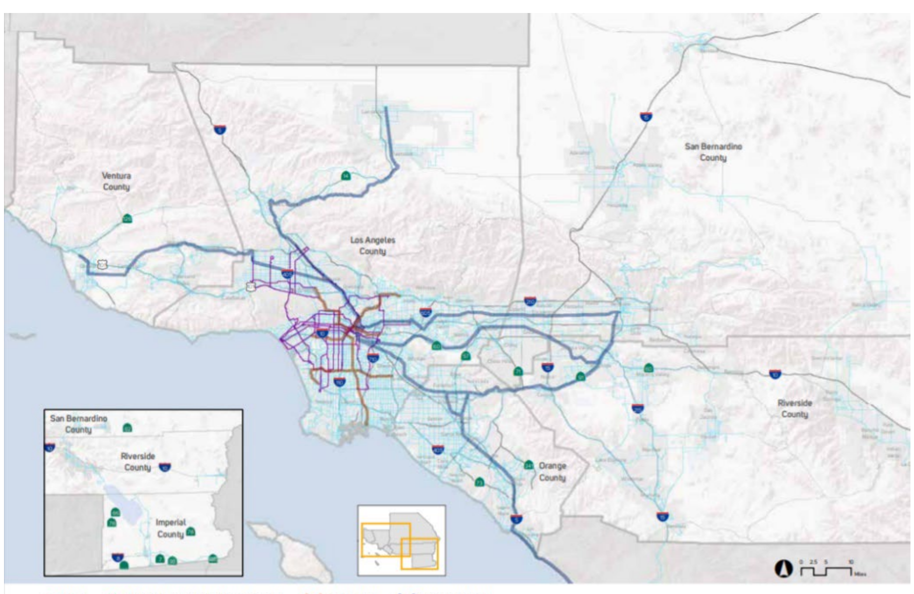
This chapter will describe the data source, the definition and descriptive statistics of each data variable, and the correlation among these variables, including the spatial relationship. This part of the research process aims to explore the most suitable data for model estimation. Before the data are fed into the model, it is critical to ensure data quality. For example, the potential multicollinearity among the explanatory variables will be examined closely. The correlation analysis between the dependent variables will also shed light on whether it is necessary to use the joint model.

4.1 Data

The data used in this research were obtained from multiple sources: Southern California Association of Governments (SCAG), Transportation Injury Mapping System (TIMS), L.A. Metro, InfoUSA, and the U.S. Census Bureau. All data were compiled into GIS databases. The dependent variables include the total number of pedestrian-involved crashes, the total number of bike-involved crashes, and the total number of vehicle-only crashes (i.e., crashes with no pedestrian or bike involved). The independent variables are categorized into various groups, including the following: socioeconomic characteristics, employment characteristics, diversity/mixed use of land, built environment/access to active transportation and transit, land development characteristics such as Transit-Oriented Development (TOD), and biking-/walking-related built environment variables.

In the SCAG region, as shown in Figure 1, various modes of transit form an extensive network.

Figure 1. Transit Network in the SCAG Region for 2012 Base Year
(Source: SCAG 2016–2040 RTP/SCS)

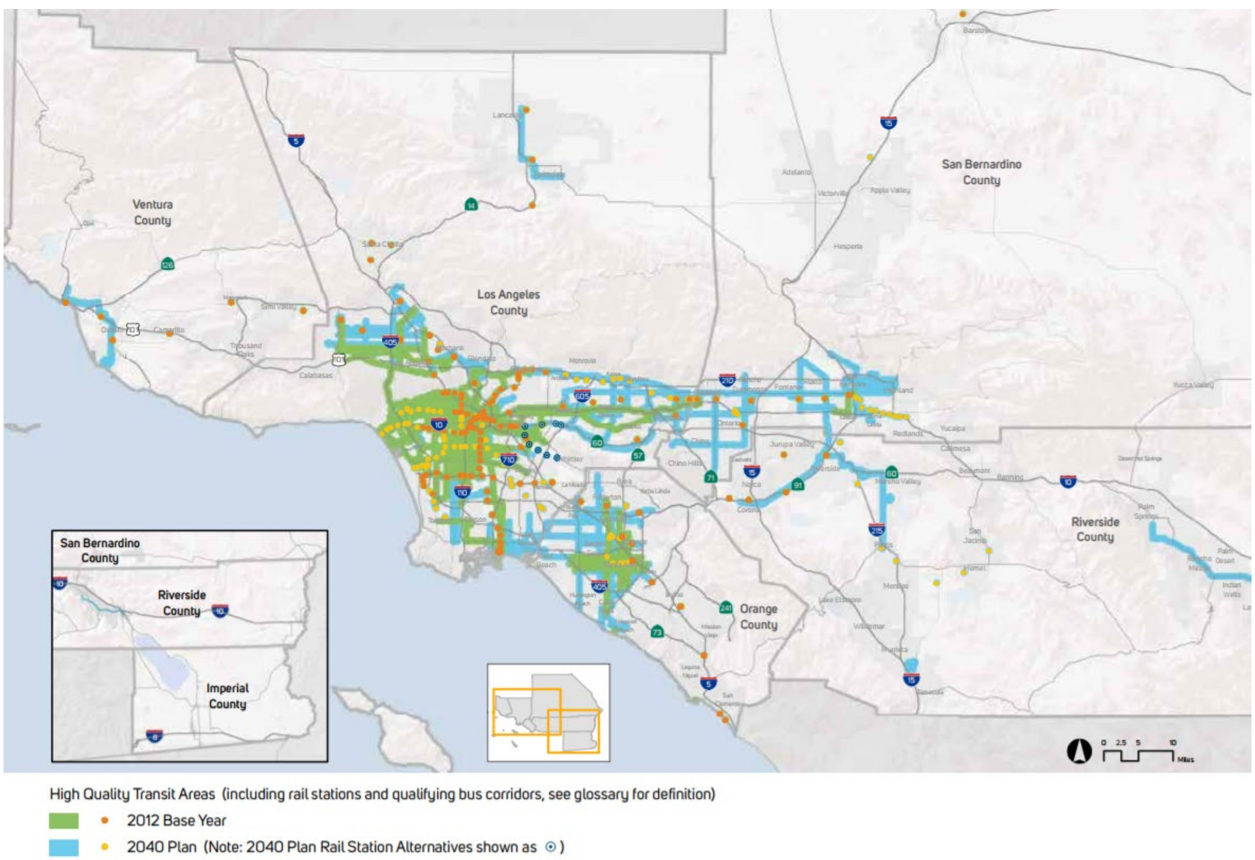


SCAG's definition of a HQTA is within one half-mile from major transit stops and high-quality transit corridors; it was developed based on the language in Senate Bill 375 (Barbour, 2016). According to SCAG, the definitions of major transit stops and high-quality transit corridors are as follows:

> *Major Transit Stop: A site containing an existing rail transit station, a ferry terminal served by either a bus or rail transit service, or the intersection of two or more major bus routes with a frequency of service interval of 15 minutes or less during the morning and afternoon peak commute periods (C.A. Public Resource Code Section 21064.3). It also includes major transit stops that are included in the applicable regional transportation.*
>
> *High-Quality Transit Corridor (HQTC): A corridor with fixed route bus service with service intervals no longer than 15 minutes during peak commute hours.* (SCAG, 2016)

Figure 2 shows the HQTAs in the SCAG Region for the 2012 Base Year and 2040 Plan Year.

Figure 2. High-Quality Transit Areas in the SCAG Region for 2012 Base Year and 2040 Plan (Source: SCAG 2016–2040 RTP/SCS)



This research is focused on major transit stops, as defined by SCAG's HQTAs (SCAG 2016–2020 RTP/SCS). The data collected cover the half-mile buffer zones around high-quality transit stops in the HQTAs in the SCAG region as defined in the base year 2016. The radius of half a mile for buffer zones was determined by SCAG, and most of the variables in these areas were derived from SCAG's travel demand models and land-use models.

There are 948 major transit stations in the SCAG region, including 155 rail stations. Figure 3 shows the location of these stations. Out of 948 stations, 870 stations are in Los Angeles County (Figure 3). This research will focus on L.A. County for research purposes. A station-to-station distance matrix is developed to reflect the spatial relationships among stations on the network. Furthermore, to avoid duplication among stations that are too close to each other because of the half-mile buffer, 293 stations that are within one mile of each other are removed in the analysis. This choice results in 655 stations in L.A. County.

Figure 3. Transit Stations (948 Major) and 11,267 TAZs (Tier 2) in the SCAG Region (Year 2016)

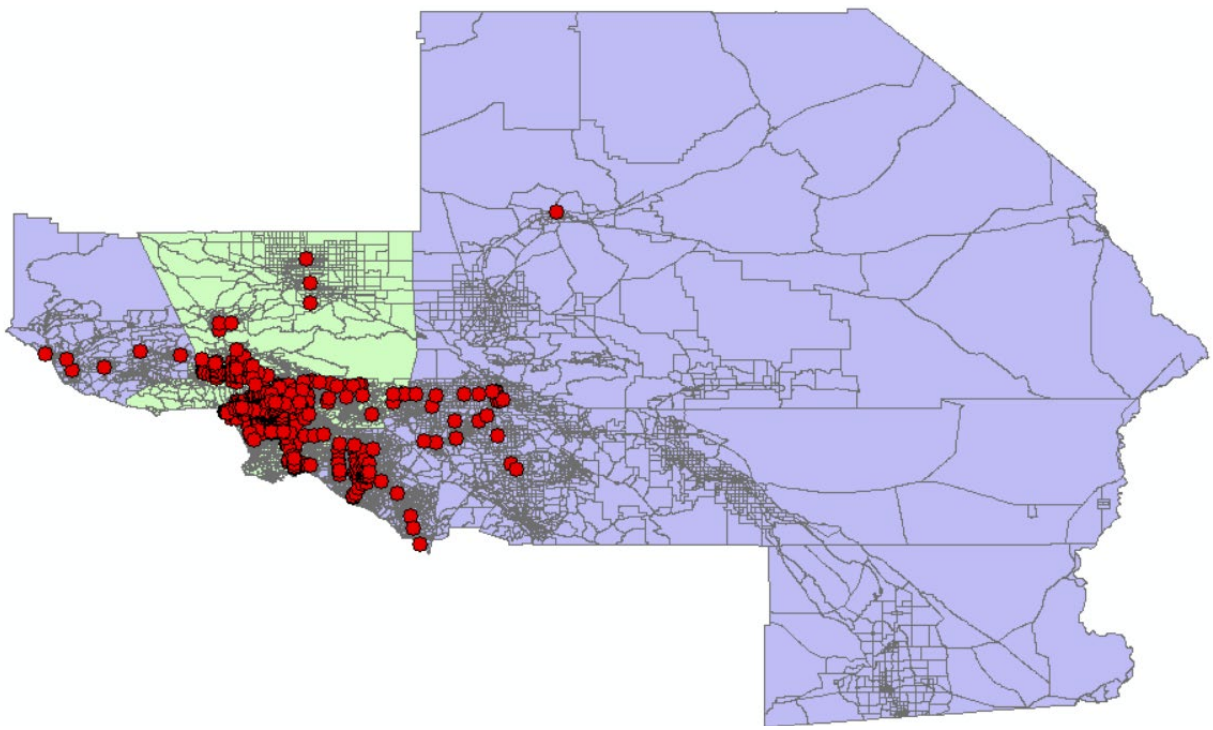


Figure 4. Transit Stations (870 Major) and 5,697 TAZs (Tier 2) in L.A. County (Year 2016)

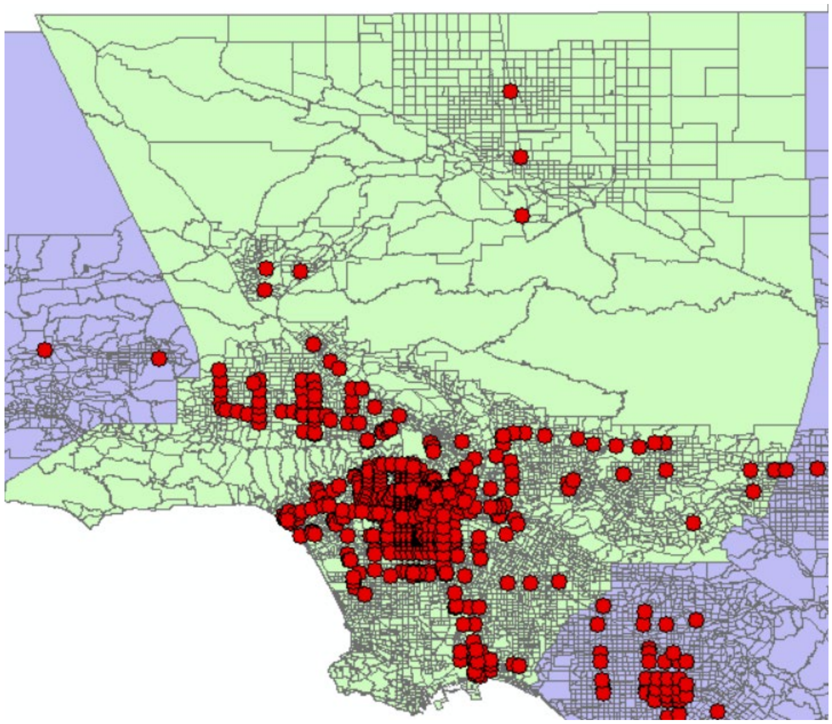
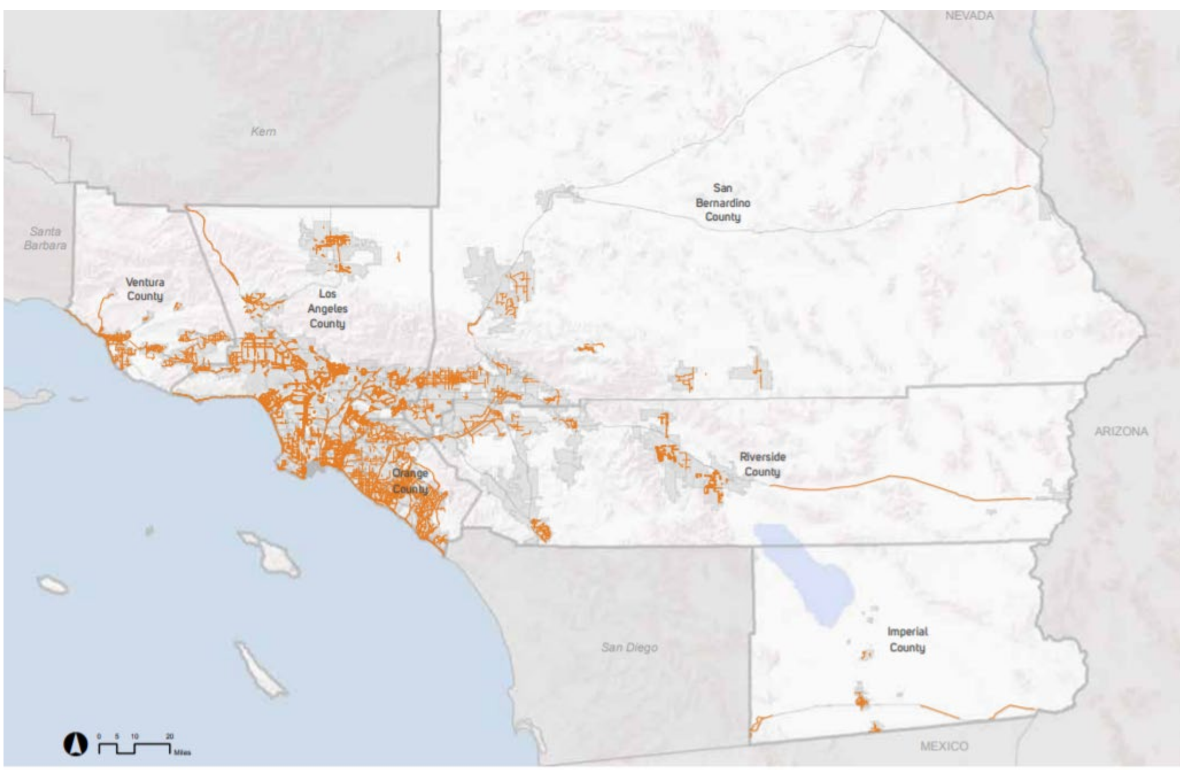


Figure 5. Bike Network in the SCAG Region (Year 2012)
(Source: SCAG 2016–2040 RTP/SCS)



4.2 Variables

A variety of variables related to the built environment were derived at SCAG's Tier 2 TAZ (traffic analysis zones) level. There are 11,267 Tier 2 TAZs in the SCAG travel demand model (Figure 3), and about half (5,697) of these TAZs are in L.A. County (Figure 4). Each TAZ has detailed information on population, households, employment, land use, and transit usage.

Variables related to bike lane density and access are also derived from the bikeway network. As an example, Figure 5 shows the extensive bikeway network in the SCAG region. Similarly, variables related to intersection density and walk accessibility are also derived and included in this study.

Table 1 shows the descriptive statistics of all variables used in this research. The bottom row indicates the station-to-station distance matrix statistics, which are used for spatial autocorrelation analysis in the model development.

Table 1. Descriptive Statistics of Collected Data

Variables	**Description**	**Min**	**Max**	**Mean**	**S.D.**
Density					
Pop_den1	Population density (persons/acre)	0.00	76.86	22.13	12.74
HH_den1	Household density (households/acre)	0.00	30.59	7.56	5.03
Emp_den1	Employment density (jobs/acre)	0.05	127.48	11.81	12.17
Ret_den1	Retail job density (jobs/acre)	0.00	7.11	1.12	1.03
RetSer_den1	Retail + Service (retail + FIRE + Arts&Food + Other Serv.) job density (jobs/acre)	0.02	26.62	3.84	4.13
Diversity / Mixed Use of Land					
Jobmix131	Employment mix (13 sectors); 1 = highest mix (i.e., jobs are equal for all sectors)	0.23	0.86	0.65	0.08
Jobmix91	Employment mix (9 sectors) ; 1 = highest mix (i.e., jobs are equal for all sectors)	0.25	0.79	0.61	0.08
Emix131	Employment mix (13 sectors); 1 = highest mix (i.e., jobs are equal for all sectors)	0.30	0.77	0.66	0.05
Emix91	Employment mix (9 sectors)); 1 = highest mix (i.e., jobs are equal for all sectors)	0.14	0.55	0.41	0.06
EH_ratio1	Job/Household ratio	0.00	10495.77	94.07	757.99
EP_ratio1	Job/Population ratio	0.00	10495.55	85.17	817.95
Built Environment / Access to Active Transportation & Transit					
Int34_Den1	Intersection density (three and four legs)	0.01	0.58	0.20	0.06
BKlnAcc1	Bike lane access (1 = if a TAZ has bike lane)	0.00	1.00	0.56	0.30

Variables	Description	Min	Max	Mean	S.D.
Built Environment / Access to Active Transportation & Transit					
Rail1	1 = at least one rail station in a TAZ	0.00	0.72	0.08	0.14
ExBus_D1	Stop density for express bus and BRT	0.00	0.71	0.03	0.05
HFLbus_D1	High-frequency bus stop density (local bus headway ≤ 20 mins)	0.00	0.47	0.05	0.05
TTbus_D1	Total bus stop density	0.00	2.71	0.42	0.29
Land Development Characteristics: TOD (HQTA/TPA)					
Mlt_pct1	Percent of households living in multiple units	0.00	0.49	0.28	0.12
HQTA_pct1	Percent of TAZ area in non-freeway HQTA (high-quality transit area)	0.00	1.00	0.89	0.23
TPA_pct1	Percent of TAZ area in TPA	0.00	1.00	0.71	0.33
Additional Biking- or Walking-Related Built Environment Variables					
BLdenIND1	Bike Lane Density Indicator = Sum (Bike Lane Density/Distance to Home TAZ within 3 miles) Bike Lane Density for Each TAZ = ((Street 15–25mph)*1 + (Street 35mph)*2 + Bike Lane Class1*3 + Bike Lane Class2*4 + Bike Lane Class3*5) / Total TAZ area (excluding streets speed > 60mph)	0.03	11.82	6.32	2.83
Blck_len1	Estimated block length = LocalSt/Int34new (Total street length / number of intersections) but freeways and state highways are excluded	0.11	1.12	0.24	0.08
WalkAcc1	Walk Accessibility (RS_den2/block_len) = (weighted retail + service job density) / estimated block length	0.01	53.41	6.94	6.75
Pct_Art1	Percent of major arterials (45–55mph) of TAZ where higher % means more difficult to cross the street (also larger block to across the street); can be used with WalkAcc	0.00	0.27	0.04	0.05
Natural Log Transformation					
L_Pden1	L_Pden = Ln (Pop_den + 0.001)	0.01	6.91	2.78	0.90
L_Hden1	L_Hden = Ln (HH_den + 0.001)	0.01	6.91	1.84	0.97
L_Eden1	L_Eden = Ln (Emp_den + 0.001)	0.06	4.13	1.81	0.94
L_REden1	L_REden = Ln (Ret_den + 0.001)	0.00	5.62	1.04	0.76
L_RSEden1	L_RSEden = Ln (RetSer_den + 0.001)	0.00	4.55	0.80	0.66
Crash Count					
Ped	Pedestrian-involved accident counts	0.00	222.00	48.12	38.50
Bike	Bike-involved accident counts	0.00	177.00	35.93	29.65
Distance					
Distance	The distance between each pair of stations (655 stations in total)	0.00	156.97	20.90	17.61

Notes: 1. S.D. represents standard deviation. 2. Variables fed into the final spatial model development are marked in bold.

A total of 250,817 non-freeway collisions in L.A. County from 2012 to 2017 were collected from the Transportation Injury Mapping System (TIMS) website, which provides quick, easy, and free access to California crash data and is maintained by the Statewide Integrated Traffic Records System (SWITRS). The multiple sources of data were integrated into GIS with the following procedures.

1. A GIS layer with the selected major transit stations was used to create the half-mile buffer zones.
2. The above-mentioned GIS layer was overlaid with SCAG's Tier 2 TAZ layer, where the built environment information is available.
3. The above two layers were overlaid with another GIS layer, which stores collision information in L.A. County from 2012 through 2017 from SWITRS (see Figure 2 for the screenshot of the collision layer overlaying transit station buffer zones).
4. The integrated dataset was cleaned and developed to contain transit-oriented crash counts and a variety of built-environment-related variables.

The integrated database connects each accident within the half-mile buffer around major transit stops in L.A. County with various built-environment-related variables (Table 1). As shown in Figures 6, 7, 8, 9, and 10, these accidents are further aggregated to the station (buffer) level. In other words, for each major transit station, a database was created with the number of accidents by mode (pedestrian, bike, auto) and variables related to the built environment.

Figure 6. Traffic Accidents (2012–2017) within Half-Mile Buffers around Major Transit Stops in L.A. County

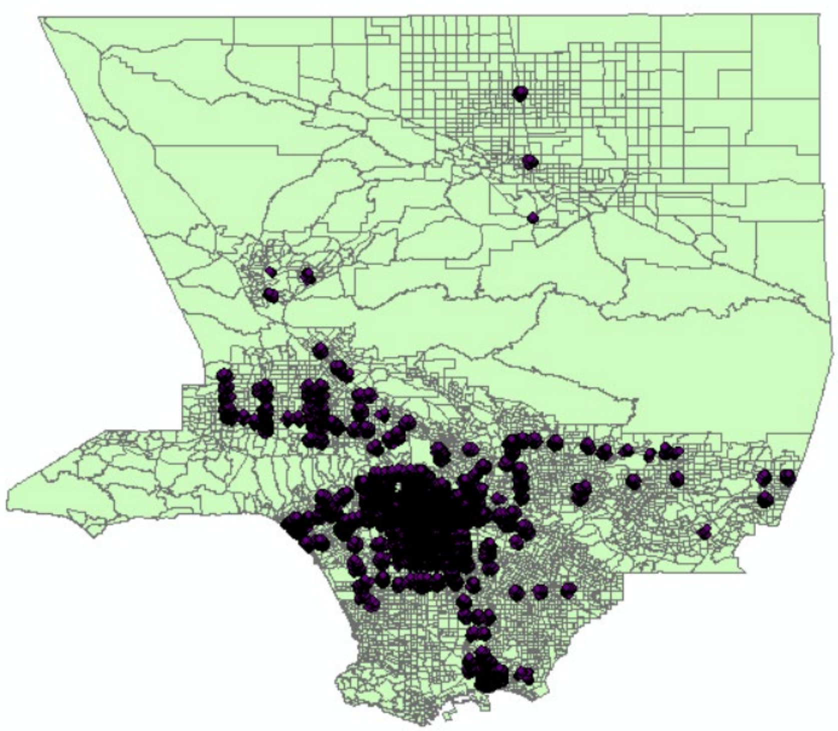


Figure 7. Traffic Accidents (2012–2017) within Half-Mile Buffers of Major Transit Stops in L.A. County (Central LA)

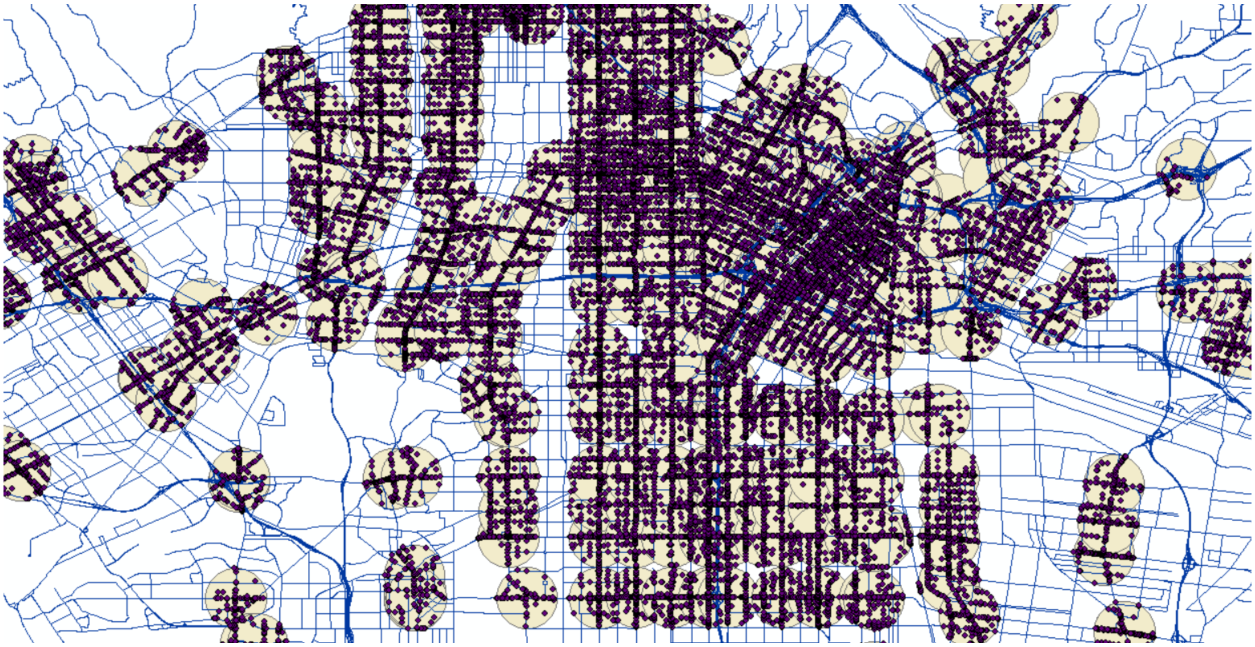


Figure 8. Traffic Accidents (2012–2017) within Half-Mile Buffers of Major Transit Stops in L.A. County (San Fernando Valley)

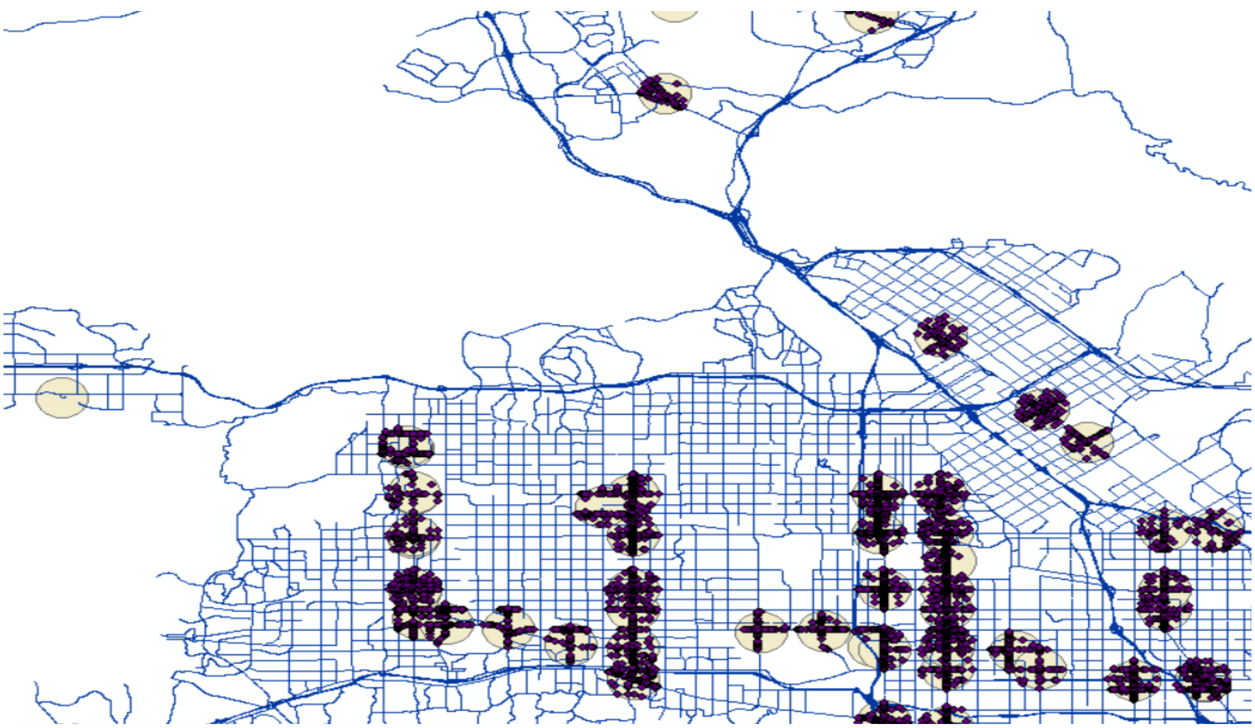


Figure 9. Traffic Accidents (2012–2017) within Half-Mile Buffers of Major Transit Stops in L.A. County (East Los Angeles and San Gabriel Valley)

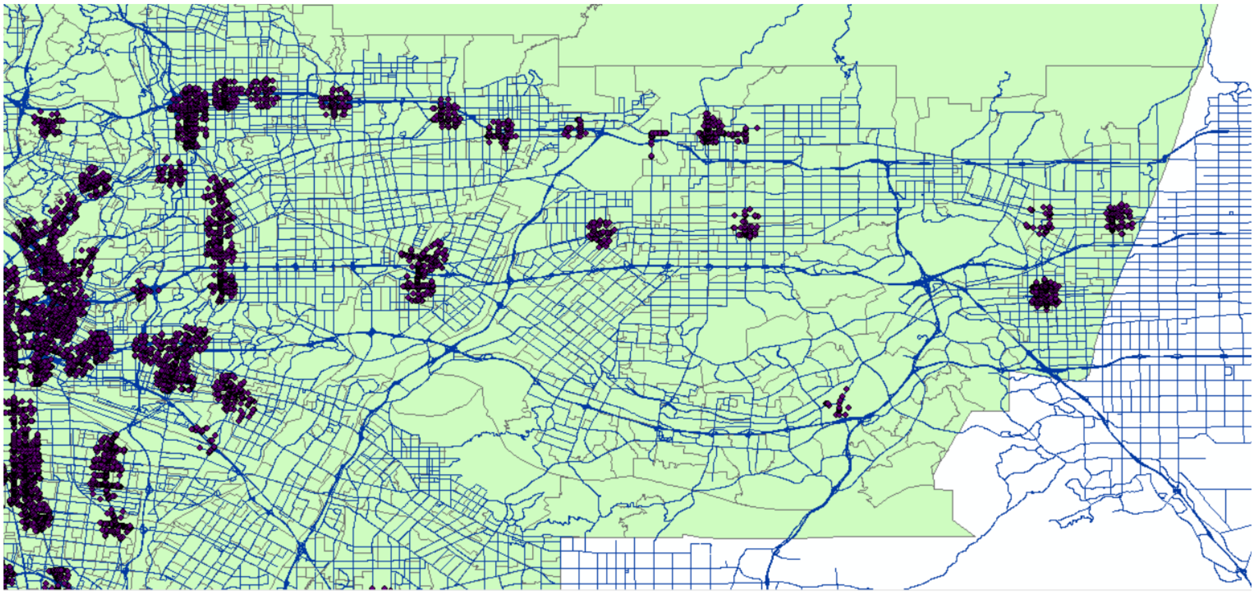


Figure 10. Traffic Accidents (2012–2017) within Half-Mile Buffers of Major Transit Stops in L.A. County (South Los Angeles County)

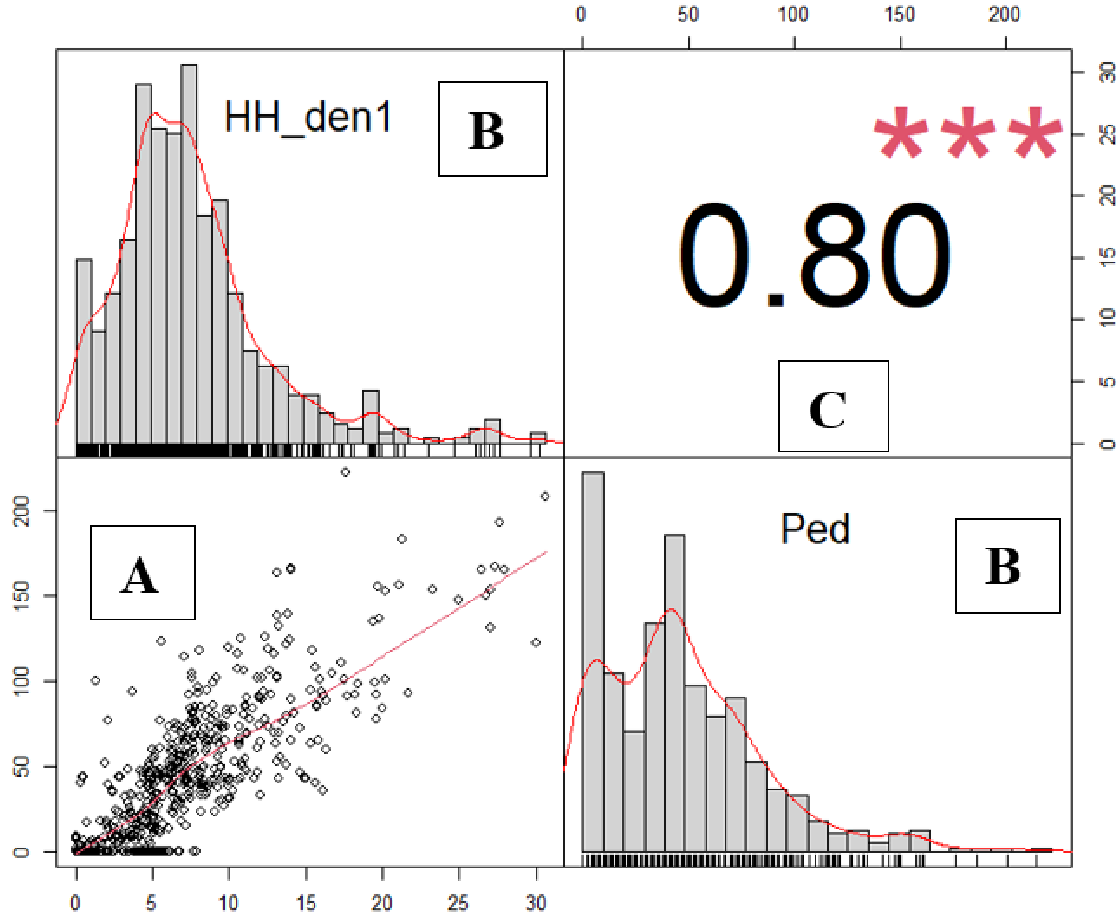
4.3 Correlation Analysis

For each variable included in this study, the authors developed a distribution plot to display a more comprehensive illustration from the statistical perspective.

To prepare the model inputs, 11 correlation matrix plots with significance levels (p-values) have been generated between all significant individual variables (and related dependent variables) from the results of the joint model. As shown in Figures 11 to 21, each individual variable's distribution is displayed on the diagonal in each plot. The top of the diagonal represents the correlation coefficient and the significance level, with stars indicating the p-values. The different symbols reveal different ranges of p-values (in particular, *** denotes 0–0.001, ** denotes 0.001–0.01, * denotes 0.01–0.05, . denotes 0.05–0.1, and a blank space denotes 0.1–1). Generally, a p-value less than 0.05 is considered statistically significant, and an absolute value of the correlation coefficient larger than 0.6 indicates that they are statistically different. If the p-value is less than 0.05 while the absolute value of the correlation coefficient is greater than 0.6, they are statistically correlated. The bottom of the diagonal represents the bivariate scatter plots with a fitted line.

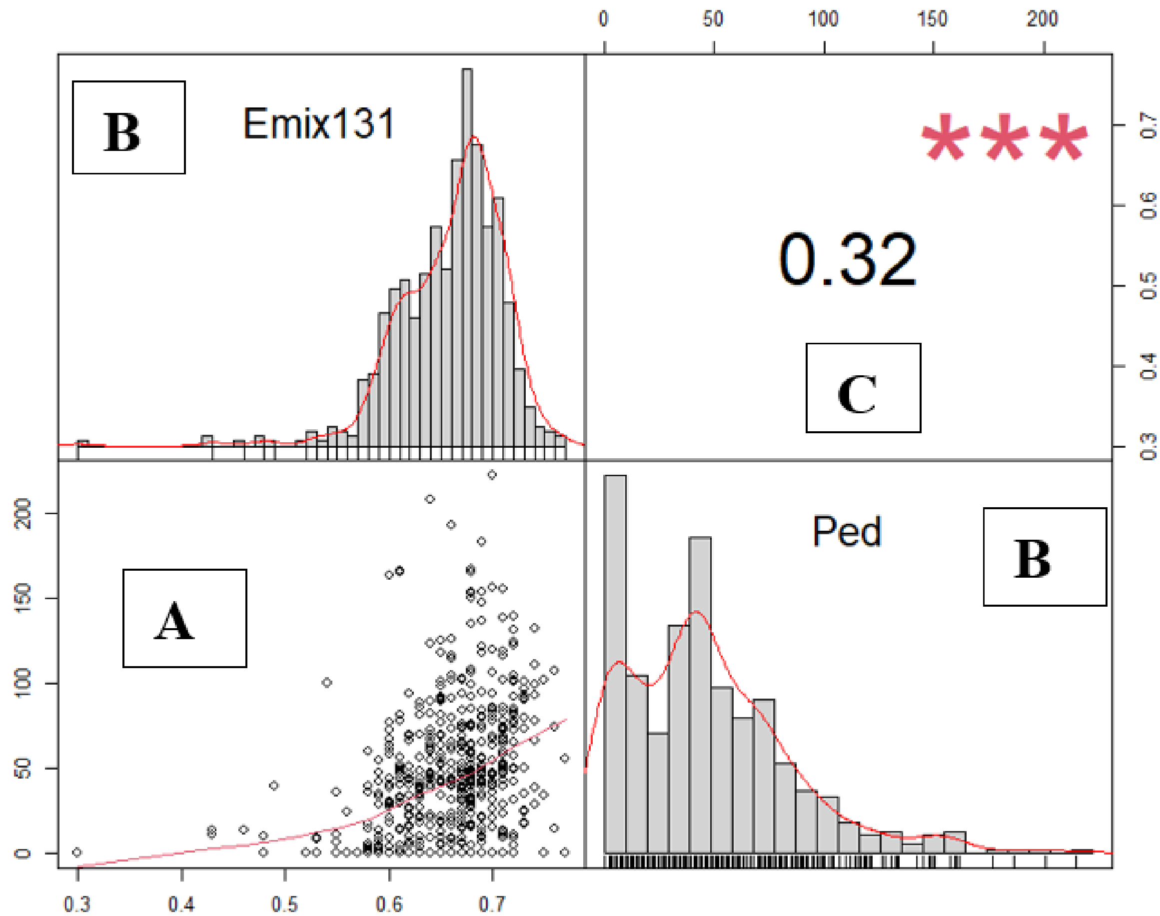
Figures 11 to 21 show the correlation matrix plot between each significant individual variable (based on the results from the joint model estimation) and related dependent variables (i.e., pedestrian-involved collision count, bike-involved collision count).

Figure 11. Numerical Correlation Test between Household Density and Pedestrian-Involved Collision Counts



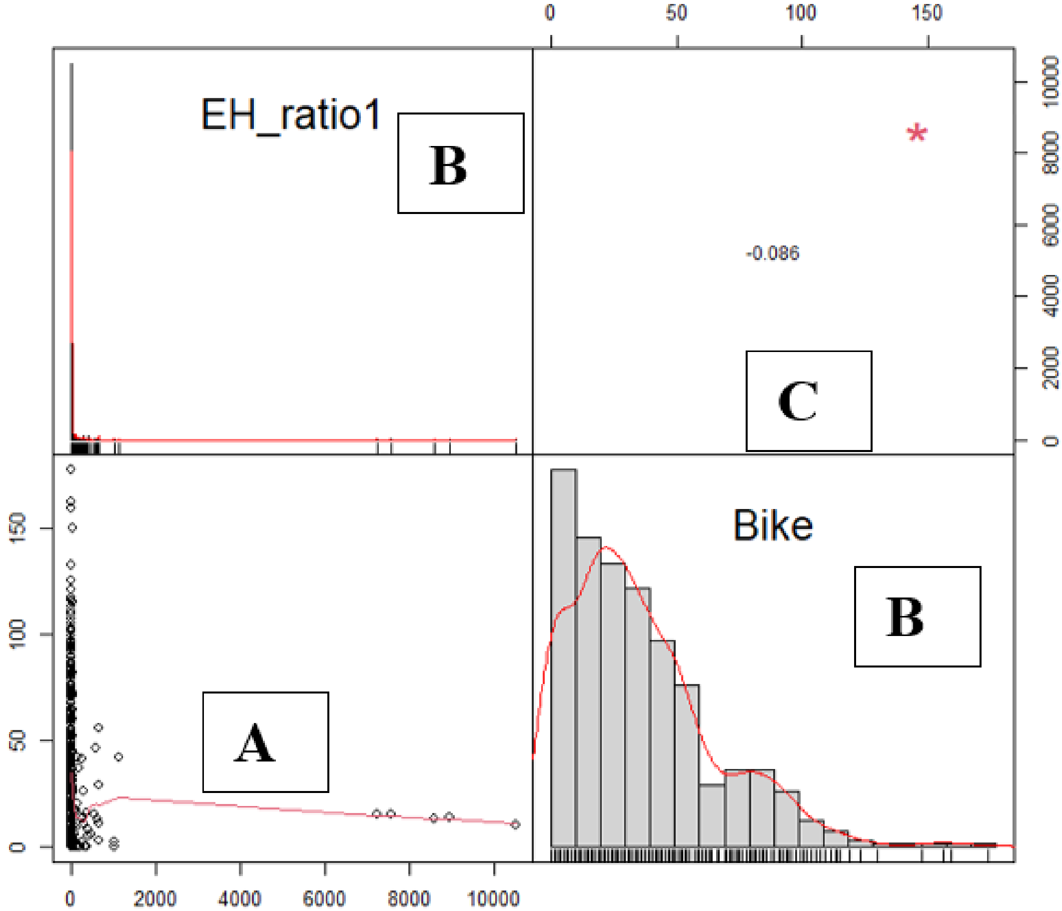
A: The bivariate scatter plot with a fitted line. B: The distribution of household density and pedestrian-involved collision counts. C: Correlation coefficient and significance level.

Figure 12. Numerical Correlation Test between Employment Mix (13 sectors) and Pedestrian-Involved Collision Counts



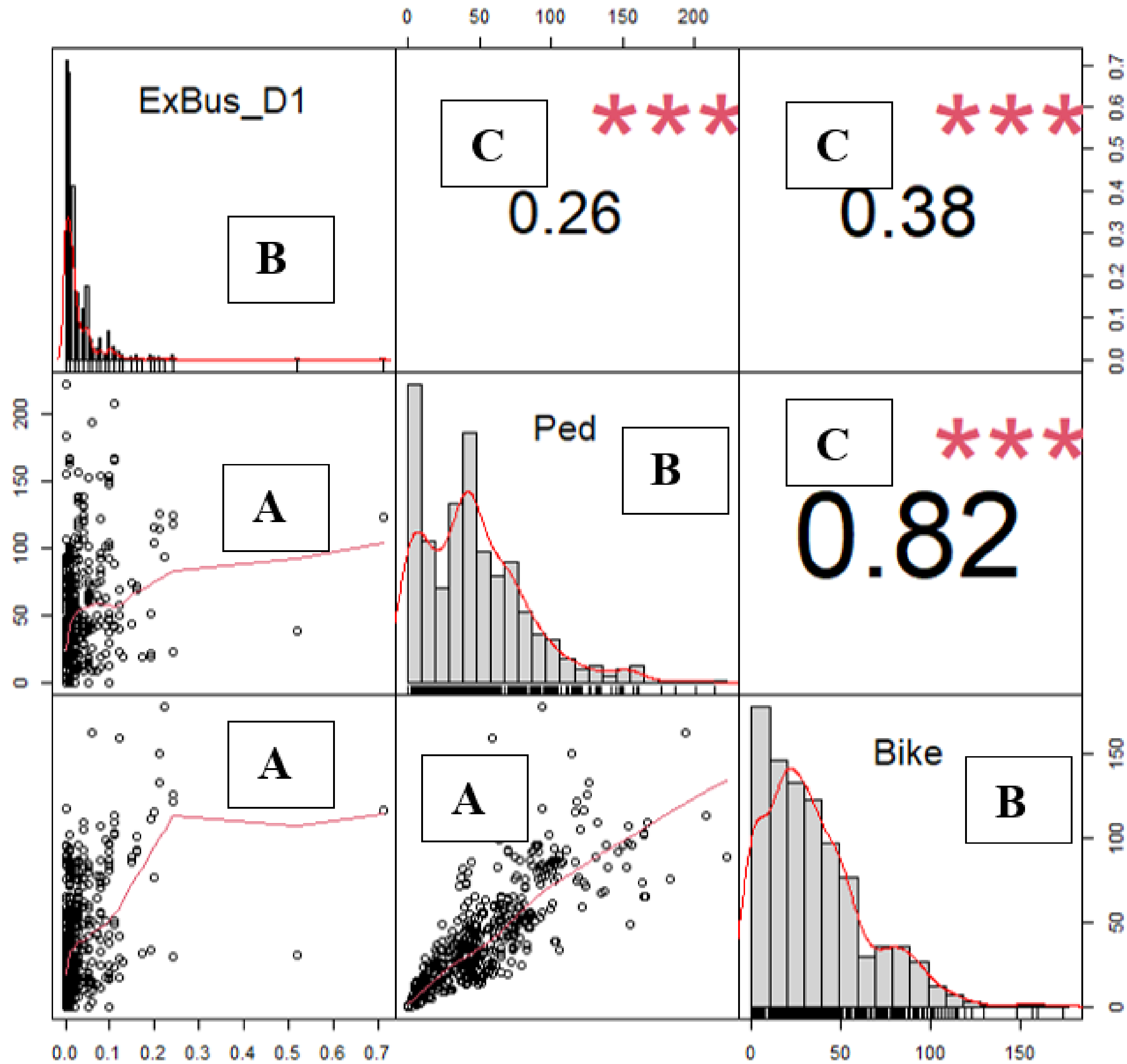
A: The bivariate scatter plot with a fitted line. B: The distribution of employment mix (13 sectors) and pedestrian-involved collision counts. C: Correlation coefficient and significance level.

Figure 13. Numerical Correlation Test between Job/Household Ratio and Bike-Involved Collision Counts



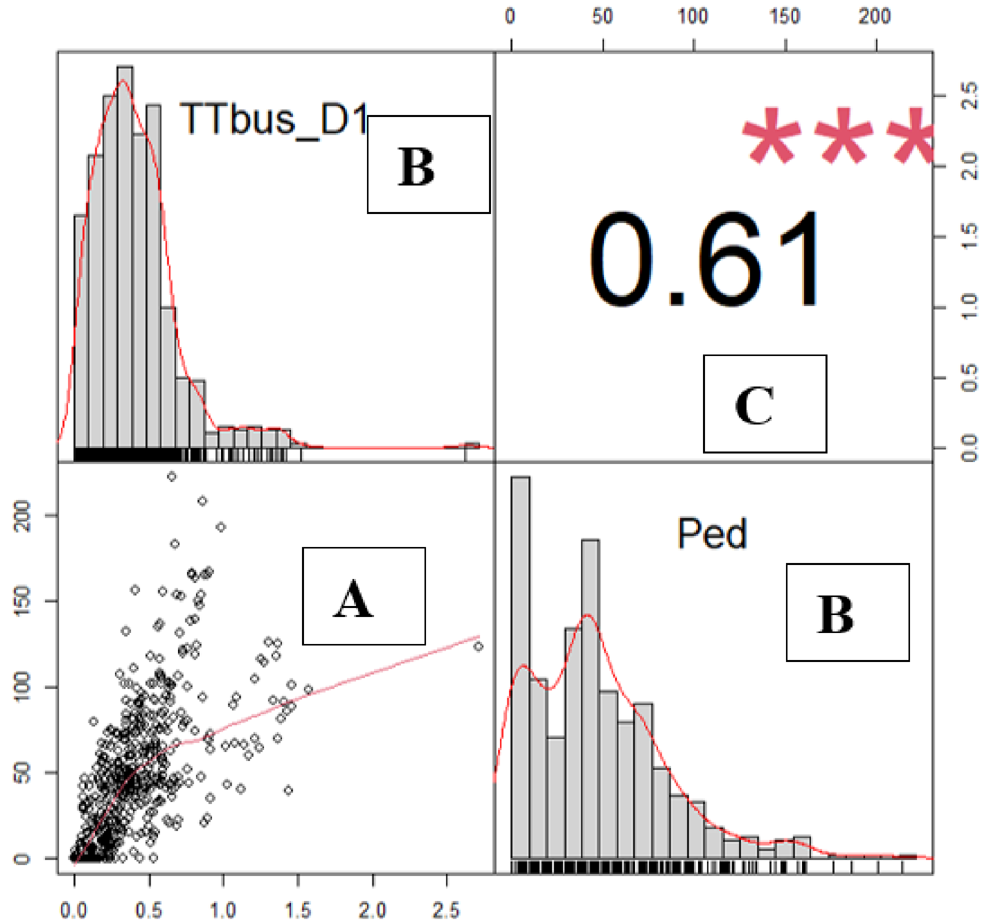
A: The bivariate scatter plot with a fitted line. B: The distribution of job/household ratio and bike-involved collision counts. C: Correlation coefficient and significance level.

Figure 14. Numerical Correlation Test between Stop Density for Express Bus and BRT and both Pedestrian- and Bike-Involved Collision Counts



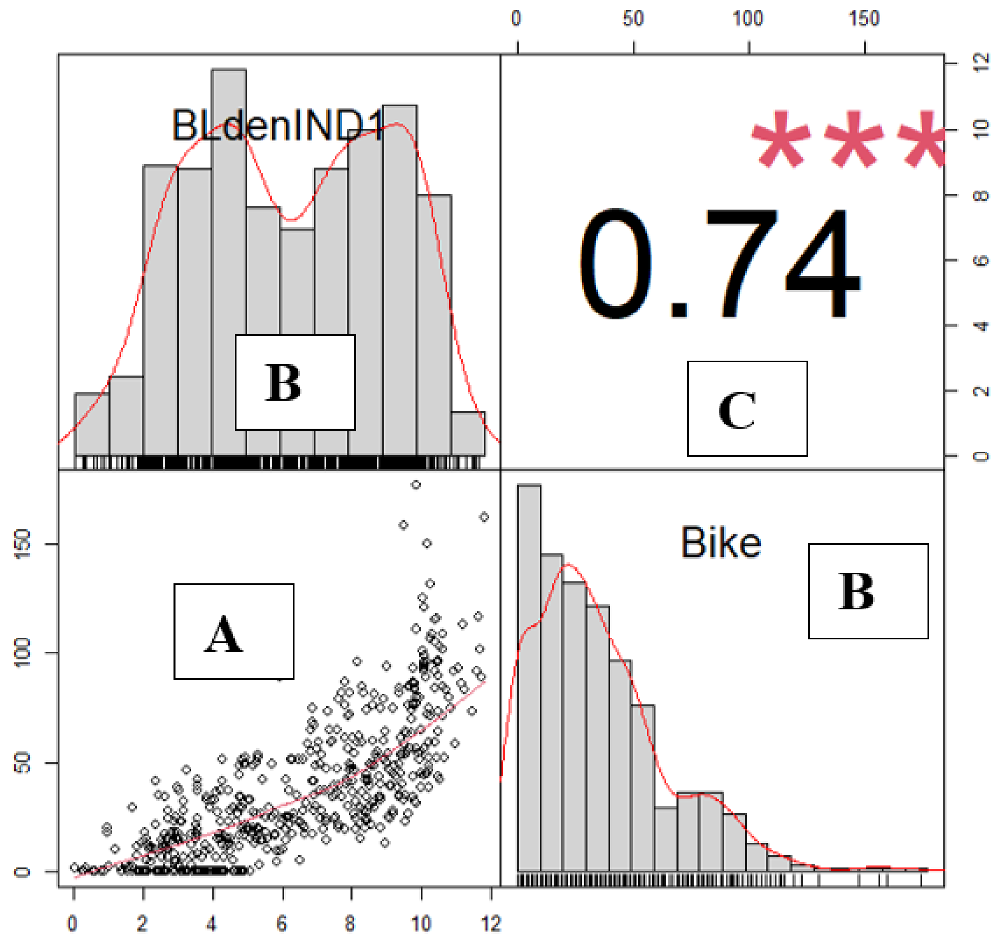
A: The bivariate scatter plots with a fitted line. B: The distribution of job/household ratio, pedestrian-involved and bike-involved collision counts. C: Correlation coefficients and significance levels.

Figure 15. Numerical Correlation Test between Total Bus Stop Density and Pedestrian-Involved Collision Counts



A: The bivariate scatter plots with a fitted line. B: The distribution of total bus stop density and pedestrian-involved collision counts. C: Correlation coefficient and significance level.

Figure 16. Numerical Correlation Test between Bike Lane Density Indicator and Bike-Involved Collision Counts



A: The bivariate scatter plots with a fitted line. B: The distribution of bike lane density indicator and bike-involved collision counts. C: Correlation coefficient and significance level.

Figure 17. Numerical Correlation Test between Estimated Block Length and Bike-Involved Collision Counts

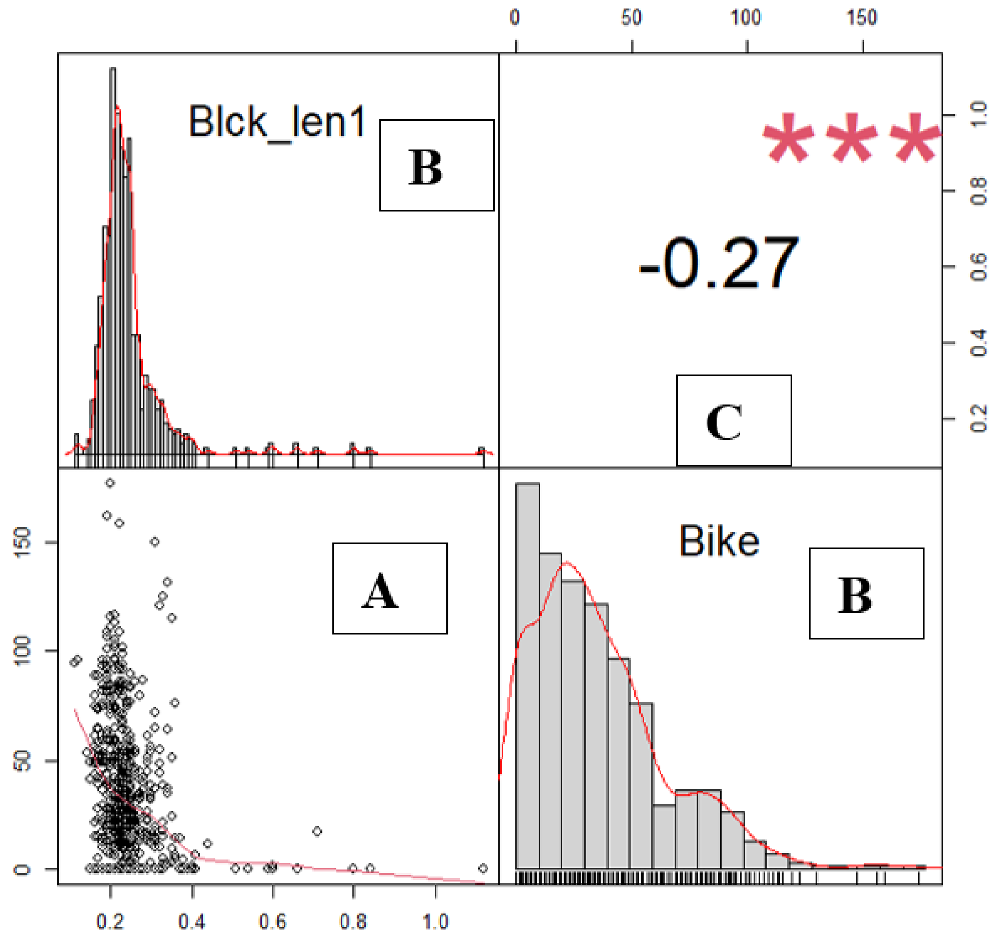
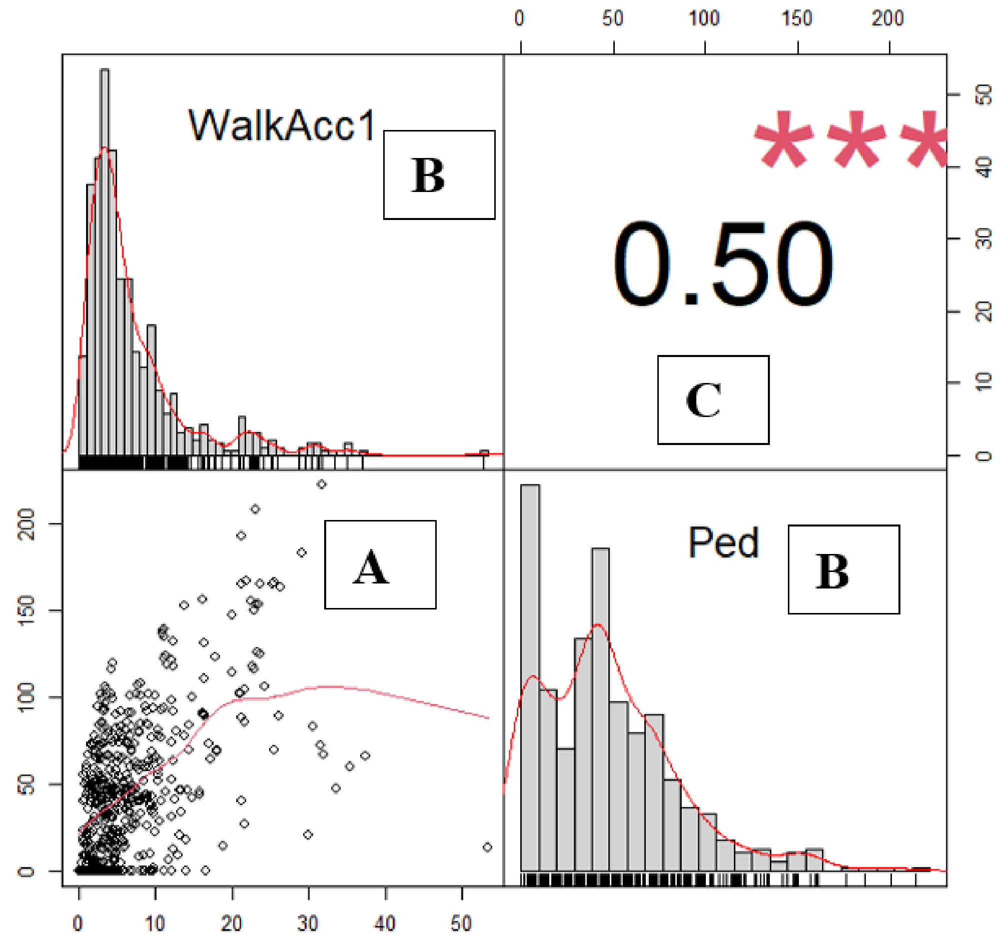


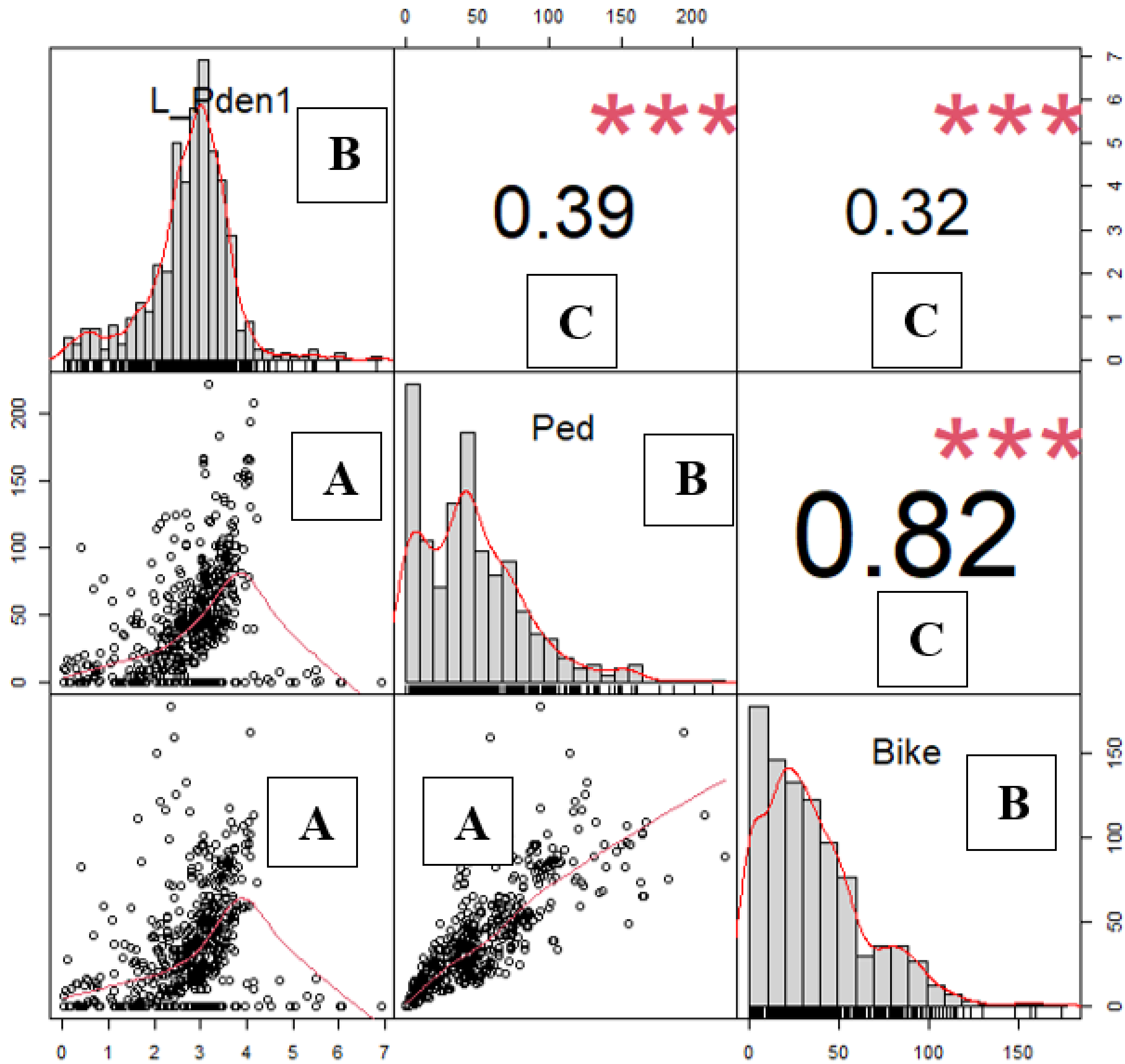
Figure 17. A: The bivariate scatter plots with a fitted line. B: The distribution of estimated block length and bike-involved collision counts. C: Correlation coefficient and significance level.

Figure 18. Numerical Correlation Test between Walk Accessibility and Pedestrian-Involved Collision Counts



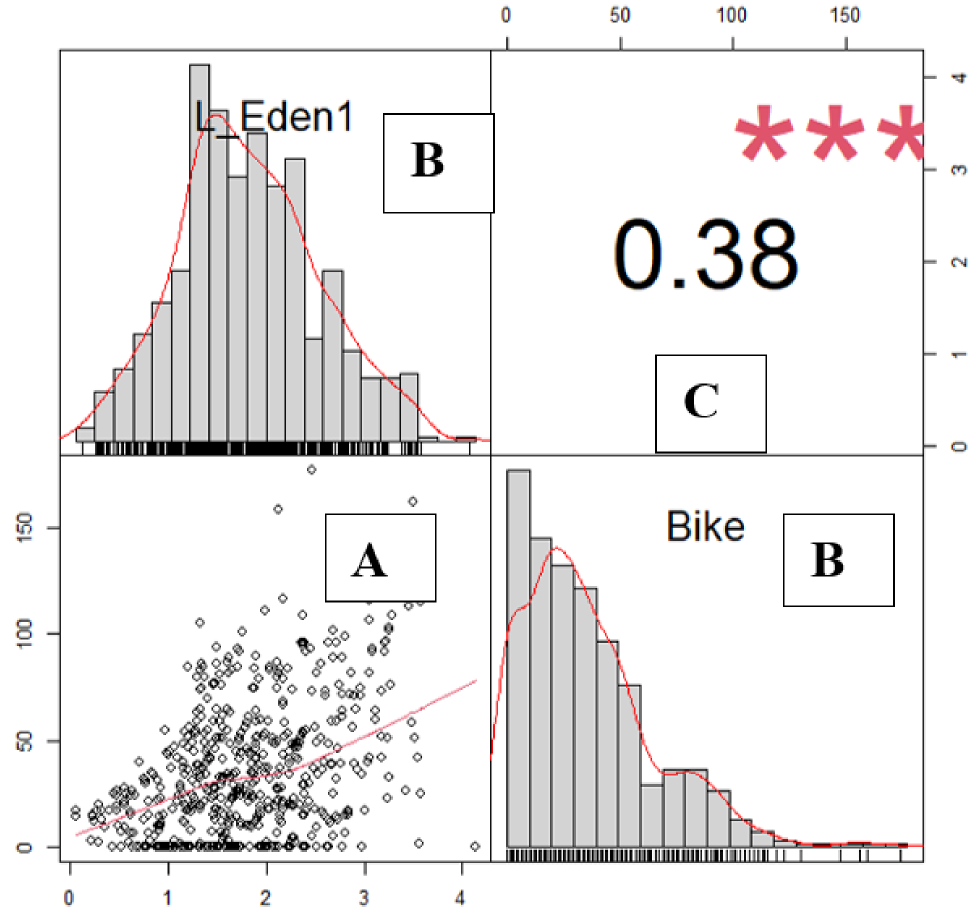
A: The bivariate scatter plots with a fitted line. B: The distribution of walk accessibility and pedestrian-involved collision counts. C: Correlation coefficient and significance level.

Figure 19. Numerical Correlation Test between Log Population Density and both Pedestrian and Bike Collision Counts



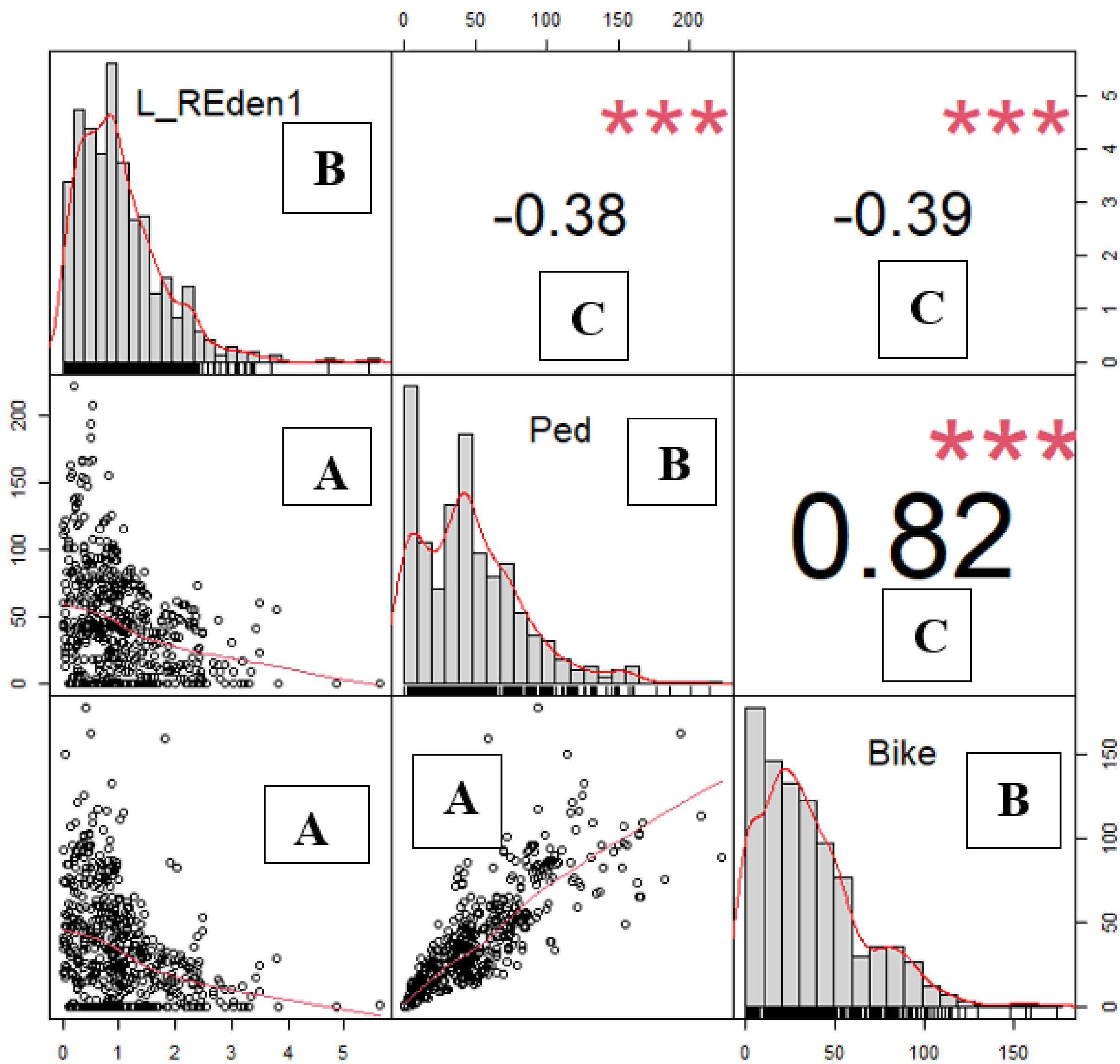
A: The bivariate scatter plots with a fitted line. B: The distribution of log population density, pedestrian-involved, and bike-involved collision counts. C: Correlation coefficients and significance levels.

Figure 20. Numerical Correlation Test between Log Employment Density and Bike-Involved Collision Counts



A: The bivariate scatter plots with a fitted line. B: The distribution of log employment density and bike-involved collision counts. C: Correlation coefficient and significance level.

Figure 21. Numerical Correlation Test between Log Retail Job Density and both Pedestrian and Bike Collision Counts



A: The bivariate scatter plots with a fitted line. B: The distribution of log retail job density, pedestrian-involved, and bike-involved collision counts. C: Correlation coefficients and significance levels.

To visualize the entire correlation matrix, the authors also created a graphical display (also known as a correlogram); it is useful to reflect the most correlated variables in one data table since the correlogram can distinguish various correlation coefficient values by color. As shown in Figure 22, the distributions of relatively highly correlated variables have been highlighted in different colors. For example, "HH_den1" (household density) and "BLdenIND1" (bike lane density indicator) are highly correlated with "Ped" (pedestrian-involved collision count).

Figure 22. Correlogram Plot of Correlation Coefficients among Individual Significant Variables from the Joint Model and Dependent Variables

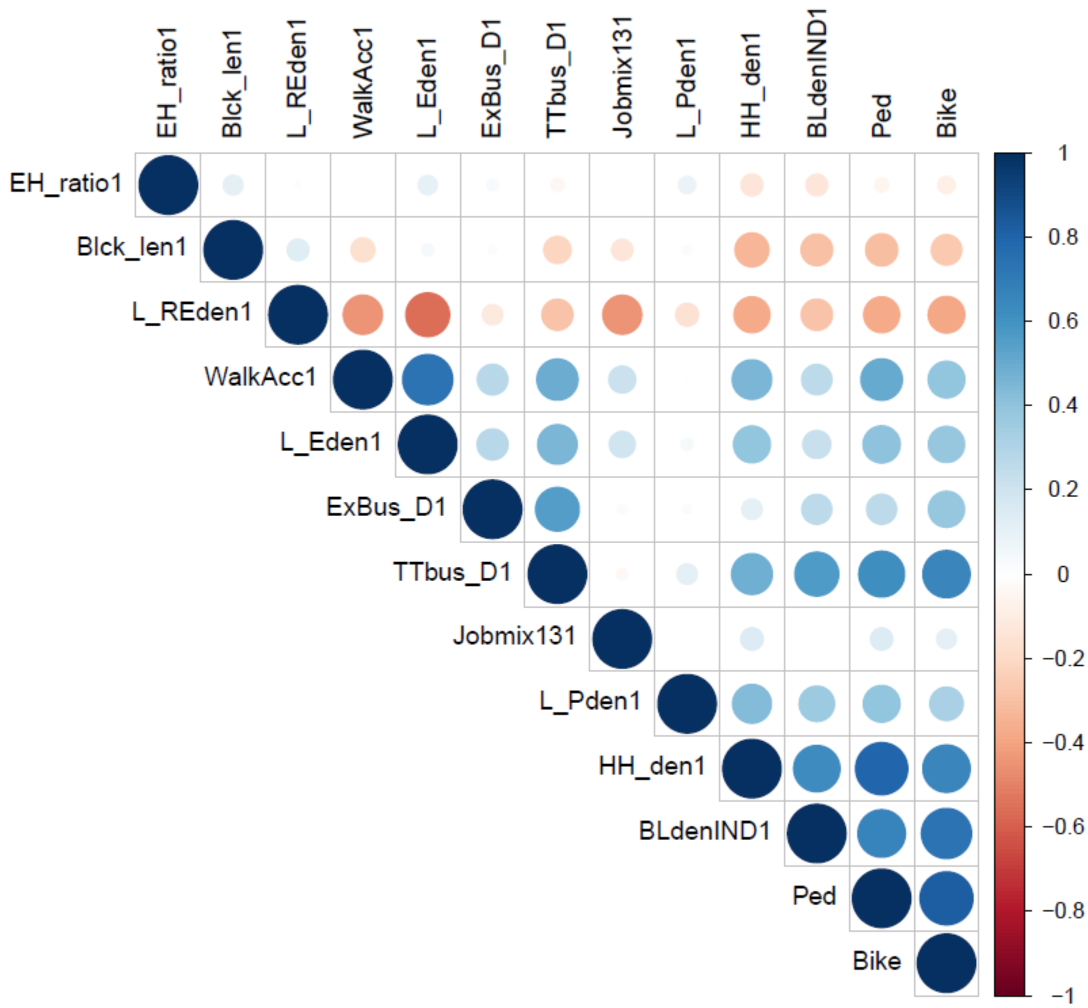


Table 2 shows the summary of correlation coefficients between significant individual variables from the joint model and dependent variables. As mentioned above, this study compares the results from joint models with the individual correlation test to explore the results of significant variables without any association between dependent variables. As revealed in Table 2, the results from the individual correlation test show that fewer variables are statistically correlated with pedestrian-involved or bike-involved crash counts. Only household density, total bus stop density, and bike lane density indicators are highly correlated with related dependent variables, namely pedestrian- and bike-involved collision counts. A possible explanation is the individual correlation test didn't include an unobserved relationship between pedestrians and cyclists in the model at the same time. This is because the individual correlation test also shows that pedestrian-involved collisions and bike-involved collisions are statistically correlated with each other (as shown in Figure 14, Figure 19, and Figure 21). The difference between the results of the individual correlation test and the joint model has crucially illustrated the reliability of the joint model since it is very flexible in selecting covariates pertaining to different dependent variables and considering correlations between dependent variables in the model as well.

Table 2. Summary of Correlation Coefficient between Individual Significant Variables and Dependent Variables

Variables	Description	Correlation Coefficient Related with Ped (Significance Level)	Correlation Coefficient Related with Bike (Significance Level)
HH_den1	Household density (households/acre)	**0.799(***)**	NA
Emix131	Employment mix (13 sectors); 1 = highest mix (jobs are equal for all sectors)	0.318(***)	NA
EH_ratio1	Job/Household ratio	NA	-0.086 (*)
ExBus_D1	Stop density for express bus and BRT	0.262(***)	0.381(***)
TTbus_D1	Total bus stop density	**0.614(***)**	NA
BLdenIND1	Bike Lane Density Indicator = Sum (Bike Lane Density/Distance to Home TAZ within three miles) Bike Lane Density for Each TAZ = ((Street15–25mph)*1 + (Street35mphj)*2 + Bike Lane Class1*3 + Bike Lane Class2*4 Bike Lane Class3*5) / Total TAZ area (excluding speed > 60mph)	NA	**0.735(***)**
Blck_len1	Estimated block length = LocalSt/Int34new (Total street length / number intersection) but highways are excluded	NA	-0.269(***)
WalkAcc1	Walk Accessibility (RS_den2/block_len) = (weighted retail + service density)/ estimated block length	0.503(***)	NA
L_Pden1	L_Pden = Ln (Pop_den + 0.001)	0.393(***)	0.325(***)
L_Eden1	L_Eden = Ln (Emp_den + 0.001)	NA	0.382(***)
L_REden1	L_REden = Ln (Ret_den + 0.001)	-0.376(***)	-0.390 (***)

V. Results and Discussion

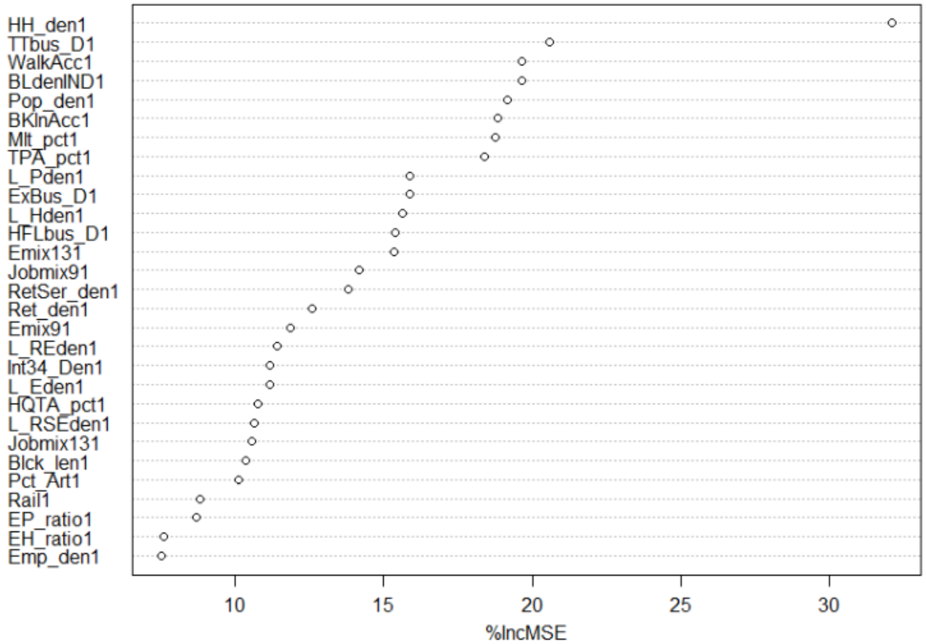
For the bivariate spatial crash prediction models, the different covariates were first selected for each of the two transportation modes: pedestrian and bicycle. The R package INLAMSM was utilized to develop the models and generate the posterior mean of the model parameters. Distinct evaluation criteria were then employed to assess the model performance.

5.1 Feature Selection

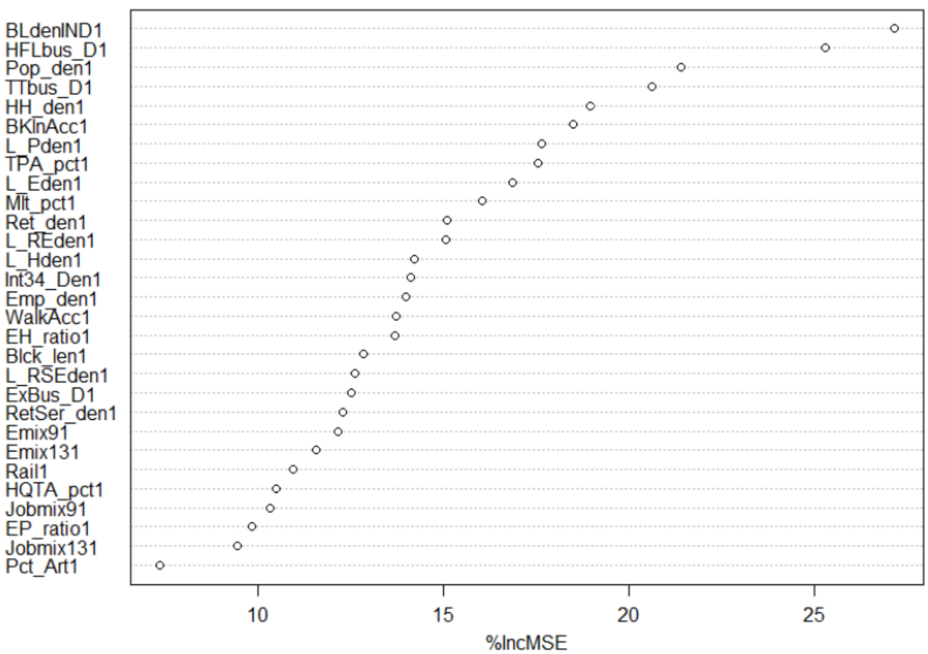
Based on the parsimony rule, it is often desirable to reduce the model data load to the fewest number of inputs with maximal predictive accuracy. The typical feature selection techniques integrated with statistical model development, like backward-forward feature selection, are usually subject to a strong assumption of a specific distribution function and cannot handle the possibility of complex variable interactions. Therefore, the present study performs feature selection using the correlation analysis and one of the ensemble techniques (random forest) which has gained greater popularity recently due to its benefits over the traditional techniques.

The random forest (RF) model was developed via the R package "randomForest" (Cutler et al., 2012). During the tree-growing stage, four predictor variables were randomly sampled as candidates at each split. The OOB error rate was found to be at a minimum of 0.264, with 63.24% of data variability being explained by the model. Once the RF model had been generated, the variable importance ranking was determined based on the contribution of variables to reducing the mean squared errors (MSE) in the OOB samples. The larger the contribution, the more important is the given variable for model development. The variable importance plots are shown in Figure 23 in decreasing order (vertically) for both pedestrian and bike counts.

Figure 23. Variable Importance Plot for (a) Pedestrian Crash Counts and (b) Bike Crash Counts



(a)



(b)

"%IncMSE" represents the percentage of the drop in mean squared errors with certain variables being excluded from the model development.

As described in detail in the previous chapter, this study also performed a correlation analysis to avoid feeding redundant information into the models. To determine whether the variables are highly correlated or not, the popular cut line of 0.6 for the correlation coefficient with the significance level of 0.05 was used. The correlated variables were removed in multiple steps using engineering judgment to avoid the exclusion of any significant variables. This procedure acts as a trade-off between omitted variable bias and multicollinearity. At last, out of thirty-two variables, fourteen variables were retained to perform modeling. It is noteworthy that different covariates are used for different transportation modes, which enhances the model flexibility and accuracy with more related salient variables being used for the respective modes.

5.2 Model Results and Discussions

Table 3 illustrates posterior estimates of model parameters of fixed effects with and without correlation of pedestrian-involved crash counts and bike-involved crash counts, as well as goodness-of-fit criteria. It can be clearly observed that the model with a correlation of pedestrian-involved crash counts and cyclist-involved crash counts highlights more significant covariates than the model without correlation. For instance, the variables "ExBus_D" (Stop density for express bus and BRT), "WalkAcc1" (Walk accessibility), "L_Pden1" (Population density, for pedestrian-involved crash counts), and "EH_ratio1" (Job/Household ratio, for cyclist-involved crash counts) were only found to be significant in the model with correlation. This finding allows the research team to address the correlation effects to obtain more insightful results explicitly. Only one variable, "L_REden1" (Retail job density), was found to be significant across both models (i.e., the presence and absence of a correlation effect). At the individual model level, for pedestrian-involved crash counts, four covariates appeared to be statistically significant across both models (i.e., the presence and absence of a correlation effect): "HH_den1" (Household density), "Emix131" (Employment mix), "TTbus_D1" (Total bus stop density), and L_REden1 (Retail job density). For bike-involved crash counts, six covariates appeared to be statistically significant across both models (i.e., the presence and absence of a correlation effect): "ExBus_D1" (Express bus stop density), "BLdenIND1" (Bike lane density indicator), "Blck_len1" (Estimated block length), "L_Pden1" (Population density), "L_Eden1" (Employment density), and "L_REden1" (Retail job density). It is worth noting that the "L_REden1" (Retail job density) variable was found to have a negative impact on both pedestrian-involved and bike-involved crashes. Further, "Blck_len1" (Estimated block length) was found to have a negative impact on the bike collision counts for both models (with and without correlation).

Table 3. Description of Model Parameter Estimates

Variables	**With Correlation of Pedestrian and Bike Collisions**				**Without Correlation of Pedestrian and Bike Collisions**			
	Pedestrian		Bike		Pedestrian		Bike	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Model Parameters								
(Intercept)	**1.572**	**0.153**	**3.788**	**0.735**	**1.604**	**0.153**	**3.792**	**0.734**
HH_den1	**0.188**	**0.028**	NA	NA	**0.453**	**0.052**	NA	NA
Jobmix91	0.015	0.044	NA	NA	0.002	0.044	NA	NA
Emix131	**0.135**	**0.024**	NA	NA	**0.286**	**0.044**	NA	NA
Emix91	NA	NA	0.021	0.040	NA	NA	0.048	0.036
EH_ratio1	NA	NA	**-0.072**	**0.019**	NA	NA	0.029	0.032
BKlnAcc1	-0.031	0.044	0.034	0.040	-0.030	0.036	0.018	0.031
ExBus_D1	**0.234**	**0.045**	**0.270**	**0.040**	-0.007	0.042	**0.130**	**0.031**
TTbus_D1	**0.095**	**0.025**	NA	NA	**0.524**	**0.050**	NA	NA
BLdenIND 1	NA	NA	**0.251**	**0.022**	NA	NA	**-0.684**	**0.038**
Blck_len1	NA	NA	**-0.086**	**0.026**	NA	NA	**-0.322**	**0.041**
WalkAcc1	**0.108**	**0.027**	NA	NA	-0.034	0.046	NA	NA
L_Pden1	**0.295**	**0.049**	**0.289**	**0.044**	0.073	0.045	**0.085**	**0.037**
L_Eden1	NA	NA	**0.100**	**0.033**	NA	NA	**0.089**	**0.043**
L_REden1	**-0.216**	**0.050**	**-0.286**	**0.050**	**-0.098**	**0.046**	**-0.139**	**0.047**
Performance Evaluation Criteria								
DIC	**8752.37**				9285.00			
WAIC	**8706.51**				9186.79			
$\overline{D}$	**7801.69**				8502.66			
P_D	950.68				**782.34**			

Notes:

1. S.D. represents standard deviation; DIC represents the deviance information criterion; WAIC represents widely applicable information criterion; $\overline{D}$ represents posterior mean deviance; P_D represents the effective number of parameters; NA means not applicable.
2. Refer to Table 1 for detailed variable descriptions.
3. For model parameters, the bold figures represent the statistically significant variables at the significance level of 0.05.
4. For performance evaluation criteria, the bold figures indicate the best performance under specific criteria.

In the following section, each statistically significant variable is examined closely to better understand its impact on pedestrian-involved and/or bike-involved accidents.

Household density (HH_den1, i.e., the number of households per acre) was observed to significantly impact the pedestrian crash counts for both cases (with and without correlation). This finding suggests that pedestrian-involved collisions are more likely to happen with increased household density. It could be attributed to the rise in pedestrian exposure elevating the risk of pedestrian crash frequency.

Employment mix across 13 sectors (Emix131) was found to be statistically significant to the increase in the pedestrian crash counts for both cases (with and without correlation). This indicates that an increase in the mix of job sectors or commercial areas increases the likelihood of pedestrian crashes. A possible reason may be that the increased amount of walking activities among job sectors elevates the pedestrian crash propensity.

Interestingly, the variable Employment/Household ratio (EH_ratio1) appeared to have a statistically negative impact on the bicyclist crash counts in models with correlation, indicating that the propensity of bicyclist crash frequency reduces with increasing employment/household ratio. A possible explanation may be due to the proximity of travel distance between the job sector and households; the commuter tends to make less use of motor vehicles, which leads to reduced vehicular-bicycle interactions. This indicates the planning and design of the built environment should be pursued in accordance with creating residential areas near the commercial or job sectors to reduce travel distance.

Stop density for express bus and BRT (ExBus_D1) was observed to be statistically significant across pedestrian and bike crash counts for both cases (with and without correlation). This suggests that the close distance between express bus or BRT stops increases the likelihood of pedestrian and bicyclist crashes. Since walking and biking activity volumes are higher at or near these transit stops, the risk of a collision between motor vehicles and pedestrians or cyclists increases.

Total bus stop density (TTbus_D1) appeared to have a statistically positive impact on pedestrian-involved crash frequency in models with and without correlation. This indicates that the high pedestrian volume or exposure at or near the bus stops tends to increase the pedestrian crashes.

As expected, the Bike lane density indicator (BLdenIND1) was found to have a statistically negative impact on bicyclists crash counts across both models with and without correlation, indicating that bike-related crashes' propensity decreases with the provision of separate infrastructure (bike lanes) to bicyclists. The bike lanes create a safer roadway environment for bicyclists and motor vehicles and encourage the use of active modes. This finding is consistent with many previous studies (McNeil et al., 2015; Park and Abdel-Aty, 2016; Basch et al., 2019)

Estimated block length (Blck_len1) seems to influence the bike crashes negatively. The phenomenon may be due to the longer reaction time and environmental adaptability given to the bicyclists by the greater street block lengths.

Walk accessibility (WalkAcc1) was found to be statistically significant for pedestrians with the consideration of correlation. This indicates that providing better walking accessibility increases pedestrian exposure and hence pedestrian crashes.

Population density (L_Pden1) was statistically significant for pedestrians and bicyclists with correlation and bicyclists without correlation. Since the population density can be easily and directly obtained from the census dataset, it is generally used to represent pedestrian exposure (Siddiqui et al., 2012; Wang et al., 2017), which increases pedestrian crash counts.

Employment density (L_Eden1) appeared to have a statistically positive impact on the bicyclist crash counts for both cases, i.e., with and without correlation. This suggests that commercial areas tend to increase bike exposure, leading to an elevation of the bike-involved crashes.

Retail job density (L_REden1) was found to have a statistically significant adverse impact across both modes and with or without correlation. The possible explanation may be that frequent walking and biking activities at retail areas reduce vehicle operating speeds and thereby decrease the likelihood of pedestrian and bicyclist collisions. This finding is consistent with the previous study (Dumbaugh and Li, 2010).

To summarize what has been discussed above, here are the major findings from the model results.

For pedestrian-involved crash counts, three covariates (Household density, Employment mix, and Total bus stop density) appear to have a statistically positive impact in both cases (with and without correlation), indicating the propensity of pedestrians to be involved in collisions with the increase of household density, employment mix, and total bus stop density, which then increases the pedestrian exposure. On the other hand, Retail job density has a negative impact on both cases.

Likewise, for bicyclist-involved crash counts, there are five statistically significant variables with consistent signs in both situations (with and without correlation): Stop density for express bus and BRT, Estimated block length, Population density, Employment density, and Retail job density. As expected, Stop density for express bus and BRT, Population density, and Employment density demonstrate positive coefficient values, suggesting that an increase in the number of express bus stops, population, or employment leads to more bicyclist-involved collisions. On the other hand, Estimated block length seems to have a statistically negative influence on the bike crash frequency. The phenomenon may be due to the longer reaction time and more environmental adaptability provided to the bicyclists by the greater street block lengths. As with pedestrian-involved accidents, Retail job density also has a negative impact on bike-involved accidents in both cases.

5.3 Model Evaluation

As previously mentioned, this study employed DIC and WAIC to assess the goodness-of-fit of different models from different perspectives. DIC, a penalized criterion, acts as a trade-off between model fit and model complexity, which are represented by its two components, posterior deviance ($\bar{D}$) and effectiveness number of parameters (P_D). WAIC, a fully Bayesian approach, was adopted as a cross-validation measure to assess the model performance from a different perspective. The models with comparatively small values of DIC and WAIC indicate better performance (Gelman et al., 2013). The results for DIC and WAIC are illustrated in Table 3. For comparison across models, the model with a correlation of pedestrian and bike counts demonstrates superior performance for DIC, WAIC, and $\overline{D}$. This superiority may be attributed to its incorporating a correlation structure that explicitly allows the flexibility to capture the spatial heterogeneity. However, the value of P_D (950.68) in the model with correlation structure is sufficiently high (with a difference of 168.34 points) compared to the model without correlation. This suggests that the inclusion of the effective number of parameters provides the flexibility to fit the data—but this advantage is accompanied by the increase in the model complexity, which translates to increased computational effort.

Table 4 shows the marginal precision and correlation coefficients between walking and biking crash counts with and without correlation. The positive marginal precision across active modes in both models indicates a positive association between pedestrian and bike crashes. The strong correlation between the transportation modes (ρ = 0.953) was observed within the correlated effect model, which shows that the close spatial proximity (0.5-mile-radius zone) may be attributed to shared unobserved factors such as road surface type, lighting condition, day/night, and weather condition between pedestrian and bicyclist crashes (Obaidat, 2012). It also shows that using the multivariate spatial framework in this study is reasonable and sensible.

Table 4. Marginal Precision and Correlation Coefficient

	With Correlation of Pedestrian-Involved Crash Counts and Cyclist-Involved Crash Counts		**Without Correlation of Pedestrian-Involved Crash Counts and Cyclist-Involved Crash Counts**	
	Pedestrian (S.D.)	**Cyclist (S.D.)**	**Pedestrian (S.D.)**	**Cyclist (S.D.)**
τ(Tau)	0.001 (4.2E-05)	0.002 (4.9E-05)	0.002 (4.2E-05)	0.003 (1.6-E04)
ρ(Rho)	0.953 (0.002)		-	

Notes:

1. τ represents the marginal precision of the values of pedestrian- and bicyclist-involved crash counts.
2. ρ illustrates the correlation coefficient between pedestrian- and cyclist-involved crash counts.
3. S.D. is the standard deviation.

VI. Conclusions

6.1 Summary

Active transportation has gained more and more attention in the past few years due to its economic, environmental, and health benefits over the typical motorized modes. However, improving the safety of non-motorists is imperative in promoting active transportation modes, especially around transit stations where different transportation modes have much higher rates of interaction. Compared with other infrastructure facilities such as intersections, sidewalks, and bike lanes, there is a lack of studies dedicated to safety analysis around transit stations. It has remained unclear which factors are the main contributors to the walking and biking safety conditions adjacent to transit stops given the complexity of the influential factors and their interactions.

To bridge this gap, the objective of the present study was to rank the importance and quantify the impact of variables pertinent to pedestrian and bicycle traffic safety near the transit stops. It is the authors' hope that the findings from this research can provide some insights for the effective design of complete streets, planning policies, and countermeasures to reduce the crashes. To address the spatial heterogeneity in the dataset surrounding high-quality transit stations, spatial models associated with multivariate settings were employed. Specifically, the researchers used joint models based on the multivariate conditionally autoregressive (MCAR) priors with a distance-oriented neighboring weight matrix. In order to take advantage of a substantial reduction in computational time and effort for estimation under a multivariate spatial framework, a package oriented to Integrated Nested Laplace Approximation (INLAMSM) was employed to carry out approximate Bayesian inference. Feature selection was also conducted using both random forest and correlation analyses, yielding different covariates to each of the two active transportation modes and increasing the model flexibility. Moreover, the statistical analysis used data centered on high-quality transit stations, including built environment characteristics, socioeconomic and demographic information, and crash data aggregated at the 0.5-mile-radius zone surrounding the stations. Finally, to assess the models' predictive accuracy, four distinct evaluation criteria were used: deviance information criterion (DIC), widely applicable information criterion (WAIC), posterior mean deviance ($\bar{D}$), and the effective number of parameters (P_D).

6.2 Conclusions

The key findings from this study are as follows.

1. Pedestrian-involved accident counts and bike-involved accident counts are highly correlated. It is better to use joint models instead of estimating the models separately.
2. The models that consider correlations of pedestrian and bike crash counts demonstrate superior performance, which may be attributed to this study's incorporation of a correlation structure that explicitly allows the flexibility to capture the spatial heterogeneity.

3. The advantage associated with models that consider correlation is accompanied by the dramatic increase in the model complexity due to the inclusion of correlation coefficients.
4. The significantly better performance accompanying models with correlation clearly justifies the benefits of addressing the transportation modes' correlation.
5. Household density, employment mix, and bus stop density positively impact pedestrian-involved crashes, indicating the propensity of pedestrians to be involved in collisions is higher with the increase of household density, employment mix, and bus stops, which then increase the pedestrian exposure. However, retail job density has a negative impact on pedestrian-involved crashes.
6. Similarly, population density, employment density, and bus stop density demonstrate a positive influence on bicyclist-involved crashes, suggesting that an increase of population, employment, or the number of bus stops leads to more bicyclist-involved collisions. On the other hand, block length has a negative influence, which may be due to the longer reaction time and more environmental adaptability given to bicyclists by the greater street block lengths. Also, retail job density has a negative impact on bike-involved crashes.

These research results can furnish transportation professionals with additional insights to create safer access to transit and thus promote active transportation modes in the State of California.

6.3 Policy Implications

The Sustainable Communities and Climate Protection Act of 2008, Senate Bill (SB) 375 of California, requires that MPOs develop a Sustainable Communities Strategy to “reduce per capita greenhouse gas emissions through integrated transportation, land use, housing and environmental planning” (SCAG 2020–2045 RTP/SCS, 2020). SB 375 creates incentives for residential or mixed-use residential projects that may be exempt from the California Environmental Quality Act (CEQA) if they are consistent with the MPO’s adopted Sustainable Communities Strategy (SCS). These “transit priority projects” must be located within half a mile of High-Quality Transit Areas (HQTAs), meaning either major transit stops or high-quality transit corridors (HQTCs). (SCAG 2020–2045 RTP/SCS, 2020). Additionally, SB 743 of California (2013) provides further opportunities for CEQA exemption and streamlining to facilitate transit-oriented development (TOD):

> *Specifically, certain types of projects within ‘transit priority areas’ (TPAs) can benefit from a CEQA exemption if they are consistent with the MPO’s adopted SCS. A TPA is an area within a half-mile of a major transit stop that is existing or planned if the planned stop is scheduled to be completed within the planning horizon included in a Federal Transportation Improvement Program (FTIP)* (SCAG 2020–2045 RTP/SCS, 2020).

For example, while HQTAs account for only 3% of total land area in the SCAG region, they are projected to accommodate 46% of the region’s future household growth and 55% of its future employment growth (SCAG, 2016). In other words, throughout the State of California, surrounding major transit stations, we should expect higher population density, employment density, and employment mix. However, as shown in this study, by increasing population density, employment density, or the employment mix, the risk of pedestrian- and bicyclist-involved accidents will increase. On the other hand, various government agencies throughout the State of California have been working hard to promote the use of various non-motorized modes, from traditional walking and biking to the trending e-bike and e-scooter. People are encouraged to use non-motorized modes to access and exit transit stations to reduce auto trips while gaining health benefits—but with more non-motorized activities, exposure to accidents is also higher.

It is clear that directing more growth around HQTAs and promoting more non-motorized travel are the right direction to go. Reducing the accidents related to these non-motorized modes while achieving the benefits associated with these adopted elements of an SCS will be a very challenging task facing various agencies.

The findings from this study reflect an improved understanding of transit-stop-related factors and their impacts on active transportation safety. It is anticipated that the present study's results will shed some light for the researchers pursuing better model development and for safety practitioners designing policies and programs regarding the safety of active transportation. Determining the significant factors affecting the safety of active modes will help the planners reduce the project costs by centralizing the resources towards potential improvements. This efficient allocation of resources will help execute effective project planning and increase the safety of active commuters for an extended period. Generally, the main focus of transportation-related agencies is directed towards automobiles and this has often limited attention paid to the safety of pedestrians and bicyclists (NACTO, 2015). This study's findings could help Department of Transportation (DOTs) and other transportation-related agencies prioritize safety-related improvements near transit stations. Such enhancements will promote public transport and make car travel less attractive in high transit areas, which will reduce pedestrian- and bicyclist-involved crashes, gas emissions, and traffic congestion—and can also help achieve substantial shifts towards active modes.

More specifically, the findings from this research show that it is essential for transportation professionals to consider walking and biking modes altogether in the planning. Despite their unique travel behavior, pedestrians and bicyclists are highly correlated, and they share the same space and time parameters when accessing transit stations.

If planned and designed well, a TOD with mixed land use can reduce auto mode share and travel distance. If people can easily access the destinations by walking/biking and transit, they don't have to drive. For example, a regular car lane can be converted to a dedicated bus-only lane to improve transit service while reducing the auto traffic around transit stations. When the demand for auto traffic is lower, more resources can be reallocated to non-motorized users, such as by widening the sidewalks and adding bike lanes, reducing the potential conflicts with vehicles further, and enhancing safety for pedestrians and bicyclists.

This research also shows when adding more jobs, particularly retail jobs, in the TODs, it is possible to reduce accidents related to non-motorized users.

It is also shown the propensity of bike-related crashes decreases by providing bike lanes to bicyclists or designing a longer street block length. The bike lanes create a safer roadway environment for bicyclists and motor vehicles and encourage active modes. While it is hard to change the street block length, physical barriers can be added to separate motor vehicle traffic from bicycle traffic if necessary.

This study also shows that the close distance between transit stations increases pedestrian and bicyclist crashes. Since walking and biking activities are higher at or near the transit stops, the crash risk between motor vehicles and active mode users increases. For transit stations with high volumes and in close proximity, transit operators need to look into the demand and capacity to provide sufficient space for pedestrians and bicyclists to wait at the transit stations safely. Transit operators may also need to coordinate the adjacent stations along busy transit corridors or stations that are very close to each other at busy intersections to ensure pedestrians and bicyclists can smoothly and safely access transit stations.

The model created in this study can be used to identify crash hot spots by ranking the frequency of the predominant crash type at each transit station. With the ranking, transit operators or city officials/planners can prioritize improvements near the stations with the greatest need. BIKESAFE has listed a total of 46 engineering, education, and enforcement countermeasures. Typical countermeasures include lighting improvements, turning restrictions for autos, pavement marking, comprehensive wayfinding system, optimizing signal timing for pedestrians and bicyclists, and so on. It is worth mentioning that in most urban intersections, signal timing is designed for vehicles only. For intersections with a high demand for pedestrian crossing, signal timing may be adjusted for pedestrians. But very rarely has the signal timing taken into consideration all users, including bicyclists. For example, pedestrian signal heads are very common, but bicycle signal heads are very rare. Bicycles (as well as e-bikes, e-scooters, scooters, skateboards) have unique operating characteristics worth paying attention to. After all, signal timings aim to provide safe crossings and minimize delays for all users.

6.4 Research Limitations

The findings from this study reflect an improved understanding of transit-stop-related factors and their impacts on active transportation safety. However, it is important to be aware of some caveats. The current findings are based on empirical results obtained from the bicycle and pedestrian crash data from Los Angeles County alone. The superiority of specific models may not hold when employed at a different spatial level. Second, investigating the effects of both spatial and temporal correlation on the count models of active modes may highlight more valuable insights. Third, other feature selection techniques may lead to different covariates and hence different coefficient values. Fourth, the zone with a fixed radius of 0.5 miles was utilized for model development and evaluation purposes. The zones with different radii or varying radii corresponding to zones' characteristics might generate different findings. Finally, only the active transportation modes were considered for the proposed research. The inclusion of other transportation modes or their associated interactions is also worthy of investigation in future research.

Abbreviations and Acronyms

AIC	Akaike Information Criterion
BRT	Bus Rapid Transit
CAR	Conditionally Autoregressive
CSUTC	*California State University Transportation Consortium*
Dbar ($\bar{D}$)	Posterior Mean Deviance
DIC	Deviance Information Criterion
DOT	Department of Transportation
FHWA	Federal Highway Administration
GIS	Geographical Information System
HQTA	High-Quality Transit Area
HQTC	High-Quality Transit Corridor
INLA	Integrated Nested Laplace Approximation
INLAMSM	Multivariate Spatial Model with "INLA"
MCAR	Multivariate Conditionally Autoregressive
MCMC	Markov Chain Monte Carlo
MPO	Metropolitan Planning Organization
MSE	Mean Squared Error
OOB	Out-of-Bag
RF	Random Forest
RTP	Regional Transportation Plan
P_D	Effective Number of Parameters
SAR	Simultaneous Autoregressive
SCAG	Southern California Association of Governments
SCS	Sustainable Communities Strategies
SWITRS	Statewide Integrated Traffic Records System

TAZ	Traffic Analysis Zone
TIMS	Transportation Injury Mapping System
TOD	Transit-Oriented Development
WAIC	Widely Applicable Information Criterion

Bibliography

Abdel-Aty, Mohamed, and Kirolos Haleem. "Analyzing Angle Crashes at Unsignalized Intersections using Machine Learning Techniques." *Accident Analysis & Prevention* 43, no. 1 (2011): 461–470. https://doi.org/10.1016/j.aap.2010.10.002

Abdel-Aty, Mohamed, and Xuesong Wang. "Crash Estimation at Signalized Intersections along Corridors: Analyzing Spatial Effect and Identifying Significant Factors." *Transportation Research Record* 1953, no. 1 (2006): 98–111. https://doi.org/10.1177/0361198106195300112

Aguero-Valverde, Jonathan. "Full Bayes Poisson Gamma, Poisson Lognormal, and Zero Inflated Random Effects Models: Comparing the Precision of Crash Frequency Estimates." *Accident Analysis & Prevention* 50 (2013): 289–297. https://doi.org/10.1016/j.aap.2012.04.019

Aguero-Valverde, Jonathon. "Multivariate Spatial Models of Excess Crash Frequency at Area Level: Case of Costa Rica." *Accident Analysis & Prevention* 59 (2013): 365–373. https://doi.org/10.1016/j.aap.2013.06.014

Aguero-Valverde, Jonathon, and Paul P. Jovanis. "Analysis of Road Crash Frequency with Spatial Models." *Transportation Research Record* 2061, no. 1 (2008): 55–63. https://doi.org/10.3141/2061-07

Anarkooli, Alireza J., Bhagwant Persaud, Mehdi Hosseinpour, and Taha Saleem. "Comparison of Univariate and Two-Stage Approaches for Estimating Crash Frequency by Severity: Case Study for Horizontal Curves on Two-Lane Rural Roads." *Accident Analysis & Prevention* 129 (2019): 382–389. https://doi.org/10.1016/j.aap.2018.08.015

Barbour, Elisa. "Evaluating Sustainability Planning under California's Senate Bill 375." *Transportation Research Record: Journal of the Transportation Research Board* 2568, no. 1 (2016): 17–25. https://doi.org/10.3141/2568-04

Barua, Sudip, Karim El-Basyouny, and Md T. Islam. "Multivariate Random Parameters Collision Count Data Models with Spatial Heterogeneity." *Analytic Methods in Accident Research* 9 (2016): 1–15. https://doi.org/10.1016/j.amar.2015.11.002

Basch, Corey, H., Danna Ethan, and Charles E. Basch. "Bike Lane Obstructions in Manhattan, New York City: Implications for Bicyclist Safety." *Journal of Community Health* 44, no. 2 (2019): 396–399. https://doi.org/10.1007/s10900-018-00596-4

Behnamian, A., Millard, K., Banks, S. N., White, L., Richardson, M., and Pasher, J. "A Systematic Approach for Variable Selection with Random Forests: Achieving Stable Variable Importance Values." *IEEE Geoscience and Remote Sensing Letters* 14, no. 11 (2017): 1988–1992.

Besag, Julian. "Spatial Interaction and the Statistical Analysis of Lattice Systems." *Journal of the Royal Statistical Society: Series B (Methodological)* 36, no. 2 (1974): 192–225. https://doi.org/10.1111/j.2517-6161.1974.tb00999.x

Besharati, Mohammad Mehdi, Ali Tavakoli Kashani, Zili Li, Simon Washington, and Carlo G. Prato. "A Bivariate Random Effects Spatial Model of Traffic Fatalities and Injuries across Provinces of Iran." *Accident Analysis & Prevention* 136 (2020), 105394. https://doi.org/10.1016/j.aap.2019.105394

Bijleveld, Frits D. "The Covariance between the Number of Accidents and the Number of Victims in Multivariate Analysis of Accident Related Outcomes." *Accident Analysis & Prevention* 37, no. 4 (2005): 591–600. https://doi.org/10.1016/j.aap.2005.01.004

BIKESAFE, Bicycle Safety Guide and Countermeasure Selection System. http://www.pedbikesafe.org/bikesafe/countermeasures.cfm

Blangiardo, Marta, Michela Cameletti, Gianluca Baio, and Havard Rue. "Spatial and Spatio-Temporal Models with R-INLA." *Spatial and Spatio-Temporal Epidemiology* 4 (2013): 33–49. https://doi.org/10.1016/j.sste.2012.12.001

Boulieri, Areti, Silvia Liverani, Kees de Hoogh, and Marta Blangiardo. "A Space-Time Multivariate Bayesian Model to Analyse Road Traffic Accidents by Severity." *Journal of the Royal Statistical Society* 180 (2016): 119–39. http://dx.doi.org/10.1111/rssa.12178

Cai, Qing, Mohamed Abdel-Aty, and Jaeyoung Lee. "Macro-Level Vulnerable Road Users Crash Analysis: A Bayesian Joint Modeling Approach of Frequency and Proportion." *Accident Analysis & Prevention* 107 (2017): 11–19. https://doi.org/10.1016/j.aap.2017.07.020

Cai, Qing, Moharmed Abdel-Aty, Jaeyoung Lee, L. Wang, and Xuesong Wang. "Developing a Grouped Random Parameters Multivariate Spatial Model to Explore Zonal Effects for Segment and Intersection Crash Modeling." *Analytic Methods in Accident Research* 19 (2018): 1–15. https://doi.org/10.1016/j.amar.2018.05.001

Cheng, Wen, Gurdiljot S. Gill, Mohan Dasu, and Xudong Jia. "An Empirical Evaluation of Multivariate Spatial Crash Frequency Models." *Accident Analysis & Prevention* 119 (2018): 290–306. https://doi.org/10.1016/j.aap.2018.07.001

Cheng, Wen, Gurdiljot Singh Gill, Taha Sakrani, Mohan Dasu, and Jiao Zhou. "Predicting Motorcycle Crash Injury Severity Using Weather Data and Alternative Bayesian Multivariate Crash Frequency Models." *Accident Analysis & Prevention* 108 (2017): 172–180. https://doi.org/10.1016/j.aap.2017.08.032

Chiou, Yu C., Chiang Fu, and Hsieh Chih-Wei. "Incorporating Spatial Dependence in Simultaneously Modeling Crash Frequency and Severity." *Analytic Methods in Accident Research* 2 (2014): 1–11. https://doi.org/10.1016/j.amar.2013.12.001

Congdon, Peter. *Bayesian Statistical Modelling* (vol. 704). John Wiley & Sons, 2007. https://books.google.com/books?hl=zh-CN&lr=&id=PthTOJyGGNIC&oi=fnd&pg=PR5&dq=Bayesian+Statistical+Modelling&ots=1a1CoddePz&sig=PcSGNsYKNS_Y3Z7uzMaXJTEaV8A#v=onepage&q&f=false

Conway, Alison, Jialei Cheng, Diniece Peters, and Nicholas Lownes. "Characteristics of Multimodal Conflicts in Urban On-Street Bicycle Lanes." *Transportation Research Record: Journal of the Transportation Research Board* 2387, no. 1 (2013): 93–101. https://doi.org/10.3141/2387-11

Cutler, Adele, D. Richard Cutler, and John R. Stevens. "Random Forests." *Ensemble Machine Learning* (2012): 157–175. https://link.springer.com/chapter/10.1007/978-1-4419-9326-7_5

Daniels, Stijin, Tom Brijs, Erik Nuyts, and Geert Wets. "Explaining Variation in Safety Performance of Roundabouts." *Accident Analysis & Prevention* 42, no. 2 (2010): 393–402. https://doi.org/10.1016/j.aap.2009.08.019

Dumbaugh, Eric, and Wenhao Li. "Designing for the Safety of Pedestrians, Cyclists, and Motorists in Urban Environments." *Journal of the American Planning Association* 77, no. 1 (2010): 69–88. https://doi.org/10.1080/01944363.2011.536101

El-Basyouny, Karim, and Tarek Sayed. "Comparison of Two Negative Binomial Regression Techniques in Developing Accident Prediction Models." *Transportation Research Record* 1950, no. 1 (2006): 9–16. https://doi.org/10.1177/0361198106195000102

Gelman, Andrew, Jessica Hwang, and Aki Vehtari. "Understanding Predictive Information Criteria for Bayesian Models." *Statistics and Computing* 24 (2013): 997–1016. https://link.springer.com/article/10.1007/s11222-013-9416-2

Hauer, Ezra. "Overdispersion in Modeling Accidents on Road Sections and in Empirical Bayes Estimation." *Accident Analysis and Prevention* 33, no. 6 (2001): 799–808. https://doi.org/10.1016/S0001-4575(00)00094-4

Heydari, Shahram, Liping Fu, Luis Miranda-Moreno, and Lawrence Joseph. "Using a Flexible Multivariate Latent Class Approach to Model Correlated Outcomes: A Joint Analysis of Pedestrian and Cyclist Injuries." *Analytic Methods in Accident Research* 13 (2017): 16–27. https://doi.org/10.1016/j.amar.2016.12.002

Hosseinpour, Mehdi, Sina Sahebi, Zamira H. Zamzuri, Ahmad S. Yahaya, and Noriszura Ismail. "Predicting Crash Frequency for Multi-Vehicle Collision Types using Multivariate Poisson-Lognormal Spatial Model: A Comparative Analysis." *Accident Analysis & Prevention* 118 (2018): 277–288. https://doi.org/10.1016/j.aap.2018.05.003

Huang, Helai, Hanchu Zhou, Jie Wang, Fangrong Chang, and Ming Ma. "A Multivariate Spatial Model of Crash Frequency by Transportation Modes for Urban Intersections." *Analytic Methods in Accident Research* 14 (2017): 10–21. https://doi.org/10.1016/j.amar.2017.01.001

Imprialou, Maria I. M., Mohammed Quddus, and David E. Pitfield. "Predicting the Safety Impact of a Speed Limit Increase Using Condition-Based Multivariate Poisson Lognormal Regression." *Transportation Planning and Technology* 39, no. 1 (2016): 3–23. https://doi.org/10.1080/03081060.2015.1108080

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning* (vol. 112, p. 18). New York: Springer, 2013. https://link.springer.com/book/10.1007%2F978-1-4614-7138-7

Jiang, Ximiao, Mohamed Abdel-Aty, Jia Hu, and Jaeyoung Lee. "Investigating Macro-Level Hotzone Identification and Variable Importance using Big Data: A Random Forest Models Approach." *Neurocomputing* 181 (2016): 53–63. https://doi.org/10.1016/j.neucom.2015.08.097

Jin, Xiaoping, Bradley P. Carlin, and Sudipto Banerjee. "Generalized Hierarchical Multivariate CAR Models for Areal Data." *Biometrics* 61, no. 4 (2005): 950–961. https://doi.org/10.1111/j.1541-0420.2005.00359.x

Jonathan, Aguero V., Kun F. Wu, and Eric T. Donnell. "A Multivariate Spatial Crash Frequency Model for Identifying Sites with Promise Based on Crash Types." *Accident Analysis & Prevention* 87 (2016): 8–16. https://doi.org/10.1016/j.aap.2015.11.006

Kim, Hoon, Dongchu Sun, and Robert K. Tsutakawa. "Lognormal vs. Gamma: Extra Variations." *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 44, no. 3 (2002): 305–323. https://doi.org/10.1002/1521-4036(200204)44:3<305::AID-BIMJ305>3.0.CO;2-J

Liaw, A., and Wiener, M. "Classification and Regression by randomForest." *R News* 2, no. 3 (2002): 18–22.

Lindgren, Finn, and Havard Rue. "Bayesian Spatial Modelling with R-INLA." *Journal of Statistical Software* 69, no. 19 (2015). https://core.ac.uk/download/pdf/161914450.pdf

Lord, Dominique. "Modeling Motor Vehicle Crashes using Poisson-Gamma Models: Examining the Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter." *Accident Analysis & Prevention* 38, no. 4 (2006): 751–766. https://doi.org/10.1016/j.aap.2006.02.001

Lord, Dominique, and James A. Bonneson. "Development of Accident Modification Factors for Rural Frontage Road Segments in Texas." *Transportation Research Record* 2023, no. 1 (2007): 20–27. https://doi.org/10.3141/2023-03

Lord, Dominique, and Fred Mannering. “The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives.” *Transportation Research Part A: Policy and Practice* 44, no. 5 (2010): 291–305. https://doi.org/10.1016/j.tra.2010.02.001

Lord, Dominique, Simon Washington, and John N. Ivan. “Further Notes on the Application of Zero-Inflated Models in Highway Safety.” *Accident Analysis & Prevention* 39, no. 1 (2007): 53–57. https://doi.org/10.1016/j.aap.2006.06.004

Lord, Dominique, Simon Washington, and John N. Ivan. “Poisson, Poisson-gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory.” *Accident Analysis & Prevention* 37, no. 1 (2005): 35–46. https://doi.org/10.1016/j.aap.2004.02.004

Lunn, David, Christopher Jackson, Nicky Best, Andrew Thomas, and David Spiegelhalter. *The BUGS Book: A Practical Introduction to Bayesian Analysis*. CRC Press, 2012. Available at: https://books.google.com/books?hl=zh-CN&lr=&id=Cthz3XMa_VQC&oi=fnd&pg=PP1&dq=The+BUGS+Book:+A+Practical+Introduction+to+Bayesian+Analysis&ots=3uCE-pU9J7&sig=Hd2LduHbtnVB-dV209YD-VLNoI0#v=onepage&q=The%20BUGS%20Book%3A%20A%20Practical%20Introduction%20to%20Bayesian%20Analysis&f=false

Ma, Jianming, and Kara M. Kockelman. “Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity.” *Transportation Research Record: Journal of the Transportation Research Board* 1950, no. 1 (2006): 24–34. https://doi.org/10.1177/0361198106195000104

Ma, Xiaoxiang, Suren Chen, and Feng Chen. “Multivariate Space-Time Modeling of Crash Frequencies by Injury Severity Levels.” *Analytic Methods in Accident Research* 15 (2017): 29–40. https://doi.org/10.1016/j.amar.2017.06.001

Mader, Emily M., and Cathleen D. Zick. “Active Transportation: Do Current Traffic Safety Policies Protect Non-Motorists?” *Accident Analysis & Prevention* 67 (2014): 7–13. https://doi.org/10.1016/j.aap.2014.01.022

Mannering, Fred, and Chandra R. Bhat. “Analytic Methods in Accident Research: Methodological Frontier and Future Directions.” *Analytic Methods in Accident Research* 1 (2014): 1–22. https://doi.org/10.1016/j.amar.2013.09.001

Mardia, Kanti. “Multi-Dimensional Multivariate Gaussian Markov Random Fields with Application to Image Processing.” *Journal of Multivariate Analysis* 24 (1998): 265–284. https://doi.org/10.1016/0047-259X(88)90040-1

Miaou, Shaw-Pin., Joon J. Song, and Bani K. Mallick. “Roadway Traffic Crash Mapping: A Space-Time Modeling Approach.” *Journal of Transportation and Statistics* 6, no. 1 (2003): 33–57. http://worldcat.org/oclc/37387952

McNeil, Nathan, Christopher M. Monsere, and Jennifer Dill. "Influence of Bike Lane Buffer Types on Perceived Comfort and Safety of Bicyclists and Potential Bicyclists." *Transportation Research Record* 2520, no. 1 (2015): 132–142. https://doi.org/10.3141/2520-15

Miranda-Moreno, Luis F. *Statistical Models and Methods for the Identification of Hazardous Locations for Safety Improvements*. Doctoral dissertation: Ph.D. thesis, Department of Civil Engineering, University of Waterloo (Ontario, Canada), 2006. https:// central.bac-lac.gc.ca/.item?id=NR23678&op=pdf&app=Library

Mothafer, Ghasak I., Toshiyuki Yamamoto, and Venkataraman N. Shankar. "Evaluating Crash Type Covariances and Roadway Geometric Marginal Effects using the Multivariate Poisson Gamma Mixture Model." *Analytic Methods in Accident Research* 9 (2016): 16–26. https://doi.org/10.1016/j.amar.2015.11.001

National Association of City Transportation Officials. Urban Street Design Guide. Washington, D.C.: National Association of City Transportation Officials, 2015. http://nacto.org/publication/ urban-street-design-guide/

Obaidat, M. T., and Ramadan, T. M. "Traffic Accidents at Hazardous Locations of Urban Roads". *Jordan Journal of Civil Engineering* 6, no. 4 (2012): 436–447.

Oh, Jutaek, Simon P. Washington, and D. Nam. "Accident Prediction Model for Railway-Highway Interfaces." *Accident Analysis & Prevention* 38, no. 2 (2006): 346–356. https://doi.org/10.1016/j.aap.2005.10.004

Osama, Ahmed, and Tarek Sayed. "Investigating the Effect of Spatial and Mode Correlations on Active Transportation Safety Modeling." *Analytic Methods in Accident Research* 16 (2017): 60–74. https://doi.org/10.1016/j.amar.2017.08.003

Palmi-Perales, Francisco, Virgilio Gomez-Rubio, and Miguel A. Martinez-Beneito. Bayesian Multivariate Spatial Models for Lattice Data with INLA. arXiv preprint, 2019. arXiv:1909.10804. https://arxiv.org/abs/1909.10804

Park, Eun S., and Dominoque Lord. "Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity." *Transportation Research Record: Journal of the Transportation Research Board* 2019, no. 1 (2007): 1–6. https://doi.org/10.3141/2019-01

Park, Juneyoung, and Mohamed Abdel-Aty. "Evaluation of safety effectiveness of multiple cross sectional features on urban arterials." *Accident Analysis & Prevention* 92 (2016): 245–255.https://doi.org/10.1016/j.aap.2016.04.017

Plummer, Martyn "Penalized Loss Functions for Bayesian Model Comparison." *Biostatistics* 9, no. 3 (2008): 523–539. https://doi.org/10.1093/biostatistics/kxm049

Quddus, Mohammed A. “Modelling Area-Wide Count Outcomes with Spatial Correlation and Heterogeneity: An Analysis of London Crash Data.” *Accident Analysis & Prevention* 40, no. 4 (2008): 1486–1497. https://doi.org/10.1016/j.aap.2008.03.009

Russo, Brendan J., Peter T. Savolainen, William H. Schneider, and Panagiotis C. Anastasopoulos. “Comparison of Factors Affecting Injury Severity in Angle Collisions by Fault Status Using a Random Parameters Bivariate Ordered Probit Model.” *Analytic Methods in Accident Research* 2 (2014): 21–29. https://doi.org/10.1016/j.amar.2014.03.001

Serhiyenko, Volodymyr. “Dynamic Modeling of Multivariate Counts-Fitting, Diagnostics, and Applications.” (2015). https://opencommons.uconn.edu/dissertations/858/

Serhiyenko, Volodymyr, Sha A. Mamun, John N. Ivan, and Nalini Ravishanker. “Fast Bayesian Inference for Modeling Multivariate Crash Counts.” *Analytic Methods in Accident Research* 9 (2016): 44–53. https://doi.org/10.1016/j.amar.2016.02.002

Siddiqui, Chowdhury, Mohamed Abdel-Aty, and Helai Huang. “Aggregate Nonparametric Safety Analysis of Traffic Zones.” *Accident Analysis & Prevention* 45 (2012): 317–325. https://doi.org/10.1016/j.aap.2011.07.019

Siddiqui, Chowdhury, Mohamed Abdel-Aty, and Keechoo Choi. "Macroscopic Spatial Analysis of Pedestrian and Bicycle Crashes." *Accident Analysis & Prevention* 45 (2012): 382–391. https://doi.org/10.1016/j.aap.2011.08.003

Smith, Robin, Sharlene Reed, and Shana Baker. “Street Design: Part 1. Complete Streets.” *Public Roads* 74, no. 1 (2010): 12–17. https://www.safetylit.org/citations/index.php?fuseaction=citations.viewdetails&citationIds[]=citjournalarticle_474698_16

Song, Joon J., Malay Ghosh, Shaw P. Miaou, and Bani K. Mallick. “Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping.” *Journal of Multivariate Analysis* 97, no. 1 (2006): 246–273. https://doi.org/10.1016/j.jmva.2005.03.007

Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn. *WinBUGS User Manual (2003)*. https://www.mrc-bsu.cam.ac.uk/wp-content/uploads/manual14.pdf

Van Der Linde, Angelika. “DIC in Variable Selection.” *Statistica Neerlandica* 59, no. 1 (2005): 45–56. https://doi.org/10.1111/j.1467-9574.2005.00278.x

Wang, Jie, Helai Huang, and Qiang Zeng. "The Effect of Zonal Factors in Estimating Crash Risks by Transportation Modes: Motor Vehicle, Bicycle and Pedestrian." *Accident Analysis & Prevention* 98 (2017): 223–231.https://doi.org/10.1016/j.aap.2016.10.018

Watanabe, Sumio. “Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory.” *Journal of Machine Learning Research* 11 (December 2010): 3571–3594. https://www.jmlr.org/papers/volume11/watanabe10a/watanabe10a.pdf

Zeng, Qiang, and Helai Huang. “Bayesian Spatial Joint Modeling of Traffic Crashes on an Urban Road Network.” *Accident Analysis & Prevention* 67 (2014): 105–112. https://doi.org/10.1016/j.aap.2014.02.018

About the Authors

Yongping Zhang

Dr. Yongping Zhang is an assistant professor of Transportation Engineering at California State Polytechnic University, Pomona. He is also a licensed professional engineer in Civil Engineering. During 2009–2015, Dr. Zhang worked as a Senior Transportation Modeler for the Southern California Association of Governments; before that, he worked as a Transportation Analyst at Wilbur Smith Associates in the Chicago area for two years.

Wen Cheng

Dr. Wen Cheng is a professor of Transportation Engineering at California State Polytechnic University, Pomona. His research focuses on highway safety, statistical modeling, traffic operation, and management. He received the 2015 Outstanding Research Award and the 2018 Outstanding Teaching Award from Cal Poly Pomona College of Engineering. Dr. Cheng has published numerous journal papers in the fields of traffic safety and operation, machine learning, and air quality.

Xudong Jia

Dr. Xudong Jia is a professor of Transportation Engineering and the Chair of the Civil Engineering Department at California State Polytechnic University, Pomona. He has been working in civil engineering and transportation engineering education, applied research, and consulting for more than 25 years. He has received the Outstanding Organization and Individual Achievement Awards from the National Transit Standards Consortium.

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Directors

MINETA TRANSPORTATION INSTITUTE