STATE OF CALIFORNIA • DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV 10/98)

Lock Data on Form

1.REPORT NUMBER	2.GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA18-3139		

4. TITLE AND SUBTITLE	5. REPORT DATE
Investigation of Multimodal Crashes using Full Bayesian Multivariate Spatial-Temporal Models	March 1, 2018
	6.PERFORMING ORGANIZATION CODE
7.AUTHOR	8.PERFORMING ORGANIZATION REPORT NO.
Wen Cheng, Xudong Jia and Gurdiljot Gill	
9.PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NUMBER
Department of Civil Engineering, California State Polytechnic University, Pomona 3801 West Temple Avenue Pomona, California 91768	11. CONTRACT OR GRANT NUMBER 65A0529 TO 052.3
12. SPONSORING AGENCY AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
California Department of Transportation Division of Research and Innovation and System Information, MS-83 1727 30thStreet Sacramento CA 95816	Final Report 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

California's Senate Bill-743 signed by the Governor on September 27, 2013 advocated the promotion of "active" transportation including bicycling, walking, etc. Despite the numerous potential advantages of indulging in multiple modes of transport, such as elevation of health and environment along with mitigation of congestion, the cyclists and pedestrians are a vulnerable segment of traveling public which is exposed to safety risks. To realize the goal of safer traffic environment, safety of commuters for all modes of transport is of paramount importance. The significance of incorporating multiple user modes of road transport in traffic safety is further catalyzed by the ongoing efforts reflected by Caltrans' *Statewide Highway Safety Plan* and *Active Transportation Program.*

However, insufficient protective infrastructures or deficiency of efficient tools to evaluate the safety and economic impacts of transport facilities on pedestrians and bicyclists is the main deterrent to such active transportation. In response to this issue, the main objective of the proposed research is to develop new multivariate crash frequency models for different modes which account for the spatial, temporal correlations, and their interactions.

17. KEY WORDS	17. DISTRIBUTION STATEMENT
Safety, bicycles, collisions, volumes	

19. SECURITY CLASSIFICATION *(of this report)*	20.NUMBER OF PAGES	21.COST OF REPORT CHARGED
Unclassified	102	N/A

DISCLAIMER STATEMENT

Investigation of Multimodal Crashes using Full Bayesian Multivariate Spatial-Temporal Models

FINAL REPORT

Prepared by:

Wen Cheng, Ph.D., P.E., Associate Professor

Xudong Jia, Ph.D., P.E., Professor

Gurdiljot Gill, MSCE, Research Associate

Department of Civil Engineering,

California State Polytechnic University, Pomona

Prepared For:

UCCONNECT and Caltrans

March, 2018

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

California's Senate Bill-743 signed by the Governor on September 27, 2013 advocated the promotion of "active" transportation including bicycling, walking, etc. Despite the numerous potential advantages of indulging in multiple modes of transport, such as elevation of health and environment along with mitigation of congestion, the cyclists and pedestrians are a vulnerable segment of traveling public which is exposed to safety risks. To realize the goal of safer traffic environment, safety of commuters for all modes of transport is of paramount importance. The significance of incorporating multiple user modes of road transport in traffic safety is further catalyzed by the ongoing efforts reflected by Caltrans' *Statewide Highway Safety Plan* and *Active Transportation Program.* However, insufficient protective infrastructures or deficiency of efficient tools to evaluate the safety and economic impacts of transport facilities on pedestrians and bicyclists is the main deterrent to such active transportation. In response to this issue, the main objective of the proposed research is to develop new multivariate crash frequency models for different modes which account for the spatial, temporal correlations, and their interactions.

For the attainment of a safer driving environment, the traffic safety management process initiates with network screening, or, hotspot identification (HSID), followed by problem diagnosis, countermeasure identification, and project prioritization. The foremost step of HSID is particularly crucial to extract the most benefit from the limited financial resources allocated towards crash remediation. Numerous methods have been proposed in the past for the purpose of HSID, ranging from the traditional ones relying on crash count and crash rate, to the more sophisticated Bayesian methods which corrected the regression-to-the-mean (RTM) bias and accommodating correlation structures usually associated with empirical crash data. Reaching towards the goal of implementation of multimodal approach, several studies investigated the factors impacting non-motorist safety on roadways such as roadway design, driver and non-motorist inebriation, low-light conditions, vehicle speed, and so on. However, these studies separately modeled the factors for cyclist, pedestrian, and motor-vehicle injury occurrence and few attempts have been made to combine these into a multimodal approach. To account for the unobserved heterogeneity shared by various transportation modes, it is essential to develop multivariate crash frequency models which can estimate crash risk of multiple modes simultaneously. Similarly, it is also necessary to accommodate the spatial and temporal correlations which are generally observed in crash data. Ignorance of such correlation structures has been illustrated to reduce the efficiency of the model due to lesser precise parameters. Recent studies have attempted to develop crash prediction models by incorporating such correlations. However, there is still no or very little research addressing traffic safety issues by employing the multivariate spatial-temporal modeling. To fill this research gap, the authors proposed different spatial-temporal models to analyze the modal crash data. Moreover, the performance of alternative models was evaluated using different criteria of varying complexity. Overall, the analyses of crash data from different levels of roadway network by employing multiple spatiotemporal models is anticipated to enhance the understanding of safety impacts from the interaction of various modes and promote the idea of active transportation. Also, it will demonstrate the use of sophisticated multivariate spatial-temporal models for research community and Caltrans engineers.

This report is structured in six chapters, where Chapter 1 offers discussion pertaining to the background of crash prediction models, Chapters 2-5 present a detailed account of the objectives, methodology, and results of the four studies conducted by the authors, and Chapter 6 offers their summary and conclusions for easier reader comprehension and comparison. Overall, the primary focus was on development of different crash prediction models which incorporate the spatial and temporal correlations by employing different specifications. Different spatial levels of roadway network were used as the areas of focus to present a relatively robust demonstration for implementation of models. The inferences from all four chapters strongly suggested the implementation of proposed methods to fully realize the power of statistical models for achieving better predictive accuracy and generate more precise estimates for causal factors. The discussion of results offers the theoretical and practical implications of proposed models and offer recommendations from planning and engineering perspectives.

CHAPTER 1: LITERATURE REVIEW

Traffic Safety Models

Roadway crashes have caused an immense burden on society with respect to emotional and financial losses. Researchers are entrusted to develop analytic approaches to gain a better understanding of the causal factors for crash occurrence and develop more accurate crash prediction models to formulate road-safety policies and engineering solutions for mitigation of crashes. However, the accuracy of inferences drawn from the statistical analysis of crashes is highly dependent on the robustness of crash data and an array of potential influential factors such as roadway geometric (number of lanes, lane width, radius of horizontal curve, etc.), traffic flow (vehicle density, volume, real-time speed, speed deviations, etc.), environment (lighting, weather), driver characteristics and mental state (gender, response time, age, etc.), among others. Unfortunately, the crash related data collected by safety agencies may be inadequate or unavailable for detailed investigation (Lord and Mannering, 2010). Hence, the researchers managed to handle this issue to study the significant factors by virtually enhancing the quantity of dataset by disaggregation over some geographical space (micro or macro level) and some specified time period (e.g. division of five year accumulated crash data into five individual subsets). The crash-frequency data are obtained in the form of non-negative integers allowing the application of count-based regression models.

These regression models (or, crash prediction models) have been used in research and practice for determination of influential factors, planning purposes, or site ranking. Models of varying complexity have been employed, ranging from very basic to sophisticated. The traditional approach to analyzing roadway crashes employed generalized linear models (McCullagh and Nelder, 1989; Zeger and Karim, 1991) to establish a linear relationship between explanatory variables and log-transformed outcomes such as crash frequencies of different severities or vehicle modes. This allowed for clear interpretation of inferences drawn from model estimates. To handle over-dispersion commonly associated with crash data, over-dispersed generalized linear models such as Poisson mixtures (e.g., negative binomial or Poisson-gamma, Poisson-lognormal, etc.) were introduced (Persaud, 1994; Hauer, 1997; Milton and Mannering, 1998; Karlaftis and Tarko, 1998). These models may not fully incorporate the unobserved heterogeneity as in count-data models, the overdispersion may be attributed to various factors, such as the grouping of data over space (segments, neighborhood, cities, regions, etc.), unaccounted temporal correlation, and model miss-specification (Gourieroux and Visser, 1997; Poormeta, 1999; Cameron and Trivedi, 1998).

Temporal

The traditional modeling approach often disaggregated the crash counts into small time intervals with different reasons such as enlarging the dataset which can circumvent the issue of small sample size or incorporating the information of time-varying factors. The disaggregation of crash data over specified time periods lead to temporal correlation as those datasets may share unobserved effects which remain constant over time. To remove the potential bias of estimated model parameters, some researchers addressed the serial correlation in crash data by employing different temporal treatments such as linear and/or quadratic trend (Andrey and Yagar, 1993; Hay and Pettitt, 2001), autoregressive correlation structure with a time step of one year (lag-1) (Huang et al., 2009; Wang et al., 2013), fixed-over-time and independent-over-time random effects (Aguero-Valverde, 2013; Jiang et al., 2014), and time-varying model coefficients and intercepts (Cheng et al., 2017b), and so on. The safety literature is replete with studies pertaining to different treatments of the temporal aspect of crash data. Wang et al. (2013) modeled temporal correlation for longitudinal crash data of 208 intersections. The longitudinal data were comprised of rear-end collisions at intersections over a 3-year period. Generalized estimating equations with negative binomial link function were utilized for the development of four models with four different types of temporal treatment: independent, exchangeable, autoregressive (AR), and unstructured. The independent approach assumed non-dependency among crash observations of an intersection over a period; the exchangeable one assumed constant correlation among two observations of an entity; the AR-1 one weighed the dependency among observations separated by a lag of one prior year; and the unstructured treatment disregarded the presence of a specific correlation and assumed different dependency among two random observations at an intersection. The comparison of different correlation structures by the method of cumulative residuals demonstrated the superior fit associated with the autoregressive structure with an estimated correlation of 0.4454 for each successive two years and the highest *p*-value of 0.86. Huang et al. (2009) proposed the Full Bayesian (FB) hierarchical approach for identification of hot spots by employing the same AR-1 model, along with five other models for intersection crash data. Three criteria were employed for assessment of model performance from different approaches: DIC (deviance information criterion); MAD (mean absolute deviance); and MSPE (mean-squared predictive error). The significant serial correlation coefficient (0.775) reflected the existence of autocorrelation for crash rate. The best fit of AR-1 model assessed by the three criteria demonstrated the benefit of inclusion of serial correlation to capture time-dependent safety effect which may have escaped from the model. The segment-level study by Aguero-Valverde (2013) compared the fixed-over-time and independent-over-time random effects based on the precision of crash frequency estimates. The notable advantage at model fit (assessed by DIC and posterior mean deviance (P_d)), more precise estimates, and consistent site ranking performance associated with fixed-over-time indicated that such temporal treatment allowed more flexibility for model parameters to 'pool strength' from the neighboring years for segment data. The subsequent study by El-Basyouny and Kwon (2012) explored the temporal aspect by developing four models which include with and without linear time trend, yearly-varying intercept, and yearly-varying coefficients. The last model demonstrated the best performance with the lowest DIC score.

Spatial

Apart from the aforementioned studies focused on temporal correlation, another dimension of crash prediction models incorporated spatial correlation structures to account for the spatial dependency of crash counts among neighboring sites (Song et al., 2006; Abdel-Aty and Wang, 2006; Guo et al., 2010). Such correlation structures improve crash estimation by allowing the flexibility to "pool strength" from neighboring entities (Aguero-Valverde and Jovanis, 2008) and also act as a surrogate for the unmeasured spatial confounding factors which are not incorporated in the model (Chiou et al., 2014). Motivated by the benefits of incorporation spatial random effects, an array of spatial levels have been explored which may be broadly classified under micro- and macro-level, where micro-level pertains to intersections, road segments, corridors, and macroscopic level comprises of areas such as block group, census tracts, traffic analysis zones (TAZs), or counties. Comparatively speaking, the microscopic analysis is primarily centered on investigating geometric or traffic characteristics which influence the safety on a network. On the other hand, macroscopic safety analysis concentrates on quantifying the impact of socioeconomic and demographic characteristics, transportation demand and network attributes so as to provide countermeasures from a planning perspective such as enactments of traffic rules, police enforcements, safety campaigns, and area-wide engineering treatments. These studies observed the superiority of inclusion of spatial correlation in crash prediction models as demonstrated by significant improvement of model fit, crash prediction, and site ranking performance.

Aguero-Valverde and Jovanis (2008) explored the effect of spatial correlation in models of crash frequency at segment level by using a Full Bayesian (FB) approach with conditional autoregressive (CAR) effects. Three adjacency-based weight matrices were developed for first, second, and third order neighbors, which showed a significantly better fit than the Poisson lognormal model which considered only heterogeneity. Guo et al. (2009) developed models to incorporate the spatial proximity at corridor level between intersections due to similarity in road design and environmental characteristics. The modeling results demonstrated that the Poisson spatial model provided the best fit. Recently, Aguero-Valverde et al. (2016) used a multivariate spatial model to account for spatial correlation among adjacent sites (road segments) to enhance model prediction for different crash types. The multivariate conditional autoregressive (MCAR) model was used with the first order adjacency-based weight matrix which was observed to have best fit due to spatial and multivariate correlation. Among the macro-level studies, Best et al. (2001) investigated the risk of leukemia in children at three different levels of data aggregation: Local Authority Districts, census wards, and 1 km^2 grid squares. They examined adjacency versus distance-based neighborhood spatial weights for each of analysis. Rhee et al. (2016) used GIS-developed spatial variables to prepare a database of traffic crashes at TAZ level to explore the significant variables influencing the crashes. The Rook adjacency-based weight matrix was used for analysis of spatial component of crash heterogeneity. Results showed that the spatial error model was better than the spatial lag model. Aguero-Valverde and Jovanis (2006) applied univariate space-time model to analyze county-level crash counts. The first-order adjacency matrix was utilized for the CAR error term. The results demonstrated the existence of spatial correlation in crash data. Huang et al. (2010) proposed a Bayesian spatial model to account for county-level variations of crash risk in Florida. A CAR prior was specified to accommodate for the spatial autocorrelations of adjacent counties. The results exhibited little difference in safety effects of risk factors on all crashes and severe crashes.

Space and time

Building on the advantages of spatial and temporal correlation structures to address the issue of unobserved heterogeneity, some studies incorporated temporal dimension for spatial models as the crash analysis is not curbed to a single time period. The study by Wang and Abdel-Aty (2006) incorporated independent spatial and temporal correlations for rear-end crashes among intersections for the 3-year longitudinal crash data. The modeling results revealed the presence of high correlations between the longitudinal (0.4454 for each successive two years) and spatially correlated (0.6316 for two nearest intersections) rear-end crashes. The study by Blazquez and Celis (2013) did a spatial and temporal analysis of child pedestrian crashes occurring during a period of nine years. The spatial autocorrelation analysis indicated that the responsibility of pedestrians is the major contributing factor for the generation of child pedestrian crashes with a tendency to cluster in space and time. Also, spatial clustering distribution of crashes in terms of time of the day was also observed. A recent study by Cheng et al. (2017) developed two spatiotemporal models, one with linear time trend and another with time-varying coefficients, for hot spot identification of intersections based on six different crash types over a ten-year period. Statistically significant spatial correlation was observed among crash types of intersections. With respect to the temporal treatment, the time-varying coefficient model performed better at model fit (DIC) but the relatively less complex linear trend model was superior at the crash prediction as assessed by the evaluation criteria of RSS (residual sum of squares), Cohen's Kappa, MCT (method consistency test), and TRD (total rank difference).

While these studies treated spatial and temporal correlations independently, some studies noted that vehicle crashes tend to cluster both spatially and temporally, hence space-time interaction specification was employed at different spatial scales. Based on the formulation proposed by Bernardinelli et al. (1995), Aguero-Valverde and Jovanis (2006) applied a spatiotemporal model to county-level fatal and injury crashes with the interaction of linear time trend and spatial correlation. This model allowed the flexibility of independent temporal trends for different entities. The random effect terms for spatial correlation, time trend, and space-time interactions were observed to be statistically significant for the injury crashes. Another spatiotemporal study by Dong et al. (2014) for hot spot identification at TAZ level analyzed the area-specific crash trends. The space-time interaction model (linear trend with spatial) significantly outperformed the independent space-time model (AR-1 with spatial) due to its capability to account for the space-time variations which deviate from expected stable patterns. A more recent study by Ma et al. (2017) explored a finer temporal scale of daily crash data for estimation of injury severity levels of segment crashes. The proposed spatiotemporal models were observed to be superior to other alternate models which accounted only for heterogeneity among crash types and spatial correlations.

Multimodal

Urban mobility and safety for all modes of transportation are key elements in the development of safer traffic environment. This goal may be realized with the implementation of multimodal approaches which allows the flexibility to simultaneously determine the injury risk of different travel modes: motorists and non-motorists. Non-motorists are defined as road users not in or upon a motor vehicle and consist of walking pedestrians, bicyclists, individuals in wheelchairs or motorized personal conveyances, skateboarders and others (NHTSA, 2012). They are a vulnerable segment of the traveling public due to the lack of a protective structure and difference in body mass between them and motor vehicles, which renders them prone to heightened injury susceptibility in case of a collision (Williams, 2013). On the other hand, active transportation provides enormous benefits for addressing the issues of congestion, health, and environment (Berrigan et al., 2006; Frank et al., 2010; Furie and Desai, 2012; Giles et al., 2010; Insall, 2013; Wanner et al., 2012). Therefore, encouraging individuals to indulge in active transportation, involving walking and bicycling, brings with it a societal obligation to protect them. In response, the traffic safety research has addressed the concerns pertaining to multimodal transportation by investigating the non-motorized crash modes (Lee and Abdel-Aty, 2005; Moudon et al., 2011; Beck et al., 2007; Wardlaw, 2002). Some studies explored the inter-relationship of pairs of motorized and non-motorized crash modes for the better understanding of influential factors and eventually design better safety improvement program, such as vehicle and pedestrian crashes (Shankar et al., 2003; Lee and Abdel-Aty, 2005; Pulugurtha and Sambhara, 2011), bicycles and vehicles (Wang and Nihan, 2004; Schepers et al., 2011; Strauss et al., 2013) and multiple vehicle crashes (Abdel-Aty and Radwan, 2000; Wang and Huang, 2016). However, the realization of safer traffic environment from the multimodal perspective demands joint estimation of multiple modes of crashes as an effort to account for the inter-relationship among crash modes. This unobserved heterogeneity may be addressed by the development of multivariate crash prediction models which reduce the bias associated with the ignorance of such correlation structures (Huang et al., 2017). The multivariate models have been extensively employed for the joint estimation of vehicle crashes on crash types (Ye et al., 2009; El-Basyouny et al., 2014; Aguero-Valverde et al., 2016; Cheng et al., 2017a) or severity levels (Park and Lord, 2007; El-Basyouny and Sayed, 2009; Aguero-Valverde and Jovanis, 2009; Gill et al., 2017a). However, the use of these models has been considerably scarce in the multimodal literature.

The study by Convay et al. (2013) employed a bivariate model to investigate the locations of conflict occurrence between bicycles and other mode users such as pedestrians, freight, passenger cars, and cabs. The study area was an urban setting as the interactions among such multiple modes are relatively more common. The characteristics which influenced the conflicts between these modes were also explored. This study recommended the development of a multivariate regression model for prediction of multimodal conflicts provided the availability of robust crash data and explanatory variables. Recently, the study by Huang et al. (2017) employed a multivariate Poisson lognormal model to jointly analyze the occurrence of motor vehicle, bicycle, and pedestrian crashes at urban intersections. This model specification allowed the flexibility to account for the unobserved heterogeneity due to the correlation among different modes involved in crashes at individual intersections. The results confirmed the presence of significant correlations among heterogeneous residuals among the crash risk of three modes considered in the study.

Similar to the multivariate specification which allows correlations among crash modes (also crash types and severity levels), spatial and serial correlations have also been explored in crash data. Accommodation of spatially unstructured (serial) correlations has been found to enhance the model fitness and precision by numerous research studies focused on vehicle crashes (Andrey & Yagar, 1993; Hay & Pettitt, 2001; Cheng et al., 2017b). Likewise, the significance of incorporating spatial correlations was also highlighted by many studies (Guo et al., 2010; Abdel-Aty & Wang, 2006) which noticed consistently superior performance of the spatial models over those accounting for heterogeneity random effect only. Some multimodal studies incorporated the spatial correlations among entities for improved estimation of crash risk as such correlation structures render the capability to "pool strength" from neighboring entities (Aguero-Valverde and Jovanis, 2008) and also act as surrogate for the unmeasured spatial confounding factors related to sites of interest which are not incorporated in the model (Chiou et al., 2014). Cai et al. (2016) incorporated the spatial spillover effects to develop dual-state models for analysis of pedestrian and bicycle crashes at the macro-level of Traffic Analysis Zones (TAZs). It was observed that consideration of spatial effects improved the performance of model while highlighting the impact of certain traffic, roadway and sociodemographic characteristics on bicycle and pedestrian crashes at the planning level. The study by Tasic and Porter (2016) incorporated spatial random effects on the level of census tracts within the negative binomial model to evaluate the relationship multimodal infrastructure and traffic safety outcomes. A significant association was observed between motorized and non-motorized crashes, and variables pertaining to socio-economic, land use, road network, travel demand. Amoh-Gyimah et al. (2016) employed the CAR (conditional autocorrelation) specification into the negative binomial model for estimation of pedestrian and bicycle crashes at the macro-level of statistical area. The study by Wang et al. (2017) focused on the motor vehicle, bicycle, and pedestrian crashes at the intersection level but incorporated the zonal factors (TAZ level) as a measure of spatial dependence. It was observed that the inclusion of zonal factors elevated the performance of the model in case of non-motorized crashes while their omission resulted in biased parameters, which indicates the role of macro-level factors for estimation of crash risk for non-motorized crash modes.

It should be noted that aforementioned multimodal spatial studies mostly focused on the macro-level as such higher levels of spatial entities better fits to account for the area-wide explanatory factors for the safety of multimodal environment (Tasic and Porter, 2016). However, the above multimodal studies did not employ a previously discussed multivariate model for simultaneous estimation of multiple crash modes although such multivariate spatial models have been employed to estimate various crash severities or outcomes (Aguero-Valverde, 2013; Cheng et al., 2017a). The recent paper by Huang et al. (2017) proposed a spatial multivariate Poisson lognormal model for joint modeling of three transportation modes at intersection level, namely: motor vehicle, bicycle, and pedestrian crashes. It was observed that the proposed model outperformed the univariate spatial and the multivariate models which accounted only for spatial correlation among sites and correlation among modes, respectively. The variance estimates for spatial correlation of the three modes were noted to be statistically significant.

Similar to the incorporation of spatial and temporal correlation structures, some studies in traffic safety observed the superiority of nonparametric and/or semiparametric models to address the unobserved heterogeneity (Heydari et al., 2016; Shirazi et al., 2016). Regarding research dedicated to active transportation, the recent study by Heydari et al. (2017) proposed the Dirichlet process mixture (Ohlssen et al., 2007) to develop a flexible latent class model for joint analysis of pedestrian and cyclist injuries at the micro-level of intersections. The authors observed that the flexible approach was advantageous as it demonstrated superior predictive performance and better capability to capture the correlated crash data which eventually provided more accurate interpretation of influential factors for improvement of safety environment.

CHAPTER 2: MULTIMODAL MULTIVARIATE SPACE-TIME MODELS WITH ALTERNATE SPATIOTEMPORAL INTERACTIONS

INTRODUCTION

Enhancement of safety for all transportation mode users plays an essential role in the implementation of multimodal transportation systems. As discussed in the literature review, the sophisticated space-time models, which combine the immense benefits associated with spatial and temporal correlations, are scarce in studies focused on vehicle crashes and virtually non-existent in multimodal literature. To fill this research gap, the authors aim to develop three multivariate spatial-temporal models to analyze the modal crash data at the macro-level of counties. The three alternate models share the common aspects of the multivariate spatial specifications but differ on the assignment of temporal trends and spatiotemporal interaction. The model formulations are presented in order of complexity: (1) the linear time trend with fixed spatial; (2) the quadratic time trend with fixed spatial; and the more sophisticated (3) the time-varying spatial model. Given that the external influential factors may not have an equal impact on different modes, the mode-varying intercept and model parameters, rather than fixed ones, were chosen which allow the flexibility to estimate different coefficients for each of the four modes (motor-vehicle only, pedestrian-involved, bicyclist-involved and motorcyclist-related). This study serves two broad objectives: (a) to examine the benefits of alternative models associated with model fit and goodness-of-fit, which is assessed by employing DIC (deviance information criterion), and LPML (log pseudo marginal likelihoods), and (b) to quantify the transferability of better model fitness and crash estimation to site ranking, which is evaluated by employing SCT (site consistency test) and MCT (method consistency test) at different threshold levels.

METHODOLOGY

This chapter analyzed four different transportation mode users-involved crashes aggregated at the 58 counties of California over a period of eight years. The models were developed assuming the Poisson-lognormal distribution, unlike the alternate Negative Binomial, as the heavier tails associated with the lognormal distribution renders the capability of better handling the small sample size (Lord and Miranda-Moreno, 2008) and overdispersion (Lord and Mannering, 2010) in crash data. To reduce the bias associated with the ignorance of dependency within crash modes, the multivariate error term was incorporated in the models for simultaneous estimation of crash rates for all four modes (Park and Lord, 2007). Finally, the random effects were incorporated to account for the spatial correlations and spatiotemporal interactions from different perspectives. To account for the unobserved heterogeneity from different perspectives, the Full Bayesian (FB) framework was employed due to its flexibility and effectiveness to incorporate complex correlations with a hierarchical structure of data (Pawlovich et al., 2006; Miranda-Moreno, 2006). The traditional approach of maximum likelihood estimation relies on the point estimates for sampling while the FB approach provides a posterior distribution of parameters from Markov-chain Monte Carlo (MCMC) simulation which samples the variables as random. This approach has been widely used for crash prediction models due to the multilevel and correlated nature of data. The three alternate models are similar based on the multivariate spatial specifications but differ on the assignment of temporal trends and spatiotemporal interaction. The model formulations are presented in order of complexity: (1) the linear time trend with fixed spatial; (2) the quadratic time trend with fixed spatial; and the more sophisticated (3) the time-varying spatial model. The following sections first explain the basic model structure and then introduce the specific formulations for the three models developed over the basic structure. (Note: the equation numbers hereafter are for within-chapter reference and they restart each chapter. There is no referencing across chapters for the equations in this report. To eliminate the cumbersome reference across chapters, each chapter's methodology is stand-alone)

Model Specification

At the first step, the FB Poisson lognormal model assumes that crash counts (y_{ijt}) at a given county i for a particular travel mode j of year t obey the Poisson distribution, while the corresponding observation specific error term ε_{ij} follows a multivariate normal distribution at the second level:

$$y_{ijt} \sim Poisson\ (\theta_{ijt} e_{it}) \tag{2.1}$$

Where θ_{ijt} is the Bayesian estimated Poisson crash rate for a mode j of year t at location i obtained by using offset of traffic exposure (e_{it}) at county i, year t. In this case, the Daily Vehicle Miles Traveled (DVMT) was utilized as the exposure term for the calculation of crash rate, which can be expressed as follows:

$$\ln(\theta_{ijt}) = \beta_{0j} + X'_{ij}\beta_j + \varepsilon_{ij} \tag{2.2}$$

Where β_{0j} is the mode-varying intercept, X' is the matrix of influential factors, β is the mode-varying vector of model parameters, and ε_{ij} is the independent random error which captures the extra-Poisson heterogeneity among locations. Given that the external influential factors may not have an equal impact on different modes, the mode varying intercept and model parameters, rather than fixed ones, were chosen which allow the flexibility to estimate different coefficients for each of the four modes. This variable coefficient approach may be regarded as random parameters in terms of modes and is expected to generate relatively precise estimates which would result in more informed inferences. The random error term (ε_{ij}) also renders flexibility to the model by accounting for the interdependency among the four crash modes and follows a multivariate normal distribution:

$$\varepsilon_{ij} \sim Normal\ (0,\ \Sigma) \tag{2.3}$$

Where $$\varepsilon_{ij} = \begin{pmatrix} \varepsilon_i^1 \\ \varepsilon_i^2 \\ \varepsilon_i^3 \\ \varepsilon_i^4 \end{pmatrix} \quad , \quad \Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{14} \\ \vdots & \ddots & \vdots \\ \sigma_{41} & \cdots & \sigma_{44} \end{pmatrix} \tag{2.4}$$

Where Σ is called the covariance matrix. The diagonal element $\boldsymbol{\sigma_{jj}}$ in the matrix represents the variance of $\boldsymbol{\varepsilon_{ij}}$, where the off-diagonal elements represent the covariance of crash counts of four crash modes considered in this study. As a normal third level of multivariate Bayesian analysis, the precision matrix may be obtained from the inverse of the covariance matrix and has the following distribution:

$$\Sigma^{-1} \sim Wishart(I, J) \tag{2.5}$$

Where Σ^{-1} is the symmetric positive definite matrix representing the precision, I is the scaler matrix with J x J dimensions (Congdon, 2006), and J is the degree of freedom. In this case, the degree of freedom was assumed to be equal to four (J=4), representing 4 crash outcomes corresponding to four different modes. The adopted values of the identity matrix were 0.1 and 0.005 for diagonal and off-diagonal elements, respectively, as recommended by Carlin and Louis (1996) and Gelman et al. (2003).

To incorporate the spatial correlation among the counties, a multivariate spatial random effect term was introduced over the model represented in Equation 2. The resultant model was obtained with the following formulation:

$$\ln(\theta_{ijt}) = \beta_{0j} + X'_{ij}\beta_j + \varepsilon_{ij} + u_{ij} \tag{2.6}$$

Where u_{ij} is the spatially structured random effect term fit by a zero-centered multivariate conditional autoregressive (MCAR) formulation (Mardia, 1988) which has a conditional normal density as follows:

$$u_i | u_k,\ \Sigma i \sim N_j(\textstyle\sum_{k\sim i} C_{ik}\, , u_k, \Sigma i\ ) \tag{2.7}$$

Subscripts i and k refer to a county and its neighbor, respectively, and $k \varepsilon N_i$ where N_i represents the set of neighbors of county i. The MCAR formulation allows the identification and number of neighbors for a particular county and incorporates the weights which may be assigned based on different approaches. This approach has been popularly employed in safety research which accommodates spatial correlation structures (Amoh-Gyimah et al., 2016; Aguero-Valverde, 2013; Cheng et al., 2017a; Huang et al., 2017). As evident from Equation 7, estimation of the risk in any site is conditional on risks in neighboring locations and the impact may be assessed by calculation of weights. In the past studies (Aguero-Valverde and Jovanis, 2010; Xu and Huang, 2015), weight structures including various adjacency-based, distance-based models, and semi-parametric geographically weighted, and so on, have been explored at different types of roadway entities. The current study employs the distance-based structure to explore the spatial correlation, which represents the pure-distance weight structure among the variety of distance-based structures (Gill et al., 2017b). The following formulation was adopted for the weight structure:

$$w_{ij} = \frac{1}{d_{ij}} \tag{2.8}$$

Where w_{ij} is the weight between counties i and j, and d_{ij} is the distance between counties i and j. With this weight structure, it is known that more weightage was assigned to the counties which are relatively close. It is noteworthy that distance-based weight matrix gives a continuous output for weight based on the mutual distance among the neighboring counties. This structure contrasts with the adjacency-based ones where the output is a binary weight governed by the proximity among sites of interest. The continuous nature of output from the distance-based approach is expected to render more flexibility for weight assignment. Equation 6 serves as the base for the three proposed models, which share the same error term and spatially structured heterogeneity and differ only in the temporal treatment and spatiotemporal interactions. The following subsections present the details of each model in order.

Model 1: Linear time trend with linear spatiotemporal interaction

In this model, the temporal aspect is accommodated by introducing a linear trend where time is regarded as a potential influential factor. Also, the interaction term was incorporated to account for the spatial-temporal interactions. Equation 6 assumes the following form:

$$\ln\left(\theta_{ijt}\right) = \beta_{0j} + X_{ij}'\beta_j + \varepsilon_{ij} + u_{ij} + \left(\alpha_j + \delta_{ij}\right) * T \tag{2.9}$$

Where α_j is the mode-varying scalar parameter for linear yearly trend T (T=1 to 8) and the product of δ_{ij} and T represents the spatiotemporal interaction as δ_{ij} is the spatial term which follows the MCAR specification given in Equation 7.

Model 2: Quadratic time trend with linear spatiotemporal interaction

In this model, a non-linear impact of time is considered. Equation 9 takes the following form:

$$\ln\left(\theta_{ijt}\right) = \beta_{0j} + X_{ij}'\beta_j + \varepsilon_{ij} + u_{ij} + \left(\alpha_j + \delta_{ij}\right) * T + \gamma_j * T^2 \tag{2.10}$$

Where γ_j is the coefficient for the quadratic trend. It is expected that this model will allow more subtle approach compared to the linear trend as it allows flexibility to fit the crash data by virtue of its additional quadratic term.

The aforementioned linear and quadratic time trend models represent the temporal treatments from an array of approaches adopted in safety literature (Cheng et al., 2017). The linear trend has been the popular choice for studies pertaining to space-time interaction while the quadratic trend has been observed to be superior to fit the crash data due to the more flexible approach. at model fit with a linear trend.

Model 3: Time-varying spatial

Model 3 combines above-mentioned linear time trend and spatiotemporal interaction into time-varying spatial random effects. In comparison with Model 1 and 2, this model allows more flexibility to the spatially structured random effect term as the spatial correlation within the segments is allowed to vary with time. Equation 6 is then modified to the following form:

$$\ln(\theta_{ijt}) = \beta_{0j} + X'_{ij}\beta_j + \varepsilon_{ij} + \delta_{ij}^t \tag{2.11}$$

Where δ_{ij}^t is the spatial correlation for each of the eight years of crash dataset considered in this study. The addition of temporal dimension to the identification of spatial patterns addresses the possibility of having different spatial trends in crash risk over different time periods.

It is worth mentioning that the current study incorporates random effects to account for the unobserved heterogeneity. This approach is based on the assumption that unobserved heterogeneity for individual roadway entities is completely unrelated to vector of covariates (Mannering et al., 2016). This approach has been extensively utilized in the safety literature dealing with panel data (Mitra and Washington, 2012; Yu et al., 2013; Deublein et al., 2013; Yu and Abdel-Aty, 2013; Aguero-Valverde, 2013; Mohammadi et al., 2014; Xie et al., 2014; Liu and Sharma, 2017; Huang et al., 2017). However, the random effects approach may be considered as a special restrictive case of a more flexible approach of full random parameters, which allows the flexibility to accommodate the site-specific unobserved heterogeneity where each entity may be assigned its own coefficient for explanatory variables (refer Mannering et al. (2016) for review of studies employing random parameters).

Evaluation Criteria for Predictive Accuracy and Goodness-of-Fit of the Models

The following sections illustrate the multiple criteria for assessment of predictive accuracy and goodness-of-fit of the three alternative models.

The Deviance Information Criterion (DIC) developed by Spiegelhalter et al. (2003) was employed to assess the complexity and fit of the models. The DIC is computed as the sum of the posterior mean deviance and estimated effective number of parameters:

$$DIC = \overline{D} + P_D \tag{2.12}$$

Where $\overline{D}$ is the sum of the posterior mean deviance which measures how well the model fits the data; the smaller the $\overline{D}$, the better the fit. P_D represents the effective number of parameters. In general, $\overline{D}$ will decrease as the number of parameters in a model increase. Therefore, the P_D term is mainly used to compensate for this effect by favoring models with a smaller number of parameters. Based on the model-selection decision criteria suggested by Lunn et al. (2012): the models with DIC value less than 5 of the 'best' model are also strongly supported (provided they do not make very different inferences), values within 5 and 10, weakly supported, and models with a DIC greater than 10 are substantially inferior. $\overline{D}$ may be regarded as the measure of training errors while DIC is the measure of indirect assessment of the test errors as it accounts for the bias due to overfitting usually resulting from more parameters (James et al., 2013).

However, due to some limitations of DIC suggested by some studies (Carlin and Louis, 2008), it may not be a true indicator of model performance due to its sensitivity towards different parameterizations (Geedipally et al., 2014). Hence, this study employed a relatively more robust conditional predictive ordinate (CPO) (Gelfand, 1996) for cross-validation based on predictive densities. Contrary to the traditional approaches of cross-validation which are prone to selection bias associated with division of data into subsets (one subset for making posterior inference and another for validation of previous estimates), the current study calculated the CPO by implementing a CV-1 (leave-one-out) technique which circumvents the issue of selection bias by employing a continuous approach of selecting all data points, except one, for model development and the left out data point to verify the prediction accuracy of the calibrated model. Under the MCMC framework, the estimate of CPO for each observation *i* can be calculated as:

$$CPO = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{f(Y_i|\beta^{(t)})}\right)^{-1} \quad (2.13)$$

Where Yi is the ith observation (i = 1, 2, 3, . . ., n) for all counties and β is the vector of estimated model parameters. This harmonic mean of density (CPO) may be extended to calculate the goodness-of-fit of models by computing the product of CPOs over all observations, which is known as the pseudo marginal likelihood. For computational convenience, the log pseudo marginal likelihoods (LPML) is calculated:

$$LPML = log\prod_{i=1}^{n} CPO_i \quad (2.14)$$

LPML may be regarded as the measure for direct assessment of test errors.

Evaluation of Site Ranking Performance

To quantify the transferability of better model fitness and crash estimation to site ranking, the models were evaluated for relative site ranking performance. As demonstrated by previous studies (Cheng and Washington, 2005; Huang et al., 2009; Jiang et al., 2014; Dong et al., 2016**)**, a set of tools have been proposed to assess the detection performance of various methods. The present study employed two of the most popular criteria which are based on the site consistency and method consistency, respectively. The details of each criterion are presented in the following subsections.

Site Consistency

Cheng and Washington (2008) first proposed the Site Consistency Test (SCT) which is used to measure the ability of a hot spot identification (HSID) method to consistently identify a site as high risk over subsequent observation periods. The underlying assumption is that a site identified as high risk during previous time period should also reveal inferior safety performance in a subsequent time period, given that no significant countermeasure treatments have been implemented at the site and given that the site is indeed high risk. Hence, the sites detected by the superior method in the before period tend to demonstrate higher crash counts or rates compared with those screened out by other methods. In equation form, SCT can be expressed as follows:

$$SCT_{i,j} = \sum_{k=n-n\alpha}^{n} CR_{k,Method=j,i+1} \tag{2.15}$$

where, n is the total number of sites being compared, CR is the crash rate for site ranked site k, α is the threshold of identified high risk sites (e.g. $\alpha = 0.05$ corresponds with top 5% of n sites identified as high-risk, and $n\alpha$ is the number of identified high risk sites), j = HSID method being compared (e.g. $j = 1$ could be Model 1, $j = 2$ Model 2, etc.), and i is observation period (the competing models are applied to before period, i to identify hazardous sites, and the crash rate of the same sites in period i +1 are summed and compared). In this test, the method j that identifies sites in a future period with the highest crash rate is the most consistent method for identifying underlying safety problems.

Method Consistency

Aside from the site consistency, method consistency is another HSID evaluation method which also assumes that sites are in the similar underlying operational and their expected safety performance remains virtually unchanged over the adjacent time periods. Under this homogeneity assumption, a superior HSID method will identify the same set of hot spots across two different periods. The more the common hot spots that are identified in both periods the more reliable and consistent is the performance of HSID method. Specifically, the criterion for this test, MCT, can be calculated using the following equation:

$$MCT_{ij} = \{k_{n-n\alpha}, k_{n-n\alpha+1}, \dots, k_n\}_{i,j} \cap \{k_{n-n\alpha}, k_{n-n\alpha+1}, \dots, k_n\}_{i+1,j} \tag{2.16}$$

Where, terms are as defined previously. Note that only sites identified in the top threshold α are compared.

Basically, MCT represents the intersection of ranked sites k identified in subsequent time periods i and i + 1 that are high risk. Even though such test is easy to implement, it has one limitation of focusing merely on the high-risk sites, while the sites that are consistent safe are ignored. To address this issue, the present paper borrows the concepts of specificity and sensitivity (Elvik, 2007) which can measure the consistency of method in identifying both safe and unsafe sites.

Under these concepts, if we assume the high-risk sites identified by Method j based on data of period i are considered as the "truly" unsafe sites, the number of common unsafe sites identified by the same method in the subsequent time period (i+1) can be considered as correct positives (CP). Likewise, the number of common safe sites flagged out in Period i+1 is considered as correct negatives. With such information given, sensitivity and specificity can be easily computed with the following form:

$$Sensitivity = \frac{CP\ (correct\ positives)}{TP\ (true\ positives)} \tag{2.17}$$

$$Specificity = \frac{CN\ (correct\ negatives)}{TN\ (true\ negatives)} \tag{2.18}$$

Where, *CP*: number of truly hazardous sites correctly identified in the subsequent period as hazardous (which is equal to MCT in Equation 16); *TP*: total number of truly hazardous sites based on before time period; *CN*: number of truly safe sites identified in the subsequent period as safe; and *TN*: total number of truly safe sites based on before time period.

DATA PREPARATION

This macro-level study focused on the crashes occurring in the 58 counties of California over an eight-year period (2006-2013). The segregated collision counts over a relatively long period were considered to closely assess the impact of different temporal treatments for crash prediction models. The multivariate specification of FB models allowed the incorporation of collisions for four modes of roadway transport: motor vehicle, pedestrian, bicycle, and motorcycle. These multimodal collision counts served as the dependent variables and were collected from Statewide Integrated Traffic Records System (SWITRS). The models were developed for estimation of crash rate, where the crash counts were offset by Daily Vehicle Miles Travel (DVMT), which is a main exposure-related factor at the macro-level (Miaou et al., 2003) and was collected from Highway Performance Monitoring System (HPMS) for the corresponding eight years. HPMS also provided the data for independent variables linked with roadways and traffic conditions such as maintain miles and travel time for work trips, respectively. The other independent variables comprised of various demographic, socioeconomic, and land use data which were expected to impact the multimodal activity in the counties and influence the collisions. The main demographic factor, population, along with other factors depicting the socioeconomic activity such as retail sales, household income, per capita income, and percent of people in poverty, employment, and land area were obtained from the California Department of Finance and the US Census Bureau, respectively. In addition, the data for the geometric centroid distance among the counties were provided by Southern California Association of Governments (SCAG), which was utilized for calculation of distance-based weights for accommodating the spatial aspect of models.

As evident from the nature of explanatory variables, DVMT is the only predictor which represents the exposure for multiple modes. The authors acknowledge that DVMT directly relates to the vehicle activity and may not be perceived as the appropriate exposure measure for other modes. It is worth mentioning that the DVMT has been utilized due to the unavailability of multimodal exposure data. Since motorized vehicles are generally the major contributors for crashes pertaining to active modes, it is hoped that the DVMT would capture the multimodal activity to some degree. Secondly, another indicator of multimodal activity, i.e. population, has been incorporated as an explanatory variable to account for other modes. Finally, it is expected that the random effect terms would reduce the errors caused due to unavailability of multimodal exposure data. Many studies focused on the analysis of multimodal crashes have adopted such practices to circumvent the issue of deficient exposure data (Haque et al., 2010; Flask & Schneider, 2013; Flask et al., 2014; Gabauer and Li, 2015; Lee et al., 2015; Zhang et al., 2015; Amoh-Gyimah et al., 2016).

Table 2.1 illustrates the summary information for all dependent and independent variables considered for model development. It should be noted that this study incorporated a mix of time-varying (yearly) and constant variables which account for the temporal trends and spatial-only covariates, respectively. This dataset replicates the real-world scenario where the possibility for the collection of a continuous set of some variables is not feasible at the macro-level. The continuous data for the given period were available for multimodal crashes, DVMT, population, and roadway miles, while rest of the variables were mostly obtained from average of data over some period. In consideration of non-varying influential factors, the authors acknowledge that the study area must have witnessed some developments during the eight-year period which may not be reflected by such factors. But due to the limited accessibility of such data, a strong assumption was made that no significant change happened during this period. Similar assumption has been adopted in literature focused on crashes as it is difficult to procure such information at the macro-level and incorporate the changes within the models (Li et al., 2007; Matkan et al., 2013). The studies which focus on crash counts on yearly-basis may be prone to erroneous inferences due to the bias induced in the model estimates by the excessive amount of zero crash counts present in the data. This issue may pose significant drawbacks for crash prediction models developed for micro-level roadway entities (intersections or segments) where crash occurrence for multiple modes in a single year is relatively less common. The current study circumvents this major issue by aggregating the multimodal collisions at a macro-level which allows significant minimization of the presence of zero crash counts for any of the four modes. As shown in Table 2.1, only the pedestrian and bicycle collisions were noted to have zero counts in a given year, with an average of two or three counties with zero collisions across the eight-year period. The significant deviation in the statistics for most of the variables represents the diversity of California counties mostly based on size and population.

As evident from the nature of independent variables, some variables were observed to be correlated and filtered out using two techniques before incorporating during the model development for crash estimation. First, the correlation tests were conducted using the Harrell Miscellaneous package in R software which allowed the calculation of Pearson correlation coefficient. The variables observed to be correlated with a significance level of 0.05 were eliminated in multiple steps using engineering judgment to prevent exclusion of any potential influential variables which would result in loss of precision of estimated parameters. In other words, the selection procedure strived to maintain a balance of omitted variable bias and multicollinearity. Second, the simulated values of all parameters were monitored using the correlation tool under the inference menu of WinBUGS which provided intuitive scatter plots and produced a matrix of cross-correlations. The second step was implemented during the model development stages where the freeware statistical software, WinBUGS (Lunn et al., 2000), was employed to generate MCMC samples for Bayesian posterior inferences.

TABLE 2.1. Descriptive Statistics of Collected Data of Various Counties

Variables	Description	Year	Minimum	Maximum	Median	Mean	S.D.
Collision	Motor Vehicle	2006	23	48,107	757	2,791	6,767
		2007	10	46,558	698	2,671	6,515
		2008	15	41,794	631	2,389	5,841
		2009	20	40,197	611	2,289	5,612
		2010	18	39,560	537	2,249	5,531
		2011	14	38,933	576	2,184	5,430
		2012	16	38,477	560	2,171	5,388
		2013	21	38,855	544	2,140	5,436
Collision	Pedestrian	2006	0	5,118	43	231	690
		2007	0	5,305	39	234	716
		2008	0	5,199	41	231	702
		2009	0	5,097	41	224	687
		2010	0	4,730	36	218	641
		2011	0	4,748	37	218	644
		2012	0	5,024	35	228	684
		2013	0	4,932	38	213	667
Collision	Bicycle	2006	0	2,935	52	179	422
		2007	0	2,929	54	183	422
		2008	0	3,348	46	203	481
		2009	0	3,747	48	208	531
		2010	1	4,226	49	219	587
		2011	0	4,788	51	236	662
		2012	0	4,955	44	241	685
		2013	0	4,682	51	230	647

Collision	Motorcycle	2006	6	2,246	54	175	341
		2007	5	2,877	54	194	414
		2008	7	3,048	66	205	436
		2009	3	2,802	60	181	399
		2010	2	2,711	60	171	388
		2011	5	3,112	57	189	443
		2012	4	3,349	54	200	483
		2013	3	3,614	65	208	516
DVMT	Daily Vehicle Miles Traveled (miles)	2006	180	217,443	5,092	15,577	32,178
		2007	162	218027	5146	15607	32,261
		2008	168	214,971	5,005	15,387	31,617
		2009	170	214,236	4,836	15,317	31,469
		2010	169	211,876	5,448	15,482	31,148
		2011	164	214,458	4,761	15,353	31,594
		2012	166	214,482	4,551	14,768	31,478
		2013	165	215,817	4,462	14,924	31,747
Pop	Population	2006	1,247	10,205,955	169,480	619,777	1,456,447
		2007	1,247	10,205,955	169,480	619,777	1,456,447
		2008	1,214	10,347,422	180,923	656,696	1,469,310
		2009	1,194	10,398,067	182,519	662,962	1,478,749
		2010	1,177	9,840,555	179,588	644,265	1,408,182
		2011	1,113	9,866,172	179,134	647,470	1,413,526
		2012	1,088	9,923,806	180,800	652,028	1,422,391
		2013	1,078	10,002,804	181,150	657,967	1,434,566
MM	Maintained Miles	2006	287	21,247	1,994	2,935	3,262
		2007	287	21,362	2,009	2,950	3,282
		2008	287	21,686	2,009	2,974	3,329
		2009	266	21,678	2,012	2,963	3,333
		2010	266	21,746	2,012	2,967	3,341
		2011	266	360,857	2,008	9,128	47,113
		2012	270	21,694	2,021	3,026	3,432
		2013	265	21,858	1,921	3,017	3,428
Retail	Total Retail Sales ($1,000)	2012	576	121,389,378	1,859,337	8,306,904	18,251,399
Travel Time	Mean Travel Time to Work (minutes)	2014	13	34	25	24	4

HH Income	Median Household Income (dollars)	2014	35,997	35,997	35,997	35,997	35,997
Income	Per Capita Income for past year (dollars)	2014	16,409	58,004	26,190	27,604	8,198
Poverty	Persons in Poverty (percentage)	2014	7	28	16	16	5
Employed	Total Employment	2014	211	3,932,904	44,911	232,458	573,978
Firms	All Firms	2012	125	1,146,701	13,613	61,845	160,522
Area	Land Area (Square miles)	2010	46	20,056	1,535	2,685	3,102
Distance	Distance among centroids of counties (miles)	N/A	25	962	227	273	176

Note: S.D. represents standard deviation; N/A means Not Applicable

RESULTS

Variable Estimates

The three alternate models were run using a statistical freeware WinBUGS package (Spiegelhalter et al., 2003) using an MCMC algorithm. For calibration of the model, first 30,000 iterations were discarded as burn-in and further 110,000 iterations were regarded for parameter estimation. To ensure convergence of models, the sample MC errors were recorded to be less than 6% of the associated standard deviation.

It should be noted that for all three models, the spatial and temporal components captured the underlying global unobserved heterogeneity across county and years (Huang et al., 2017). All three models explicitly account for the spatial and temporal effects, but this approach may not be able to completely capture the temporal instability. Even though this study utilized the crash data aggregated at yearly-basis (a relatively short time period), some studies have found statistical evidence for the presence of temporal instability even in crash data aggregated at short time periods (Malyshkina and Mannering, 2010; Xiong et al., 2014; Behnood and Mannering, 2015; Venkataraman et al., 2016). The recent paper by Mannering (2018) extensively discusses the issue of temporal instability in crash data, which stems from the "individual driver decision-making and the potential evolution of driver decision-making over time". It is known that the unobserved heterogeneity models developed in this study (where the random effects approach of this study may also be considered as heterogeneity models, given random effects is a special case of random parameters) may not be able to capture the distinction between impact of global time trend on crashes and the impact of other sources which may or may not be incorporated in model development (such as human behavior and physiology, vehicle and roadway characters, driving environment and so on).

The temporal instability may play a vital role in this study as the area of focus is an eight-year period (2006-2013) for multimodal crashes aggregated at county level. This period may be highly susceptible to temporal instability due to significant changes in terms of economics and changing accident rates, since the 2007–2009 years witnessed the global recession and some studies have found significant temporal instability during such economic downturns (Maheshri and Winston, 2016; Behnood and Mannering, 2016). However, the temporal instability has limitations for heterogeneity models as well as the data driven models even though advanced methodologies are adopted to explicitly account for temporal elements (Mannering, 2018). Hence, careful consideration should be exercised during interpretation of results from parameter estimates so as to refrain from making erroneous conclusions regarding the impact of causal factors.

For all models, most of the explanatory variables were observed to be statistically significant, as indicated by the absence of zero at the 95% posterior credible region. As shown in Table 2.2, the estimated coefficients for influential variables are observed to be dissimilar across the four modes, which is intuitively reasonable as the external factors may not be expected to have an equivalent impact on the collisions of different modes. For example, the population of a county has a weaker association with vehicle crashes, compared to other modes, which seems logical as population is mostly regarded as a surrogate to reflect the activity of other modes. Apart from the varying quantitative impact, a particular explanatory variable may have a contrasting correlation with crashes of different modes. For example, the median house income of a county tends to increase the crash risk for motorcycle and bicycles while an opposite trend is observed for other two modes. This suggests that areas with higher median income may have population more inclined towards motorcycle and bicycle driving, which directly translates to higher crash risk.

These findings justify the employment of mode-varying coefficients for the covariates as more intuitive and meaningful inferences may be drawn by investigating the segregated impact of factors on different crash modes. The study included two time-varying parameters, population and maintain miles, which are observed to be statistically significant for majority of the cases. This validates the employment of temporal models for the crash dataset at county level.

The statistically significant correlation of linear trend with all crash modes may suggest the presence of temporal instability. Even though this study did not incorporate time-varying coefficients, nevertheless, the significant time trend indicates that the coefficients of explanatory variables may be shifting over time, as expected in the case of an economic downturn. Additionally, the temporal trend may also be capturing the heterogeneity of unobserved factors pertaining to variations in driving behavior or human response in selection of different modes. The heterogeneity models employed in this study, and in general, are limited in establishing a distinction between the unobserved heterogeneity due to temporal variations or due to other factors not incorporated in the model (Mannering, 2018).

TABLE 2.2. Model Estimated Coefficients Statistics

Crash Mode	Variable Coefficient	Model 1		Model 2		Model 3	
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Motorcycle	Intercept	-5.8550	1.4370	-4.2510	0.3128	-8.5980	0.6589
	Population	0.0926	0.0329	0.0801	0.0362	*-0.0179*	*0.0234*
	Maintain Mile	-0.1828	0.0612	-0.2213	0.0450	*0.0012*	*0.0689*
	Total Retail Sales	-0.1683	0.0297	-0.1441	0.0203	-0.1296	0.0207
	Mean Travel Time	0.7039	0.0941	0.7852	0.1689	0.5989	0.0760
	Median House Income	0.1898	0.1138	*0.0293*	*0.0421*	0.4070	0.0661
	Linear Trend	-0.0173	0.0048	*-0.0247*	*0.0219*	NA	NA
	Quadratic Trend	NA	NA	*0.0008*	*0.0024*	NA	NA
Bicycle	Intercept	-7.9690	0.3718	-7.5560	0.9123	-6.2290	0.5310
	Population	0.0766	0.0173	0.0694	0.0262	0.1026	0.0214
	Maintain Mile	-0.5867	0.0623	-0.6074	0.0528	-0.6502	0.0564
	Total Retail Sales	0.3313	0.0277	0.3480	0.0221	0.3825	0.0363
	Mean Travel Time	-0.9509	0.1225	-0.8535	0.1310	-0.9905	0.0775
	Median House Income	0.4547	0.0573	0.3820	0.0922	0.2544	0.0544
	Linear Trend	*0.0006*	*0.0044*	0.0437	0.0200	NA	NA
	Quadratic Trend	NA	NA	-0.0047	0.0022	NA	NA
Pedestrian	Intercept	-6.4660	0.2939	-5.1650	0.6163	-3.1250	0.7451
	Population	0.0952	0.0227	0.0931	0.0399	0.0884	0.0227
	Maintain Mile	-0.4849	0.0738	-0.5467	0.0910	-0.5517	0.0896
	Total Retail Sales	0.3130	0.0370	0.3479	0.0229	0.3772	0.0404
	Mean Travel Time	*0.2324*	*0.1836*	0.3874	0.1291	-0.2362	0.1045
	Median House Income	-0.0909	0.0655	-0.2592	0.0479	-0.3002	0.0755
	Linear Trend	-0.0244	0.0046	*-0.0083*	*0.0221*	NA	NA
	Quadratic Trend	NA	NA	*-0.0016*	*0.0023*	NA	NA
Vehicle	Intercept	-1.8600	0.1181	-2.2250	0.3285	-1.8890	0.1443
	Population	0.0578	0.0115	0.0906	0.0141	0.0683	0.0172
	Maintain Mile	-0.0742	0.0166	-0.0810	0.0266	-0.1045	0.0319
	Total Retail Sales	0.0464	0.0130	*0.0151*	*0.0084*	*0.0386*	*0.0226*
	Mean Travel Time	*0.1175*	*0.0972*	*0.1634*	*0.1104*	*0.1384*	*0.1141*
	Median House Income	-0.1100	0.0329	-0.0734	0.0117	-0.1097	0.0278
	Linear Trend	-0.0393	0.0025	-0.0840	0.0140	NA	NA

	Quadratic Trend	NA	NA	0.0049	0.0015	NA	NA

Notes: 1. The estimates in italics represent the variables which were nonsignificant at 95% level.

2. NA refers to Not Applicable

3. SD refers to Standard Deviation

Model Evaluation

As previously mentioned, DIC is a penalized criterion which balances the model complexity and posterior deviance to eventually give a measure of overall fit of the model. It is worth mentioning that Model 1 and Model 2 differ based on the temporal treatment while Model 3 differs from the former two models by the assignment of spatiotemporal interaction. The results shown in Table 2.3 demonstrate that the essential differences in model specifications are carried over to their performance at model fit. Model 3 leads to the best fit with the lowest value of Dbar at 13,685, which is recorded to be lower from Model 2 and Model 1 by a difference of almost 100 points, as latter two models have minor score difference. This significant difference may be attributed to the inclusion of varying spatial term for each year that allows the flexibility to capture the interactions. However, such immense benefit at fit is accompanied by a significant increase in the complexity due to the addition of effective number of parameters, as reflected by the highest value of Pd=1,455. The model complexity is sufficiently high (with a difference of 220 and 239 points from Model 1 and Model 2, respectively) that the performance of Model 3 at good fit (Dbar) is not able to compensate the remarkable difference of complexity. This eventually leads to the worst performance at DIC (which reflects overall fit), with a difference of 133 points from the best performing Model 2 (DIC=15,007). These findings suggest that the inclusion of varying spatiotemporal interaction term renders greater flexibility to fit the crash data (compared to linear and quadratic trends), but this superiority comes at the cost of increased complexity which translates to increased computational effort and inferior performance at overall fit (DIC). The model complexity tends to be the governing factor for assessment of overall fit while comparing the models with alternate specifications. In comparison among the linear and quadratic trend models, Model 2 is mostly observed to be superior with lower values of Pd (19-point difference) and DIC (12-point difference), with the exception of Dbar (6-point difference). The comparatively less difference for Dbar values reflects that both models have similar posterior deviance, while the difference for the value of Pd reflects that model complexity serves to be the influential criterion for overall fit of the model as Model 2 is noted the least value of DIC (15,004) highlighting its superiority. These findings point towards the advantages associated with the inclusion of a quadratic trend over the linear as it allows the flexibility to fit the crash data better due to a subtler approach without sacrificing the computational ease as the complexity was even lower than the linear model.

TABLE 2.3. Goodness-of-fit Criteria Results

Criterion	Model 1	Model 2	Model 3
Dbar	13,784.4	13,790.7	**13,685.8**
Pd	1,235.27	**1,216.58**	1,455.09
DIC	15,019.7	**15,007.3**	15,140.9
LPML	-6,857	-6,860	**-6,822**

Note: The bold cells represent the best performance for that criterion

The LPML was adopted as a cross-validation measure for assessment of the goodness-of-fit of the models. The higher value of LPML reflects relative superiority at model fit property. As shown in Table 2.3, Model 3 is observed to have the highest value of LPML (-6,822) followed by Model 2 and Model 1. The difference of LPML values among two competing models is referred to as log pseudo Bayes factors (LPBF) (Basu and Chib, 2003), where an LPBF greater than 5 points for a particular model indicates its superior fit (Ntzoufras, 2009). The best performing model (Model 3) is observed to have an average LPBF of 37 points from the other models which reflects the exceptional advantage associated with the incorporation of time-varying spatiotemporal interaction. Unlike the comparison between Model 3 and temporal trend models (Model 1 and Model 2), the difference of model fit is not so pronounced in case of comparison among temporal trend models as reflected by the LPBF of 3 points which is lower than the threshold of 5 points. This indicates that the inclusion of quadratic trend allows relatively more flexibility to fit the data but may not replicate the significant improvement in model fit which is contributed by the inclusion of time-varying spatiotemporal interactions. It is worth mentioning that the superior performance of such complex specification (Model 3) at model fit points towards similar advantage at cross-validated predictive performance (CPO), which formed the basis for computation of LPML.

Site Ranking Performance Evaluation Results

In order for the implementation of the previously mentioned SCT and MCT, the data were divided into two groups of equal time period, Period 1 (2006-2009, "before" period) and Period 2 (2010-2013, "after" period). The three models were used to sort the sites (or, counties) in descending order in terms of crash rate.

Site Consistency Test Results

The larger SCT score in the After Period indicates that the HSID method is better in screening out the counties that consistently demonstrate the inferior performance. To represent different real-life budgeting scenarios, both top 5% and 10% ranked counties were filtered for further analysis. The detailed test results associated with each model are shown in Table 2.4 which represent the accumulated results of the four transportation modes.

TABLE 2.4: The Accumulated SCT Results of Four Transportation Modes

Sites	Model 1	Model 2	Model 3
Top 5% Counties	7.541	7.513	7.554
Top 10% Counties	13.717	13.717	13.717

Table 2.4 exhibits the sum of crash rates of the four transportation modes for both top 5% and top 10% identified counties. In case of top 5%, the counties identified by Model 3 yield the largest total crash rates (7.554), followed by Model 1 and Model 2. However, in case of top 10%, the three models show exactly the same performance. Overall, based on the two cases combined, it is concluded that Model 3 possesses slightly better performance given the small difference across all models.

Method Consistency Test Results

In addition to the site consistency, the study also checked the method consistency of competing models in identifying the same safe or unsafe sites across adjacent time periods. To yield more reliable trend of result findings, the authors employed a continuous set of thresholds which performed the sensitivity/specificity analysis in steps (with 5% as the interval), starting with top 5% of sites and progressing towards the lower bound of top 50% of all counties as hotpots for different transportation modes. The detailed results are graphically illustrated in Figures 2.1 and 2.2, respectively.

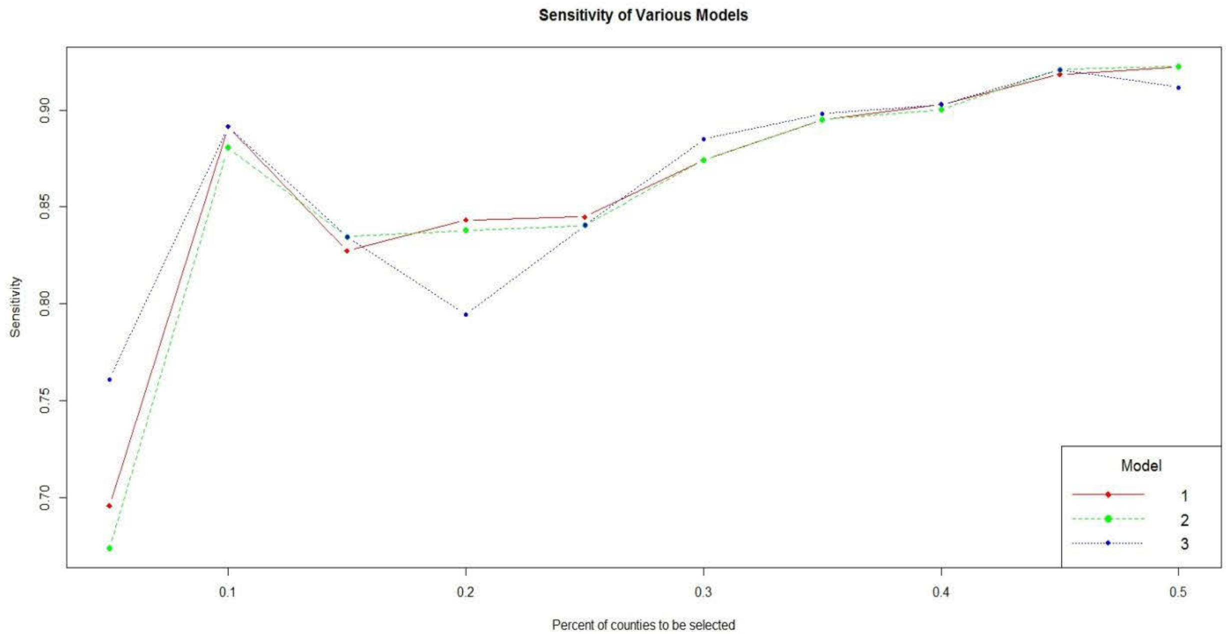


FIGURE 2.1: Sensitivity of Identifying Truly Hazardous Counties in the Subsequent Period for Various Models

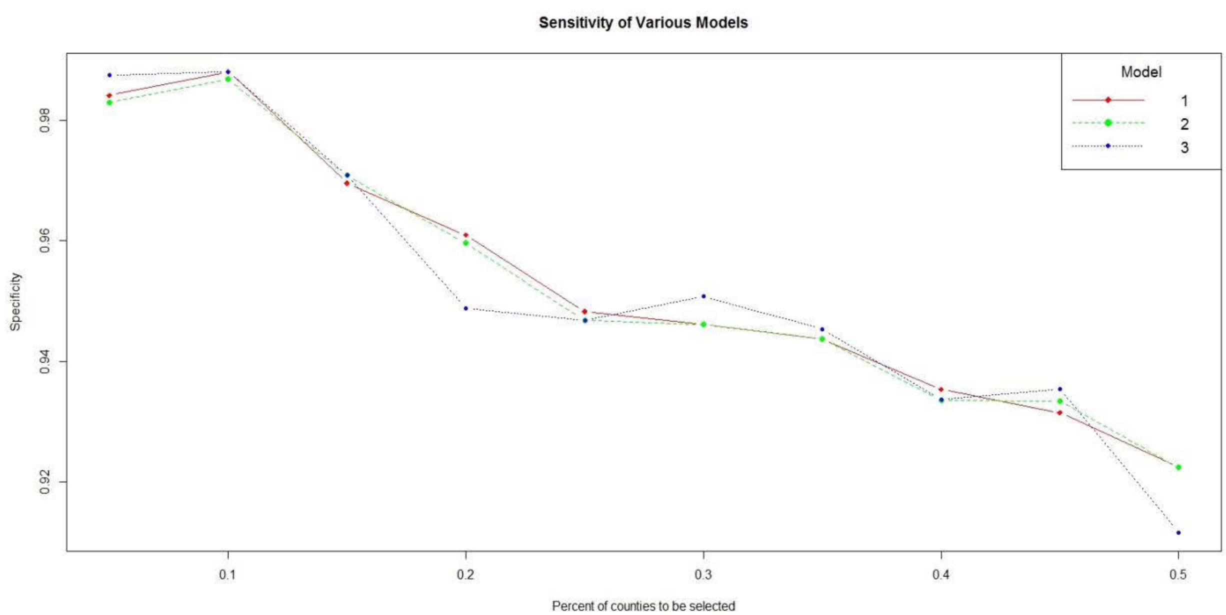


FIGURE 2.2: Specificity of Identifying Truly Hazardous Counties in the Subsequent Period for Various Models

Inspection of Figure 2.3 illustrates that Model 3 has the dominant best performance in consistently identifying top 5%, 10% and 15% counties in both time periods. However, with the threshold of identifying the high-risk sites being dropped, Model 3 tends to have mixed performance compared with other two models. Out of the remaining seven cases, Model 3 shows the best or better performance in 4 situations while performs the worst in detecting top 20%, 25%, and 50%, respectively. In case of specificity analysis, Model 3 exhibits the similar trend by claiming the first place in the first three cases while performing slightly better than other two in the remaining seven cases when lower threshold values are used.

In all, Model 3 appears to have the best performance in consistently identifying the both high and low risk counties across the adjacent time periods and perform slightly better than the other two models in terms of site consistency. In other words, the superiority of the model's predictive performance can be transferred to yield more accurate result of site ranking.

CHAPTER 3: BIVARIATE DIRICHLET PROCESS MIXTURE SPATIAL MODEL FOR ACTIVE MODES CRASHES

INTRODUCTION

This chapter presents the comprehensive analysis of bivariate Dirichlet process mixture spatial model for estimation of pedestrian and bicycle crash counts. The literature review presented many studies which observed the advantages associated with accommodation of different structured and unstructured correlations such as spatial, temporal, and multivariate, for estimating multimodal crashes. Conversely, it also illustrated the limited use of semi- or non-parametric models for simultaneous analysis of active transportation mode crashes. In effect, to the knowledge of the authors, the research for comprehensive analysis of flexible multivariate spatial models focusing on active transportation is non-existent in the safety literature. To fill this research gap, the authors adopted semi-parametric formulation that accounts for the unobserved heterogeneity by combining the strengths of incorporating bivariate specification of dependency among crash modes (pedestrian and bicyclists), spatial random effects for the impact of neighboring areas, and Dirichlet process mixture for random intercepts. Four alternate models were developed for comparison based on the goodness-of-fit and predictive accuracy. The models were evaluated by employing different criteria, namely: LPML (log pseudo marginal likelihood), MSPE (mean-squared predictive error), the R_p^2 statistic, the G^2 statistic, and RSS (residual sum of squares).

METHODOLOGY

Model Specification

The Full Bayesian (FB) framework was employed for estimation of six-year bicyclist and pedestrian crashes aggregated at the Traffic Analysis Zone (TAZ) level. The FB approach has been widely used for crash prediction models due to its capability to handle the multilevel and correlated nature of crash data. Four crash frequency models were developed. The general functional form of the models is given in the following subsections while progressing from simple to sophisticated.

Model 1: Bivariate

This model assumes that crash count of certain modal crash ***j*** at a given location ***i***, $\mathbf{y_{ij}}$, obeys Poisson distribution, while the corresponding observation specific error term ε_{ij} follows a bivariate normal distribution:

$$y_{ij}|\lambda_{ij} \sim Poisson\ (\lambda_{ij}) \tag{3.1}$$

$$\ln(\lambda_{ij}) = X'_{ij}\beta + \varepsilon_{ij} \tag{3.2}$$

$$\varepsilon_{ij} \sim MVN\ (0,\ \Sigma) \tag{3.3}$$

Where $$y_{ij} = \begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix}, \quad \lambda_{ij} = \begin{pmatrix} \lambda_{i1} \\ \lambda_{i2} \end{pmatrix}, \quad \varepsilon_{ij} = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \tag{3.4}$$

In above equations, *X'* is the matrix of risk factors, β is the vector of model parameters, ε_{ij} is the independent random effect which captures the extra-Poisson heterogeneity among locations. Σ is called the covariance matrix. The diagonal element σ_{jj} in the matrix represents the variance of ε_{ij}, where the off-diagonal elements represent the covariance of crash counts of different modes. The inverse of the covariance matrix represents the precision matrix and has the following distribution:

$$\Sigma^{-1} \sim Wishart(I, J) \tag{3.5}$$

Where I is the J x J identity matrix (Congdon, 2006), and J is the degree of freedom, J=2 herein representing two crash outcomes corresponding to bicyclist and pedestrians crashes.

Model 2: Bivariate Spatial

Under Model 2, the spatial random effects were incorporated over the model represented in Equation 2. The final model takes the following form to account for spatial correlations among the TAZs:

$$\ln(\lambda_{ij}) = X'_{ij}\beta + \varepsilon_{ij} + u_{ij} \tag{3.6}$$

Where u_{ij} is the spatially structured random effect which follows the MCAR (multivariate conditional autoregressive) (Mardia, 1998) formulation to incorporate the spatial correlation among crashes occurring at neighboring TAZs.

$$u_i|u_k,\ \Sigma i\ \sim N_j(\textstyle\sum_{k\sim i} C_{ik}\,, u_k, \Sigma i\ ) \tag{3.7}$$

Where each Σ_i is a positive definite matrix representing the conditional variance matrix, and the adjacency matrix C_{ij} is of the same dimension with Σ_i (Jonathan et al., 2016). The precision matrix Σ^{-1} follows the Wishart distribution as shown in Equation 5.

As we can see from the above equations, estimation of the risk in any site is conditional on risks in neighboring locations. Subscripts i and k refer to a TAZ and its neighbor, respectively, and k belongs to N_i where N_i represents the set of neighbors of TAZ i. Besides the identification of neighbors, the assigned weights also affect the risk estimation. In the past studies (Aguero-Valverde and Jovanis, 2009; Xu and Huang, 2015), weight structures including various adjacency-based, distance-based models, and semi-parametric geographically weighted, and so on, have been explored. The current study employs the commonly used distance-based structure to explore the spatial correlations with the following formulation:

$$w_{ij} = \frac{1}{d_{ij}} \tag{3.8}$$

Where w_{ij} is the weight between TAZ i and j, and d_{ij} is the distance between TAZ i and j. With this weight structure, it is known that more weightage was assigned to TAZs which are relatively close.

Model 3: Bivariate Dirichlet Process Mixture

The parametric model specification of the aforementioned models assumed the distribution of the parameters to be specific (normal in this study) across all concerned sites. However, the nonparametric specification removes such constraints by employing a flexible approach of the Dirichlet process that allows the incorporation of unknown random density for the parameters. The current study employs a semi-parametric approach which relaxes the restrictive distributional assumption for the intercept only, instead of all of the parameters. The removal of constraints for the intercept to follow a specific distribution represents a plausible scenario where the TAZs are not expected to have a normal distribution. This flexible approach is expected to capture the extra variability which may escape the error terms introduced in parametric models. Equation 2 was modified to use Dirichlet process mixture over the intercept as follows (Heydari et al., 2016):

$$\ln(\lambda_{ij}) = \beta_{0rj} + X_i'\beta \tag{3.9}$$

$$\beta_{0rj} \approx \sum_{n=1}^{C} p_n I_{\theta_{z_i}} \sim TDP\left(kG_{0j}\right),\ \ z_i = n\ with\ probability\ of\ p_n \tag{3.10}$$

$$G_0 \sim MVN\ (\mu_{G_{0j}},\ \Sigma) \tag{3.11}$$

Where β_{0rj} is the intercept for cluster r (r ranges from 1 to C) of mode j, k is the precision parameter, and G_0 is the baseline distribution for β_{0r} which follows a bivariate normal distribution with mean μ_{G_0} and variance Σ, which also follows the Wishart distribution. β_{0rj} essentially represents a vector of probabilities over the space of concerned entities (203 TAZs) and follows a

Truncated Dirichlet Process (TDP) with a vector of parameters represented by kG_{0j}. The precision parameter k indicates the variability of the Dirichlet process around G_{0j}. The intercept draws random points (θ_{Z_i}) and the associated probabilities (p_n) can be obtained through the stick-breaking procedure (Heydari et al., 2016; Ohlssen et al., 2007). If one cluster is occupied, the indicator function ($I_{\theta_{z_i}}$) at θ_{Z_i} will take the value of 1, otherwise it would be 0. The number of latent clusters (r) in β_{0rj} could range from 1 to infinity, which requires immense computational effort. To reduce the computational complexity by obtaining finite dimensional approximation, a truncated Dirichlet process is utilized to fix the maximum number of possible clusters to C, where C is governed by the precision parameter k and is estimated by $5k$+2 (Ohlssen et al., 2007). As the prior distribution for precision parameter k was assumed to be $k \sim$ uniform (0.3, 9), so eventually the number of clusters were limited to be maximum of 47. The value of C used in the study can be considered in a normal range given the different C values utilized previously such as 5 (Ghosh and Norris, 2005), 10 (Erkanli et al., 2006), and 52 (Heydari et al., 2017).

Model 4: Bivariate Dirichlet Process Mixture Spatial

Model 4 is distinct from Model 3 by incorporating the spatial random effects to account for the correlation among the neighboring TAZs. The model in Equation 9 takes the following form:

$$\ln(\lambda_{ij}) = \beta_{0rj} + X_i'\beta + u_{ij} \tag{3.12}$$

Where all terms are defined as previously.

Comparison of Models Based on Cross-Validation

Many traditional approaches of cross-validation are prone to overestimation due to double usage of data, once during model development and then again for model checking. The approach of cross-validatory predictive densities was proposed to tackle this issue (Gelfand et al., 1992) where the full set of data was divided into two subsets (one subset for development and the other for checking). However, the splitting of two subsets posed a major problem as the selection of different subsets provides varying results. This was resolved by implementing a CV-1 (leave-one-out) technique to estimate the cross-validatory conditional predictive ordinate (CPO) (Gelfand, 1996). This technique removed the selection bias by employing a continuous approach of selecting all data points, except one, for model development and the left out data point to verify the prediction accuracy of the calibrated model. Under the MCMC (Markov Chain Monte Carlo) framework, the estimate of CPO for each observation *i* can be calculated as:

$$CPO = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{f(Y_i|\beta^{(t)})}\right)^{-1} \tag{3.13}$$

Where Y_i is the *i*th observation (i = 1, 2, 3, . . ., n) for all 203 TAZs and β is the vector of estimated model parameters. This harmonic mean of density (CPO) may be extended to calculate the goodness-of-fit of models by computing the product of CPOs over all observations, which is known as the pseudo marginal likelihood. For computational convenience, the log pseudo marginal likelihoods (LPML) is calculated (Heydari et al., 2016; Cheng et al., 2018):

$$LPML = \sum_{i=1}^{n}\log(CPO_i) \tag{3.14}$$

Evaluation Criteria for Predictive Accuracy

In this study, the four competing models were also evaluated based on some criteria used from previous studies: MSPE (mean-squared predictive error, Narayanamoorthy et al., 2013), the G^2 statistic (Cheng and Washington, 2005), the R_p^2 statistic (Washington et al., 2003), the Chi-squared Residual Sum of Square (RSS, Earnest et al., 2007). The details of each criterion are shown in the following subsections.

MSPE

As indicated by the name, such criterion is related with the average squared deviations, or, the predictive errors. Specifically, the MSPE was calculated as follows:

$$\mathrm{MSPE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\lambda_i - y_i)^2 \tag{3.15}$$

Where λ_i is the Bayesian estimated crash frequency for zone *i* while y_i is the observed crash counts of the same zone. The smaller MSPE is preferred which indicates a better prediction performance.

RSS

MSPE is based on the deviations. A potential issue is that the larger estimated count of one zone might mask the smaller ones of multiple TAZs. To address this issue, we also calculated the chi-squared residual sum of squares to determine the deviation standardized by the estimated number of crash counts:

$$RSS = \sum_{i=1}^{n} \frac{(\lambda_i - y_i)^2}{\lambda_i} \tag{3.16}$$

The model with a smaller value of RSS tends to have more predictive capabilities.

The R_p^2 Statistic

The typical R-square in ordinary linear regression cannot be directly applied to the crash frequency model due to the nonlinearity of conditional mean (E[y|X]) and heteroscedasticity associated with the Poisson models. Therefore, we adopted an equivalent measure, R_p^2, which is based on standardized residuals:

$$R_p^2 = 1 - \frac{\sum_{i=1}^{n}\left[\frac{y_i - \lambda_i}{\sqrt{\lambda_i}}\right]^2}{\sum_{i=1}^{n}\left[\frac{y_i - \bar{y}}{\sqrt{\bar{y}}}\right]^2} \tag{3.17}$$

Where $\bar{y}$ represents the mean value of the observed counts. Similar to R-square, a smaller R_p^2value indicates the inferior performance.

The G^2 Statistic

The sum of model deviances, G^2, is zero for a model with a perfect fit. The G^2 statistic is given as:

$$G^2 = 2\sum_{i=1}^{n} y_i LN\left(\frac{y_i}{\lambda_i}\right) \tag{3.18}$$

A large G^2 deviating from zero indicates that the model fits poorly as compared to the saturated model.

DATA PREPARATION

Pedestrian and bicyclist crashes which occurred in the City of Irvine from 2007 to 2012 were analyzed for the study. Like many other research studies, TAZs were selected as the base units, and the crash data were aggregated at the TAZ-level. Overall, there are 203 TAZs in the City. The map in Figure 3.1 displays the distribution of all TAZs and associated crash counts. The two transportation mode-related crashes were collected from SWITRS (California Statewide Integrated Traffic Records System) Shape file of TAZ boundary and TAZ characteristics were provided by SCAG (Southern California Association of Governments).

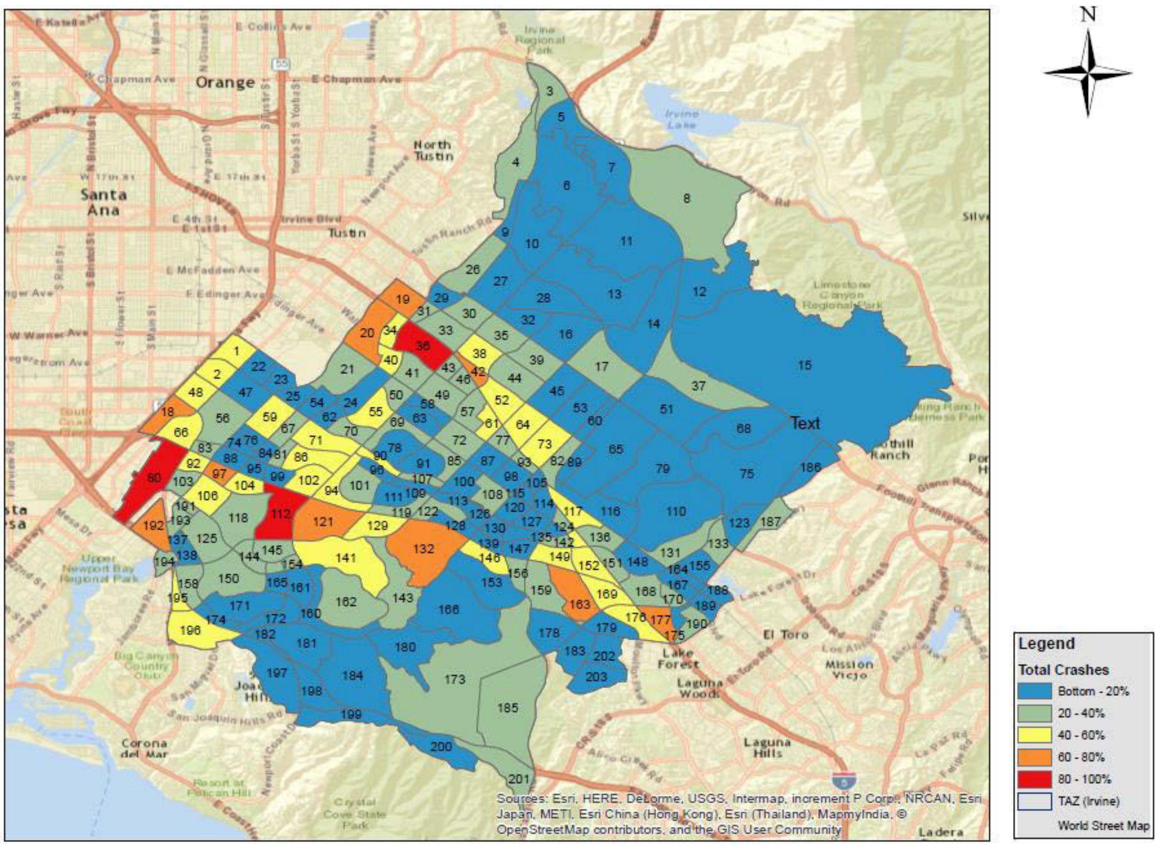


FIGURE 3.1. TAZ map with crash distributions in the city of Irvine, California.

The variables used for model development and the associated descriptive statistics are shown in Table 3.1. The six-year aggregated pedestrian and bicyclist crashes were used as the dependent variables. DVMT acted as a measure of exposure. The explanatory variables were the predictors commonly used in previous regional safety analyses which include socioeconomic, transportation-related, and environment-related factors, and so on. It is worth mentioning that the data from 2008 were available for explanatory variables due to less frequent collection by the agencies and hence it is used for model development. Also, the distance matrix containing distances among various TAZ centroids were also collected from SCAG for the estimation of distance-based spatial random effect. Their descriptive statistics can be found in Table 3.1 as well.

TABLE 3.1. Summary Statistics of Variables for TAZ's of the City of Irvine

Variables	Description	Mean	SD	Minimum	Maximum

Bike	Total bike-involved crashes (2007-2012)	1.82	2.45	0	12
Ped	Total pedestrian-involved crashes (2007-2012)	0.81	1.33	0	8
DVMT	Daily vehicle miles traveled	5,4262.44	56,156.84	112.57	276,079.90
Acre	TAZ Area in acre	282.90	431.75	0.69	5,062.95
Median	Median house income ($)	48,440.78	50,635.10	0	183,347
Pop_den	Population density by area	6.18	7.96	0	32.40
HH_den	Household density by area	2.34	3.15	0	13.62
Emp_den	Employment density by area	10.34	17.43	0	121.10
Ret_den	Retail job density	0.79	2.02	0	17.45
% age 5_17	% of population age 5-17	8.64%	8.78%	0	27%
% age 18_24	% of population age 18-24	5.79%	7.42%	0	40%
% age 24_64	% of population age 24-64	38.35%	36.12%	0	95%
% age 65+	% of population age 65 or older	6.25%	10.21%	0	83%
K12	K12 student enrollment	0.39	1.00	0	5.52
College	College student enrollment	0.11	1.00	0	12.59
Int34_den	Intersection density (3- and 4- legs)	0.12	0.12	0	0.62
BKlnACC	Bike lane access (1=if a TAZ has bike lane)	0.92	0.28	0	1
BL_den	Bike lane density	3.40	1.80	0	7.26
Rail	1=at least one rail station in a TAZ	0.01	0.10	0	1
TTbus_D	Total Bus Stop Density	0.05	0.09	0	0.53
Exbus_D	Stop density for Express Bus and BRT	0.002	0.007	0	0.06
HFLbus_D	High-Frequency Bus Stop Density (local bus headway <= 20 mins)	0.001	0.004	0	0.03
WalkAcc	Walk Accessibility	3.87	9.46	0	74.53
% Arterial	Percent of main arterial (45-55mph) of TAZ	10.61%	17.33%	0	80%
Distance	Distance among TAZ centroids (in miles)	4.06	2.09	0.16	11.78

Note: SD refers to standard deviation.

RESULTS

The crash prediction models were estimated with the freeware statistical package WinBUGS (Spiegelhalter et al., 2003). A total of 10,000 MCMC iterations were utilized for parameter estimation after discarding first 1,000 iterations as burn-in. The convergence was ensured by employing different approaches such as visual inspection of history plots, trace plots, and Gelman-Rubin diagram (Gelman and Rubin, 1992). The Pearson correlation coefficient was calculated and the covariates correlated at a significance level of 0.05 were subsequently eliminated.

Modeling Results

As shown in Table 3.2, the posterior inferences for influential factors for all four models demonstrate their robustness to fit the multimodal crash data at the TAZ spatial scale. All four models identify similar significant factors that affect crash frequency for a particular mode. In the case of bicycle crashes, three variables are observed to be statistically significant, namely: K12 student enrollment, percentage of arterials, and bike-lane density for the TAZ. The TAZs with higher K12 student enrollment increases the crash risk as the instances of interaction of bicyclists with other modes increases due to more exposure. However, the similar positive correlation for bike-lane density seems counter-intuitive since the presence of bike lanes is expected to facilitate more usage of bicycles due to lower perceived risk of interaction with other modes. The possible rationale for this finding may be explained by the lower perceived risk which may encourage bicyclists to ride more in such areas, while conversely increasing the crash risk due to higher exposure of bicyclists to vehicular traffic. The negative relationship among percentage of arterial roads and bicycle crashes indicates that maybe the bicyclists tend to travel less in areas with more arterials. For the crashes pertaining to pedestrians, the college enrollment is also observed to be influential, along with other three factors shared with bicycle crashes. The increase in student population in the colleges of TAZs is noted to be negatively linked with pedestrian crashes, though the increased pedestrian activity usually associated with the presence of college students was expected to increase crash occurrence. The probable justification may be that the known presence of students influences the vehicle drivers to be more cautious and drive sensitively, or the vehicular activity may be minimal in such areas which may help significantly reduce the possibility of interaction with pedestrians. The common significant factors (K12 student enrollment, percentage of arterials, and bike-lane density) responsible for bicycle and pedestrian crashes support the joint estimation of such modes which are most vulnerable and impacted by similar characteristics. As shown in Table 3.3, the heterogeneity error term demonstrates the presence of statistically significant correlation among the bicycle and pedestrian crashes which further justifies the employment of bivariate structure for joint estimation of crashes. However, the spatial random effect term exhibits the absence of a significant correlation, as indicated by the covariance matrix. It may be possible that the explanatory variables incorporated for model development are sufficiently robust to account for the spatial characteristics that influence crash occurrence for the particular modes.

TABLE 3.2. Posterior Inference for Bicyclist and Pedestrian-involved Crash Counts

Count Type	Variables	Model 1	Model 2	Model 3	Model 4
Bicyclist	Intercept	**-10.860 (0.243)**	**-10.880 (0.246)**	**-10.780 (0.248)**	**-10.790 (0.234)**
	% age 65+	1.532 (0.922)	1.467 (0.895)	1.413 (0.830)	1.401 (0.798)
	K12	**0.203 (0.088)**	**0.203 (0.091)**	**0.213 (0.079)**	**0.211 (0.074)**
	College	-0.013 (0.078)	-0.015 (0.077)	-0.014 (0.079)	-0.012 (0.075)
	WalkAcc	-0.007 (0.010)	-0.008 (0.010)	-0.006 (0.010)	-0.007 (0.010)
	% Arterial	**-3.517 (0.674)**	**-3.529 (0.685)**	**-3.472 (0.691)**	**-3.399 (0.655)**
	BL_den	**0.260 (0.056)**	**0.271 (0.057)**	**0.245 (0.056)**	**0.246 (0.056)**
Pedestrian	Intercept	**-12.390 (0.326)**	**-12.430 (0.357)**	**-12.360 (0.340)**	**-12.380 (0.346)**
	% age 65+	1.205 (1.145)	1.192 (1.101)	1.097 (1.074)	1.131 (1.009)
	K12	**0.280 (0.104)**	**0.280 (0.106)**	**0.291 (0.095)**	**0.291 (0.094)**
	College	**-0.976 (0.567)**	**-0.968 (0.563)**	**-0.962 (0.562)**	**-0.957 (0.558)**
	WalkAcc	0.009 (0.010)	0.008 (0.010)	0.010 (0.010)	0.009 (0.010)
	% Arterial	**-3.826 (0.989)**	**-3.805 (0.985)**	**-3.727 (0.991)**	**-3.658 (0.996)**
	BL_den	**0.384 (0.068)**	**0.397 (0.075)**	**0.374 (0.069)**	**0.375 (0.074)**

Notes: 1. Intercept for Dirichlet Process models indicates the intercept mean from mixture points.

2. Refer to Table 3.1 for detailed description of variables.

3. Numbers in parentheses represent uncertainty estimates, or, posterior standard deviations.

4. The statistically significant variable coefficients are shown in bold.

5. Model 1: bivariate; Model 2: bivariate spatial; Model 3: bivariate dirichlet process mixture; Model 4: bivariate dirichlet process mixture spatial.

TABLE 3.3. Covariance Matrices for the Four Alternative Models

Models	Modes	Heterogeneity (ε_{ij})		Spatial (u_{ij})	
		Bicycle	Pedestrian	Bicycle	Pedestrian
Model 1	Bicycle	**0.896 (0.166)**	**0.854 (0.166)**		
	Pedestrian	**0.854 (0.166)**	**0.890 (0.237)**		
Model 2	Bicycle	**0.860 (0.168)**	**0.827 (0.153)**	**0.001 (2.2x10^{-4})**	6.7x10^{-5} (1.5x10$^{-4)}$
	Pedestrian	**0.827 (0.153)**	**0.856 (0.213)**	6.7x10^{-5} (1.5x10$^{-4)}$	**0.001 (2.2x10^{-4})**

Model 3	Bicycle	**0.602 (0.200)**	**0.538 (0.182)**		
	Pedestrian	**0.538 (0.182)**	**0.561 (0.226)**		
Model 4	Bicycle	**0.507 (0.231)**	**0.461 (0.234)**	**0.001 (2.1x10^{-4})**	7.4x10^{-5} (1.5x10^{-4})
	Pedestrian	**0.461 (0.234)**	**0.503 (0.270)**	7.4x10^{-5} (1.5x10^{-4})	**0.001 (2.2x10^{-4})**

Notes: 1. Numbers in parentheses represent posterior standard deviations.

2. The statistically significant covariance values are shown in bold.

3. Model 1: bivariate; Model 2: bivariate spatial; Model 3: bivariate dirichlet process mixture; Model 4: bivariate dirichlet process mixture spatial

Evaluation Results

As previously stated, the four models are evaluated from different perspectives using five evaluation criteria. Unlike the traditional parametric models which usually employ DIC (deviance information criterion) for model comparison, LPML is adopted in this study as DIC is not generated by the WinBUGS due to its sensitivity to different parameterizations (*28,49*). The higher value of LPML is desirable as it reflects relatively superior model fit property and a difference of more than 5 points among two competing models help identify the model of interest (*50*). As shown in Table 3.4, the LPML values of all four models are close enough to not cross the threshold of 5 points for identification of the model of interest. However, the sample size also impacts the numerical value of LPML. Hence it may be worthwhile to record the model with highest LPML value and compare the observation with other criteria. As evident from the evaluation results, Model 3 demonstrates the best fit based on relatively large LPML (-474.433), closely followed by Model 4. A similar trend is observed for all other criteria suggesting the strong correlation among the capability of a model to fit crash data and its performance at crash predictive accuracy.

Further inspection of the evaluation results reveals that the models which account for spatial correlations (Models 2 and 4) have consistently inferior performance to those with spatially unstructured heterogeneity (Models 1 and 3). Such phenomenon suggests that the inclusion of spatial correlation structures raises the model complexity without notable advantage at crash prediction, which is usually expected in such cases as reduced posterior deviance compensates the increased complexity. The potential reason might be due to the insignificant spatial dependency among the two modal crashes as shown in Table 3.3. Clearly, the Dirichlet models (Models 3 and 4) outperform the non-Dirichlet ones (Model 1 and 2) based on all five criteria suggesting the use of such flexible framework.

Apart from the above findings, the logical aspect of employing the flexible approach should also be given consideration. For a given crash dataset, the parametric approach assumes a restrictive stationary distribution of explanatory variables across all the sites under focus. As discussed in previous studies of semi-parametric models (*25, 26*), the Dirichlet formulation allows the examination of the adequacy of standard parametric assumption. As clearly shown in Figure 3.2, the kernel posterior density plots of Dirichlet precision parameter k illustrate the closeness of the peak towards zero which reflects that the unknown density (G) of non-parametric intercept is far from the baseline distribution (G_0). Similar plots for both Dirichlet models suggest their robustness and indicate that the normal assumption of intercept associated with traditional parametric models does not hold true for the TAZ level crash dataset of the current study. This indicates that the 203 intercepts associated with the TAZs are not normally distributed and the standard parametric approach does not hold true for the concerned pedestrian and bicycle crashes at the planning level. This finding seems plausible since the safety mechanisms which impact the pedestrian and bicycle crashes may vary across different TAZs due to diverse factors (such as driving behavior, road environment, and so forth) which may not be captured in the explanatory variables. These findings also suggest the presence of distinct subpopulations among the crash data which was confirmed from the histogram of posterior number of latent clusters with a median of 2 clusters for most of the data. This capability of Dirichlet models to identify the latent subpopulations may prove highly beneficial for the safety agencies to investigate similarities in the safety issues among different sites and allocate funding for dedicated countermeasures (*26*). This is achieved by calculating the expected probabilities of sites to fall into same clusters, which allows detection of the degree of similarity or dissimilarity among sites based on the crash risk (*25*).

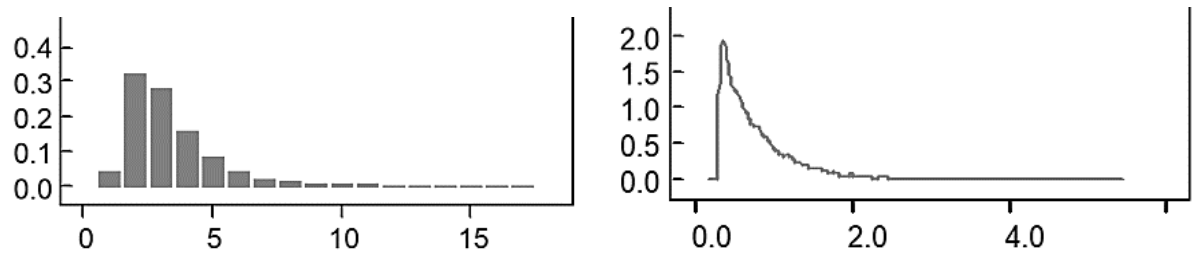
The aforementioned advantages justify the use of Dirichlet process mixture with a flexible intercept as such model specification helps more precise estimation leading to better inferences. Contrary to the parametric models which restrict the priors to a specific distribution fixed across all entities, the latent clusters capture the multimodality due to unconstrained nature.

TABLE 3.4. Evaluation Results for Alternative Models

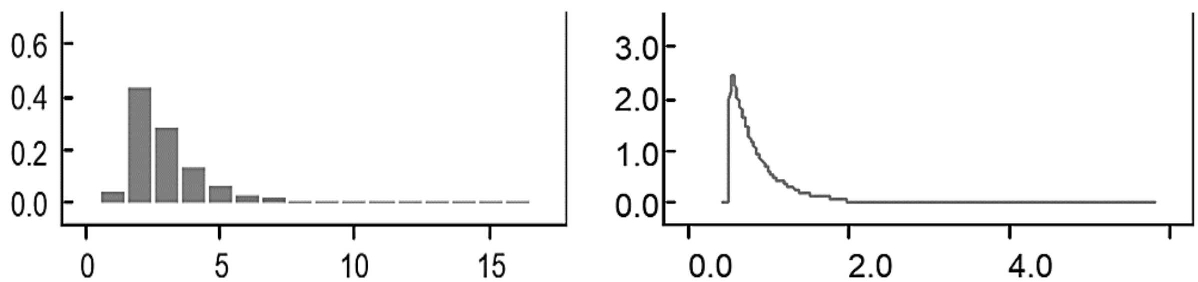
Model	LPML	MSPE	R_p^2	G^2	RSS
Model 1	-476.753	0.690	0.786	177.995	272.367

Model 2	-477.492	0.691	0.781	179.544	278.749
Model 3	**-474.433**	**0.682**	**0.823**	**169.137**	**225.018**
Model 4	-474.831	0.687	**0.823**	169.998	225.291

Notes: Model 1: bivariate; Model 2: bivariate spatial; Model 3: bivariate dirichlet process mixture; Model 4: bivariate dirichlet process mixture spatial.



(a) Kernel densities for Dirichlet Spatial (Model 4)



(b) Kernel Densities for Dirichlet without Spatial (Model 3)

FIGURE 3.2. Kernel density plots for precision parameter and latent clusters.

CHAPTER 4: COMPREHENSIVE ASSESSMENT OF TEMPORAL TREATMENTS IN CRASH PREDICTION MODELS

INTRODUCTION

As evident from the literature review, as an important feature associated with crash data, the temporal correlation issue in the crash frequency models has been the focus of a fairly large number of studies. Although some studies attempted the comparison among few temporal models, a comprehensive comparison of temporal treatments of crash data is not well documented and hence much needed. To this end, the authors proposed different models which represent alternate ways of addressing the serial correlations in the crash prediction models. Specifically, the models were classified into nine groups: (1) independent-over-time random effect (serving as the base); (2) linear time trend; (3) quadratic time trend; (4) yearly-varying intercept; (5) Autoregressive-1 (AR-1); (6) Autoregressive-2 (AR-2); (7) moving average-1 (MA-1); (8) moving average-2 (MA-2); and (9) time adjacency. These models (except the last one) were selected as representative of the temporal models commonly employed in the safety literature. It is important to note that the time adjacency model is first proposed in the traffic safety field, which is borrowed from a previous disease mapping study (Abellan et al., 2008). This model marks the innovative contribution of this study towards the field of traffic safety and is proposed with the intention to compare its performance relative to the prevailing temporal models. Essentially, it contributes a novel approach to incorporate the temporal aspect of crash data by employing CAR (conditional autoregressive) specification for assigning different weights, an approach usually employed to obtain weight matrices for correlations in spatial models. Additionally, since the serial correlations have been handled in different situations in terms of spatial dependency, the research also developed three types of models for each temporal treatment group which contains: (1) temporal correlation only; (2) independent temporal and spatial correlations without spatiotemporal interactions; (3) temporal and spatial correlations with their interactions. Overall, there are 27 models being developed for the assessment purpose. Furthermore, in order to the reveal the modeling performance from different angles, ten different evaluation criteria were utilized which include: (1) $\overline{D}$ (the posterior mean deviance); (2) P_D(the effective number of parameters); (3) DIC (Deviance Information Criterion, which is the sum of $\overline{D}$ and P_D); (4) LPML (log pseudo marginal likelihood); (5) MAD (Mean Absolute Deviation); (6) MSPE (Mean Square Predictive Error); (7) The R_p^2 statistic; (8) The G^2 statistic; (9) Chi-squared Residual RSS (Sum of Squares); (10) TRD (Total Rank Difference). The array of evaluation criteria have been implemented in previous studies and are anticpated to demonstrate the models' capabilities from various aspects. However, there are very few studies dedicated to the correlation among these criteria. Building upon the large numer of models proposed (twenty seven), the study also intends to explore the dependency among the selected evaluation measures. It is hoped that such correlation analysis would equip safety professionals with more confidence when selecting the models based on certain assessment criteria. Finally, all models and evaluations were performed on county-level collision data in California using Bayesian hierarchical framework.

METHODOLOGY

One of the benefits of the modern Monte Carlo Markov Chain (MCMC) sampling methods is the ease with which full marginal densities of parameters may be obtained. Based on the iterative MCMC methods, Bayesian techniques substantially improve on traditional multiple integrations or analytical approximation methods that are infeasible with a large number of parameters (Congdon, 2005). As a result, the use of Bayesian methods for the analysis of data has grown significantly in areas such as traffic safety where a wide range of modeling formulations have been proposed in the past. Specifically, for crash frequency data, various analytical approaches including Poisson, Poisson-Gamma, Negative Binomial, Poisson-lognormal, Poisson-Weibull models, and so on, have been implemented. This present study chose Bayesian Hierarchical Poisson-Lognormal model given its capability of better handling of small sample size due to the heavier tails associated with Lognormal distribution.

As mentioned before, there are 27 models (nine groups and three types) being implemented which allow comprehensive evaluation of different ways of treating serial correlations in the crash frequency models. For ease of description, the following subsections first present the general model specification for the Bayesian Hierarchical model, then cover the details of groups of temporal treatments, followed by difference of the three types of models with respect to spatial dependency. Finally, the analytics of various evaluation criteria are outlined.

Bayesian Hierarchical Model Specification

In the Bayesian Hierarchical Poisson Lognormal framework, at the first level, the authors defined a Poisson model for the within-county variability of crash counts:

$$y_{it} \sim Poisson\ (e_{it}\lambda_{it}) \tag{4.1}$$

Where y_{it} is the observed crash count for county i and year t, λ_{it} is associated mean expected crash rate, and e_{it} is the corresponding exposure, which is the total daily vehicle-miles (DVMT) by county in this study.

At the second level, λ_{it} can be split into portions explained by overall risk (β_0), main temporal effects (ξ_t) and spatial effects (π_i, applies merely to Model Types 2 & 3), space-time interactions (υ_{it}, applies only to Model Type 3) and random errors (ε_{it}), which can be assumed with the noninformative normal priors. As a result, λ_{it} can be expressed as follows:

$$\begin{cases} \ln(\lambda_{it}) = \beta_0 + \xi_t + \pi_i + \upsilon_{it} + \varepsilon_{it} \\ \varepsilon_{it} \sim Normal\ (0, \tau^2\ ) \end{cases} \tag{4.2}$$

The detailed distributions of ξ_t, π_i, and υ_{it} are provided in subsections of Nine Groups of Temporal Treatments and Three Type of Models. It should be noted that usually the model development incorporates some potential influential factors. However, the model in this study does not incorporate any covariates with multiple reasons. First, the major focus of this study is to investigate the impact of different ways of addressing temporal effects on modeling performance, the inclusion of various factors might blur such comparison. Second, some spatial-only covariates (e.g. the geographic areas which are constant throughout time) would have different impacts on the models with and without spatially structured random effects and thus potentially yield biased comparative results. Third, the exclusion of covariates would allow enough heterogeneity to support the estimation of various random effects, especially the space-time interaction one.

At the third level, the variance parameters that are involved in the equation of second-level (Equation 2) are generally treated with unknown and hyper-prior distributions. As with many other Bayesian analyses, the authors chose inverse gamma for τ^2 with parameters 0.5 and 0.0005.

Nine Groups of Temporal Treatments

As previously stated, there are many ways of addressing the main temporal effects (ξ_t) of the crash data. The study explores nine different temporal treatments and the associated model formulations are presented in order of complexity, from the independent-over-time random effect to more sophisticated ones.

Group 1: Independent-over-time Random Effect (Base)

In this group, all crash counts of different years are assumed to be independent of each other, without any applicable serial trends or correlations (*28*). Under this condition, Equation 2 is reformed using the following expression and would serve as the “base” scenario for other temporal treatments.

$$\ln(\lambda_{it}) = \beta_0 + \pi_i + \upsilon_{it} + \varepsilon_{it} \tag{4.3}$$

Group 2: Linear Trend

Under this group, a global linear temporal trend is applied to crash counts of all counties. Equation 2 is modified to the following form:

$$\ln(\lambda_{it}) = \beta_0 + \beta_{t1} t + \pi_i + \upsilon_{it} + \varepsilon_{it} \tag{4.4}$$

Where β_{t1} is the the scalar parameter for linear yearly trend, and t represents various years. The crash counts of counties are anticipated to be decreased or increased with the same rate with time going.

Group 3: Quadratic Trend

Within this group, an additional quadratic term is applied to crash counts in addition to the linear trend. The underlying premise is that the quadratic term can capture some heterogeneity which escapes from the linear term. The resulting expression is shown as follows.

$$\ln(\lambda_{it}) = \beta_0 + \beta_{t1}t + \beta_{t2}t^2 + \pi_i + \upsilon_{it} + \varepsilon_{it} \tag{4.5}$$

Where β_{t2} is the scalar parameter for quadratic yearly trend.

Group 4: Time-varying Intercepts

This group assumes that the crashes of differing years have a time-changing overall risk (intercepts). The corresponding formulation is expressed in Equation 6:

$$\ln(\lambda_{it}) = \beta_{0t} + \pi_i + \upsilon_{it} + \varepsilon_{it} \tag{4.6}$$

Where β_{0t} is the vector of yearly varying intercepts. Essentially, it can be regarded as a "random parameter" model in terms of intercepts. It is worth mentioning that other covariates are not utilized. Otherwise, the associated variable coefficients might be time-varying as well.

Group 5: First-order Autoregressive Process (AR-1)

This group takes into the consideration the autoregressive safety effect by relating the crashes of current year (t) to the ones from previous year (t-1) with one certain coefficient. The study chose AR-1 based on the assumption of stationarity restriction. Under this formulation, the distributions of ε_{it} in Equation 2 are given by:

$$\begin{cases} \varepsilon_{it} \sim normal\left(0, \frac{\sigma_{it}^2}{(1-\emptyset_1^2)}\right) \\ \varepsilon_{it} \sim normal(\emptyset_1 \varepsilon_{i,t-1}, \sigma_{it}^2), t > 1 \end{cases} \tag{4.7}$$

Where $\emptyset_1$ is the autocorrelation coefficient with the range of $-1 < \emptyset_1 < 1$ (*39*).

Group 6: Second-order Autoregressive Process (AR-2)

Different from AR-1, AR-2 assumes that the crashes of the present year are dependent on the ones of two previous years. Correspondingly, the distributions of ε_{it} are shown in the following expression:

$$\begin{cases} \varepsilon_{it} \sim normal\left(0, \frac{\sigma_{it}^2}{(1-\emptyset_2)^2-\emptyset_1^2} \cdot \frac{1-\emptyset_2}{1+\emptyset_2}\right) \\ \varepsilon_{it} \sim normal(\emptyset_1 \varepsilon_{i,t-1} + \emptyset_2 \varepsilon_{i,t-2}, \sigma_{it}^2), t > 2 \end{cases} \tag{4.8}$$

Where $\emptyset_1$ and $\emptyset_2$ are the autocorrelation coefficients. An AR-2 process is stationary provided $|\emptyset_2| < 1, and\ |\emptyset_1| + \emptyset_2 < 1$ (Jung et al., 2006).

Group 7: First-order Moving Average Process (MA-1)

Under this group, the crashes of the present year are a linear function of errors of past year. In specific, MA-1 takes the form as follows:

$$\begin{cases} \ln(\lambda_{it}) = \beta_0 + \pi_i + \upsilon_{it} + \varepsilon_{it} + \theta_1 \cdot \varepsilon_{i,t-1} \\ \varepsilon_{it},\ \varepsilon_{i,t-1} \sim Normal\ (0, \sigma^2) \end{cases} \tag{4.9}$$

Where θ_1 is the coefficient associated with error of past year. The error terms of current and past years are assumed to be independently distributed with a normal distribution with mean zero and a constant variance σ^2. MA-1 process is invertible provided $|\theta_1| < 1$ (Barron, 1992).

Group 8: Second-order Moving Average Process (MA-2)

An MA-1 process can be easily extended to MA-2 where crashes of present year are related with errors of past two years. The resultant expression is shown as below:

$$\begin{cases} \ln(\lambda_{it}) = \beta_0 + \pi_i + \upsilon_{it} + \varepsilon_{it} + \theta_1 \cdot \varepsilon_{i,t-1} + \theta_2 \cdot \varepsilon_{i,t-2} \\ \varepsilon_{it},\ \varepsilon_{i,t-1}, \varepsilon_{i,t-2} \sim Normal\ (0, \sigma^2) \end{cases} \tag{4.10}$$

Where θ_1 and θ_2 are the coefficients associated with errors of past two years. Again, the error terms of current and past two years are assumed to be independently distributed with a normal distribution with mean zero and a constant variance σ^2. MA-2 process is invertible provided $\left|\theta_j\right| < 1, j = 1,2$ (Barron, 1992).

Group 9: Time Adjacency

As indicated by the name, this group assumes strong influence of the adjacent years on crash risk, similar to the dependency of adjacent roadway entities in the case of spatial correlation. Compared to the AR or MA models which capture the impact of crash risk of the year(s) prior to the concerned time, this model also accounts for the potential impact of the subsequent year(s). To the best knowledge of authors, such formulation has not been applied in the traffic safety field yet. Such treatment is borrowed from a previous study regarding public health issues (Abellan et al., 2008) with the intention to explore its capabilities in analyzing safety data.

Similar to many other spatial correlation studies, the authors estimated temporal effects (ξ_t) in Equation 2 by applying CAR (conditional autoregressive) formulation which incorporates the temporal correlation among crashes occurring at neighboring years. The details of CAR can be found in the following subsection. It is important to note that the relatively small sample size (t=8) might harm the performance of this group to some degree.

Three Types of Models

Aside from the temporal treatment, the 27 models can also be divided into three types in terms of different ways of handling the spatial correlations among counties. Each of the model types is applied to all groups of temporal treatments. The detailed description is provided below in order.

Type 1: Temporal Correlation only

This type of models assumes that the spatial feature of the crash risk at other counties has ignorable impacts on the crashes of concerned county. In response, Equation 2 is modified by removing the terms of π_i and υ_{it}:

$$\ln(\lambda_{it}) = \beta_0 + \xi_t + \varepsilon_{it} \tag{4.11}$$

Type 2: Space-Time Model without Spatiotemporal Interactions

Different from Type 1, this model type also accounts for the spatial correlations among the crash risks of counties, but the associated spatial and temporal interactions are not considered. Hence, the updated Equation 2 is shown as below for this type.

$$\ln(\lambda_{it}) = \beta_0 + \xi_t + \pi_i + \varepsilon_{it} \tag{4.12}$$

Alternative autoregressive models were used to estimate the spatially structured random effect such as the conditional autoregressive model (CAR) and the simultaneously autoregressive model (SAR, Wall, 2004). In this study, the authors selected the commonly used CAR in the safety field with the distance-based weight structure. The specific formulation is shown in Equation 13:

$$\pi_i \sim N\left(\frac{\overline{\pi_i},1}{\tau_i}\right) \qquad \overline{\pi_i} = \frac{\sum_{i\neq k}\pi_k w_{ik}}{\sum_{i\neq k} w_{ik}} \qquad \tau_i = \frac{\tau_c}{\sum_{i\neq k} w_{ik}} \quad and \quad w_{ij} = \frac{1}{d_{ij}} \tag{4.13}$$

Where w_{ij} is the proximity weight matrix, τ_c is the precision parameter in the CAR prior, and d_{ij} is the distance between counties. As shown in the above equation, estimation of the risk in any county is conditional on risks in other locations. Subscripts i and k refer to a county and its neighbor, respectively, and $k \varepsilon N_i$, where N_i represents the set of neighbors of county i. The current study employs the distance-based structure to explore the spatial correlation, where all other counties are considered neighboring counties and relative influence is inversely related with distances in between. It is noteworthy that, for the serial correlation in Time Adjacency Group, the authors chose Adjacency-based weight structure where the weights are given by $(w_{ij}) = 1$ if counties i, j are adjacent, and 0 otherwise. Such weight matrix gives a dichotomous output for weight with only two responses, zero for non-neighbors and one for neighbors. Hence, the proximity of immediate neighbors establishes a temporal dependency while more distant years are considered not to have a significant impact.

Once π_i is estimated, it is also interesting to calculate the percentage of extra-Poisson variability that is due to spatial clustering:

$$\alpha = \frac{sd\ (\pi)}{sd\ (\pi) + sd(\varepsilon)} \tag{4.14}$$

Where α is defined as fraction, and sd is the marginal standard deviation function. The larger the fraction value, the more variability explained by the spatially structured random effects.

Type 3: Space-Time Model with Spatiotemporal Interactions

Type 3 is distinct from Type 2 with the inclusion of interactions between temporal and spatial correlations. Different formats of interactions have been found in the literature which includes linear, time-varying spatial (Lawson et al., 2003) and the one represented by mixed components (Abellan et al., 2008). This research chose the linear interaction that has been used in the previous safety study (Aguero-Valverde and Jovanis, 2006). In this case, Equation 2 is changed with the following form:

$$\ln(\lambda_{it}) = \beta_0 + \xi_{\boldsymbol{t}} + \pi_i + \delta_i \cdot \boldsymbol{T} + \varepsilon_{it} \tag{4.15}$$

Where $\boldsymbol{T}$ is the vector representing various time periods and δ_i represents the interacted spatial correlations which are also estimated with CAR formulation. The interaction term allows different temporal trend in crash risk for different spatial locations.

Analogously, the updated formula for fraction is shown in Equation 16.

$$\alpha = \frac{\mathrm{sd}\ (\pi) + \mathrm{sd}\ (\delta)}{\mathrm{sd}\ (\pi) + \mathrm{sd}\ (\delta) + \mathrm{sd}(\varepsilon)} \tag{4.16}$$

The fraction, in this case, indicates the proportion of crash variability explained by both spatial effects: $\pi_i$and δ_i.

Evaluation Criteria

In addition to various groups of temporal treatment and types of models, the study also assesses the model performance by using ten different evaluation criteria which were found in the past research. The authors intend to evaluate the performance of the array of models from varying angles, and also to explore whether there exist strong correlations among various criteria. The brief background and analytics of each criterion are provided as follows.

Criteria 1~3: DIC and Its Components

Similar to Akaike information criterion (AIC), the Deviance Information Criterion (DIC) is a penalized fit criterion developed by Spiegelhalter et al. (2003). Compared with AIC and BIC, the DIC is more easily calculated from the samples generated by an MCMC simulation. Usually, the estimation of DIC can be conducted through the following expression.

$$\begin{cases} DIC = \bar{D} + P_D \\ D(\theta) = -2 \cdot \log(p(y|\theta)) \\ \bar{D} = E^{\theta}[D(\theta)], \;\; P_D = \bar{D} - D(\bar{\theta}) \end{cases} \tag{4.17}$$

Where, y are the data; θ are the unknown parameters of the model; $p(y|\theta)$ is the likelihood function; $\bar{\theta}$ is the expectation of θ; $\bar{D}$ is the posterior mean deviance, or the expectation of the deviance, $D(\theta)$; $\bar{D}$ measures how well the model fits the data; the smaller the $\bar{D}$, the better the fit. P_D represents the effective number of parameters. In general, $\bar{D}$ 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 number of parameters. Usually, smaller values of DIC are preferred. Some general guidelines are also suggested by Lunn et al. (2012): models with DIC values within 5 of the 'best' model are comparable and also strongly supported, values within 5 and 10, weakly supported, and models with a DIC greater than 10 are substantially inferior. Therefore, it could be misleading to only report the model with lowest or highest DIC. For clarification, $\bar{D}$, P_D, and DIC are considered as criterion 1, 2, and 3, respectively.

Criterion 4: Log Pseudo Marginal Likelihoods (LPML)

Such criterion checks the predictive capability of models based on a special case of cross-validation, leave-one-out cross-validation, where the number of folds equals the number of instances in the dataset. It has the benefit of avoiding selection bias in the sense that it treats each of the data observation as a single-item test set, while using all other instances as a training set. This process is closely related to the statistical method of jack-knife estimation (Efron, 1982). In general, Under the MCMC framework, calculation of LPML can be done through the following expression:

$$\begin{cases} CPO = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{p(Y_i|\beta^{(t)})}\right)^{-1} \\ LPML = \sum_{i=1}^{n}\log(CPO_i) \end{cases} \tag{4.19}$$

Where CPO is conditional predictive ordinate; Y_i is the ith observation (i = 1, 2, 3, . . ., n) for all counties; β is the vector of estimated model parameters; $p(Y_i|\beta^{(t)})$ is the likelihood; t represents interation. The larger the LPML value, the better the predictive performance tends to be.

Criterion 5: Mean Absolute Deviation (MAD)

MAD is also frequently used by researchers to check the fitness of data irrespective of the data distribution. It is simply based on the model deviation or residue. The corresponding expression is shown in Equation 20.

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|Y_i - O_i| \tag{4.20}$$

Where Y_i is the Bayesian-estimated crash frequency and O_i is the observed crash count for county i by a model during the same time period. The smaller the MAD value, the better fitness to the data.

Criterion 6: Mean Square Predictive Error (MSPE)

MSPE differs from MAD by calculating the square of deviation, rather than the corresponding absolute value. Specifically, MSPE assumes the form shown as below:

$$MSPE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - O_i)^2 \tag{4.21}$$

Where terms are as defined previously. It is expected that MSPE assigns more penalties to the counties whose deviations are larger due to the squared deviation. Again, the larger MSPE indicates an inferior performance.

Criterion 7: The R_p^2 statistic

A Poisson regression model is usually associated with the nonlinearity of conditional mean (E[y|**X**]) and heteroscedasticity (Washington et al., 2003). Therefore, R-square in ordinary linear regression cannot be directly applied to the crash frequency model. Instead, the present research chose an equivalent measure, R_p^2, which is based on standardized residuals and computed as follows:

$$R_p^2 = 1 - \frac{\sum_{i=1}^{n}\left[\frac{O_i - Y_i}{\sqrt{Y_i}}\right]^2}{\sum_{i=1}^{n}\left[\frac{O_i - \bar{O}}{\sqrt{\bar{O}}}\right]^2} \tag{4.22}$$

Where Y_i and O_i are as defined previously, and $\bar{O}$ represents the mean value of the observed counts. Similar to R-square, a larger R_p^2value is preferred.

Criterion 8: The G^2 statistic

The sum of model deviances, G^2, gives a test of whether the model gives an adequate explanation of the data relative to the saturated model (Cheng and Washington, 2005). The G^2 statistic is given as:

$$G^2 = 2\sum_{i=1}^{n} O_i LN\left(\frac{O_i}{Y_i}\right) \tag{4.23}$$

A large G^2 indicates that the model fits poorly as compared to the saturated model.

Criterion 9: Chi-squared Residual Sum of Squares (RSS)

RSS is defined as:

$$RSS = \sum_{i=1}^{n} \frac{(O_i - Y_i)^2}{Y_i} \tag{4.24}$$

Under MAD and MSPE, the larger counties are expected to subject to more penalties due to greater counts and residues. RSS tends to remove such bias by calculating the squared residual relative to the estimated number of crashes. A particular model is considered more reliable if smaller RSS value is observed.

Criterion 10: Total Rank Difference (TRD)

Finally, rather than use the magnitude of residue value, TRD accounts for the rank deviations based on the observed and estimated crash counts (Cheng et al., 2017). The rank difference is calculated by using the following equation.

$$TRD = \sum_{i=1}^{n} |R(i_o) - R(i_Y)| \tag{4.25}$$

Where $R(i_o)$ is the observed data rank at county i and $R(i_Y)$ is the rank based on estimated crash counts for the same time period. A model is considered superior if smaller TRD value is revealed, which signifies that the specific model assigns rankings close to the observed crash counts.

DATA PREPARATION

As previously stated, for ease of comparison of various temporal treatments and types of models, only the number of fatal and injury vehicle-involved collisions and a main exposure-related yearly-varying factor of county safety performance, that is, Daily Vehicle Miles Travel (DVMT), were collected for analysis. The first one was obtained from the website of TIMS, which provides tools for accessing and mapping collision data from the California Statewide Integrated Traffic Records System (SWITRS). The information of DVMT was collected from Highway Performance Monitoring System (HPMS). Eight-year of data were used for both variables which span from 2006 to 2013. In addition, for implementation of distance-based CAR priors for spatial correlations, a matrix containing distances among centroids of all counties in California were developed based on the information provided by SCAG (Southern California of Association of Governments). Since the State has 58 counties, so the distance matrix has the size of 58x57 with a mean of 273 miles. The descriptive statics of variables are illustrated in Table 4.1.

TABLE 4.1 Descriptive Statistics of Variables for Various Counties in California

Variables	Description	Year	Min	Max	Median	Mean	S.D.
Collision	Total Annual Fatal and Injury Vehicle-involved Collisions	2006	23	48,107	758	2,791	6,767
		2007	10	46,558	699	2,672	6,515
		2008	15	41,794	631	2,389	5,841
		2009	20	40,197	611	2,290	5,612
		2010	18	39,560	538	2,249	5,531
		2011	14	38,933	576	2,185	5,430
		2012	16	38,477	560	2,171	5,389
		2013	21	38,855	544	2,141	5,437
DVMT	Daily Vehicle Miles Travelled (in thousand)	2006	180	217,444	5,092	15,577	32,178
		2007	162	218,027	5,146	15,608	32,261
		2008	168	214,971	5,005	15,387	31,894
		2009	171	214,237	4,837	15,318	31,744
		2010	169	211,877	5,449	15,483	31,421
		2011	165	214,458	4,762	15,353	31,871

		2012	167	214,482	4,789	15,438	31,927
		2013	165	215,818	4,791	15,549	32,162
Distance	Distance among Centroids of Counties (unit: miles)	N/A	25	962	227	273	176

Note: S.D. represents standard deviation; min refers to minimum; max refers to maximum; N/A means Not Applicable

RESULTS

The three sets and nine groups of models were implemented in freeware WinBUGS package which employs an MCMC algorithm for estimation of parameters. For the model calibration, two Markov chains were utilized to visually inspect trace plots of posterior estimates for convergence, where a wide range of burn-in iterations (5,000 to 15,000) and total interactions (20,000 to 25,000) was used in the calculation of posterior estimates due to the varying complexity of different models. The convergence was further checked by using the Gelman-Rubin convergence statistic (*51*) and ensuring the sample Monte Carlo errors to be less than 5% of the associated standard deviation. It is important to reiterate that all models did not incorporate any covariates as the primary focus was on the comparison of temporal treatment, with and without spatial correlations, though DVMT was used to represent traffic exposure and offset the crash frequency.

Modeling Results

The posterior mean and standard deviation for the different model coefficient estimates are illustrated in Table 4.2. The intercept term (β_0) was observed to be statistically significant for all the groups, and the impact of inclusion of spatial correlations may be noted on closer scrutiny with the standard deviation for β_0 consistently decreasing across all groups for Type 2 and Type 3, compared with Type 1. The proportion of variability due to spatial clustering was also estimated to further study the impact of inclusion of spatial correlation. The fraction α signified the heterogeneity explained by the spatial random effects and was observed to be statistically significant for all spatiotemporal models which reflected the presence of spatial clustering of crashes. This also bolsters the loss of precision noticed for intercept of Type 1 across all groups. Apart from the spatial correlation, the time dependency appeared to be statistically significant as well by observation of the different temporal treatment model terms such as β_{t1}, β_{t2}, φ_1, φ_2, θ_1, θ_2 , etc. It illustrated the strong serial correlation among crashes of successive years and the tendency of crashes to cluster both spatially and temporally as reflected by the space-time interaction of Type 3. The importance of accounting for temporal changes was also evident from the time-varying intercept models (Group 4) which demonstrated the different magnitude of intercepts for eight years. The parameter values in Table 4.2 can also be used to predict the annual crash counts of various counties. For example, the eight-year DVMT (in thousand) for Amador County are (1193.12, 1223.70, 1204.87, 1201.17, 1067.05, 1058.22, 1027.92, 1035.65), and the associated observed crash counts are (265, 233, 203, 204, 212, 180, 161, 142). Under the Type 1, Group 1 Model, β_0 =-2.068, the predicted crash counts ($DVMT * e^{\beta_0}$) would be (150.9, 154.8, 152.4, 152.0, 135.0, 133.9, 130.0, 131.0). However, if the estimated independent-over-time random effect, ε_{it}= (0.5384, 0.3885, 0.2696, 0.2767, 0.4273, 0.2772, 0.1978, 0.0716), is also included for calculation, the predicted values ($DVMT * e^{(\beta_0+\varepsilon_{it})}$) would be (258.5, 228.7, 200.1, 200.9, 207.4, 177.1, 158.9, 141.2), which are much closer to the observed counts. Such phenomenon indicates the importance of random effect for crash count estimation.

TABLE 4.2. Parameter Estimates and associated Standard Deviation for Various Models

Parameters	Type 1	Type 2	Type 3
Group 1: Independent-over-time random effect (Base)			
β_0	-2.068 (0.0140)	-2.072 (0.0068)	-2.114 (0.0119)
α		0.678 (0.0082)	0.654 (0.0095)
Group 2: Linear Trend			
β_0	-1.893 (0.0304)	-1.902 (0.0111)	-1.901 (0.0109)
β_{t1}	-0.0379 (0.0061)	-0.0379 (0.0022)	-0.0382 (0.0020)
α		0.754 (0.0075)	0.754 (0.0076)
Group 3: Quadratic Trend			
β_0	-1.871 (0.0368)	-1.825 (0.0207)	-1.838 (0.0186)
β_{t1}	-0.053 (0.0167)	-0.084 (0.0107)	-0.076 (0.0097)
β_{t2}	0.002 (0.0018)	0.005 (0.0012)	0.004 (0.0011)
α		0.759 (0.0075)	0.781 (0.0106)
Group 4: Time-varying Intercept			
β_0^1	-1.907 (0.0383)	-1.914 (0.0132)	-1.913 (0.0127)
β_0^2	-1.941 (0.0386)	-1.950 (0.0129)	-1.950 (0.0124)
β_0^3	-1.941 (0.0386)	-1.950 (0.0129)	-1.950 (0.0124)
β_0^4	-2.053 (0.0384)	-2.069 (0.0137)	-2.070 (0.0132)
β_0^5	-2.115 (0.0410)	-2.153 (0.0144)	-2.152 (0.0132)
β_0^6	-2.113 (0.0374)	-2.123 (0.0134)	-2.123 (0.0131)
β_0^7	-2.134 (0.0384)	-2.143 (0.0143)	-2.144 (0.0128)
β_0^8	-2.169 (0.0422)	-2.187 (0.0137)	-2.189 (0.0129)

α		0.763 (0.0075)	0.787 (0.0109)
Group 5: AR-1			
β_0	-2.028 (0.0147)	-2.057 (0.0064)	-2.052 (0.0132)
φ_1	-0.401 (0.0221)	-0.491 (0.0384)	-0.456 (0.0392)
α		0.677 (0.0078)	0.618 (0.0161)
Group 6: AR-2			
β_0	-2.093 (0.0203)	-2.051 (0.0118)	-2.075 (0.0116)
φ_1	1.013 (0.0519)	1.103 (0.0438)	1.191 (0.0239)
φ_2	-0.527 (0.0242)	-0.418 (0.0438)	-0.540 (0.0294)
α		0.631 (0.0438)	0.621 (0.0219)
Group 7: MA-1			
β_0	-2.065 (0.0171)	-2.071 (0.0089)	-2.122 (0.0146)
θ_1	0.924 (0.0530)	0.700 (0.0915)	0.849 (0.0765)
α		0.721 (0.0122)	0.728 (0.0119)
Group 8: MA-2			
β_0	-2.049 (0.0182)	-2.067 (0.0106)	-2.135 (0.0189)
θ_1	0.976 (0.0219)	0.830 (0.0787)	0.955 (0.0331)
θ_2	0.627 (0.0449)	0.428 (0.0700)	0.514 (0.0394)
α		0.728 (0.0126)	0.753 (0.0126)
Group 9: Time Adjacency			
β_0	-2.068 (0.0160)	-2.073 (0.0093)	-2.084 (0.0062)
α		0.762 (0.0075)	0.786 (0.0101)

Notes: 1. the shaded cell indicates non-significant coefficient at 95% confidence level.

2. The standard deviations are shown in the parentheses.

3. Refer to Methodology Section for details of various parameters and groups/types of models.

Evaluation Results

An array of evaluation criteria was employed to assess the performance of models from different perspectives. Generally speaking, $\overline{D}$, LPML, MAD, MSPE, R_P^2, G^2, RSS, and TRD are akin to each other as they are centered on the predictive accuracies of models, while P_d is mainly focused on the model complexity. DIC is the penalized criterion which considers the combined effects of both modeling prediction and complexity. Table 4.3 demonstrated the evaluation results of each of the 27 individual models. For ease of description, G#T# will be used hereafter to denote the model of certain group and type. For example, G3T2 represents model of Group 3 and Type 2. Inspection of Table 4.3 reveals some interesting findings: (1) In terms of the prediction-pertinent criteria, G6T1 appeared to have the worst performance while G5T3 tended to be ranked first in most cases. (2) In terms of model complexity (P_d), G6T1 performed the best while G1T3 was placed last. (3) In terms of the penalized criterion (DIC), G6T2, G6T3 and G9T2 were observed with the superiority while G6T1 had the disadvantage compared with others. Such phenomena indicated that (a) the various measures for prediction accuracies tended to be highly correlated with each other; (b) the models' predication capabilities may be inconsistent with their efficiency in reducing the effective number of parameters; (c) under the DIC which combines both effects, the models with spatial random effects (either independent or interactive ones) seemed to have better performance than the ones with temporal random effects only under most conditions.

For easier understanding the impact of temporal treatment and spatial correlations, Table 4.4 demonstrates the aggregated model performance based on two perspectives: across different temporal treatments and the three model types. In the case of aggregation by types, the proposed Group 9 models recorded the lowest DIC value (4753.36) while Group 8 models exhibited the worst fit. The base models may be expected to demonstrate the overall best predictions with the best rankings in most prediction-related measures, but the model complexity turned out to be the governing factor, which was the highest, resulting in overall inferior performance. This finding may indicate another perspective to quantify the impact of the inclusion of temporal treatment as compared to the base group in the sense that the other model groups exhibited considerably lower complexity. The best performance of Group 9 (albeit with a low sample of 8 years) in terms of DIC suggested that this kind of temporal treatment is worth further exploration in the safety field in the future. The possible explanation for best predictions associated with the base models might be due to the myriad of influential factors of the safety (e.g., policy, population, weather, etc.) in the large spatial unit, county, which makes it difficult to predict the county-level crashes by using a universal temporal trend or correlation.

In the case of aggregation by temporal treatments, Type 2 models were observed to have best overall fit as reflected by the lowest DIC (4759.82), followed by Type 3 while Type 1 recorded the highest DIC. Once again, the governing factor was the model complexity which was significantly decreased by the inclusion of spatial correlation structures. These findings suggest that inclusion of space and time at the same time provides substantial efficiency in reducing the number of effective parameters, which also indicates the tendency of crash data to cluster spatially as well as temporally. It is noteworthy to mention that among the spatiotemporal models (Types 2 and 3), the space-time interaction tends to increase the complexity due to the addition of more effective parameters which eventually corresponds to increase in DIC. With respect to the prediction capabilities, Type 1seemed to enjoy the best performance in most of the pertinent criteria (except MSPE). It follows that it might be difficult to predict the crashes of the large spatial unit by using a certain type of spatial correlation (distance-based in our case) applicable to all counties due to the enormous impacts from a very large number of factors.

In summary, under the predictive evaluation criteria, both aggregated results indicated the best performance associated with the simpler modeling formulation (without spatial and/or temporal correlations). It is possible that the counties, compared to smaller spatial units, are subject to much more influential safety factors which render the difficulty of predicting county-level crashes via certain types of temporal or spatial random effects. However, the results of individual models illustrated that G5T3 was superior to others in 5 out of 7 prediction cases. Such phenomenon suggests that it is still worth utilizing the certain serial or spatial random effects for better prediction, even though care should be taken for selection of the temporal treatment as different treatments lead to a large variability in terms of prediction. In terms of P_d and DIC, the spatio-temporal models (earth Type 2 or 3) exhibited the best performances in most cases based on both aggregated and individual modeling results.

Finally, some of the previous notable trends within the results motivated the authors to conduct a correlation analysis of the ten criteria. It is expected that such correlation result would shed some light on the modeling selection for safety professionals especially when the limited safety resources or man hours deter the comprehensive evaluation of models. As shown in Table 4.5, statistically significant correlations were observed among a set of evaluation criteria including $\overline{D}$, LPML, MAD, MSPE, R_P^2, G^2, RSS and TRD. Such finding is illuminating given that these measures check predictive capabilities using different ways based on cross validation, ranking difference, and deviation relative to the estimated crash counts, and so on. P_d seemed to have strong reverse correlations with all these prediction-related criteria. As a combination of P_d and $\overline{D}$, DIC did not demonstrate strong correlations with all criteria for predictive capabilities, except for MSPE and TRD. Given that DIC and P_d have a statistically significant positive correlation (0.38), it appears that P_d has a heavier weight for the determination of the DIC for the county situation, as also observed in the previous discussion of evaluation results. To summarize, these trends suggest that model complexity (P_d) may have a strong influence on the overall goodness-of-fit (DIC), and $\overline{D}$ could be regarded as an excellent indication for the crash prediction capability because it is consistently correlated with all other prediction criteria (LPML, MAD, MSPE, R_P^2, G^2, RSS & TRD).

TABLE 4.3. Evaluation Results of Various Models under Different Criteria

Group	Type	$\overline{D}$	P_d	DIC	LPML	MAD	MSPE	R_p^2	G^2	RSS	TRD
1	1	4328	443.5	4771	-2090.0	2.611	**10.814**	0.994	27.2	37.1	382
	2	4342.3	426.4	4768	-2094.4	4.755	42.192	0.989	35.4	68.5	571
	3	4328.5	465.3	4793	-2088.6	2.284	18.359	0.998	22.2	12.3	216
2	1	4331.3	442.2	4773	-2091.1	2.684	11.780	0.993	11.3	41.7	399
	2	4360.1	392.4	4752	-2102.2	5.974	76.870	0.982	102	116.7	862

	3	4369.6	393.4	4763	-2106.2	5.992	76.448	0.982	111	116.2	862
3	1	4331.5	442.3	4773	-2091.2	2.685	11.763	0.993	26.8	41.8	402
	2	4358.9	389.3	4748	-2102.0	6.119	80.481	0.982	99.8	118.5	864
	3	4373.0	390.1	4763	-2106.0	6.202	87.479	0.98	182.1	130.5	918
4	1	4331.2	442.2	4773	-2091.0	2.697	11.855	0.993	61.1	41.1	401
	2	4358.3	388.6	4746	-2101.9	6.263	83.228	0.981	112.7	119.0	871
	3	4370.4	387.7	4758	-2105.4	6.327	90.072	0.98	70.9	131.3	908
5	1	4350.7	450.3	4801	-2092.0	2.685	14.752	0.994	9.3	39.5	383
	2	4355.1	420.4	4775	-2094.2	5.464	57.852	0.989	102	72.6	611
	3	**4322.8**	451.2	4774	**-2084.2**	**1.858**	15.378	**0.998**	53.1	**6.9**	**163**
6	1	4491.1	**332.7**	4823	-2142.7	16.41	2727.5	0.956	274.1	286.6	1097
	2	4397.9	337.3	**4735**	-2110.2	9.429	215.03	0.971	131.6	188.4	1079
	3	4388.4	346.9	**4735**	-2106.8	8.865	193.64	0.974	155.9	167.8	1033
7	1	4346.1	413.4	4759	-2091.7	3.897	29.842	0.99	26.2	65.3	586
	2	4384.1	394.3	4778	-2103.2	6.242	82.645	0.982	117.8	117.9	821
	3	4365.8	429.9	4795	-2095.1	4.370	44.389	0.991	55.7	58.2	577
8	1	4361.6	398.4	4760	-2092.4	4.342	41.117	0.987	**-8.4**	80.6	676
	2	4405.6	386.7	4792	-2105.4	6.405	93.739	0.979	108	132.3	861
	3	4377.7	434.5	4812	-2094.3	3.967	39.525	0.991	27	58.7	563
9	1	4330.7	442.3	4773	-2090.9	2.690	11.878	0.993	96.3	40.6	394
	2	4354.1	385.7	**4739**	-2100.9	6.161	81.223	0.982	62	118.1	867
	3	4364.8	382.2	4747	-2104.1	6.277	88.442	0.98	130.1	130	902

Notes: 1. The shaded cells indicate the worst performance of models under different evaluation criteria.

2. The bold fonts indicate the best performance of models under different evaluation criteria.

3. There are three models sharing the best performance in terms of DIC since the associated difference in DIC is less than 5.

4. Refer to Table 4.2 for Group names.

TABLE 4.4. Aggregated Evaluation Results under Different Criteria

	$\bar{D}$	P_d	DIC	LPML	MAD	MSPE	R_p^2	G^2	RSS	TRD
Models	**Aggregation by Nine Groups of Temporal Treatments**									

1	**4355.8**	423.0	4778	**-2097.04**	**4.52**	319.04	**0.989**	**58.259**	**74.96**	**524**
2	4368.5	**391.2**	**4759**	-2101.64	6.31	90.36	0.982	96.854	116.94	823
3	4362.3	409.0	4771	-2099.01	5.12	**72.63**	0.986	89.835	90.27	682
Groups	**Aggregation by Three Types of Models**									
1	**4332.9**	445.1	4778	-2091.0	**3.21**	**23.78**	**0.994**	**28.325**	**39.35**	389
2	4353.7	409.3	4763	-2099.8	4.88	55.03	0.986	74.817	91.59	707
3	4354.5	407.2	4761	-2099.8	5	59.90	0.985	102.935	97.01	728
4	4353.3	406.2	4759	-2099.4	5.09	61.71	0.985	81.618	97.19	726
5	4342.8	440.6	4783	**-2090.1**	3.33	29.32	0.994	54.858	39.7	**385**
6	4425.8	**338.9**	4764	-2119.9	11.57	1045.41	0.967	187.263	214.3	1069
7	4365.4	412.5	4777	-2096.6	4.83	52.29	0.988	66.625	80.49	661
8	4381.6	406.5	4788	-2097.4	4.90	58.12	0.986	42.231	90.59	700
9	4349.8	403.4	**4753**	-2098.6	5.04	60.51	0.985	96.175	96.31	721

Notes: 1. The shaded cells indicate the worst performance of models under different evaluation criteria.

2. The bold fonts indicate the best performance of models under different evaluation criteria.

TABLE 4.5. The Estimate of the Correlation Coefficients among Various Evaluation Criteria

	$\bar{D}$	P_d	DIC	LPML	MAD	MSPE	R_p^2	G^2	RSS	TRD
$\bar{D}$	1.00									
P_d	-0.79	1.00								
DIC	*0.26*	0.38	1.00							
LPML	-0.93	0.85	*-0.08*	1.00						
MAD	0.93	-0.90	*-0.01*	-0.97	1.00					
MSPE	0.80	-0.50	0.42	-0.83	0.81	1.00				
R_p^2	-0.90	0.96	*0.15*	0.96	-0.98	-0.70	1.00			
G^2	0.77	-0.76	*-0.04*	-0.87	0.85	0.67	-0.85	1.00		
RSS	0.90	-0.96	*-0.15*	-0.96	0.98	0.70	-1.00	0.85	1.00	
TRD	0.75	-0.96	-0.38	-0.82	0.84	0.39	-0.92	0.73	0.92	1.00

Notes: 1. The text in italics signifies non-significance at 95% level.

CHAPTER 5: SPATIOTEMPORAL MODELS WITH MIXTURE COMPONENTS FOR SPACE-TIME INTERACTIONS

INTRODUCTION

The literature review illustrated notable benefits at various fronts associated with inclusion of spatial and temporal correlations and their interactions. However, in comparison with other types of models, a very limited body of research dedicated to spatiotemporal models exists in the field of traffic safety. Moreover, most of the current limited spatiotemporal models in the field assume a linear temporal trend and linear space-time interaction which may be seen as a restrictive assumption (Lawson and Clark, 2002; Lawson et al., 2003). For example, the temporal random effects may take on nonlinear shape or have autocorrelation with previous crash counts. In addition, the temporal trend might have a non-linear change across the spatial units. To add to the current literature with more spatiotemporal models, this chapter aimed to develop four alternative spatial-temporal models which employ different temporal treatments with the varying complexity of random effects: (I) linear time trend; (II) quadratic time trend; (III) Autoregressive-1 (AR-1); and (IV) time adjacency. Furthermore, instead of using linear space-time interaction, the authors borrowed a flexible two-component-mixture interaction from one previous disease-mapping study (Abellan et al., 2008). Such mixture can easily capture the space-time trends that depart from the predictable patterns of overall temporal and spatial risk surface as it allows the smoothness as well as discontinuities in the space-time variations within the roadway entities. The interested readers can be referred to this study for more details of the mixture components. The study results demonstrated a number of benefits associated with the proposed mixture model. However, the performance of the mixture model in traffic safety area is unknown and is therefore worth studying. In addition, in order for a comprehensive comparison of the predictive accuracy of the four models, five evaluation criteria were utilized which include log pseudo marginal likelihoods (LPML), mean square predictive error (MSPE), mean absolute deviation (MAD), residual sum of squares (RSS), and total rank difference (TRD). LPML assesses the predictive capability using cross-validation while the rest utilize the same dataset for model development. Three different approaches (model replicated, model predicted, and Bayesian estimated) were used to generate the crash dataset for evaluation as an effort to quantify the prediction performance and understand the benefits of concerned models from different perspectives. Finally, a residual analysis was also performed to observe the impact of the inclusion of random effects on the prediction accuracy.

METHODOLOGY

This section is divided into two main subsections: Model Development and Evaluation Criteria. The former subsection first describes the formulation of Poisson lognormal model which was employed for the estimation of crash counts; then introduces the Base model specification which is usually employed in conventional safety studies focused on space-time interaction and discusses the advantage of Mixture model over the Base; and finally, four different types of temporal treatments are introduced which represent the temporal specifications employed in different studies. In the second subsection (Evaluation Criteria), different criteria are discussed which were used for assessment of model goodness-of-fit and comparison of models' predictive accuracy based on the cross-validation for out-of-sample and in-sample crash data.

Model Specification

The Full Bayesian (FB) framework was employed to estimate the crash frequency of vehicles at the segment level. To account for the unobserved heterogeneity from different perspectives, the FB approach is preferable due to its flexibility and effectiveness to incorporate complex correlations with a hierarchical structure of data. FB approach provides a posterior distribution of parameters from Markov-chain Monte Carlo (MCMC) simulation which samples the variables as random, unlike the point estimates generated by the traditional approach of maximum likelihood estimation. This approach has been widely used for crash prediction models due to the multilevel and correlated nature of data (refer Lord and Mannering, 2010 for detailed review). Under the FB approach, to account for the overdispersion usually associated with crash data, the basic model formulation was usually adopted from the traditional approach of generalized linear models with a random error term to account for the overdispersion. Instead of using the Negative Binomial model, the present study chose Poisson-Lognormal model given its capability of better handling of low sample mean and small sample size due to the heavier tails associated with Lognormal distribution (Lord and Miranda-Moreno, 2008). This model assumes that crash count (y_i) at a given segment i, obeys Poisson distribution, while the corresponding observation specific error term ε_i follows a normal distribution:

$$y_i|\lambda_i \sim Poisson\ (\lambda_i) \tag{5.1}$$

Where λ_i is the expected Poisson count for location i, and can be expressed as follows:

$$\ln(\lambda_i) = \beta_0 + X_i'\beta + \varepsilon_i \tag{5.2}$$

Where β_0 is the intercept, X' is the matrix of risk factors, β is the vector of model parameters, ε_i is the independent random effect which captures the extra-Poisson heterogeneity among locations. ε_i can be assumed with the following noninformative normal priors:

$$\varepsilon_i \sim Normal\ (0, \tau^2\ ) \tag{5.3}$$

Where τ^2 is the variance of the normal distribution for ε_i. The inverse of τ^2 is called precision and it can be modeled using the following gamma prior with prior mean equal to one and its prior variance large (equal to one thousand), representing high uncertainty or prior ignorance:

$$\tau^{2^{-1}} \sim Gamma(0.001, 0.001) \tag{5.4}$$

The model presented in Equation 2 serves as the basic formulation for crash prediction in most safety studies. The current study also employs this model as the basis for development of spatiotemporal models.

Two Groups of Models

This section first introduces the Base model specification which serves to represent the assignment of spatial, temporal, and space-time interactions by conventional traffic safety studies. The limitation associated with such specification is discussed and then the mixture model specification is introduced.

Base

As previously mentioned, the spatial random effects were incorporated over the model represented in Equation 2. The subsequent model which accounts for the spatial correlation among neighboring segments is obtained with the following formulation:

$$\ln(\lambda_i) = \beta_0 + X_i'\beta + \varepsilon_i + u_i \tag{5.5}$$

Where u_i is the spatially structured random effect which follows the CAR (conditional autoregressive) formulation to incorporate the spatial correlation among crashes occurring at neighboring segments.

$$u_i \sim N\left(\frac{\overline{u_i}, 1}{\tau_i}\right) \qquad \overline{u_i} = \frac{\sum_{i \neq k} u_k w_{ik}}{\sum_{i \neq k} w_{ik}} \qquad \tau_i = \frac{\tau_c}{\sum_{i \neq k} w_{ik}} \tag{5.6}$$

Where w_{ik} is the proximity weight matrix and τ_c is the precision parameter in the CAR prior. As evident from the above equation, estimation of the risk in any site is conditional on risks in neighboring locations. Subscripts i and k refer to a segment and its neighbor, respectively, and $k\ \varepsilon\ N_i$, where N_i represents the set of neighbors of segment i. Besides the identification of neighbors, the assigned weights also affect the risk estimation. In the past studies (Aguero-Valverde and Jovanis, 2010; Xu and Huang, 2015), weight structures including various adjacency-based, distance-based models, and semi-parametric geographically weighted, and so on, have been explored for incorporating the spatial correlations at different levels of roadway entities. The current study employs the adjacency-based structure to explore the spatial correlation, which is commonly used for a micro-level entity of segments. The weights for the adjacency matrix are given by (w_{ik}) = 1 if segments i, k are adjacent, and 0 otherwise. It is noteworthy that this adjacency-based weight matrix gives a dichotomous output for weight with only two responses, zero for non-neighbors and one for neighbors, hence the proximity of immediate neighbors establishes a spatial dependency while the impact of more distant neighbors is assumed to be negligible.

If temporal treatment also needs to be considered, then Equation 5 can be easily modified and extended to account for serial correlations and associated space-time interaction. The resultant formulation can be expressed in Equation 7.

$$\ln(\lambda_{it}) = \beta_0 + X_i'\beta + f(t) + u_i + v_{it} \tag{5.7}$$

Where $f(t)$ is the temporal treatment, and v_{it} is the space-time interaction parameter. v_{it} can take different forms such as linear (Aguero-Valverde and Jovanis, 2006) and time-varying spatial correlation (Lawson et al., 2003). However, the existing space-time interaction models have a limitation in the essence that they assume all the sites under focus experience a global space-time pattern where the error (i.e. the part not accounted by isolated spatial and temporal terms) falls into one distribution common to all sites. This may be interpreted as the inefficiency to distinguish between the stable and unstable underlying risk pattern, where the stable (or predictable) patterns are represented by the overall time trend ($f(t)$) and spatial crash risk (u_i), and unstable (or unpredictable) patterns are the atypical departures from the stable pattern. This study introduces the mixture component for accommodating the large fluctuations of the space-time interaction which allows the identification of unstable locations which do not conform to the global crash risk pattern of space and time variability.

Mixture

The mixture component for space-time interaction was introduced by Abellan et al., (2008) to "uncover the full space-time profile of risks" for an epidemiologic study. The concerned study proposed a 2-component mixture which combines variabilities of different levels to represent the space-time interaction. The accommodation of different levels of variabilities is required to distinguish the stable and unstable pattern. Specifically, v_{it} from Equation 7 takes the following form:

$$v_{it} \sim p\text{Normal}(0, \tau_1^2) + (1-p)\text{Normal}(0, \tau_2^2) \tag{5.8}$$

The prior for p is uniform on [0, 1] through a Dirichlet distribution. This varied prior distribution renders the flexibility to accommodate discreet and continuous space-time variations in crash data, where the discreet variations refer to the outstanding or outlier sites in the plan of space-time interaction which deviate from the continuous global pattern of space-time variations. The first component captures the small residual noise which is generally adequately captured by the overall time trend and spatial component. The second component serves the purpose of capturing the "true" departure from global pattern by accommodating the large fluctuations which indicate the higher risk relative to other sites. Half-normal hyper-prior distributions for the standard deviations τ_k, (k = 1, 2), are specified which reflect that τ_1 (corresponding to "stable" risk pattern) has to be small to effect shrinkage so as to capture the noise, whereas the prior for τ_2 (corresponding to "unstable" risk pattern) allows a large range of values for this parameter to capture the large fluctuations. The different range of variance is shown in the following equations:

$$\tau_1 \sim Normal\ (0, 0.01).\, \mathrm{I}_{(0,+\infty)} \tag{5.9}$$

$$\tau_2 \sim Normal\ (0, 0.01).\, \mathrm{I}_{(0,+\infty)} + Normal\ (0, 100).\, \mathrm{I}_{(0,+\infty)} \tag{5.10}$$

Based on both simulation data and real case study, the authors revealed a number of benefits of the mixture model in terms of interpretation and potential for detection of localized excess for diseases. As a result, the present study borrowed such structure for the space-time interaction and tested its capability in addressing the prediction accuracy for the traffic crash data. Additionally, another characteristic of the current research lies in the different ways of handling temporal treatment (i.e., $f(t)$ in Equation 7).

Four Types of Temporal Treatments

The following subsections present the corresponding model formulations in order of complexity, from the linear time trend to more sophisticated ones. These models were developed using both the Base and Mixture specifications. It is noteworthy that all models vary on the basis of the assignment of different temporal treatments over the Poisson lognormal models which incorporate the spatial correlation and space-time interaction.

Model 1: Linear time trend

In this model, a linear trend is introduced where time is regarded as a potential influential covariate and the model estimates the coefficient for the trend. Equation 7 assumes the following form:

$$\ln(\lambda_{it}) = \beta_0 + X_i'\beta + u_i + v_{it} + \beta_t \ * T \tag{5.11}$$

Where β_t is the scalar parameter for linear yearly trend T (T=1 to 6 in the present study). It is expected that the passage of time will impact the crash risk on the segments which would be captured by the model.

Model 2: Quadratic time trend

In this model, a non-linear impact of time is considered. Equation 7 takes the following form:

$$\ln(\lambda_{it}) = \beta_0 + X_i'\beta + u_i + v_{it} + \beta_t \ * T + \beta_{t2} \ * T^2 \tag{5.12}$$

Where β_{t2} is the coefficient for the quadratic trend term. It is expected that this model will allow more subtle approach compared to the linear trend as it allows flexibility to fit the crash data by virtue of its additional quadratic trend.

Model 3: Autoregressive-1 (AR-1)

This model accounts for the autoregressive safety effect by specifying the distribution of ε_{it} as a lag-1 dependence in errors, where lag-1 means that the time is varying yearly. It incorporates the weighted sum of the past one year of values together with a random term. It is important to note that, in addition to AR-1, there are a large number of models involving autoregressive error process (Miaou and Song, 2005), such as higher order AR model AR(p), autoregressive–moving-average (ARMA), Integer valued autoregressive (INAR), Autoregressive Conditional Heteroscedastic (ARCH), and Generalized Autoregressive Conditional Heteroscedastic (GARCH), and so on. The authors opted for the popular AR-1 specification to represent such models based on the assumptions of stationarity restriction. Under this model, Equation 7 assumes the following form:

$$\ln(\lambda_{it}) = \beta_0 + X_i'\beta + \varepsilon_{it} + u_i + v_{it} \tag{5.13}$$

The weighted sum is fixed and the random terms change at every time step following the same distribution, which means this model is homoscedastic. The distributions are given by:

$$\varepsilon_{it} \sim normal\left(0, \frac{\sigma_{it}^2}{(1-\gamma^2)}\right) \tag{5.14}$$

$$\varepsilon_{it} \sim normal(\gamma\varepsilon_{i,t-1}, \sigma_{it}^2) \quad \text{for t >1} \tag{5.15}$$

Where γ is the autocorrelation coefficient with the range of $0 < \gamma < 1$. This model addresses the potential correlation between successive time periods and is expected to deliver a more precise estimation of the model parameters.

Model 4: Time Adjacency

In this model, a CAR formulation is employed for the temporal aspect of the data, similar to the spatial aspect discussed previously. In this case, Equation 7 assumes the following form:

$$\ln(\lambda_{it}) = \beta_0 + X_i'\beta + u_i + v_{it} + \alpha_t \tag{5.16}$$

Where α_t is the temporal correlation term which follows the CAR formulation given in Equation 6. The six year time period of this study assumes strong influence of the adjacent years on crash risk, similar to the dependence of adjacent segments in case of spatial correlation. Compared to the AR-1 model which captures the impact of one year prior to the concerned time, this model also accounts for the potential impact of the subsequent year. The similar formation of such temporal treatment can be found in a previous study (Abellan et al., 2008).

Evaluation Criteria

This study assessed the model performance by employing different evaluation criteria which were found in the past research. The authors intend to evaluate the model performance from different perspectives such as goodness-of-fit (while accounting for complexity), and in-sample and out-of-sample cross validation predictive accuracy for crash counts. The brief background and analytics of each criterion are provided as follows.

DIC and Its Components

The Deviance Information Criterion (DIC) (Spiegelhalter et al., 2003) served as the measure for assessment of goodness-of-fit for the models. It is a penalized fit criterion which can be calculated from the samples generated by an MCMC simulation. Usually, the estimation of DIC can be conducted through the following expression.

$$\begin{cases} DIC = \bar{D} + P_D \\ D(\theta) = -2 \cdot \log(p(y|\theta)) \\ \bar{D} = E^{\theta}[D(\theta)],\ \ P_D = \bar{D} - D(\bar{\theta}) \end{cases} \tag{5.17}$$

Where, y are the data; θ are the unknown parameters of the model; $p(y|\theta)$ is the likelihood function; $\bar{\theta}$ is the expectation of θ; $\bar{D}$ is the posterior mean deviance, or the expectation of the deviance, $D(\theta)$; $\bar{D}$ measures how well the model fits the data; the smaller the $\bar{D}$, the better the fit. P_D represents the effective number of parameters. Usually, smaller values of DIC are preferred. Some general guidelines are also suggested by Lunn et al. (2012): models with DIC values within 5 of the 'best' model are comparable and also strongly supported, values within 5 and 10, weakly supported, and models with a DIC greater than 10 are substantially inferior.

Comparison of models based on out-of-sample cross-validatory predictive density

This study employed the conditional predictive ordinate (CPO) (Gelfand, 1996) for cross validation based on predictive densities. To prevent potential bias due to random selection of data division into subsets (one subset for development and another for checking), the current study calculated the CPO by implementing a CV-1 (leave-one-out) distribution which removed the selection bias by employing a continuous approach of selecting all data points, except one, for model development and the left out data point to verify the prediction accuracy of the calibrated model. Such implementation also has benefits over traditional modeling evaluations which are prone to overestimation due to double usage of data (once during model development and then again for model checking). Under the MCMC framework, the estimate of CPO for each observation i can be calculated as:

$$CPO = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{f(Y_i|\beta^{(t)})}\right)^{-1} \tag{5.18}$$

Where Yi is the ith observation (i = 1, 2, 3, . . ., n) for all segments and β is the vector of estimated model parameters. This harmonic mean of density (CPO) may be extended to calculate the goodness-of-fit of models by computing the product of CPOs over all observations, which is known as the pseudo marginal likelihood. For computational convenience, the log pseudo marginal likelihoods (LPML) is calculated:

$$LPML = \sum_{i=1}^{n} \log(CPO_i) \tag{5.19}$$

Evaluation criteria for predictive accuracy based on the same dataset

In this study, to obtain a more comprehensive comparison of four alternate models, the traditional evaluation criteria such as MSPE, MAD, RSS, and TRD were also calculated to assess the crash prediction accuracy of the models while treating the observed crash counts as the reference. These criteria are expected to quantify the variability in the prediction performance of models, which will help understand the benefits of concerned models from different perspectives. The evaluation was conducted by employing three different approaches, essentially governed by the selection of datasets for comparison with the observed crash counts. In the first case, the Bayesian estimated crash frequency was utilized which was obtained from Equations 11, 12, 13, and 16 for corresponding models. The difference of Bayesian estimated and observed crash counts is referred as Bayesian residuals (Lawson et al., 2003). In the second case, the model predicted crash frequency was compared with the observed counts. To obtain the model predicted counts, Equation 7 may be modified as follows:

$$\ln(PD_{it}) = \beta_0 + X_i'\beta \tag{5.20}$$

Where PD_{it} is regarded as the model predicted crash count. The comparison of model predicted and observed crash counts is referred as normal residuals. This predicted count differs from the Bayesian estimated as no random effect terms are incorporated, hence the comparison of Bayesian and normal residuals is expected to quantify the variability in predictive performance of models without the impact of spatial and temporal correlation structures. In the third case, the model replicated crash counts are utilized for comparison with observed counts. The replicated counts were generated from the posterior predictive distribution by introducing a simple step in MCMC sampler as follows:

$$y(rep)_i|\lambda(rep)_i \sim Poisson\ (\lambda_i) \tag{5.21}$$

Where *λ(rep)* refers to the expected crash counts estimated based on the assumption that adopted model truly fits the observed crash data. The closeness of *λ(rep)* and λ_i signifies that the model was able to precisely replicate the crash counts based on the observed ones and hence may be employed for accurate representation of crash risk in the future (Lawson et al., 2003). For the sake of simplicity in the description of evaluation criteria in the subsequent paragraph, the word "estimated" will be used for Bayesian estimated, model predicted, and model replicated crash counts.

As previously stated, four evaluation criteria were employed to assess prediction accuracy of the four competing models from different perspectives based on three cases discussed above, namely: mean square predictive error (MSPE), mean absolute deviation (MAD), residual sum of squares (RSS), and total rank difference (TRD). Specifically, the MSPE was calculated as follows:

$$MSPE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - O_i)^2 \tag{5.22}$$

Where Y_i is the estimated crash frequency and O_i is the observed crash count for entity *i* by a model during the same time period. The smaller value of MSPE for a model indicates that the discrepancy between the estimated crash counts for the sites is relatively lower which demonstrates the accuracy of the model at prediction. The second criterion that was employed for assessment of prediction accuracy is MAD, which is defined as:

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|Y_i - O_i| \tag{5.23}$$

Where terms are as defined previously. The third criterion, Chi-squared RSS (Residual Sum of Squares) is defined as:

$$RSS = \sum_{i=1}^{n}\frac{(O_i - Y_i)^2}{Y_i} \tag{5.24}$$

Finally, TRD was employed as a more sensitive criterion to account for the rank deviations based on the observed and estimated crash counts (Cheng et al., 2017). The rank difference is calculated by using the following equation.

$$TRD = \sum_{i=1}^{n}|R(i_o) - R(i_e)| \tag{5.25}$$

Where $R(i_o)$ is the observed data rank at entity *i* and $R(i_e)$ is the rank based on estimated crash counts of each of the three approaches for the same time period. A particular model is considered more reliable if smaller TRD value is observed, which signifies that the specific model assigns rankings close to the observed crash counts.

DATA PREPEARATION

The data used for this study were provided by HSIS (Highway Safety Information System) which collected the data in form of different raw files from California TASAS (Traffic Accident Surveillance and Analysis System) and NHTSA (National Highway Traffic Safety Administration). The data from each year had four files pertaining to different types of data linked with Road, Occupancy, Vehicle, and Crash characteristics. The information extracted from these four files had crash number along with other factors like geometric (lane width, number of lanes, median type), traffic (Average Annual Daily Traffic), design speed, and so on. The crash data for 279 segments in one rural roadway section of California (State Route 23) was considered for model development over a period of six years (2007-2012). The 11- mile roadway section runs from the City of Moorpark at the south to the City of Fillmore at the north. The roadway has 2-lane in each direction and mainly follows within a mountainous area. This relatively long period was required to fully incorporate the temporal trends and correlation structures for different treatments of time in four alternate models. The summary statistics of crash counts and roadway pertinent covariates are shown in Table 5.1.

TABLE 5.1. Summary Statistics of Variables for segments.

Variable	Mean	SD	Minimum	Maximum
Total vehicle crash counts per year	0.34	0.81	0	8
Lane width 9 and 10 feet (yes or no)	0.33	0.47	0	1
Lane width 11 feet (yes or no)	0.10	0.30	0	1
Lane width 12 feet (yes or no)	0.47	0.50	0	1
Lane width greater than 12 feet (yes or no)	Ref			
Speed limit 25-40 mph (yes or no)	0.247	0.431	0	1
Roadway surface width (ft)	23.35	3.22	20	48
Speed limit 45 mph (yes or no)	0.616	0.486	0	1
Speed limit 50-65 mph (yes or no)	Ref			
Left Shoulder width (ft)	2.824	2.163	0	12
Right Shoulder width (ft)	2.800	2.270	0	12
AADT	3,149.80	3,818	610	21,241
Plain cement concrete Bridge Deck (yes or no)	0.12	0.32	0	1
Asphalt concrete 7 inches thick (yes or no)	0.34	0.47	0	1
Asphalt Concrete (Oiled Earth-Gravel) (yes or no)	Ref			

Notes: 1. SD refers to standard deviation

2. Ref signifies the variables which act as reference for dummy variables

RESULTS

Modeling Results

The models were developed in the statistical software WinBUGS (Lunn et al., 2000) to generate MCMC samples for Bayesian posterior inferences. For the model calibration, two chains of 13,500 iterations were set up. Convergence was ensured by visual inspection of chains and observing the desired threshold condition of MC errors to be lower than 5% of the standard deviation of parameters. After ensuring the convergence, first 9,000 samples were discarded as adaptation and burn-in and rest of the samples were used to draw parameter estimates. The model complexity was observed to have a direct impact on model running time and this criterion may influence the selection of models by safety agencies in terms of the trade-off between efficiency and precision. AR-1 (Model 3) and time adjacency (Model 4) took the highest running time for the same number of iterations, which was almost 7% higher than the most efficient model (Model 1). In case of the independent variables incorporated for model development, it is important to note that indicator variables (categorical) were used for lane width, speed limit, and surface type, with the baseline or reference as lane width more than 12 feet, speed limit 50-65 mph, and Asphalt Concrete (Oiled Earth-Gravel), respectively. The variables selected for model development were checked for multi-collinearity and correlation issues and only the satisfying variables were incorporated as an effort to arrive at potentially less biased inferences drawn from relatively more precise estimates. The right shoulder width was therefore excluded from model development due to its high correlation with the left should width, so was roadway surface width due to high collinearity with lane width. As shown in Table 5.2, all four models identified the vehicle volume (AADT) to be statistically significant along with Speed limit of 25-40 mph for two models and width of left shoulder and Plain cement concrete Bridge Deck type for Model 4. The parameter estimates illustrate the robustness of the developed models while employing different specifications of temporal aspect of crash data. In specific, the larger AADT would lead to more crash occurrences, while the narrower left shoulder width tends to increase crash probabilities. Interestingly, in comparison with the reference speed limit (50-65 mph), the segments with smaller speed limits (such as 25-40 mph, 45 mph) are inclined to generate vehicle crashes. The potential explanation might be due to worse driving conditions (e.g., sharp curves, limited visions) associated with the segments of lower speed limits in the mountainous area. It should be noted that the focus of the current study is on the model comparison based on predictive accuracy rather than the ability to filter influential factors, hence the explanation of results is focused in that direction hereafter.

TABLE 5.2. Statistics of estimated model parameters

Model	Variable Coefficient	Mean	SD	MC error	2.50%	97.50%
Model 1	Intercept	**-6.734**	0.5839	0.0473	-7.888	-5.662
	Lane width 9 and 10 feet	0.0903	0.2444	0.0148	-0.3561	0.6136
	Lane width 11 feet	0.0096	0.2546	0.0138	-0.4578	0.5429
	Lane width 12 feet	-0.1744	0.2232	0.0114	-0.5953	0.2797
	Speed limit 25-40 mph	**0.4311**	0.1939	0.0094	0.0611	0.8273
	Speed limit 45 mph	0.1452	0.1838	0.0093	-0.1954	0.5268
	Left Shoulder width	-0.0242	0.0259	0.0011	-0.0769	0.0248
	AADT	**0.8724**	0.0699	0.0057	0.7392	1.0200
	Plain cement concrete Bridge Deck	-0.1862	0.3201	0.0172	-0.8171	0.4323
	Asphalt concrete 7 inches thick	-0.0032	0.2107	0.0110	-0.4280	0.4028
	Linear trend	-0.0309	0.0288	0.0011	-0.0895	0.0237
	P1	**0.74**	0.118	0.009	0.49	0.97
	P2	**0.25**	0.11	0.009	0.02	0.5
Model 2	Intercept	**-6.6830**	0.7520	0.0627	-8.1280	-5.4420
	Lane width 9 and 10 feet	0.0981	0.2497	0.0144	-0.3734	0.5923
	Lane width 11 feet	0.0069	0.2603	0.0140	-0.4874	0.5036
	Lane width 12 feet	-0.2159	0.2360	0.0126	-0.6474	0.2825
	Speed limit 25-40 mph	0.3429	0.1853	0.0067	-0.0148	0.7016
	Speed limit 45 mph	0.0710	0.1757	0.0087	-0.2740	0.4136
	Left Shoulder width	-0.0217	0.0254	0.0011	-0.0727	0.0276
	AADT	**0.8486**	0.0760	0.0063	0.7064	0.9906
	Plain cement concrete Bridge Deck	-0.1082	0.3083	0.0159	-0.7176	0.4813
	Asphalt concrete 7 inches thick	0.0157	0.2045	0.0104	-0.3860	0.4203
	Linear trend term	0.1472	0.1426	0.0114	-0.1145	0.4175
	Quadratic trend term	-0.0254	0.0201	0.0016	-0.0639	0.0118
	P1	**0.68**	0.18	0.015	0.29	0.9
	P2	**0.31**	0.18	0.015	0.08	0.7
Model 3	Intercept	**-7.1500**	0.6263	0.0519	-8.1440	-5.7500
	Lane width 9 and 10 feet	0.1459	0.2327	0.0136	-0.2932	0.6201
	Lane width 11 feet	0.1094	0.2459	0.0132	-0.3539	0.6080
	Lane width 12 feet	-0.0883	0.2137	0.0108	-0.5102	0.3393
	Speed limit 25-40 mph	**0.3721**	0.1820	0.0082	0.0167	0.7380
	Speed limit 45 mph	0.1264	0.1687	0.0079	-0.1999	0.4653
	Left Shoulder width	-0.0336	0.0249	0.0010	-0.0826	0.0159
	AADT	**0.9138**	0.0736	0.0061	0.7486	1.0340
	Plain cement concrete Bridge Deck	-0.2443	0.3006	0.0144	-0.8394	0.3468
	Asphalt concrete 7 inches thick	0.0138	0.1963	0.0090	-0.3756	0.3844

	P1	**0.58**	0.15	0.013	0.37	0.94
	P2	**0.41**	0.15	0.013	0.05	0.62
Model 4	Intercept	**-6.6440**	1.6520	0.1419	-9.4350	-3.5490
	Lane width 9 and 10 feet	**0.5564**	0.3060	0.0203	0.0103	1.1950
	Lane width 11 feet	0.2971	0.3034	0.0173	-0.2715	0.9211
	Lane width 12 feet	0.5269	0.3245	0.0211	-0.0917	1.1900
	Speed limit 25-40 mph	0.5439	0.3546	0.0252	-0.2143	1.2060
	Speed limit 45 mph	0.1020	0.3401	0.0248	-0.6385	0.7376
	Left Shoulder width	**-0.0803**	0.0350	0.0019	-0.1516	-0.0140
	AADT	**0.8001**	0.2074	0.0178	0.3885	1.1320
	Plain cement concrete Bridge Deck	**-0.7369**	0.3883	0.0203	-1.5300	-0.0213
	Asphalt concrete 7 inches thick	-0.0322	0.2465	0.0140	-0.5198	0.4533
	P1	**0.89**	0.09	0.008	0.68	0.99
	P2	**0.1**	0.09	0.008	0.005	0.31

Notes: Bold text refers to statistically significant variables at 95% confidence level.

Goodness-of-fit

This study employed DIC and its components for assessment of goodness-of-fit of alternate models. DIC was selected for the model comparison as it is a penalized criterion which acts as a trade-off between model complexity and model fit which are represented by the effective number of parameters (P_d) and posterior deviance ($\overline{D}$). It was expected that mixture models will generate better fit for the crash data due to the addition of more parameters compared to the Base model. However, the advantage of model fit was expected to be coupled with higher complexity. As shown in Table3, the mixture models consistently exhibited higher complexity for each model, as demonstrated by the variation in values of Pd. The highest difference was depicted by Model 1 where the Pd value seems to indicate that effective number of parameters employed by the mixture model may be three times the corresponding Base model (Pd for mixture was 174 while for Base it was 54). The least discrepancy between complexity of Base and mixture was observed for Model 4, which also demonstrated the only case of comparable DIC (difference of only 1 point). The understandable higher complexity associated with the mixture models was compensated by the significantly lower (relative to Base) posterior deviance across all models which subsequently resulted in the consistently lower DIC values. These findings clearly demonstrate that the addition of mixture component for spatiotemporal models provides the advantage of superiority at fitting the observed crash data. The significant differences in the DIC values of mixture and Base models, which were the largest for the simplest Model 1 and slowly reduced which progression of model complexity (Model 4), clearly conveys the flexibility of mixture components. In terms of comparison between different temporal treatments, Model 4 demonstrated the best overall fit with the lowest DIC value of 2,043, which was 84 points lower than Model 3 (2,127). Even though the Pd value for Model 4 indicates the higher complexity but the significant superiority for posterior deviance generated the advantage at overall fit. This may be explained by the fact that AR-1 employs the lag-1 technique to borrow information from the past year which may not be as efficient as consideration of temporal correlation among all years due to the time matrix employed by Model 4. Based solely on the goodness-of-fit criteria, the time-matrix specification for addressing the temporal aspect performed the best, followed by linear trend and quadratic, while the AR-1 specification demonstrated the worst performance.

TABLE 5.3. Goodness-of-fit of alternate models

Temporal Types	Model Types	$\bar{D}$	Pd	DIC
Model 1 (linear)	Base	2121	54	2175
	Mixture	1891	174	2065
Model 2 (quadratic)	Base	2040	94	2134
	Mixture	1888	184	2072
Model 3 (AR)	Base	2122	52	2175
	Mixture	1996	131	2127
Model 4 (time matrix)	Base	1953	90	2044
	Mixture	1883	160	2043

Predictive Accuracy

Out-of-Sample Cross-Validation

The cross-validation was adopted as one approach to compute the conditional predictive ordinate (CPO) and eventually calculate log pseudo marginal likelihood (LPML) for comparison of model fit. The higher value of LPML reflects relatively superior model fit property. The difference between the LPML values of two concerned models is referred to as the log pseudo Bayes factor (LPBF) (Basu and Chib, 2003), where a LPBF greater than 5 reflects the superiority of the model of interest (Ntzoufras, 2009). As shown in Table 5.4, Model 1 was observed to have the best performance with the highest value of LPML (-862.86), which indicates that consideration of time as a linear covariate offers the best fit with crash data coupled with the benefit of least computational effort. This superior performance by Model 1 was followed by Model 4 (-876.08) and Model 2 (-886.05). The worst performance was exhibited by Model 3 which employed the AR-1 specification with an LPML value of -918.06, which had a very significant LPBF of 55.2 when measured against the best performing Model 1. It was expected that Model 3 would borrow information from the prior one year for estimation of crash risk for the concerned time period but the results seem to demonstrate that such advantage, when used along with 2-component mixture for space-time interaction, may not be transferred to cross-validated predictive performance (CPO), which served as the basis for computation of LPML. In terms of comparison between mixture and Base models, the mixture models consistently demonstrated superior prediction accuracy across all models. The largest discrepancy of 127 points of LPML was observed for Model 1 where the Base model has (-989) while mixture model was superior with (-862). These findings corroborate the previously observed trends in case of model fit, hence indicating that the advantages of mixture models at model fit are transferable to superior performance at prediction of crash counts.

In-Sample Cross-Validation

Apart from the out -of-sample cross-validation evaluation, four criteria (MAD, MSPE, RSS, and TRD) were employed to evaluate the alternate models for predictive accuracy based on the same dataset that was utilized for model development. As shown in Table 5.4, the best performing model at cross-validation predictive capability (Model 1) was also advantageous for nine out of twelve cases, except two cases of RSS and one case of MAD. This phenomenon indicates the high correlation between cross evaluation and the typical evaluations. Comparatively speaking, the other three models had similar performance, but all performed much worse than Model 1 with each of them claiming the leading position in just one or two situations. This finding is somewhat counterintuitive as Model 1 actually has the simplest temporal treatment. The potential reason might be due to the flexible mixture spatiotemporal interaction which can efficiently absorb the residual variability escaping from the predictable part of the model (Abellan et al., 2008). Such fact might help explain why Model 1 even performed worst when normal residual (i.e., Predicted vs. Observed) was used for MAD, while it performs best under all four criteria when random effects were included for Bayesian residual (i.e., Bayesian vs. Observed). In other words, the powerful mixture components might offset the mediocre performance associated with the less sophisticated temporal treatment such as linear one. Overall, the trends discussed above warrant the sensitive consideration of inferences drawn from the evaluation criteria which are employed for assessment of prediction capabilities as the advantages of certain model specifications may not be fully exploited for varied crash datasets. For example, the possible justification for inferior performance of Model 4 may be attributed to the inclusion of a small sample size with respect to the number of years of crash data which eventually resulted in a relatively small matrix (6 x 6) for the temporal CAR specification.

In terms of comparison between the Base and mixture models, the previously discussed goodness-of-fit and predictive accuracy based on the out-of-sample data clearly established the consistent superiority of mixture models. The authors were also interested to see if such advantage also extends to the prediction accuracy based on the in-sample crash data. Hence, the Base models were also evaluated for prediction accuracy and the results are summarized in Table 5.4. The dominance of the mixture models is evident from the observed results as they consistently generated more precisive crash estimates across all temporal models as well as three different datasets.

The authors were also interested in quantifying the influence of the inclusion of random effects at predictive accuracy. The impact of incorporation of random effects to account for spatial and temporal correlation structures, and the flexibility of 2-component mixture may be observed by comparing the Bayesian residuals (Bayesian vs Observed) and normal residuals (Predicted vs Observed), which reflect the predictive performance of random effect models and normal models, respectively. Table 5.4 demonstrates the significant superiority of sophisticated models which employed random effects as the Bayesian residuals exhibited consistently more accurate predictions of crash counts for all competing models as determined through the four evaluation criteria. The Bayesian residuals exhibited lower value for each evaluation criterion, though the differences were subtle or dramatic depending on the nature of criterion. In the case of MAD, the maximum difference observed was 0.7 for Model 1 while MSPE demonstrated a maximum difference of almost double points in case of Model 1 where normal residuals (0.46) recorded a 0.21 points difference from Bayesian residuals (0.25). A similar difference of 0.21 points was also observed for Model 5 where the Bayesian and normal residuals were relatively high compared to Model 1 and Model 2. The disparity of predictive performance was notably more in the case of TRD but very significant deviations were recorded for RSS where the Bayesian residuals exhibited eight or nine times lesser values. This consistent superior trend of random effect models reinforces the advantages of incorporation of spatially structured and unstructured correlations to account for the unobserved heterogeneity which benefits the model accuracy at the prediction of crash counts.

TABLE 5.4. Evaluation results of predictive performance of alternate models

Evaluation Criteria	**Comparison Dataset**	**Model 1**		**Model 2**		**Model 3**		**Model 4**	
		Mixture	**Base**	**Mixture**	**Base**	**Mixture**	**Base**	**Mixture**	**Base**
LPML	NA	**-862.86**	-989.2	-886.65	-932.3	-918.06	-1007	-876.08	***-929.45***
MAD	Replicated vs. Observed	**0.32**	0.37	0.33	***0.35***	0.33	0.37	0.34	***0.35***
	Predicted vs. Observed	0.39	0.43	**0.37**	***0.39***	0.38	0.40	0.38	***0.39***
	Bayesian vs. Observed	**0.32**	0.37	0.33	***0.35***	0.33	0.37	0.33	***0.35***
MSPE	Replicated vs. Observed	**0.26**	0.41	0.29	0.38	0.28	0.41	0.31	***0.37***
	Predicted vs. Observed	**0.46**	0.52	**0.46**	***0.50***	0.47	***0.50***	0.49	0.51
	Bayesian vs. Observed	**0.25**	0.40	0.28	0.38	0.29	0.41	0.28	***0.37***

RSS	Replicated vs. Observed	814.32	4530.6	814.02	1777.71	**813.15**	7057.45	813.74	***1766.99***
	Predicted vs. Observed	7334.60	***4912.7***	9043.77	5685.39	7898.23	6092.34	**7056.49**	6689.79
	Bayesian vs. Observed	**813.74**	4548.9	847.92	1769.50	4467.75	5259.25	912.78	***1670.22***
TRD	Replicated vs. Observed	**546652**	591886	554405	***579092***	556374	598923	567821	579691
	Predicted vs. Observed	**618411**	***622451***	618623	627074	618563	623069	621886	636765
	Bayesian vs. Observed	**546128**	591852	554587	579114	559146	598809	555804	***578820***

Note: 1. Bold text represents the best performance for that category.

2. Bold text in *Italics* represents the best performance for the Base models among that category.

3. To prevent the infinity value obtained for RSS, a small value of 10^-4 was added to the denominator in Equation 24 for RSS.

CHAPTER 6: SUMMARY AND CONCLUSIONS

This report has presented the four chapters pertaining to different study objectives. Within each chapter, the specific objectives were introduced and the contributions to the safety literature were stated. This was followed by description of model specifications and the data utilized for model development. Finally, the discussion for observed model estimates and comparison between alternative models was offered. This chapter summarizes the previous chapters and offers conclusions and recommendations.

Conclusions for Chapter 2

This study focused on the comparison of goodness-of-fit of three alternate spatiotemporal models. The models were developed for estimation of crash rate for four modes of travel: vehicle, motorcycle, bicycle, and pedestrian. Full Bayesian framework was employed while assuming Poisson lognormal distribution for the crash counts aggregated at the macro-level of counties. The daily vehicle miles traveled was utilized as the traffic exposure for estimation of crash risk based on crash rate for all four modes. All three models incorporated different types of structured and unstructured correlations to account for the unobserved heterogeneity that may escape the explanatory variables incorporated for model development. In terms of unstructured correlations, a multivariate error term was incorporated for all models to obtain more precise estimates by removing the potential bias associated with the ignorance of correlation among different crash modes. In terms of spatially structured correlation, random effects were incorporated into all models which accounted for the distance-based dependency of crash risk among neighboring counties. The aforementioned specifications were common to all three models while the treatment of temporal aspect and the spatiotemporal interaction essentially distinguished them. The first model (Model 1) assigned a linear trend with a spatiotemporal interaction which remained fixed throughout the time period considered; the second model (Model 2) enhanced the former by inclusion of a quadratic trend; and the third model (Model 3) developed over the first model by allowing the flexibility to incorporate a yearly-varying spatiotemporal interaction term. To obtain robust understanding regarding the implications of the different treatments of time and its interaction with space, an eight-year period was considered for the crashes aggregated at the 58 counties of California, and different socioeconomic and traffic-related influential variables were incorporated.

The model estimates justified the use of mode-varying coefficients for explanatory variables as the impact of these factors varied across different crash modes. Largely a similar set of influential covariates was generated by the three models which indicate their robustness. However, notable differences were observed from the assessment of goodness-of-fit criteria employed in the study. In case of comparison between the model with time-varying spatial interactions (Model 3) and temporal trend models (Model 1 and Model 2), the flexibility of Model 3 to capture the unobserved heterogeneity and spatially structured correlation allowed superior performance at posterior deviance (Dbar), and LPML. Model 3 was penalized for the increased complexity which was a consequence of the addition of eight different spatial random effects corresponding to the time period of this study. Due to this significant increase in the effective number of parameters that were utilized for model development, Model 3 was inferior to competing models at DIC. It is worth mentioning that Model 3 was significantly better than competing models as assessed by the aforementioned criteria. This is contrary to the marginally better performance of Model 2 over Model 1 in case of comparison among the temporal trend models. Another important finding from the goodness-of-fit criteria was the correlation between the posterior deviance (Dbar) of a model and its performance at rest of the criteria which incorporated the random effects. Clearly, Model 3 exhibited remarkably lower posterior deviance which was transferred to equivalent superior performance across other criteria. This correlation illustrated the advantages of yearly-varying space-time interaction which increased the model complexity but provided a significant improvement in model fit, as assessed by different criteria employed in the study. For comparison of site ranking performance, overall, Model 3 appears to have the best performance in consistently identifying the both high and low risk counties across the adjacent time periods and perform slightly better than the other two models in terms of site consistency. In other words, the superiority of the model's predictive performance can be transferred to yield more accurate result of site ranking.

Although the research here reflects an improved understanding of how various multivariate space-time models perform, further work is still needed. First, the results here depended on county-level multimodal crashes over a period of eight years. The crash counts of other spatial units such as traffic analysis zone, roadway segment, intersection, etc. might lead to different findings as compared with the current study. Second, all three models were developed by using vehicle DVMT as the exposure for all four modes. The models may benefit with the inclusion of mode-specific exposure variables and future research is recommended to incorporate such surrogates of multimodal activity if available. Third, for the temporal aspect, linear and quadratic trends were employed to fit the crash counts. Other temporal treatments such as time-varying coefficients, autoregressive error process, moving average, and others, are also worth exploring. Future studies are also advised to place careful considerations for drawing inferences from model estimates due to the presence of temporal instability, which may be more pronounced depending on the explanatory variables and time period of study. Fourth, the spatial correlation was incorporated using the MCAR specification. Future studies may explore more sophisticated approaches such as spatial autoregressive lag structure, which accommodates spatial dependency by allowing for varying spatial autoregressive parameters across count or mode categories (Bhat et al., 2017). Fifth, the alternative models evaluated here follow the typical parametric models which restrict the priors to a specific distribution fixed across all entities. It is possible to improve upon the models by using flexible multivariate latent class approaches. Such an approach may be combined with random parameters (if the crash data shows significant differences for parameters across sites) to better account for unobserved heterogeneity (Mannering et al., 2016).

Conclusions for Chapter 3

The current study contributes to the safety literature by proposing a bivariate Dirichlet process mixture spatial model and comparing its performance for crash predictions with other three competing models. The proposed semi-parametric model accounted for the unobserved heterogeneity by combining the strengths of incorporating bivariate specification to accommodate correlation among crash modes, spatial random effects for the impact of neighboring TAZs, and Dirichlet process mixture for random intercept. The present model structure allowed the flexibility to infer stochastic parameter from the crash data instead of a prespecified distribution.

All four models shared similar influential factors across both crash modes which indicated the robustness of the models. For crashes pertaining to bicycles, K12 student enrollment, percentage of arterials, and bike-lane density for the TAZ were observed to be statistically significant at the 95% confidence interval. The positive correlation of K12 student enrollment with crash risk suggests the increased risk due to higher chances of physical interactions of bicyclists/pedestrians with other modes due to more exposure. However, the perceived risk appears to be the governing factor in the case of positive correlation for bike-lane density, which seems counter-intuitive. The lower perceived risk may encourage bicyclists to ride more in such areas and therefore yield higher chances of the exposure of bicyclists to vehicular traffic. A negative correlation was observed for percentage of arterial roads and bicycle crashes which suggests a lesser tendency of bicyclists to travel in areas with more arterials, hence reducing the exposure to possible interactions. The pedestrian crashes were observed to reduce with an increase in student population in the colleges of TAZs. This may be justified by the policies implemented in these areas for reduced vehicular traffic which eventually reduces the possibility of interaction with pedestrians.

The heterogeneity error term demonstrated the presence of statistically significant correlation among the bicycle and pedestrian crashes while the spatial random effect term exhibited the absence of a significant correlation, which might explain why models considering the spatial random effects did not yield the expected advantages compared with their non-spatial counterparts. In the comparison between Dirichlet and non-Dirichlet models, the former ones were consistently superior to typical bivariate ones under all criteria. These findings demonstrate the advantages associated with consideration of flexible approach, Dirichlet process mixture in the current study, based on the goodness-of-fit and predictive accuracy of estimated crash counts. Moreover, the Dirichlet models exhibited the capability to identify the latent distinct subpopulations and suggested the that the normal assumption of intercept associated with traditional parametric models does not hold true for the TAZ level crash dataset of the current study. These findings justify the development of sophisticated flexible models which generate more precise estimate due to the unrestrictive approach which eventually leads to better inferences.

Based on the results, this study recommends careful consideration of spatial correlations at the macro-level of TAZs as they increased the complexity without any significant advantage at model fit or predictive accuracy. The authors also recommend exploring other spatial levels and observe if the results of the current study hold true or if the spatial random effects prove beneficial. Similar to other studies that focus on crashes pertaining to modes of active transportation, it should be noted that both the pedestrian and bicycle crashes have been modeled by utilizing the exposure of vehicles, rather than pedestrians and bikes, due to the unavailability of exposure data for the concerned modes. It is recommended that novel methods may be explored to account for the exposure data such as using bike mode share, or calibrating the exposure from socio-economic factors related to such modes (e.g. number of employees walking or cycling to work). Finally, the crash dataset utilized for model development was aggregated for a six-year period and future studies may incorporate temporal correlations and adopt disaggregated crash counts (*51*).

Conclusions for Chapter 4

The primary objective of this study was the comprehensive comparison of temporal crash prediction models from different perspectives. Nine groups of methodological approaches were employed which differed on the basis of treatments of serial correlations. A novel approach of time-adjacency was borrowed from the medical field to analyze the crash data and compare its performance with the prevalent temporal models in traffic safety research. Moreover, three types of models were proposed for each group to assess the modeling performance under various conditions in terms of spatial random effects. Finally, ten different evaluation criteria were utilized for modeling assessment. It is worth mentioning that the multiple temporal models (except for the time adjacency models) and performance-checking criteria selected for the study were not exhaustive but rather a representation of the models and measures commonly employed in safety literature. The Full Bayesian framework was employed for the accommodation of complex correlations structures.

The posterior model estimates demonstrated the notable impact of inclusion of spatial correlation on improving the precision of intercept across all groups of temporal treatments. Both spatial and temporal random effects were noted to be statistically significant reflecting the presence of spatial clustering and temporal dependency of crash counts which accounted for the variability in models. The changing coefficients of intercept for the random intercept model further demonstrated the impact of serial changes on crash risk.

In terms of prediction capabilities, the aggregation results by model types reveal that the base model without considering serial correlations outperforms others, while the aggregated results by temporal treatments indicate the top-ranked performance associated with Type 1, or, the one without considering spatial dependency. Such phenomena suggest that, due to the myriad of influential factors, it is difficult to predict county-level crash counts by using a certain type of temporal and spatial correlations applicable to all counties, even though such correlations are statistically significant. However, the individual modeling results show the profound predictive benefits associated with G5T3. It follows that it is still worth addressing the serial and spatial dependency for the crash prediction of the large spatial unit, county, while caution should be exercised as there is large variability of predictive capabilities among the different temporal treatments. Although the simpler model specifications are desirable for practitioners due to computational challenges posed by sophisticated models, but given the advantages associated with more complex approach of AR-1 space-time interaction, it would be beneficial to employ such a model for the given crash dataset which exhibit spatiotemporal clustering.

In review of P_d, both aggregated and individual results indicate the spatial-temporal models (either Type 2 or Type 3) seem to be more efficient in reducing the effective number of parameters than Type 1, and the models accounting for temporal dependency appear to perform better than the base models in most scenarios. Interestingly, similar trends are found in the criterion of DIC, which suggest that P_d tends to have a larger impact on the DIC than does $\overline{D}$. It is noteworthy that, based on aggregation result, Group 9 (time adjacency models) are observed to have the best performance in terms of DIC compared with other groups. Hence, it is recommended that this group should be further explored in the safety field.

Building upon the large number of models being developed, the correlation analysis among all ten criteria confirms that P_d exerts more influence on DIC than does $\overline{D}$, while the latter one demonstrates the statistically significant correlations with a set of criteria which checks the predictive accuracy from different aspects. It is anticipated that such correlation analysis should render safety researchers or practitioners more confidence for modeling selection, especially when the safety resources are limited which prevent a comprehensive evaluation of models.

The results observed in this paper require some caveats. First, only DVMT was included in the modeling development which allows an easier comparison of the wide range of models. The results with other covariates being included might be different. Second, eight years of data being utilized seem to be adequate for safety study since the objective was the comprehensive comparison of temporal models and eight-year period seemed optimum to replicate the safety literature (*52*). However, the sample size of 8 might be relatively small for some models especially the time adjacency ones and therefore impact their performances to some degree. Future research may be focused on different time-adjacency models with a greater time period. Third, the aggregated and individual modeling results demonstrate different characteristics with respect to the prediction-related criteria. The discrepancy might be due to the myriad of safety factors influencing the counties. Studies focusing on smaller spatial units (such as intersections, segments, Traffic Analysis Zones, census tracts, Local Authority Districts) are highly recommended for verification of the results. Since the crash-related factors may interact differently at smaller spatial levels, some deviations from the performance of current models is anticipated. Finally, only the distance-based weight structures were used for implementing the spatial priors. Other types of spatial correlations merit further exploration.

Conclusions for Chapter 5

This study introduced the mixture models for space-time interaction in the safety research. The models with the mixture component allow the accommodation of global space-time patterns, which are referred as "stable", while also capturing the atypical departures from the stable pattern which may not be captured by the conventional spatiotemporal models in safety research. This was possible by incorporating two components for the space-time interaction term where one component allowed the accommodation of noise to convey stable pattern while the second component captured the larger fluctuations. The mixture component was explored by collaboration with four different temporal treatments and the models were evaluated based on the goodness-of-fit, and in-sample and out-of-sample cross-validation for predictive accuracy. Full Bayesian framework was employed to build Poisson lognormal models where conditional autoregressive specification was employed to incorporate spatially structured random effects using adjacency-based weight matrix, and the space-time interaction term was allowed the flexibility to capture the discrete and continuous space-time variations of the crash data that may have escaped from the predictable patterns of overall temporal and spatial risk surface. This flexible framework served as the base for development of four models which differed on the treatment of temporal aspect: (I) linear time trend; (II) quadratic time trend; (III) Autoregressive-1 (AR-1); and (IV) time adjacency. Moreover, five evaluation criteria were employed for assessment of the predictive capability of alternate models from different perspectives. The Base models were also developed to serve as a reference for demonstrating the advantages of mixture models associated with model fit and prediction capabilities.

Using six-year crash data from various roadway segments of one state route in California, the modeling results demonstrated the robustness of the proposed modeling framework with the vehicle volume being identified as statistically significant for all four models. Plus, Speed limit of 25-40 mph, width of left shoulder and Plain cement concrete Bridge Deck type were also flagged out as highly influential factors of crash occurrence for some of the models. The penalized criterion of DIC was employed for assessment of goodness-of-fit while accounting for the model complexity. Understandably, the mixture models depicted larger complexity associated with the incorporation of higher number of parameters due to the two-component specification. However, due to the remarkable reduction in posterior deviance, the mixture models subsequently demonstrated lower (compared to corresponding Base models) values of DIC across all four models. The discrepancy of DIC values among Base and mixture models was observed to be largest for the simplest temporal model (linear trend) and experienced constant reduction as the complexity of temporal treatment increased, with the least discrepancy for the time matrix model. The assessment of model fit clearly established the advantages associated with the mixture models. Considering the temporal treatment, the model with time matrix was the best as it seemed to borrow strength from the neighboring years for fitting the crash data.

For comprehensive comparison of predictive accuracy of model estimated crash counts, this study employed an array of evaluation measures based on out-of-sample and in-sample cross validation. In the former case, leave-one-our cross-validation (CV-1) technique was utilized to avoid the potential bias of results due to different splits of data. The resultant value of log pseudo marginal likelihood (LPML) illustrated that Model 1 (i.e., the model with the linear time trend treatment) was the best model, whose log pseudo Bayes factor (LPBF) values against competing ones ranged from 4.2 to 55.2. Four different evaluation criteria, including MAD, MSPE, RSS, and TRD, were used for the typical in-sample validation. Under each criterion, the corresponding statistics were obtained by comparing the observed crash counts with three different types of data, respectively, which contain Bayesian estimated counts (or, the predication with random effects), the normal predicted counts (the prediction without random effects), and the model replicated ones. Out of the 12 scenarios, Model 1 again claimed the first place in 9 of them, indicating somewhat high correlation between cross validation and typical validation results. The remaining three models demonstrated similar performance even though mixed rankings were observed under different conditions. Further examination of results illustrated that Model 1 was placed 2nd and last for two cases when the prediction without random effects was used for evaluation, while its superior performance is consistent across all situations involving prediction with random effects. Such phenomenon indicated that the least sophisticated temporal treatment, linear one, exhibited merely mediocre performance when only the predictable part of the models were considered. However, such limitation was masked when the 2-component-mixtrure spatiotemporal interaction was employed. Indeed, the inclusion of such mixture rendered Model 1 the leading positions in most cases, which might attest to the high efficiency of the mixture components in capturing the unobserved heterogeneity escaping from the risk explained by the covariates and temporal treatment. Moreover, the residual analysis revealed the consistent superior trend of random effect models which bolsters the incorporation of spatially structured and unstructured correlations to account for the unobserved heterogeneity which benefits the model accuracy at prediction of crash counts. In terms of comparison between Base and mixture models on the basis of predictive accuracy for both in-sample and out-of-sample data, the mixture models demonstrated consistent superiority across all models and crash datasets. This finding supplements the dominance of mixture models observed for model fit and establishes them to be beneficial from different perspectives required in studies of traffic safety (superior fit to crash data and higher precision for crash estimation).

The present study conducts a comprehensive evaluation of four spatiotemporal models with different temporal treatments. However, the results observed in this paper require some caveats. First, the roadway segment crash data are subject to a low-sample mean issue, even though the chosen Poisson-Lognormal model could better handle low sample mean and small sample size as mentioned previously. The data collected from larger spatial units such as traffic analysis zone, census tract, city, or county, are recommended to verify the results obtained hereby. Second, the mixture components are employed for the space-time interactions. The relative performances of models might change when other forms of interactions are used which include linear, quadratic, and time-varying spatial correlation, and so on. Third, only six years of data are used for performance evaluation of various models. It is, therefore, worth exploring longer time periods which might lead to different findings, especially for time adjacency model which is subject to limited sample size in terms of the number of years in the study. Fourth, the comparison is done on the basis of univariate models. The analysis is also recommended for the multivariate settings. Finally, the promising results demonstrated by the mixture models at goodness-of-fit and predictive accuracy warrant the need for exploring their advantages at another important use of crash prediction models, i.e. site ranking or hot spot identification. Since the mixture component allows the detection of higher fluctuations (variability) in the space-time profile of sites under focus, hence it will be beneficial at identification of sites which deviate from the global space-time pattern observed for all the sites. Such high-risk sites may remain hidden in the traditional spatiotemporal models. The practical applications at the segment level could be the detection of sites with underlying risk which might not be incorporated directly from the data. The possible examples could be some short-term change in traffic behavior such as a musical concert which attracted more activity pertaining to pedestrians, the impact of sun glare during specific months on a section of roadway which may increase crash risk, the construction work for repair of pavement or new construction which may not be evident in data, the issues with line of sight for drivers due to the presence of some obstruction such as a tree, and so on. The mixture model may help highlight such sites with underlying crash risk, which may not be evident in traditional models.

ACKNOWLEDGEMENT

Support for this study was provided by 2017 UCCONNECT Research Program (Grant number: 65A0529-052.3). The authors would also like to thank the California Department of Transportation for their support of this project. We especially acknowledge the consistent assistance and advice of Mr. Jerry Kwong and Dr. John L. Ensch.

REFERENCES

1. Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention 32 (5), 633–642.
2. Abdel-Aty, M., Wang, X., 2006. Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors. In: Transportation Research Record: Journal of the Transportation Research Board, No. 1953. Transportation Research Board of the National Academies, Washington, DC, pp. 98–111.
3. Abellan, J.J., Richardson, S. and Best, N. Use of space-time models to investigate the stability of patterns of disease. *Environmental Health Perspectives, 116*(8), 2008 p.1111.
4. Aguero-Valverde, J., Wu, K.F., Donnell, E.T., 2016. A multivariate spatial crash frequency model for identifying sites with promise based on crash types. Accident Analysis and Prevention 87, 8–16.
5. Aguero-Valverde, J. 2013. Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates. Accident Analysis & Prevention, 50, 289-297.
6. Aguero-Valverde, J., and Jovanis, P. 2006. Spatial Analysis of Fatal and Injury Crashes in Pennsylvania. Accident Analysis and Prevention. Vol. 38, No. 3, pp. 618–625.
7. Aguero-Valverde, J., Jovanis, P.P., 2009. Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. Transportation Research Record 2136, 82–91.
8. Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of road crash frequency with spatial models. Transportation Research Record 2061, 55–63.
9. Aguero-Valverde, J. and Jovanis, P., 2010. Spatial correlation in multilevel crash frequency models: Effects of different neighboring structures. Transportation Research Record: Journal of the Transportation Research Board, (2165), pp.21-32.

10. Amoh-Gyimah, R., Saberi, M. and Sarvi, M., 2016. Macroscopic modeling of pedestrian and bicycle crashes: a cross-comparison of estimation methods. Accident Analysis & Prevention, 93, pp.147-159.
11. Andrey, J., & Yagar, S. 1993. A temporal analysis of rain-related crash risk. Accident Analysis & Prevention, 25(4), 465-472.
12. Barron, D. N. The analysis of count data: Overdispersion and autocorrelation. *Sociological methodology*, 1992, pp. 179-220.
13. Basu, S. and Chib, S., 2003. Marginal likelihood and Bayes factors for Dirichlet process mixture models. Journal of the American Statistical Association, 98(461), pp.224-235.
14. Beck, L.F., Dellinger, A.M., O'Neil, M.E., 2007. Motor vehicle crash injury rates by modeof travel United States: using exposure-based methods to quantify differences. Am. J. Epidemiol., 166.
15. Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., & Songini, M. Bayesian analysis of space—time variation in disease risk. *Statistics in medicine*, *14*(21-22), 1995, 2433-2443.
16. Berrigan, D., R. Troiano, T. McNeel, C. DiSogra, and R. Ballard-Barbash. Active Transportation Increases Adherence to Activity Recommendations. *American Journal of Preventive Medicine*, Vol. 31, No. 3, 2006, pp. 210-216.
17. Bhat, C.R., Astroza, S. and Lavieri, P.S., 2017. A new spatial and flexible multivariate random-coefficients model for the analysis of pedestrian injury counts by severity level. *Analytic Methods in Accident Research*, *16*, pp.1-22.
18. Blazquez, C.A. and Celis, M.S., 2013. A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile. Accident Analysis & Prevention, 50, pp.304-311.
19. Cai, Q., Lee, J., Eluru, N., & Abdel-Aty, M. 2016. Macro-level pedestrian and bicycle crash analysis: incorporating spatial spillover effects in dual state count models. Accident Analysis & Prevention, 93, 14-22.
20. Carlin, B.P. and Louis, T.A., 1996. Bayes and empirical Bayes methods for data analysis. Volume 69 of Monographs on Statistics and Applied Probability.
21. Carlin, B.P. and Louis, T.A., 2008. Bayesian methods for data analysis. CRC Press.
22. Cheng, L, Geedipally, S., Lord, D. The Poisson-Weibull generalized linear model for analyzing motor vehicle crash data. *Safety Science 54*, 2013, pp.38-42.

23. Cheng, W., Gill, G.S., Dasu, R., Xie, M., Jia, X. and Zhou, J., 2017a. Comparison of Multivariate Poisson lognormal spatial and temporal crash models to identify hot spots of intersections based on crash types. Accident Analysis & Prevention, 99, pp.330-341.
24. Cheng, W., Gill, G.S., Sakrani, T., Dasu, M. and Zhou, J., 2017b. Predicting motorcycle crash injury severity using weather data and alternative Bayesian multivariate crash frequency models. Accident Analysis & Prevention, 108, pp.172-180.
25. Cheng, W., G.S. Gill, S. Choi, J. Zhou, X. Jia, and M. Xie. Comparative Evaluation of Temporal Correlation Treatment in Crash Frequency Modelling. *Transportmetrica A: Transport Science*, 2017c, pp.1-40. https://doi.org/10.1080/23249935.2017.1418458
26. Cheng, W., Gill, G. S., Sakrani, T., Dasu, M., & Zhou, J., 2017d. Predicting motorcycle crash injury severity using weather data and alternative Bayesian multivariate crash frequency models. Accident Analysis & Prevention, 108, 172-180.
27. Cheng, W., G. S. Gill, Y. Zhang, and Z. Cao. Bayesian spatiotemporal crash frequency models with mixture components for space-time interactions. *Accident Analysis & Prevention*, *112*, 2018a. pp.84-93.
28. Cheng, W., Gill, G.S., Ensch, J.L., Kwong, J. and Jia, X., 2018b. Multimodal crash frequency modeling: Multivariate space-time models with alternate spatiotemporal interactions. *Accident; analysis and prevention*, *113*, pp.159-170.
29. Cheng, W., & Washington, S. P. (2005). Experimental evaluation of hotspot identification methods. *Accident Analysis & Prevention*, *37*(5), 870-881.
30. Cheng, W., & Washington, S. 2008. New criteria for evaluating methods of identifying hot spots. Transportation Research Record: Journal of the Transportation Research Board, (2083), 76-85.
31. Chiou, Y.C., Fu, C., Chih-Wei, H., 2014. Incorporating spatial dependence in simultaneously modeling crash frequency and severity. Analytic Methods in Accident Research 2, 1–11.
32. Congdon, P. *Bayesian models for categorical data*. John Wiley & Sons, 2005.
33. Congdon, P., 2006. Bayesian Statistical Modeling. 2nd ed. Wiley, New York.
34. Conway, A., Cheng, J., Peters, D., & Lownes, N. 2013. Characteristics of Multimodal Conflicts in Urban On-Street Bicycle Lanes. Transportation Research Record: Journal of the Transportation Research Board, (2387), 93-101.

35. Dong, N., Huang, H., Lee, J., Gao, M., & Abdel-Aty, M. (2016). Macroscopic hotspots identification: a Bayesian spatio-temporal interaction approach. Accident Analysis & Prevention, 92, 256-264.
36. Dong, N., Huang, H., Xu, P., Ding, Z. and Wang, D. Evaluating spatial-proximity structures in crash prediction models at the level of traffic analysis zones. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2432, 2014, pp.46-52.
37. Dong, N., Huang, H. and Zheng, L., 2015. Support vector machine in crash prediction at the level of traffic analysis zones: assessing the spatial proximity effects. *Accident Analysis & Prevention*, *82*, pp.192-198.
38. Earnest, A., Morgan, G., Mengersen, K., Ryan, L., Summerhayes, R., & Beard, J. Evaluating the effect of neighbourhood weight matrices on smoothing properties of Conditional Autoregressive (CAR) models. *International journal of health geographics, 6*(1), 2007, 54.
39. Efron, B. The Jackknife, the Bootstrap and other resampling plans. *CBMS-NSF regional conference series in applied mathematics* 1982. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM).
40. El-Basyouny, K., & Kwon, D. W. Assessing time and weather effects on collision frequency by severity in Edmonton using multivariate safety performance functions. *Transportation Research Board 91st Annual Meeting,* 2012, (No. 12-0494).
41. El-Basyouny, K., Barua, S., & Islam, M. T. 2014. Investigation of time and weather effects on crash types using full Bayesian multivariate Poisson lognormal models. Accident Analysis and Prevention, 73, 91-99
42. El-Basyouny, K., Sayed, T., 2009. Collision prediction models using multivariate Poisson-lognormal regression. Accident Analysis and Prevention 41 (4), 820– 828.
43. Elvik, R. (2007) State-of-the-Art Approaches to Road Accident Black Spot Management and Safety Analysis of Road Networks. Report 883. Institute of Transport Economics, Oslo, Norway.
44. Erkanli, A., M. Sung, E. Jane Costello, and A. Angold. Bayesian semi-parametric ROC analysis. *Statistics in Medicine*, Vol. 25, No. 22, 2006, pp. 3905-3928.
45. Flask, T., & Schneider, W. (2013). A Bayesian analysis of multi-level spatial correlation in single vehicle motorcycle crashes in Ohio. Safety science, 53, 1-10.

46. Flask, T., Schneider, W. H., & Lord, D. (2014). A segment level analysis of multi-vehicle motorcycle crashes in Ohio using Bayesian multi-level mixed effects models. Safety science, 66, 47-53.
47. Flegal, J. M., Haran, M., & Jones, G. L. Markov chain Monte Carlo: Can we trust the third significant figure?. *Statistical Science,* 2008, pp.250-260.
48. Frank, L., M. Greenwald, S. Winkelman, J. Chapman, and S. Kavage. Carbonless footprints: Promoting health and climate stabilization through active transportation. *Preventive Medicine*, Vol. 50, 2010, pp. S99-S105.
49. Furie, G., and M. Desai. Active Transportation and Cardiovascular Disease Risk Factors in U.S. Adults. American Journal of Preventive Medicine, Vol. 43, No. 6, 2012, pp. 621-628.
50. Gabauer, D. J., & Li, X. (2015). Influence of horizontally curved roadway section characteristics on motorcycle-to-barrier crash frequency. Accident Analysis & Prevention, 77, 105-112.
51. Geedipally, S.R., Lord, D. and Dhavala, S.S., 2014. A caution about using deviance information criterion while modeling traffic crashes. Safety science, 62, pp.495-498.
52. Gelfand, A. E., D. K Dey, and H. Chang. *Model determination using predictive distributions with implementation via sampling-based methods* (No. TR-462). STANFORD UNIV CA DEPT OF STATISTICS, 1992.
53. Gelfand, A. E. Model determination using sampling-based methods. *Markov chain Monte Carlo in practice*, 1996, 145-161.
54. Gelman, A., Carlin, J., Stern, H.S., Rubin, D.B., 2003. Bayesian Data Analysis, second ed. Chapman & Hall/CRC
55. Ghosh, S., and J. Norris. Bayesian capture-recapture analysis and model selection allowing for heterogeneity and behavioral effects. *Journal of Agricultural, Biological, and Environmental Statistics*, Vol. 10, No. 1, 2005, pp. 35-49.
56. Giles-Corti, B., S. Foster, T. Shilton, and R. Falconer. The co-benefits for health of investing in active transportation. *New South Wales Public Health Bulletin*, Vol. 21, No. 6, 2010, p. 122.
57. Gill, G.S., Cheng, W., Zhou, J. and Park, V.S., 2017a. Comparative analysis of cost-weighted site ranking using alternate distance-based neighboring structures for spatial crash frequency modeling. Journal of Transportation Safety & Security, pp.1-23.
58. Gill, G., W. Cheng, M. Xie, T. Vo, X. Jia, J. Zhou. 2017b. Evaluating the Influence of Neighboring Structures on Spatial Crash Frequency Modeling and Site Ranking Performance,

Transportation Research Record: Journal of the Transportation Research Board, (2659), pp.117-126.

59. Guo, F.,Wang, X., Abdel-Aty,M. A., 2010. Modeling signalized intersection safety with corridor-level spatial correlations. Accident Analysis and Prevention. 42(1), 84-92.
60. Haque, M. M., Chin, H. C., & Huang, H. (2010). Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. Accident Analysis & Prevention, 42(1), 203-212.
61. Hay, J. L., & Pettitt, A. N. 2001. Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease. Biostatistics, 2(4), 433-444.
62. Heydari, S., L. Fu, D. Lord, and B. Mallick. Multilevel Dirichlet process mixture analysis of railway grade crossing crash data. *Analytic Methods in Accident Research*, Vol. 9, 2016, pp. 27-43.
63. Heydari, S., L. Fu, L. Miranda-Moreno, and L. Jopseph. Using a flexible multivariate latent class approach to model correlated outcomes: A joint analysis of pedestrian and cyclist injuries. *Analytic Methods in Accident Research*, Vol. 13, 2017, pp. 16-27.
64. Huang, H., Chin, H. and Haque, M., 2009. Empirical evaluation of alternative approaches in identifying crash hot spots: naive ranking, empirical Bayes, and full Bayes methods. *Transportation Research Record: Journal of the Transportation Research Board*, (2103), pp.32-41.
65. Huang, H., Zhou, H., Wang, J., Chang, F. and Ma, M., 2017. A multivariate spatial model of crash frequency by transportation modes for urban intersections. Analytic methods in accident research, 14, pp.10-21.
66. Insall, P., 2013. Active travel: transport policy and practice for health. Nutr. Bull. 38 (1), 61–69.
67. Ishwaran, H., and L. James. Gibbs Sampling Methods for Stick-Breaking Priors. *Journal of the American Statistical Association*, Vol. 96, No. 453, 2001, pp. 161-173.
68. James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. An introduction to statistical learning. Vol. 112. New York: springer, 2013.
69. Jiang, X., Abdel-Aty, M., & Alamili, S. (2014). Application of Poisson random effect models for highway network screening. Accident Analysis & Prevention, 63, 74-82.
70. Jung, R. C., Kukuk, M., & Liesenfeld, R. Time series of count data: modeling, estimation and diagnostics. *Computational Statistics & Data Analysis, 51*(4), 2006, pp. 2350-2364.

71. Lawson, Andrew B., William J. Browne, and Carmen L. Vidal Rodeiro. *Disease mapping with WinBUGS and MLwiN.* Vol. 11. John Wiley & Sons, 2003
72. Lee, C., and M. Abdel-Aty. Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida. *Accident Analysis & Prevention*, Vol. 37, No. 4, 2005, pp. 775-786.
73. Lee, J., Abdel-Aty, M., Jiang, X., 2015. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. Accident Analysis and Prevention 78, 146–154.
74. Levine, N., K. E. Kim, and L. H. Nitz. Spatial Analysis of Honolulu Motor Vehicle Crashes: I. Spatial Patterns. *Accident Analysis and Prevention,* Vol. 27, No. 5, 1995, pp. 663–674.
75. Li, L., Zhu, L. and Sui, D.Z., 2007. A GIS-based Bayesian approach for analyzing spatial–temporal patterns of intra-city motor vehicle crashes. Journal of Transport Geography, 15(4), pp.274-285.
76. Lord, D. and Miranda-Moreno, L.F., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective. Safety Science, 46(5), pp.751-770.
77. Lord, D. and Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice, 44(5), pp.291-305.
78. Liu, C. and Sharma, A., 2017. Exploring spatio-temporal effects in traffic crash trend analysis. *Analytic Methods in Accident Research*, *16*, pp.104-116.
79. Lunn, D., Jackson, C., Best, N., Thomas, A. and Spiegelhalter, D., 2012. The BUGS book: A practical introduction to Bayesian analysis. CRC press.
80. Ma, X., Chen, S. and Chen, F. Multivariate space-time modeling of crash frequencies by injury severity levels. *Analytic Methods in Accident Research,* 15, 2017, pp.29-40.
81. Ma, P. E. Bayesian analysis of underreporting Poisson regression model with an application to traffic crashes on two-lane highways. *Transportation Research Board 88th Annual Meeting*, 2009 (No. 09-3192).
82. MacNab, Ying C. Bayesian spatial and ecological models for small-area accident and injury analysis. *Accident Analysis & Prevention* 36, no. 6, 2004: 1019-1028.
83. Mannering, F., 2018. Temporal instability and the analysis of highway accident data. *Analytic Methods in Accident Research*, *17*, pp.1-13.

84. Mannering, F.L., Shankar, V. and Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic methods in accident research*, *11*, pp.1-16.
85. Mardia, K.V., 1988. Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. Journal of Multivariate Analysis, 24(2), pp.265-284.
86. Matkan, A.A., Mohaymany, A.S., Shahri, M. and Mirbagheri, B., 2013. Detecting the spatial–temporal autocorrelation among crash frequencies in urban areas. Canadian Journal of Civil Engineering, 40(3), pp.195-203.
87. McCullagh, P., Nelder, J.A. *Generalized Linear Models (2nd ed.),* Chapman & Hall. London, England, 1989.
88. Miaou, S.P., Song, J.J. and Mallick, B.K., 2003. Roadway traffic crash mapping: a space-time modeling approach. Journal of Transportation and Statistics, 6, pp.33-58.
89. Miranda-Moreno, L.F., 2006. Statistical Models and Methods for Identifying Hazardous Locations for Safety Improvements. University of Waterloo.
90. Moudon, A., L. Lin, J. Jiao, P. Hurvitz, and P. Reeves. The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington. *Accident Analysis & Prevention*, Vol. 43, No. 1, 2011, pp. 11-24.
91. National Highway Traffic Safety Administration, (2012b). Traffic Safety Facts: Bicyclists and Other Cyclists. United States Department of Transportation (DOT HS 811 624).
92. Narayanamoorthy, S., Paleti, R., & Bhat, C. R. On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. *Transportation research part B: methodological*, *55*, 2013, 245-264.
93. Nashad, T., S. Yasmin, N. Eluru, J. Lee, and M. Abdel-Aty. Joint Modeling of Pedestrian and Bicycle Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2601, 2016, pp. 119-127.
94. Noland, R.B. and Quddus, M.A. A spatially disaggregate analysis of road casualties in England. *Accident Analysis & Prevention*, *36*(6), 2004, pp.973-984.
95. Ntzoufras, I., 2009. Bayesian Modeling Using WinBUGS. John Wiley & Sons.
96. Ohlssen, D., L. Sharples, and D. Spiegelhalter. Flexible random-effects models using Bayesian semi-parametric models: applications to institutional comparisons. *Statistics in Medicine*, Vol. 26, No. 9, 2007, pp. 2088-2112.

97. Park, E., & Lord, D. 2007. Multivariate poisson-lognormal models for jointly modelling crash frequency by severity. Transportation Research Record, 2019(2019), 1-6
98. Pawlovich, M.D., W. Li, A. Carriquiry and T. Welch. 2006. Iowa's Experience with "Road Diet" Measures: Impacts on Crash Frequencies and Crash Rates Assessed Following a Bayesian Approach. Transportation Research Record: Journal of the Transportation Research Board 1953, TRB, National Research Council, Washington, DC, pp.163–171.
99. Posada, D., & Buckley, T. R. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic biology, 53*(5), 2004, pp. 793-808.
100. Pulugurtha, S.S., Sambhara, V.R., 2011. Pedestrian crash estimation models for signalized intersections. Accident Analysis and Prevention 43 (1), 439–446.
101. Quddus, M. A. Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. *Accident Analysis & Prevention*, *40*(4), 2008, 1486-1497.
102. Schepers, J.P., Kroeze, P.A., Sweers, W., Wüst, J.C., 2011. Road factors and bicycle–motor vehicle crashes at unsignalized priority intersections. Accident Analysis and Prevention 43 (3), 853–861.
103. Shankar, V.N., Ulfarsson, G.F., Pendyala, R.M., Nebergall, M.L.B., 2003. Modeling crashes involving pedestrians and motorized traffic. Safety Science 41 (7), 627–640.
104. Shirazi, M., D. Lord, S. Dhavala, and S. Geedipally. A semiparametric negative binomial generalized linear model for modeling over-dispersed count data with a heavy tail: Characteristics and applications to crash data. *Accident Analysis & Prevention*, Vol. 91, 2016, pp. 10-18.
105. Song, J. J., M. Ghosh, S. Miaou, and B. Mallick. Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping. *Journal of Multivariate Analysis,* Vol. 97, 2006, pp. 246–273.
106. Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. "WinBUGS Version 1.4 User Manual." MRC Biostatistics Unit, Cambridge, http://www.mrc-cam.ac.uk/bugs
107. Strauss, J., Miranda-Moreno, L.F., Morency, P., 2013. Cyclist activity and injury risk analysis at signalized intersections: a Bayesian modeling approach. Accident Analysis and Prevention 59, 9–17.
108. Tasic, I. and Porter, R.J., 2016. Modeling spatial relationships between multimodal transportation infrastructure and traffic safety outcomes in urban environments. Safety science, 82, pp.325-337.

109. Wang, Y., Nihan, N., 2004. Estimating the risk of collisions between bicycles and motor vehicles at signalized intersections. Accident Analysis and Prevention 36 (3), 313–321.

110. Wang, J., Huang, H., 2016. Road network safety evaluation using Bayesian hierarchical joint model. Accident Analysis and Prevention 90, 152–158.

111. Wang, J., Huang, H. and Zeng, Q., 2017. The effect of zonal factors in estimating crash risks by transportation modes: Motor vehicle, bicycle and pedestrian. Accident Analysis & Prevention, 98, pp.223-231.

112. Wanner, M., T. Götschi, E. Martin-Diener, S. Kahlmeier, and B. Martin. Active Transport, Physical Activity, and Body Weight in Adults. *American Journal of Preventive Medicine*, Vol. 42, No. 5, 2012, pp. 493-502.

113. Wardlaw, M., 2002. Assessing the actual risks faced by cyclists. Traffic Eng. Control 352–356.

114. Washington, S., M. G. Karlaftis, and F. L. Mannering 2003. Statistical and Econometric Methods for Transportation Data Analysis, Chapman & Hall.

115. Weakliem, D. L. A critique of the Bayesian information criterion for model selection. *Sociological Methods & Research, 27*(3), 1999, pp. 359-397.

116. Williams, A.F. Protecting Pedestrians and Bicyclists: Some Observations and Research Opportunities. Insurance Institute for Highway Safety, 2013, http://www.iihs. org/frontend/iihs/documents/masterfiledocs.ashx

117. Xu, P. and Huang, H., 2015. Modeling crash spatial heterogeneity: random parameter versus geographically weighting. Accident Analysis & Prevention, 75, pp.16-25.

118. Ye, X., Pendyala, R.M., Washington, S.P., Konduri, K., Oh, J., 2009. A simultaneous equations model of crash frequency by collision type for rural intersections. Safety Science 47 (3), 443–452.

119. Zeger S.L., Karim, M.R. Generalized linear models with random effects; A Gibbs sampling approach. *Journal of the American Statistical Association 86* (413), 1992, pp.79-86