STATE OF CALIFORNIA • DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV 10/98)

1. REPORT NUMBER
CA16-3006

2. GOVERNMENT ASSOCIATION NUMBER

3. RECIPIENT'S CATALOG NUMBER

4. TITLE AND SUBTITLE
UCCONNECT Summer Research Program at Cal Poly Pomona

5. REPORT DATE
December 2016

6. PERFORMING ORGANIZATION CODE

7. AUTHOR
Principal Investigator, Wen Cheng, Associate Professor, Civil Engineering Department

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Civil Engineering Department
Cal Poly University
3801 W. Temple Ave.
Pomona, CA 91768

10. WORK UNIT NUMBER

11. CONTRACT OR GRANT NUMBER
65A0528 TO 022

12. SPONSORING AGENCY AND ADDRESS
Sponsored Projects Office
University of California Transportation Center (UCTC)
2150 Shattuck Avenue, Suite #300
Berkeley, CA 94704-5940

13. TYPE OF REPORT AND PERIOD COVERED
Final Report, April 1, 2016 - October 31, 2016

14. SPONSORING AGENCY CODE
Caltrans DRISI

15. SUPPLEMENTARY NOTES
The summer program seeks to attract student applicants with interests in transportation research from underserved and underrepresented demographics (e.g. African-American, Latino, Native Americans and Pacific Islanders). An additional goal is to increase the number of students to choose transportation as their specialty and pursue graduate-level studies at one of the Region 9 campuses. The program aims to prepare and train young professionals for the transition from academia to industry practice.

16. ABSTRACT
The project scope of services and related outcomes are as follows: 1) Recruitment and Student/Faculty Selection. Outcome: A website designed to display the program-related information and facilitate the student application and recruiting process; 2) Development of Curricula for Training Seminars. Outcome: A set of specially designed training modules focused on multiple transportation modes and areas such as planning, design and operation that are related to industry practice; 3) Implementation of the UCCONNECT Summer Research Program. Outcome: About 30 young professionals getting prepared for their future graduate study and/or industry practice. The reports documenting the application of classroom learning to solve real-world issues. Potential papers submitted for presentation or publication in conferences or journals; 4) Conduction of Follow-up Survey of Participating Students/Faculty. Outcome: A survey questionnaire specially designed to evaluate the efficiency of the Program and obtain recommendations for future ones.

17. KEY WORDS
2016 UCCONNECT Summer Research Conference, Cal Poly Pomona
Wen Cheng

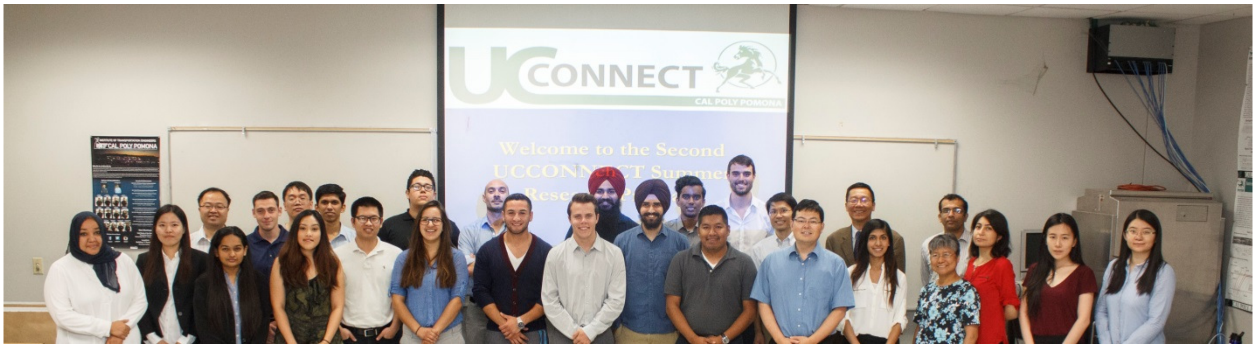
18. DISTRIBUTION STATEMENT
The readers can freely refer to and distribute this report. If there is any questions, please contact one of the authors.

19. SECURITY CLASSIFICATION *(of this report)*
No security issues

20. NUMBER OF PAGES
143

21. COST OF REPORT CHARGED
Free for E-copy

DISCLAIMER STATEMENT



Final Report:

2016 UCCONNECT Summer Research program at Cal Poly Pomona

Prepared by: Wen Cheng, Ph.D., P.E.

Associate Professor, Civil Engineering Department

Cal Poly Pomona

December, 2016

Participating Student Profile (Total 20 students)

- Gender: Female (9); Male (11);
- School: UCLA (1); UCI (1); CSU Fullerton (3); UCR (1); CPP(14)
- Academic Status: Undergraduate (10); Graduate (10)
- Major: CE (10); EE (2); Computer Science (3); Transportation System Engineering (2); Construction Engineering (2); Environmental Engineering (1)
- Ethnicities: Chicano/Latino (3), Pacific Islander (1), African American (1), White (1), Hispanic/Middle Eastern (2) and Asian (11).

Participating Faculty Profile (Total 14 faculty)

- Gender: Female (3); Male (11);
- School: CSU Fullerton (1); UCR (1); CPP(10)
- Specialty: Electrical Engineering (1); Computer Science (1); Transportation System Engineering (5); Structural Engineering (1) Construction Engineering (3); Geospatial Engineering (2); Water Treatment (1)

List of Achievements and Products

- 20 presentation made by 20 students at Symposium on August 17, 2016
- 20 final reports by 20 students on August 25,2016
- Comments and feedback provided by more than 15 students.
- 6 Peer-reviewed Journal Papers:

1. Cheng, W., G. Gill, R. Dasu, M. Xie, X. Jia, J. Zhou, “Comparison of Alternative Multivariate Poisson Lognormal Crash Frequency Models to Identify Hot Spots of Intersections based on Crash Types”, **Journal of Accident Analysis and Prevention**, (Accepted for Publication)
2. Cheng, W., X. Jiang, W. Lin, X. Wu, X. Jia, J. Zhou. “Ranking Cities for Safety Investigation by Potential for Safety Improvement”. **Journal of Transportation Safety and Security**, (Under 2nd round of review)
3. Gill, G., W. Cheng, M. Xie, T. Vo, X. Jia, J. Zhou. “Evaluating the Influence of Neighboring Structures 1 on Spatial Crash Frequency Modeling and Site Ranking Performance”, Journal of Transportation Research Record, (Accepted for Publication in 2017 TRB Annual meeting, Under 2nd round of review for publication at **Journal of Transportation Research Record**)
4. Cheng, W., G. Gill, R. Falahati, X. Jia, J. Zhou, T. Vo “Alternative Multivariate Multimodal Crash Frequency Models”, **ASCE Journal of Transportation Engineering**, (Under the first round of review)
5. Xie, M., W. Cheng, G. Gill, J. Zhou, X. Jia, S. Choi. “Comparison of Alternative Multivariate Poisson Lognormal Crash Frequency Models to Identify Hot Spots of Intersections based on Crash Types”, (Accepted for Publication in 2017 TRB Annual meeting, Under 2nd round of review for publication at **Journal of Traffic Injury and Prevention**)
6. Cheng, W., G. Gill, S. Choi, X. Jia, J. Zhou, M. Xie. “A New Approach to Addressing Temporal Correlation in Crash Frequency Modeling: Combination of Time-varying Coefficients and Autoregressive Process”, (Accepted for Publication in 2017 TRB Annual meeting, Under 2nd round of review for publication at **Journal of Traffic Injury and Prevention**)

Appendix: List of Peer-reviewed journal Papers

Paper #1: Application of Multivariate Poisson Lognormal Spatial Crash Count Model to Identify Hot Spots of Intersections based on Crash Types

Wen Cheng[a,*], Gurdiljot Singh Gill[a], Ravi Dasu[b], Meiquan Xie[a,c], Xudong Jia[a], Jiao Zhou[a]

[a] *Department of Civil Engineering, California State Polytechnic University, United States*
[b] *California Department of Public Health, United States*
[c] *School of Traffic and Transportation Engineering, Central South University, PR China*

Wen Cheng* Associate Professor, Department of Civil Engineering, California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Tel: (909) 869-2957
Email:wcheng@cpp.edu

Gurdiljot Singh Gill Graduate Research Assistant, Department of Civil Engineering, California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Tel: (909) 908-7084
Email:gurdiljotg@cpp.edu

Ravi Dasu, Research Scientist, California Department of Public Health#
Sacramento, CA 95899-7377
Tel: (916)552-9835
Email: rkmdasu@gmail.com

Meiquan Xie, Lecturer, School of Traffic and Transportation Engineering, Central South University c
Changsha, Hunan 410075, PR China
Visiting Scholar, Department of Civil Engineering, California State Polytechnic University, Pomona a

3801 W. Temple Ave., Pomona, CA 91768
Tel: (909)282-9142
Email: mxie@cpp.edu

Xudong Jia, Professor, Department of Civil Engineering, California State Polytechnic University, Pomona

3801 W. Temple Ave., Pomona, CA 91768
Tel: (909) 869-4312
Email:xjia@cpp.edu

Jiao Zhou, Graduate Student, Department of Civil Engineering, California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Tel: (714) 488-7965
Email:jiaozhou@cpp.edu

* Corresponding author

Affiliation for this author is given for information purposes only. CDPH does not validate or authenticate the views/data presented in the paper and is the sole and only view of the author.

ABSTRACT

Most of the hot spot identification (HSID) studies are focused on total crash counts with considerably less research dedicated to different crash types. This study compares four crash type-based count models with and without the multivariate and spatial correlations for HSID purpose. It is anticipated that comparison of the ranking results of the four models would identify the impact of crash type and spatial random effects on the HSID. The data over a six-year time period (2004-2009) of a set of intersections in the City of Corona, California were selected for the analysis. The crash types collected in this study include: rear end, head on, side swipe, broad side and hit object. Four evaluation tests which contain the Site Consistency Test, the Method Consistency Test, the Total Rank Difference Test, and the Total Performance Difference Test were applied to evaluate the performance of the four models. Moreover, two cutoff levels for hot spots were explored with the aim to represent different real world financial situations. Two goodness-of-fit measurements of models suggest that the strong correlations exist not only in different crash types but also in neighboring intersection across crash types. The four evaluation test results reveal that modeling performance of the four models is generally in line with the corresponding HSID performance. However, sometimes the heterogeneity uncaptured by the models might play a more important role in representing the safety condition of sites under investigation. Overall, it is suggested to develop sophisticated crash prediction models for HSID such as the model which accounts for both crash type and spatial correlations.

Keywords: Hot Spot Identification, Multivariate, Spatial Correlations, Evaluation Tests, Crash Count Models

1. INTRODUCTION AND BACKGROUND

During the year of 2014, 32,675 fatalities occurred on the US roads and the number of injuries and trauma sufferers is far greater at 2,338,000 annual injuries. In addition, road accidents were the leading cause of death among ages 16 through 24 in 2014 (NHTSA, 2016). The fatalities reflect a significant proportion of healthy lives which could have been saved by the application of appropriate safety countermeasure treatments. The traffic management processes which address safety issues include network screening, problem diagnosis, countermeasure identification, and project prioritization. Among these processes, detection of high risk sites (also called hotspots, black spots, site with promise, etc.) is of paramount importance for the improvement of driving environment from safety perspective. The consequences of inaccurate identification would result in two scenarios. First, the screening process may detect truly safe sites as unsafe. Second, truly unsafe sites are not detected, and thus the opportunity to treat the real hotspots is missed.

In general, the network screening follows into two categories: the Systemic Approach and Spot Location Approach (Preston et al. 2013). Comparatively speaking, the latter one is more traditional and relies heavily on the crash history to screen out the most unsafe locations which need remediation. Under Spot Location approach, upon completion of screening, the next step of problem diagnosis is conducted on the identified locations where site issues are usually revealed through the overrepresentation of certain crash outcomes such as rear-end, head-on, and others. Further, safety countermeasures are implemented to enhance the roadway safety situation. The effectiveness of such countermeasures is normally assessed on the basis of their benefit of crash reduction and the deployment cost. The Spot Location approach has been very popular among researchers and widely used in practice. The hot spot identification (HSID) methods of this type range from classical crash count (Deacon et al., 1975) and crash rate (Norden, 1956) methods to more sophisticated ones including Empirical Bayes (Hauer et al., 2002; Cheng & Washington, 2005; Persaud et al.,2010; Wu et al., 2014) and Full Bayesian approaches (Davis and Yang, 2001; Washington and Oh, 2006; Huang et al., 2009;Lan et al., 2009; Persaud et al., 2010), which can obliterate the Regression to the Mean (RTM) bias (Hauer, 1986; Hauer, 1996; Persaud, 1988; Hauer, 1997; Carriquiry and Pawlovich, 2004) associated with observed crash count data. Some researchers flag out the hazardous locations based on potential safety improvement or "excess" crashes (Jiang et al., 2014), while others conduct HSID through the Level of Service of Safety (Kononov & Allery, 2003 and 2004). Finally, a study by Miranda-Moreno et al. (2009) recommends incorporating crash severity and occupancy into site ranking. One condition of the success of the above mentioned Spot Location HSID methods is the availability of crash history for sites under investigation. This may become an issue in some situations. For example, there is non-availability of robust crash data for lots of rural areas, especially for the occurrence of severe crashes with typically low density (Preston et al. 2013). In such instances, the traditional Spot Location approach sometimes tends to underperform in HSID and the procedure may result in low safety benefits (Caltrans, 2015). This issue can be addressed by the Systemic approach, which is relatively new and tend to bridge the gap between hotspot detection and countermeasures implementation (Sawyer et al., 2011). Rather than filter out the sites based on crash history, this method is somewhat proactive and targets the sites lacking safety measures to prevent a specific type of crash. It mainly involves the implementation of remedial safety countermeasures, which are previously proven efficient for certain crash types such as run-off road crashes, at multiple crash locations, corridors, or geographic areas (Wang et al, 2014). In many cases, this method is more cost efficient than the Spot Location one due to the large scale impact. A major characteristic of the Systemic Approach is the crash type-oriented HSID, and a clear understanding of the interaction between crash count of various types and their causal factors is important for the successful implementation of such approach.

Most of the studies are focused on the general crashes or total crash counts while considerably less research has been dedicated to different crash types. Qin et al. (2005) using Poisson regression models developed using Markov Chain Monte Carlo methods found a nonlinear relation between crashes and daily volume, and variation in the relationship for different crash types: single-vehicle, multivehicle same direction, multivehicle opposite direction, and multivehicle intersecting. . Kim et al. (2006) used univariate Poisson and Negative Binomial models for crash counts of different types at 160 rural intersections. Data suggests that different pre-crash conditions were linked with crash types and models based on prediction of total crash frequency may fail to identify pertinent countermeasures. Subsequently, Kim et al. (2007) used Binomial multilevel modeling techniques to validate the presence of hierarchical structure in crash data which points towards the causal mechanisms in vehicular crashes (Angle, head-on, rear-end, and sideswipe) due to their relationship with roadway, environmental, and traffic factors. The effects of time and weather on crash types were explored by El-Basyouny et al. (2014) using Bayesian multivariate Poisson lognormal models for the prediction of seven crash types (Follow-Too-Close, Failure-To-Observe-Traffic Signal, Stop-Sign-Violation, Left-Turn-Across-path, Improper-Lane-Change, Struck-Parked-Vehicle, and Ran-Off-Road). This study established the strong significance of temperature, snowfall, and day of week on occurrence of different types of crashes. More recently, Jonathan et al. (2016) applied Bayesian multivariate Poisson lognormal spatial model to a group of two-lane highway segments in rural areas of Pennsylvania for HSID and compared its ranking performance to three competing models. Four categories of crashes were analyzed which include same-direction, opposite-direction, angle and hit fixed-object. Their results show that the model that considers both multivariate and spatial correlation has the best fit.

The primary goal of the present study is to compare four models of crash type based HSID methods with and without the multivariate and spatial correlations. Additionally, this study also demonstrates unique contributions and key differences from Jonathan et al. (2016). First, the study is targeted in analyzing the data of intersections rather than road segments. This would serve as an important addition as intersections are more prone to a diverse nature of crash types due to a variety of reasons (geometric limitations and interaction between pedestrians, bicyclists and vehicles, and so on). Second, instead of treating the data as a singular unit, this study divides crash dataset into two time periods of the same size, which allows us to cross validate the relative ranking performances in terms of before and after periods. Third, based on the two subgroups of data, four previously proposed HSID evaluation tests which include namely Site Consistency Test (SCT), Method Consistency Test (MCT), Total Performance difference test (TPDT) and Total Rank difference test (TRDT), are employed to assess the performance of alternative methods from different angles. Fourth, two cutoff levels for hot spots are explored which contain both top 5% and 10% of intersections. This aims to represent different real world financial situations. Lastly, five different crash types are examined: Rear end, Head on. Side Swipe, Broad Side, and Hit Object. It is expected that the inclusion of crash counts of different types would help to understand the different impacts of geometric, traffic, and environmental factors on crash type, and catch some hazardous locations which might escape the total count-based HSID methods.

The remainder of this paper first describes the four hierarchical Bayesian models four HSID purpose. These models predict the different crash outcomes happening at intersections with and without accounting for the crash type and spatial random effects. Then, we describe the four HSID evaluation tests which compare the performance of the four models from different perspectives. Following this, the data preparation and the crash prediction modeling results are then described. We then present the results of the four evaluation tests. Conclusions and recommendations follow.

2. METHODOLOGY

This study analyzed five types of crashes happened at the intersections of a city. The process involved development of four hierarchical Bayesian models for estimation of crash types and evaluation of site ranking performance based on two time periods. The four models are: univariate, multivariate, univariate spatial and multivariate spatial. The results from these methods were evaluated by four previously proposed tests to distinguish the best approach for HSID based on crash types at intersections. It is anticipated that comparison of the ranking results of the four models would identify the impact of heterogeneity and spatial random effects on the HSID. The details of different models and tests are described below:

2.1 Full Bayesian Method

A full Bayesian approach is based upon Bayes' theorem. Similar to the Empirical Bayesian (EB) method, the Full Bayesian (FB) method has been widely used in traffic safety analysis (Davis and Yang, 2001; Washington and Oh, 2006). In this study, an FB was selected over EB for three reasons. First, since the sample size in the current study is relatively small and considering the impact of the size of reference population on the validity of EB results (Lan et al., 2009), we chose FB method. Second, FB can produce a smoother integration of prior information and all available data into a posterior distribution, rather than point estimates provided by EB. Third, FB allows more complicated model specifications such as the multivariate conditional autoregressive for spatial correlations analysis, as required by this study.

2.2 Univariate Poisson-Lognormal Model (UVPLN)

This models rests on the assumption that crash occurrence of certain type j at a given location i in time period t, y_{ijt}, obeys Poisson distribution, while the corresponding observation specific error term ε_{ijt} follows normal distribution. In comparison with the normal Negative Binomial model (Poisson Gamma), this model was better suited for this study as it could better handle the low sample mean and small sample size due to the heavier tails associated with Lognormal distribution(Lord and Miranda-Moreno, 2008), The model specification of PLN can be expressed as:

$$\begin{cases} y_{ijt}|\lambda_{ijt} \sim Poisson\ (\lambda_{ijt}) \\ \ln(\lambda_{ijt}) = \boldsymbol{X'_{ijt}\beta} + \varepsilon_{ijt} \\ \varepsilon_{ijt} \sim Normal\ (0,\ \tau^2\ ) \end{cases} \tag{1}$$

Where i is the site index, j is the crash type, t is the time period index, y is the recorded crash number, λ is the expected crash number, $\boldsymbol{X}'$ is the matrix of risk factors, $\boldsymbol{\beta}$ is the vector of model parameters, ε_{ijt} is the independent random effects and τ^2 is the variance of the normal distribution for ε_{ijt}.. These random effects capture the extra-Poisson heterogeneity among intersections. The inverse of variance is known as precision and it has a gamma prior:

$$\tau^{2^{-1}} \sim Gamma(0.01, 0.01) \tag{2}$$

with prior mean equal to one and its prior variance large (equal to one hundred), representing high uncertainty or prior ignorance.

2.3 Multivariate Poisson-Lognormal Model (MVPLN)

The major difference between MVPLN and UVPLN lies in the error term ε. Rather than assume an independent ε_{ijt} across crash types as shown in Equation 1, MVPLN employs an error term $\boldsymbol{\varepsilon_{ij}}$ which is assumed to follow a multivariate normal distribution with the following expression:

$$\begin{cases} y_{ijt}|\lambda_{ijt} \sim Poisson\ (\lambda_{ijt}) \\ \ln(\lambda_{ijt}) = \boldsymbol{X'_{ijt}\beta} + \boldsymbol{\varepsilon_{ij}} \\ \boldsymbol{\varepsilon_{ij}} \sim Normal\ (\mathbf{0},\ \Sigma) \end{cases} \tag{3}$$

Where $\boldsymbol{\varepsilon_{ij}} = \begin{pmatrix} \varepsilon_{it}^{1} \\ \varepsilon_{it}^{2} \\ \varepsilon_{it}^{3} \\ \varepsilon_{it}^{4} \\ \varepsilon_{it}^{5} \end{pmatrix}$, $\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{15} \\ \vdots & \ddots & \vdots \\ \sigma_{51} & \cdots & \sigma_{55} \end{pmatrix}$ (4)

The diagonal element $\boldsymbol{\sigma_{jj}}$ in the covariance matrix of Equation 4 represents the variance of $\boldsymbol{\varepsilon_{ij}}$, where the off-diagonal elements represents the covariance of different crash types. The inverse of the covariance matrix represent the precision matrix and has the following distribution:

$$\Sigma^{-1} \sim Wishart(I, J) \tag{5}$$

Where I is the J x J identify matrix (Congdon, 2006), and J is the degree of freedom, J=5.

This model specification allows simultaneous processing of various crash types and takes into consideration the correlation between the dependent variables.

2.4 Univariate Poisson-Lognormal Spatial Model (UVPLNS)

UVPLNS is very similar to UVPLN with the exception of the additional spatial random effect, $\boldsymbol{u_{ij}}$.

Equation 1 is slightly modified to represent the model specification of UVPLNS shown as follows.

$$\begin{cases} y_{ijt}|\lambda_{ijt} \sim Poisson\ (\lambda_{ijt}) \\ \ln(\lambda_{ijt}) = \boldsymbol{X'_{ijt}\beta} + \varepsilon_{ijt} + u_{ij} \\ \varepsilon_{ijt} \sim Normal\ (0,\ \tau^{2}\ ) \end{cases} \tag{6}$$

Where u_{ij} is fit by the conditional auto-regressive model (CAR) originally proposed by Besag (1974) which can be expressed with the following distribution:

$$u_{ij}|u_{kj},\ \emptyset_j \sim N\left(\frac{\sum_{i\sim k} u_{jk}W_{ik}}{\sum_{i\sim k} W_{ik}}, \frac{\emptyset_j^{-1}}{\sum_{i\sim k} W_{ik}}\right) \tag{7}$$

Where i~k represents the neighbors of intersections i, W_{ik} is the weight intersection k has on intersection i, and Φ_j is the precision for each type j. The same gamma prior used in Equation 2 was employed for Φ_j. Various weight structures have been explored in previous studies (Xu and Huang, 2015; Aguero-Valverde and Jovanis, 2010; Guo et al., 2010) containing adjacency-based, corridor-based, distance order, distance exponential decay, semi-parametric geographically weighted, and so on. For the present analysis which focuses on the evaluation of various HSID methods, the adjacency-based first order structure was used. In other words, if i and k are adjacent, W_{ik} = 1, otherwise, W_{ik} =0.

2.5 Multivariate Poisson-Lognormal Spatial Model (UVPLNS)

Under this model, a spatial error term $\boldsymbol{u_i}$ is added to Equation 3 which leads to the following expression:

$$\begin{cases} y_{ijt}|\lambda_{ijt} \sim Poisson\ (\lambda_{ijt}) \\ \ln(\lambda_{ijt}) = \boldsymbol{X'_{ijt}\beta} + \boldsymbol{\varepsilon_{ij}} + \boldsymbol{u_i} \\ \boldsymbol{\varepsilon_{ij}} \sim Normal\ (\boldsymbol{0},\ \textstyle\sum) \end{cases} \tag{8}$$

Where $\boldsymbol{u_i}$ is fit by a zero-centered multivariate conditional auto-regressive model (MCAR, Mardia, 1998) which has a conditional normal density shown as follows:

$$u_i|u_k,\ \textstyle\sum i\ \sim N_j(\sum_{k\sim i} C_{ik}\,, u_k, \sum i\ ) \tag{9}$$

Where each $\sum i$ is a positive definite matrix representing the conditional variance matrix, and the adjacency matrix C_{ik} is of the same dimension with $\sum i$. $\sum i$ also follows the Wishart distribution as shown in Equation 5.

2.6 Goodness-of-Fit of the Models

The deviance information criterion (DIC) was used as a measure to assess the goodness of fit of the models. DIC (Spiegehalter et al., 2003) is a hierarchical modeling generalization of the AIC (Akaike information criterion). Specifically, DIC is defined as:

$$DIC = D(\bar{\theta}) + 2p_D = \overline{D(\theta)} + p_D \tag{10}$$

Where $D(\bar{\theta})$ is the deviance evaluated at the posterior means of estimated unknowns $(\bar{\theta})$, and posterior mean deviance $\overline{D(\theta)}$ can be taken as a Bayesian measure of fit or “adequacy”. p_D is motivated as a complexity measure for the effective number of parameters in a model, as the difference between $D(\bar{\theta})$ and $\overline{D(\theta)}$, i.e., mean deviance minus the deviance of the means. Similar to the AIC, DIC gives a measure of fitness of the model with the actual data. Models with comparatively lower DIC values indicate a better fit which in turn indicates that the model closely replicates the real data. As a general guideline by Spiegehalter et al., (2003), a difference of 7+ points in the DIC is treated as significant for modeling performance.

2.7 HSID EVALUATION

2.7.1 Evaluation Criteria of HSID Performance

The above four models under the FB framework were applied to the group of intersections for HSID purpose. The sites were ranked in decreasing order of the posterior mean of crash count for each crash type, i.e., λ_{ijt}. Four previously proposed evaluation tests were then employed to quantify the superiority among these methods which include the Site Consistency Test (SCT), Method Consistency Test (MCT), Total Rank Differences Test (TRDT) and Total Performance

Difference Test (TPDT). The reader wishing more detail on these tests can refer to the studies (Cheng and Washington, 2008; Huang et al., 2009; Montella, 2010; Jiang et al., 2014). The brief description of each test is presented below.

2.7.2 Site consistency test (SCT)

This test bears the assumption that an unsafe site would remain hazardous in the future if there is no safety treatment delivered. The test is conducted by computing the sum of posterior mean of crash counts, λ_{ijt}, in time period i+1 for a certain number of hotspots that were identified by various methods in previous period *i*. The larger SCT score indicates a better HSID method. The expression of SCT is shown as follows:

$$SCT_j = \sum_{k=1}^{n} Crash\ Statistic_{k(i),method=j,i+1} \tag{11}$$

Where $Crash\ Statistic_{k(i),method=j,i+1}$ is the crash statistic in time period i+1 for a site that is ranked *k* in time period *i* as identified by the HSID method *j*. The crash statistic varies from method to method. To ensure the comparability of all HSID methods under this test, the study used the relative difference of SCT which can be expressed by the following equation:

$$Relative\ Difference\ (SCT_j) = \frac{\sum_{k=1}^{n} Crash\ Statistic_{k(i),method=j,} - \sum_{k=1}^{n} Crash\ Statistic_{k(i),method=j,i+1}}{\sum_{k=1}^{n} Crash\ Statistic_{k(i),method=j,i}} \tag{12}$$

The smaller value of the relative difference of SCT indicates more reliability of the corresponding method.

2.7.3 Method consistency test (MCT)

This test shares the same premise as SCT, i.e. the sites which are actually unsafe would rank higher in both periods provided that no safety treatment is applied. It relies on the number of common sites that make it to the top ranks in consecutive time periods. A larger MCT score indicates a more preferred HSID method. By definition, MCT is expressed as:

$$MCT_j = \{k_1, k_2, \dots, k_n\}_i \cap \{k_1, k_2, \dots, k_n\}_{i+1} \tag{13}$$

Where i, k have the same definition as shown in Equation 11. Likewise, the percentage of common sites relative to the total number of hot spots can also be calculated.

2.7.4 Total rank difference test (TRDT)

This test develops upon the limitation of MCT by taking into account the rank difference of an identified hotspot during successive periods of time. The rank difference is calculated by using the following equation.

$$TRDT_j = \sum_{k=1}^{n} \left| R\left(k_{j,i}\right) - R(k_{j,i+1}) \right| \tag{14}$$

Where $R\left(k_{j,i}\right)$ is the rank of site k in time period i identified by the HSID method j.

The smaller value indicates that the particular HSID method assigns nearly same rankings to the same hotspots during successive time periods, and therefore, is more reliable.

2.7.5 Total performance difference test (TPDT)

This test is somewhat similar to total rank difference test. This test assumes that crash statistic of the same site across different time periods should remain close. The computation of TPDT is shown in Equation 15.

$$TPDT_j = \sum_{k=1}^{n} \left| (Crash\ Statistic_{k(i),method=j,i+1} - Crash\ Statistic_{k(i),method=j,i}) \right| \tag{15}$$

Where Crash statistic, i, j, k have the same definitions as shown in Equation 11.

Likewise, the crash statistic is different for different HSID methods. In order for comparable results, the study used the relative difference of TPDT which has the following expression:

$$Relative\ Difference\ (TPDT_j) = \frac{TPDT_j}{\sum_{k=1}^{n} Crash\ Statistic_{k(i),method=j,i}} \tag{16}$$

The relative difference is a percent value. The smaller the relative difference of TPDT, the better performance the method tends to have.

3. DATA DESCRIPTION

In order to capture a relationship between crashes of different types and covariates with spatial correlations being considered, a total of 137 intersections in the City of Corona, California were randomly selected for analysis. The crash severity and type were outputted for years 2004 to 2009 using Crossroads Collision Database software. The crash types collected in this study include: rear end, head on, side swipe, broad side and hit object. In addition to crash type, the crashes were divided into five crash severities: fatal, severe injury, other visible injury, complaint of pain, and non-injury. Aerial photographs and GIS maps were used to collect traffic and roadway information which include: (1) Major road speed limit, (2) minor road speed limit, (3) major road ADT, (4) minor road ADT, (5) signalized intersection, (6) at least one exclusive right turn lane on major road, (7) at least one exclusive left turn lane on major road, (8) at least one exclusive right turn lane on minor road, (9) at least one exclusive left turn lane on minor road, (10) number of lanes on major road in both directions, (11) number of lanes on minor road in both directions, (12) presence of pedestrian crossing at least one leg of major road, (13) presence of pedestrian crossing at least one leg of minor road, (14) presence of a T or three way intersection, (15) presence of a four way intersection, (16) number of driveways that are within 250' radius of center of intersection of major road, and (17) number of driveways that are within 250' radius of center of intersection of minor road. Relevant information about the various variables is shown below in Table 1.

Table 1: Variables Used in the Study

Variables	Definition	Mean	Min	Max
RE	Total number of Rear End crashes at an intersection	6.15	0	52
HEO	Total number of Head On crashes at an intersection	0.64	0	5
SS	Total number of Side Swipe crashes at an intersection	2.19	0	24
BS	Total number of Broad Side crashes at an intersection	5.43	0	64
HIO	Total number of Hit Object crashes at an intersection	0.84	0	7
AADTMAJ	Average Annual Daily Traffic on Major Road	18290	2700	49100
AADTMIN	Average Annual Daily Traffic on Minor Road	7756	1300	30200
TINT	Intersection type (1 if T or 3-way intersection, 0 if 4-way intersection)	0.34	0	1
SIGNAL	Intersection type (0 if non-signalized intersection, 1 if signalized intersection)	0.75	0	1
RTLMAJ	Right-Turn Lane Indicator (1 if at least one right-turn lane on the major road, 0 otherwise)	0.328	0	1
LTLMAJ	Left-Turn Lane Indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	0.854	0	1
RTLMIN	Right-Turn Lane Indicator (1 if at least one right-turn lane on the minor road, 0 otherwise)	0.51	0	1
LTLMIN	Left-Turn Lane Indicator (1 if at least one left-turn lane on the minor road, 0 otherwise)	0.74	0	1
DRWYMAJ	Number of Driveways on Major road within 250ft of the intersection center	1.46	0	9
DRWYMIN	Number of Driveways on Minor road within 250ft of the intersection center	1.56	0	11
SPDLIMAJ	Speed Limit on Major road in mph	40.54	35	45
SPDLIMIN	Speed Limit on Minor road in mph	38.9	25	50
PEDMAJ	Pedestrian crossing indicator (1 if at least one pedestrian crossing on the major road, 0 otherwise)	0.73	0	1
PEDMIN	Pedestrian crossing indicator (1 if at least one pedestrian crossing on the minor road, 0 otherwise)	0.85	0	1
NUMMAJ	Number of lanes on major road (both direction)	3.94	2	6
NUMMIN	Number of lanes on minor road (both directions)	2.62	1	6

4. Description of Modeling Results

The aforementioned four models were tested in freeware WinBUGS version 1.4.3 package (Spiegelhalter et al., 2003). In model calibration of MVPLN, UVPLNS, and UVPLN models, two chains of 20,000 iterations were set up for each model. After ensuring the convergence, first 5,000 samples were discarded as adaptation and burn-in. However, 50,000 iterations were used for MVPLNS since this one has substantially more complex random effects and hence require more simulations for satisfaction of the desired threshold of MC errors lower than 5% of the standard deviation of the parameters. It is noteworthy that these models should not be judged on their ability to explain the causal factors related to crash occurrence. The main purpose for developing these functions is to provide the expected crash counts for various crash types that are required for site ranking—thus the focus is on crash prediction, not explanation.

Modeling results of these models including, DIC, Dbar, the mean, standard deviation, and MC error of the posterior distribution, are presented in Table 2. Both the DIC and Dbar values indicate that the MVPLNS model is superior to the other models in fitting the crash data with the smallest DIC and Dbar values of 3,025 and 2,559, respectively. The MVPLN model places second with DIC 120 points higher than that of MVPLNS. The two univariate models are very similar and perform the worst with the DIC and Dbar values significantly larger than their Multivariate counterparts. This is consistent with previous research (Jonathan et al., 2016) and suggests that strong correlations exist among different crash types and the multivariate random errors are significantly spatially correlated.

In terms of variable significance and coefficients, it is important to note that both statistically significant variables and corresponding coefficient values vary across various crash types and models. The potential reason might be due to relative small sample size (137 intersections by 3 years) used. In comparison with other variables, the AADT for major streets (β_3) has significant impacts on crash types in most cases with different coefficient values, indicating AADT exerting varied influences on the crash types.

Table 2: Description of Results of Various Models

Types	Variables	MVPLNS			MVPLN			UVPLNS			UVPLN		
		mean	sd	MC error	mean	sd	MC error	mean	sd	MC error	mean	sd	MC error
Rear End (1)	β_0							-24.24	6.91	0.8215	-21.57	8.50	1.01
	β_3	0.69	0.26	0.0302	1.26	0.23	0.0275	1.60	0.21	0.0250	1.81	0.22	0.0253
	β_{12}	1.59	0.42	0.0373							1.01	0.47	0.4424
	β_{14}	-15.24	1.50	0.1727	-22.18	4.80	0.5698	-7.00	1.86	0.2166			
	β_{15}	-15.26	1.52	0.1755	-21.99	4.77	0.5667	-7.83	1.54	0.1772			
Head On (2)	β_0	-18.36	5.59	0.6463				-55.32	17.63	2.0930	-51.91	13.35	1.5810
	β_3										1.73	0.67	0.0788
	β_4	1.98	0.52	0.0601	1.29	0.56	0.0642	4.07	1.64	0.1948			
	β_6	-2.45	1.02	0.0491	-1.76	0.95	0.0353				-2.35	1.16	0.0516
	β_{12}	13.22	6.27	0.7327				15.79	4.12	0.4852	12.14	5.72	0.6627
	β_{14}	-44.19	7.56	0.8849	-32.94	10.12	1.1940						
	β_{15}	-44.93	7.34	0.8639	-33.14	10.66	1.1872	-36.77	20.49	2.4340			
Side Swipe (3)	β_0				-17.80	4.69	0.0721	36.31	7.44	0.8847			
	β_3	1.18	0.34	0.0398							0.43	0.20	0.0337
	β_4				1.01	0.31	0.0356						
	β_8	1.68	0.39	0.0265							0.93	0.40	0.0234
	β_{14}	-12.29	5.43	0.6434				-36.38	6.05	0.7178			
Broad Side (4)	β_0	-17.29	2.46	0.2891									
	β_3	1.16	0.26	0.0310	0.47	0.19	0.0224	0.69	0.19	0.0206	0.79	0.31	0.0370
	β_4				0.44	0.19	0.0218				0.51	0.28	0.0328
	β_8	-0.98	0.30	0.0236	-0.65	0.29	0.0198	-1.457	0.39	0.0350	-0.73	0.29	0.0172
	β_{14}				-11.00	2.18	0.2543				-14.11	6.14	0.7283
	β_{15}				-10.59	2.20	0.2571				-13.77	6.13	0.7283
Hit Object (5)	β_0	-15.53	9.25	1.095				-12.8	6.61	0.7770			
	β_1							0.19	0.08	0.0096	0.07	0.06	0.0070
	β_3	1.26	0.96	0.1132	1.86	0.59	0.0694	2.93	0.85	0.1003	1.84	0.66	0.0781
	β_{14}				-25.79	7.09	0.8374						
	β_{15}				-25.81	6.99	0.8247						
Goodness-of-fit	Dbar	2,559			2,685			2,801			2,774		
	p_D	466			460			390			421		
	DIC	3,025			3,145			3,191			3,195		

Notes: 1. Only the explanatory variables that are statistically significant at the 95% significance level are shown in the table. The blank cells indicate the variables are not statistically significant for the corresponding models. 2. β_0-Intercept; β_1-coefficient for "MAJORSPEED"; β_3-coefficient for "MAJORADT"; β_4-coefficient for "MINORADT"; β_6-coefficient for "MAJRTURN"; β_8-coefficient for

"MINRTURN"; β_{12}-coefficient for "PEDXMAJ"; β_{14}-coefficient for "TINTERSECTION"; β_{15}-coefficient for "FOURWAYINT". Refer to Table 1 for full list and descriptions of explanatory variables.

5. EVALUATION TEST RESULTS

Four evaluations tests (SCT, MCT, TRDT and TPDT) were used to analyze the relative superiority in HSID of four hierarchical Bayesian crash count models namely, MVPLNS, MVPLN, UVPLNS and UVPLN. It is expected that the method(s) with better evaluation test results would be the preferred ones(s). Following steps were followed for this evaluation procedure:

1. The dataset was evenly divided into two periods, Period 1 (2004-2006, "before" period) and Period 2 (2007-2009, "after" period).
2. Each of the four crash count models were developed for each of the five crash types, based on both before and after time periods.
3. For each intersection, the average of the three-year Bayesian estimated crash counts for both Period 1 and 2 was calculated for application of four HSID evaluation tests.
4. For each test, both top 5% and 10% were used as the cutoff level for HSID.

The detailed test results for each crash type and the aggregate one are described in the following subsections.

5.1 Site Consistency Test Results

Table 3: Site Consistency Test Results of Various Crash Count Models

	MVPLNS		MVPLN		UVPLNS		UVPLN	
CRASH TYPE	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**
Rear End (1)	**3.58%**	**-11.43%**	13.07%	-7.20%	17.29%	7.11%	35.47%	23.87%
Head On (2)	**92.75%**	**90.13%**	99.19%	98.58%	93.01%	93.44%	99.61%	91.77%
Side Swipe (3)	46.78%	54.09%	**35.59%**	**29.68%**	35.94%	41.13%	53.95%	55.79%

Broad Side (4)	**42.07%**	**36.42%**	45.56%	41.19%	47.52%	43.88%	50.85%	45.17%
Hit Object (5)	**28.81%**	**28.92%**	37.86%	32.09%	68.56%	61.33%	54.95%	62.71%
Accumulated	**42.80%**	39.63%	46.26%	**38.87%**	52.47%	49.38%	58.97%	55.86%

Note: The bold text represents the best performance in different cases. If different HSID methods share the same best performance, then each of the HSID method is highlighted with bold text.

Table 3 exhibits the relative difference of SCT from Period 1 to Period 2 following Equation 12. If a particular model is better for HSID, then the corresponding SCT will have small percentage change across the two time periods. Table 3 clearly demonstrates that MVPLNS is significantly consistent in identification of hotspots in consecutive periods as it has the lowest SCT percent change, for both 5% and 10% thresholds, in eight out of ten cases. For the Crash type 3, MVPLN performs best with the percent changes of 35.59% and 29.68%, respectively. The univariate models perform the worst in all situations. This evaluation test reflects that the multivariate models performed better than the univariate models. This trend is similar for both 5% and 10% hotspots. This indicates that there is a need to consider correlation among various crash types at the same sites for HSID. Given the MVPLNS has the best test results in most cases, it implies that accounting for spatial correlation among intersections also enhances the HSID performances.

5.2 The Method Consistency Test Results

Table 4: Method Consistency Test Results of Various Crash Count Models

	MVPLNS		MVPLN		UVPLNS		UVPLN	
CRASH TYPE	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**
Rear End (1)	**42.86%**	71.43%	**42.86%**	**78.57%**	**42.86%**	57.14%	**42.86%**	42.86%
Head On (2)	0.00%	0.00%	0.00%	0.00%	0.00%	**7.14%**	0.00%	**7.14%**
Side Swipe (3)	14.29%	**35.71%**	14.29%	21.43%	**28.57%**	28.57%	**28.57%**	28.57%
Broad Side (4)	**42.86%**	28.57%	**42.86%**	21.43%	**42.86%**	**35.71%**	28.57%	**35.71%**

Hit Object (5)	**28.57%**	**28.57%**	**28.57%**	**28.57%**	14.29%	**28.57%**	**28.57%**	14.29%
Accumulated	**25.71%**	**32.86%**	**25.71%**	30.00%	**25.71%**	31.43%	**25.71%**	25.71%

Note: The bold text represents the best performance in different cases. If different HSID methods share the same best performance, then each of the HSID method is highlighted with bold text.

Table 4 records the percentage of common sites identified in both periods relative to the total number of hot spots. The larger MCT percentage values indicate a more consistent HSID performance. Different than Table 3, it can be seen from Table 4 that the performance of alternative results is truly mixed across different crash types. For the accumulated result of top 5% sites, every model performed almost the same. In case of accumulated result of top 10%, the MVPLNS is the best method with highest percentage value of 32.86%, followed by UVPLNS and MVPLN. UVPLN again ranks the last place.

5.3 Total Rank Difference Test Results

Table 5: Total Rank Difference Test Results of Various Crash Count Models

	MVPLNS		MVPLN		UVPLNS		UVPLN	
CRASH TYPE	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**
Rear End (1)	89	131	**81**	**110**	137	250	264	384
Head On (2)	494	838	615	908	**322**	**635**	705	1128
Side Swipe (3)	201	563	250	516	**184**	553	277	**418**
Broad Side (4)	**204**	**436**	265	451	310	587	385	672
Hit Object (5)	**187**	421	221	**398**	368	599	393	623
Accumulated	**235**	478	286	**477**	264	525	405	645

Note: The bold text represents the best performance in different cases. If different HSID methods share the same best performance, then each of the HSID method is highlighted with bold text.

Table 5 shows the total rank difference test results using Equation 14. The smaller TRDT score signifies a better HSID method. The Accumulated section of Table 5 shows that MVPLNS has the lowest score of 235 for top 5%. For 10%, MVPLN and MVPLNS are very close with the MVPLN performing best with slightly lower score of 477. In terms of the different crash types, the four models again have mixed performances in different cases. For example, UVPLNS ranks the first in identifying both top 5% and 10% of head on crashes, while UVPLN outperforms others in the case of top 10% of side swipe crashes.

5.4 Total Performance Difference Test Results

Table 6: Total Performance Test Results of Various Crash Count Models

	MVPLNS		MVPLN		UVPLNS		UVPLN	
CRASH TYPE	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**
Rear End (1)	32.65%	36.90%	**28.39%**	**34.95%**	42.38%	48.68%	54.14%	56.48%
Head On (2)	**92.83%**	**90.15%**	99.02%	98.53%	92.99%	93.44%	99.60%	99.64%
Side Swipe (3)	74.13%	72.48%	80.80%	84.65%	78.47%	81.19%	**67.03%**	**64.30%**
Broad Side (4)	**46.64%**	**44.65%**	50.70%	44.78%	49.75%	51.04%	50.85%	52.03%
Hit Object (5)	65.93%	67.24%	**46.76%**	**56.88%**	86.29%	80.95%	81.43%	79.71%
Accumulated	62.44%	**62.28%**	**61.13%**	63.96%	69.98%	71.06%	70.61%	70.43%

Note: The bold text represents the best performance in different cases. If different HSID methods share the same best performance, then each of the HSID method is highlighted with bold text.

Table 6 presents the relative difference of TPDT according to Equation 16. The smaller TPDT percent value suggests that the particular HSID method is relatively better. It is known from Table 6 that multivariate models perform better than their univariate counterparts in eight out of 10 cases which include the four crash types for 5% and 10% thresholds. The accumulated results also show similar trend. In cases of crash types 1 and 5, the percent values of univariate models are almost twice the ones of MVPLN. This trend clearly depicts that multivariate models prove to be superior while handling HSID in case of intersections experiencing different types of crashes.

Overall, review of Tables 2~6 reveals that modeling performance of the four models is generally in line with the corresponding HSID performance, especially for the accumulated evaluation results, where MVPLNS and MVPLN claim the top places under all conditions. However, if we consider the cases of individual crash types, the four models have mixed HSID performance with UVPLNS and UVPLN showing the superior performances in some cases, albeit much less than those of the multivariate ones. Such phenomena indicate the models with better crash prediction performances have an overall better HSID performance. However, sometimes the heterogeneity uncaptured by the models might play a more important role in representing the safety condition of sites under investigation.

6. Conclusions and Recommendations

In this study, we applied multivariate passion log normal spatial model (MVPLNS) to a group of intersections at the City of Corona in California and compared its ranking performance with three other models including MVPLN, UVPLNS and UVPLN similar to a previous study on site ranking by crash types (Jonathan et al. 2016). Our study adds to the current literature on crash type-based ranking methods and thus provides additional tools for the recently proposed Systemic Approach for network screening of roadways (Preston et al. 2013).

Two goodness-of-fit measurements (Dbar and DIC) were used to assess the performance of the four models in fitting the crash count of various types. The findings indicate that the MVPLNS model performs the best followed by the MVPLN model. Both univariate models have similar performances that are below expectation. Our results are in agreement with the previous study

(Jonathan et al. 2016) and suggest that the strong correlations exist not only in different crash types but also in neighboring intersection across crash types.

Four previously proposed evaluation tests containing SCT, MCT, TRDT, and TPDT were applied for the assessment of alternative models via the cross validation based on before and after time periods. In terms of accumulated results which combine all crash types, the two multivariate models consistently outperform their univariate counterparts. Specifically, in the case of top 5% hot spots, MVPLNS has the greatest advantage in 3 out 4 tests, while under the condition of top 10% sites, MVPLNS and MVPLN share the best performance with each of them claiming the first place in 2 tests. It follows that the better models fitting the crash count also leads to better performance in hot spot identification. However, it is also important to note that the four models have mixed performance in terms of identifying hot spots of different crash types. For example, under the TPDT test, UVPLN outperforms others in identifying both top 5% and 10% sites of Side Swipe, while in the TRDT test, the UVPLNS is superior to others in flagging out both top 5% and 10% intersections of head on. Such phenomenon indicates the benefits associated with MVPLNS and MVPLN in the HSID are not as prominent as in the crash prediction functions. The main reason might be due to the two kinds of safety clues of each intersection under investigation (Hauer et al., 2002). The first type of clues is represented by the common traits of reference populations as shown in the crash prediction models. The second kind goes to the crash history of individual sites which may contain the unobserved heterogeneity associated with omitted variables such as pavement conditions, enforcement levels, etc. Sometimes such heterogeneity uncaptured by the models plays a significant role in representing the crash safety and therefore offset the benefits of better crash prediction models to some degree. Overall, based on the aggregate results, it is suggested to develop sophisticated crash prediction models for HSID such as MVPLNS which accounts for both crash type and spatial correlations.

Even though the study complemented the previous research through the use of intersection data, different cut-off levels for hot spots and various ranking performance evaluation tests, it is important to note that the results presented here carry some caveats. First, the sample size of the data used is relatively small (137 intersections of 3 years before and after data), and the relative performances of HSID methods may change when using crash data of larger size (this result is possible but not expected). Second, only adjacency-based first order weight matrix is used in the research when modeling the spatial random effects. Other weight structures such as corridor-based, distance order, distance exponential decay, semi-parametric geographically weighted, etc. are works in progress for further evaluation of the current paper's findings.

ACKNOWLEDGMENTS

This research received the grant from UCCONNECT summer research program.

REFERENCES

Aguero-Valverde,J., Paul, J., 2010. Spatial correlation in multilevel crash frequency models: Effects of different neighboring structures. Transp. Res. Rec.: J. Transp. Res. Board 2165, 21-32.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc. Ser. B 36 (2), 192–236.

Carriquiry, A.L., Pawlovich, M.,2004. From Empirical Bayes to Full Bayes: Methods for Analyzing Traffic Safety Data, White Paper. Iowa Department of Transportation, Ames, Iowa.

Cheng, W., Washington, S. P., 2005. Experimental evaluation of hot spot identification methods. Accid. Anal. Prev. 37(5), 870-881.

Cheng, W., Washington, S., 2008. New criteria for evaluating methods of identifying hot spots. Transp. Res. Rec. 2083, 76–85.

Congdon, P., 2006. Bayesian Statistical Modeling, 2nd ed. Wiley, New York.
Davis, G.A., Yang, S., 2001. Bayesian identification of high-risk intersections for older drivers via Gibbs sampling. Transp.Res.Rec.1746, 84–89.
Deacon, J.A., Zegeer, C.V., Deen, R.C., 1975. Identification of Hazardous Rural Highway Locations. Transp.Res.Rec.543, 16-33.
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. Accid.Anal.Prev.73, 91-99.
Guo, F.,Wang, X., Abdel-Aty,M. A., 2010. Modeling signalized intersection safety with corridor-level spatial correlations. Accid. Anal. Prev. 42(1),84-92.
Hauer, E., 1986. On the estimation of the expected number of accidents. Accid. Anal. Prev. 18 (1), 1–12.
Hauer, E., 1996. Identification of sites with promise. Transp. Res. Rec. 1542, 54–60.
Hauer, E., 1997. Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety. Pergamon Press.
Hauer, E., Harwood, D.W., Council, F.M., Griffith, M.S., 2002.Estimating Safety by the Empirical Bayes Method: A Tutorial. Transp. Res. Rec. 1784, 126-131.
Huang,H., Chin, H., Haque, M., 2009. Empirical evaluation of alternative approaches in identifying crash hot spots: naïve ranking, empirical Bayes, and full Bayes methods. Transp. Res. Rec.: J. Transp. Res. Board 2103, 32-41.
Jiang, X.M., Abdel-Aty, M., Alamili, S., 2014. Application of Poisson Random Effect Models for Highway Network Screening. Accid. Anal. Prev. 63, 74-82.
Jonathan, A. Wu, V., Donnell, K. F.K., Donnell, E.T., 2016. A multivariate spatial crash frequency model for identifying sites with promise based on crash types. Accid.Anal.Prev.87, 8-16.
Kim, D., G., Washington, S., Oh, J., 2006.Modeling crash types: New insights into the effects of covariates on crashes at rural intersections. J. Transp. Eng.132 (4), 282-292.
Kim, D., G., Lee, Y., Washington, S., Choi, K., 2007. Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. Accid. Anal. Prev. 39(1), 125-134.
Kononov, J.,Bryan, A., 2003, Level of service of safety: Conceptual blueprint and analytical framework. Transp. Res. Rec.: J. Transp. Res. Board 1840, 57-66.
Kononov, J.,Bryan, A., 2004. Explicit consideration of safety in transportation planning and project scoping. Transp. Res. Rec.: J. Transp. Res. Board 1897, 116-125.
Lan, B., Persaud, B., Lyon, C., Bhim, R., 2009. Validation of a full Bayes methodology for observational before–after road safety studies and application to evaluation of rural signal conversions. Accid. Anal. Prev. 41 (3), 574–580.
Local Roadway Safety, A., 2015. Manual for California's Local Road Owners. Caltrans, Version, 1.2.
Lord, D., 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), 751-770.
Mardia, K.V., 1988. Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. J. Multivar. Anal. 24 (2), 265–268.
Miranda-Moreno, Fu,L., Ukkusuri, S., Lord, D., 2009. How to Incorporate Accident Severity and Vehicle Occupancy into the Hot Spot Identification Process? Transp. Res. Rec.: J. Transp. Res. Board 2102, 53-60.

Montella, A., 2010. A comparative analysis of hotspot identification methods. Accid. Anal. Prev. 42(2), 571-581.
National Highway Traffic Safety Administration. 2016. NHTSA. Quick facts 2014.(Report DOT HS 812234).
Norden, M., Orlansky, J., Jacobs, H., 1956. Application of Statistical Quality-Control Techniques to Analysis of Highway-Accident Data. Bulletin 117, HRB, National Research Council, Washington, D.C.: 17-31.
Persaud, B.N., 1988. Do traffic signals affect safety? Some methodological issues. Transp. Res. Rec. 1185, 37–46.
Persaud, B., Lan, B., Lyon, C., Bhim, R., 2010. Comparison of empirical Bayes and full Bayes approaches for before-after road safety evaluations. Accid. Anal. Prev.42, 38–43.
Preston, H., Storm, R., Bennett, J.D., Wemple, B., 2013. Systemic Safety Project Selection Tool (No.FHWA-SA-13-019).
Qin, X.,Ivan, J.N., Ravishanker,N., Liu, J., 2005. Hierarchical Bayesian Estimation of Safety Performance Functions for Two-Lane Highways Using Markov Chain Monte Carlo Modeling. J. Transp. Eng. 131(5), 345–351.
Sawyer, M., Fiedler, R., Nabors, D., McGee, H., Miller, J., Barrington, M., 2011. A Focused Approach to Safety Guidebook, Report FHWA-SA-11-44. FHWA.
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
Wang, X., Xie, K., Abdel-Aty, M., Chen, X., Tremont, P., 2014. Systematic Approach to Hazardous-Intersection Identification and Countermeasure Development. J. Transp. Eng.140 (6), 1-12.
Washington, S., Oh, J., 2006. Bayesian Methodology Incorporating Expert Judgment for Ranking Countermeasures Effectiveness under Uncertainty: Example Applied to at Grade Railroad Crossings in Korea. Accid. Anal. Prev. 38, 234–247.
Wu, L., Zou, Y., Lord, D., 2014. Comparison of sichel and negative binomial models in hot spot identification. Transp. Res. Rec.: J. Transp. Res. Board 2460, 107-116.
Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. Accid. Anal. Prev. 75, 16-25.

Paper #2: Ranking Cities for Safety Investigation by Potential for Safety Improvement

Wen Cheng* Associate Professor, Department of Civil Engineering, California State Polytechnic University, Pomona

3801 W. Temple Ave., Pomona, CA 91768
Tel: (909) 869-2957
Email:wcheng@cpp.edu

Wei Hua Lin, Associate Professor, Department of Systems and Industrial Engineering and Engineering Management, University of Arizona

1127 E. James E. Rogers Way, Tucson, AZ
Tel: (520)621-6553
Email:weilin@sie.arizona.edu

Xudong Jia, Professor, Department of Civil Engineering, California State Polytechnic University, Pomona

3801 W. Temple Ave., Pomona, CA 91768
Tel: (909) 869-4312
Email:xjia@cpp.edu

Xinkai Wu, Assistant Professor, Department of Civil Engineering, California State Polytechnic University, Pomona

3801 W. Temple Ave., Pomona, CA 91768
Tel: (909) 869-3330
Email:xinkaiwu@cpp.edu

Jiao Zhou, Graduate Studnet, Department of Civil Engineering, California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Tel: (909) 869-4312
Email:jiaozhou@cpp.edu

* Corresponding Author

ABSTRACTS

The authors performed a city-level hotspot identification by using the four-year data of 265 cities in California. It is intended to equip road safety professionals with more useful information to compare the safety performance of city as a whole. Potential for Safety Improvement (PSI) was adopted as a measure of crash risk to compare alternate HSID methods, including the Empirical Bayes and three full Bayesian alternatives, Negative-Binomial Poisson Log-Normal, and the Poisson Temporal Random Effect, for ranking the safety performance of cities. Five evaluation tests which contain the Site Consistency Test, the Method Consistency Test, the Total Rank Difference Test, the Total Performance Difference Test and the Total Score Test were applied to evaluate the performance of the four HSID methods. Moreover, two cutoff levels, top 5% and 10% cities, were employed for more reliable results.

Overall, the study results are consistent with the results of previous quantitative evaluations focused on micro-level HSID. The three FB approaches significantly outperform the EB counterpart. The method accounting for temporal random effect produces more reliable HSID results than those without considering the serial correlations in collision counts.

Keywords: Hotspot Identification, the empirical Bayes, the full Bayesian, city-level

1. INTRODUCTION

Identification of hot spot (HSID), also known as site with promise, black spots, or accident-prone locations, is an important task in road and traffic safety which seeks to screen out the hazardous locations in a roadway network for further improvement. The importance of this task has been echoed in various transportation bills including the Intermodal Surface Transportation Efficiency Act (ISTEA), the subsequent Transportation Efficiency Act for the 21st Century (TEA-21), the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU), and currently the Moving Ahead for Progress in the 21st Century Act (MAP 21). The Federal transportation legislation requires each state to develop a work plan outlining strategies to implement Safety Management Systems (2003) and submit an annual report describing at least 5% of their highway locations demonstrating the most severe safety needs.

In the last several decades, there has been a fairly extensive literature focused on methods for ranking sites for further investigation. There are papers that discuss methods based on crash count or frequency (Deacon et al., 1975), papers that employ crash rate and rate-quality control (Norden, 1956; Stokes and Mutabazi, 1996).To correct for the regression-to-the-mean (RTM) bias associated with typical HSID methods (Hauer, 2002), some researchers have suggested using the Empirical Bayes (EB) techniques (Hauer, 1986; Hauer and Persaud, 1987; Hauer et al., 1988; Hauer et al., 1991). The EB method combines clues from both the crash history of a specific site and expected safety of similar sites, and has the advantage of revealing underlying safety problems which otherwise would not be detected. However, it is also revealed that EB has a limitation of ignoring the uncertainty in the variances of the sites to be studied and the reference population (Carlin and Louis, 2000). Therefore, more recently, scholars started to employ the Full Bayesian (FB) models for hotspot identification which include, but not limited to, Bayesian multivariate Poisson Log-Normal models by Aguero-Valverde and Jovanis (2009), alternative FB models by Huang et al. (2009) and Poisson random effect models by Jiang et al (2014). Rather than using overall crash frequencies at sites, some researchers have suggested using the potential for safety improvement to identify hot spots (Hakkert and Mahalel, 1978; McGuigan, 1981 & 1982; Persaud, 1999, hereafter referred to as PSI for Potential for Safety Improvement). The PSI-based methods rest on the premise that only "excess" crashes over those expected from similar sites can be prevented by applying appropriate treatments, and thus the potential for reduction is a better method for identifying sites with promise. Finally, there are papers that emphasize the importance of crash severity and costs (Tamburri and Smith, 1970; Taylor and Thompson, 2006).

One prominent characteristic existing in above methods is that they were all applied to micro-level locations such as intersections and roadway segments for network screening purpose. As for macro-level safety analysis, majority of the research studies are dedicated to crash modeling development aiming to incorporate safety into transportation planning or to link aggregate crash counts with various variables such as exposure, socioeconomic and demographic factors. For instance, Lovegrove and Sayed (2006) developed 47 community-based collision prediction models, each significantly associated with one or more of 22 explanatory variables. Kim et al. (2006) used various linear regression models to explore the relationships between land use, population, employment by sector, economic output, and motor vehicle crashes in a uniform 0.1-mi^2 (0.259-km^2) grid structure in Hawaii. There are also many studies focused on the development of zonal models of crash frequency at Traffic Analysis Zone (TAZ) level. Ladron de Guevara et al. (2004) applied simultaneous Negative Binomial (NB) crash model to demonstrate that planning-level data or traffic analysis zone information in Tucson, AZ, including population density, employment, and other information, could be significantly related to crashes. Abdel-Aty et al. (2011) investigated the association using NB model between crash frequencies and various types of trip productions and attractions in combination with the road characteristics of 1,349 TAZs of four counties in the state of Florida. Hadayeghi et al. (2007) developed a series of zonal-level collision prediction models using a generalized linear regression modeling approach to explore relationships between collision frequency in a planning zone and some explanatory variables such as traffic intensity, land use, and traffic demand measures. Yet Aguero-Valverde (2013) adopted Full Bayes hierarchical approach to estimate the models of crash frequency at canton level for Costa Rica. An intrinsic multivariate conditional autoregressive model was used for modeling spatial random effects. Finally, there is research centered on modeling development of various safety performance measures at state level considering temporal random effect (Kweon, 2008).

Compared with the large number of studies focused on the development of various macrolevel crash models, considerably less research has been dedicated to performing hotspot identification using pertinent models at such levels. Miaou and Song (2005) used Bayesian generalized linear mixed models with multivariate spatial random effects to rank sites by crash cost rate at the county level. Three crash severities were analyzed: fatal, incapacitating injury and non-incapacitating injury crashes. Subsequently, based on 35 previously developed zonal collision prediction models, Lovegrove and Sayed (2007) conducted a black spot study with data from 577 urban and rural neighborhoods across Greater Vancouver in British Columbia, Canada. Several collision-prone zones were identified and ranked for diagnosis. The identification criterion selected is the probability that EB safety estimate of specific location exceeds the regional average or norm for locations with identical traits. Finally, two zones were analyzed in detail and revealed several potential enhancements to conventional methods. Therefore, they recommended that macroreactive use has the potential to complement traditional road safety improvement programs.

In order to add more research to the current limited literature centered on macrolevel hotspot identification, this paper investigates macrolelvel crash modeling use in hotspot identification and represents a natural continuation of the above two methods, with a number of important differences and unique contributions. First, the network screening was conducted at city level, rather than county or zonal level in previous research. Second, a set of popular and comparable HSID methods, or, EB, FB with NB models, FB with Poisson Lognormal (PLN) models, and FB with Poisson temporal random effect (PTRE) models, were investigated. Third, PSI was chosen as the measure of crash risk to conduct the network screening. Finally, the performance of various alternative HSID methods were evaluated by various criteria which include the Site Consistency Test (SCT), the Total Performance Difference Test (TPDT), the Method Consistency Test (MCT), the Total Rank Difference Test (TRDT), and the Total Score Test (TST). The research results are anticipated to equip road safety professionals with additional information for comparing and assessing the safety performance of city as a whole such as the HSID method selection, performance evaluation criteria choice, etc., and therefore aid the states in allocating the appropriate proportion of federal safety funds to various cities with confidence.

The remainder of this paper first describes the HSID methods to be compared in the analysis. Then, we describe and develop the analytics of the five performance criteria. The data preparation and the crash prediction modeling development are then described. We then present the results of a comprehensive test of the HSID methods using the 5 criteria described. Conclusions and recommendations follow.

2. METHODOLOGIES

2.1 Empirical Bayesian Method

EB technique was originally applied to traffic safety in 1980's due to its great advantage of addressing the well-known regression to the mean (RTM) issue of recorded crash statistics. This method rests on two assumptions: crash occurrence at a given location obeys the Poisson probability law, while probability distribution of the expected safety of the population of sites is gamma distributed. On the basis of these assumptions, the probability that a site has a random number of crashes is approximated by the negative binomial (NB) probability distribution. This method combines clues from both the crash history of a specific site and expected safety of similar sites, and has the advantage of revealing underlying safety problems which otherwise would not be detected. With EB method gradually becoming the standard and staple of professional practice, Hauer *et al*.(2002) provided a detailed tutorial on EB which features a series of application examples. The readers can refer to the paper for the EB details. Shown below are the basics of the EB method.

In the EB method, the expected safety of a site λ_i is expressed as follows:

$$\lambda_i = wE[\lambda_i] + (1 - w)x_i \quad (1)$$

Where w is a weight factor, $E[\lambda_i]$ is the expected safety of a reference population of the specific location, and x_i is the observed count history for site i. The w (weigh factor) can be calculated through the following equation:

$$w = E[\lambda_i]/\{E[\lambda_i] + VAR[\lambda_i]\} \quad (2)$$

Where $VAR[\lambda_i]$ is the corresponding variance of the expected safety of a reference population. If a safety performance function (SPF), or, crash prediction model, for the reference population which relate crashes to covariates can be developed, w can be rewritten as follows (Hauer et al., 2002):

$$w = [1 + (\mu * Y)/\varphi]^{-1} \quad (3)$$

Where μ is expected number of crashes/km-year on similar segments or crashes/year expected on similar intersections, Y is the number of years of crash count data used, and φ is the overdispersion parameter which is a constant for the SPF and is derived during the regression calibration process.

By definition, PSI for the EB method can be derived as:

$$PSI_i = \lambda_i - E[\lambda_i] \quad (4)$$

One point worth mentioning is that the study is dedicated to city-level analysis. With crashes being spatially highly aggregated, the advantage of EB in removing the RTM issue at the city-level might not be so propounded as compared to micro-level analysis. However, considering the over-dispersion existing in city-level crashes and as an excellent counterpart of FB, the EB is still considered in the study as some others (e.g., Lovegrove and Sayed, 2007) where EB is adopted for macro-level analysis.

2.2 Full Bayesian Methods

A full Bayesian approach is the other method under Bayes' theorem. Similar to the EB method, the FB method has also enjoyed wide applications in safety analysis (Davis and Yang, 2001; Aguero-Valverde and Jovanis, 2009; Lan et al., 2009; Washington and Oh, 2006), especially with the availability of the software package WinBUGS (Spiegelhalter et al., 2003). Even though numerous studies have illustrated favorable results yielded by the EB method (Higle and Hecht, 1989; Maher and Mountain, 1988; Cheng and Washington, 2005 & 2008), some researchers also noticed the limitations associated with the EB approach (Huang et al., 2009; Persaud et al., 2010). In EB analysis, an external SPF has to be calibrated based on the locations of similar traits. Sometimes the limited reference samples can significantly impact the validity of the analysis results. Another criticism goes to the EB's inadequate capability to explicitly account for the "uncertainty" of model parameters and coefficients. Once the SPF is developed, all the model parameters and coefficients are treated as constant values and then are incorporated into the point estimates of the long-term safety of candidate sites. Through empirical analyses and/or comparisons, a set of studies (Miranda-Moreno and Fu, 2007; Miaou and Lord, 2003; Pawlovich et al., 2006) revealed the potential advantages of the FB approach relative to the EB one: its capability to seamlessly integrate prior information and all available data into a posterior distribution (rather than point estimate), its capability to provide more valid safety estimates in smaller data samples, its capability to allow more complicated model specifications. In addition to the normal Poisson-Gamma distribution, the FB models are also capable of accommodating the Poisson-Lognormal distribution and various Hierarchical Poisson distributions which can address the serial and spatial correlations among the sites. In this study, HSID using the FB approach with alternative model specifications including Poisson-Gamma (or, NB) model, Poisson Lognormal (PLN) model, and Poisson temporal random effect (PTRE) model, were assessed. The details are shown in the following subsections.

2.2.1 Model 1: Poisson-Gamma Model/Negative Binomial Model (NB)

Under the Poisson–Gamma assumption, crash occurrence at a given location i in time period t, y_{it}, obeys Poisson distribution, while the associated observation specific error term ε follows gamma distribution. The framework of FB NB regression model can be expressed as Eq. 5.

$$\begin{cases} y_{it}|\lambda_{it} \sim Poisson\ (\lambda_{it}) \\ \ln(\lambda_{it}) = \boldsymbol{X}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \\ \varepsilon_{it} \sim Gamma\ (\alpha, \frac{1}{\alpha}) \end{cases} \tag{5}$$

Where i is the site index, t is the time period index, y is the recorded crash number, λ is the expected crash number, $\boldsymbol{X}'$ is the matrix of risk factors such as Daily Vehicle Miles Traveled (DVMT) and population, $\boldsymbol{\beta}$ is the vector of model parameters, ε_{it} is the random effects, and α is the hyper-parameter of the model.

2.2.2 Model 2: Poisson-Lognormal Model (PLN)

Under the Poisson–Lognormal assumption, everything remains the same as in NB model except the error term ε_{it} follows normal distribution. In comparison with NB model, Lord et al. (Lord and Miranda-Moreno, 2008) found that PLN model could be a better alternative in case of low sample mean and small sample size. The potential reason is due to the heavier tails associated with Lognormal distribution compared to those of the Gamma distribution. Since it is not always clear to choose one model over the other, it is advisable to select model by comprehensive model diagnostics. The model specification of PLN can be expressed as:

$$\begin{cases} y_{it}|\lambda_{it} \sim Poisson\ (\lambda_{it}) \\ \ln(\lambda_{it}) = \boldsymbol{X}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \\ \varepsilon_{it} \sim Normal\ (0,\ \tau^2\ ) \end{cases} \tag{6}$$

Where i, t, y, λ, $\boldsymbol{X}, \boldsymbol{\beta}$, ε_{it} have the same definition as shown in Equation 5, and τ^2 is the variance of the normal distribution for ε_{it}.

2.2.3 Model 3: Poisson Temporal Random Effect Model (PTRE)

As shown in Equation 6, ε_{it} in PLN model varies across different sites and time periods. Therefore, like NB model, PLN can also address the over-dispersion issue in the regular Poisson model. However, someone might argue that the same city shares identical unobserved features across various time periods. In other words, the error term ε in the above equation should change merely over locations, rather than across years. This is the so-called temporal correlation. Under this assumption, the model specification can be expressed as follows:

$$\begin{cases} y_{it}|\lambda_{it} \sim Poisson\ (\lambda_{it}) \\ \ln(\lambda_{it}) = \boldsymbol{X'_{it}\beta} + \varepsilon_i \\ \varepsilon_i \sim Normal\ (0,\ \tau^2\ ) \end{cases} \quad (7)$$

Where *i, t, y,* λ, ***X, β,*** and τ^2 have the same definition as shown in Equation 6, and ε_i is a city specific random effect term. For each city, this element is generated independently from a Normal distribution, hence it is not correlated with explanatory variables. For details of random effect models, the readers can refer to Greene (2010).

Accordingly, the PSI under these models (NB, PLN, and PTRE) can be calculated as follows:

$$\begin{cases} PSI_{it} = e^{x'\beta} * (e^{\varepsilon_{it}} - 1),\ for\ NB\ and\ PLN\ models \\ PSI_{it} = e^{x'\beta} * (e^{\varepsilon_i} - 1),\ for\ PTRE\ model \end{cases} \quad (8)$$

2.2.4 Goodness-of-Fit of the Models

Two measures were selected to assess the good ness of fit of such models: The deviance information criterion (DIC) and mean squared predictive error (MSPE).

DIC (Spiegehalter et al., 2003) is a hierarchical modeling generalization of the AIC (Akaike information criterion) and BIC (Bayesian information criterion). It is particularly useful in Bayesian model selection problems where the posterior distributions of the models have been obtained by Markov chain Monte Carlo (MCMC) simulation. Specifically, DIC is defined as:

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

Where $D(\bar{\theta})$ is the deviance evaluated at the posterior means of estimated unknowns ($\bar{\theta}$), and posterior mean deviance $\overline{D(\theta)}$ can be taken as a Bayesian measure of fit or "adequacy". p_D is motivated as a complexity measure for the effective number of parameters in a model, as the difference between $D(\bar{\theta})$ and $\overline{D(\theta)}$, i.e., mean deviance minus the deviance of the means. Similar to the AIC, DIC gives a measure for how well each model fits the data and penalties for the number of parameters. Models with lower DIC values provide a better fit.

MSPE is the alternative goodness-of-fit measure and can be used to assess the model efficiency. Similar to DIC, smaller MSPE value indicates a preferred model fit. In specific, MSPE is defined as shown in Eq. 10.

$$MSPE = \frac{1}{n}\sum\left(y_{it}^{pred} - y_{it}^{obs}\right)^2 \quad (10)$$

Where y_{it}^{pred} is the model prediction of expected crash number at site *i* in time period *t* and y_{it}^{obs} is the observed crash number for the specific site.

2.3 Evaluation Criteria of HSID Performance

Compared with the large number of studies focused on the development of various HSID methods, considerably less research has been dedicated to devising the evaluation criteria for comparing the performance of various methods. Hauer and Persaud (1984) proposed the use of false identifications, consisting of false negatives and false positives, to measure the performances of various methods for HSID. Based on these two statistics, Elvik (2007) presented two diagnostic criteria including sensitivity and specificity. Subsequently Cheng and Washington (37) developed four new evaluation criteria containing the Site Consistency Test (SCT), the Method Consistency Test (MCT), the Total Rank Difference Test (TRDT), and the Poisson Mean Difference Test (PMDT), where PMDT can be conducted only when the "true" Poisson mean of crash history is known. On the basis of this study, later Montella (2010) proposed a new criterion called Total Score Test (TST), which is a weighted combination of the SCT, MCT and TRDT criteria. Finally, Jiang et al. (2014) modified these criteria to make them suitable for the PSI-centered HSID methods. In addition, they also proposed a new evaluation criterion entitled Total Performance Difference Test (TPDT). Since the paper utilized PSI as a measurement of crash risk, the five criteria developed or modified by Jiang et al., or, SCT, TPDT, MCT, TRDT, and TST, were employed in the study to assess the HSID performance. The readers wishing details like assumptions, procedures and advantages associated with each test, can refer to the pertinent papers. The following subsections present the succinct description of these tests.

2.3.1 Total Rank Difference Test (TRDT)

This test relies on site ranking to evaluate the performance of HSID methods. The test is conducted by calculating the sum of total rank differences (absolute value) of the hazardous road sections identified in successive time periods. The smaller is the total rank difference, the more reliable the HSID method is. TRDT test is expressed as:

$$TRDT_j = \sum_{k=1}^{n} \left| R\left(k_{j,i}\right) - R(k_{j,i+1}) \right| \tag{11}$$

Where $R\left(k_{j,i}\right)$ is the rank of site k in time period i identified by the HSID method j.

2.3.2 Method Consistency Test (MCT)

The test relies on the number of common sites that are identified in both time periods i and $i+1$ to evaluate the performance of HSID method. The underlying assumption is that a site identified as high risk in previous period should also reveal inferior safety performance should the crash determinants be not significantly changed. The greater number of sites consistently identified in successive periods, the more consistent and reliable the HSID method is. In specific, MCT is expressed as:

$$MCT_j = \{k_1, k_2, \dots, k_n\}_i \cap \{k_1, k_2, \dots, k_n\}_{i+1} \tag{12}$$

Where i, k have the same definition as shown in Equation 11.

2.3.3 Site Consistency Test (SCT)

The SCT is used to measure the ability of a HSID method to consistently identify a site as high risk over subsequent observation periods. It has the same premise as does MCT. The test is conducted by computing the sum of PSI in time period i+1 for a certain number of hotspots that were identified by various methods in previous period i. The larger the SCT score, the more reliable the HSID method in capturing sites expected to have crashes in the future. By definition, the expression of SCT is shown as follows:

$$SCT_j = \sum_{k=1}^{n} PSI_{k(i),method=j,i+1} \tag{13}$$

Where $PSI_{k(i),method=j,i+1}$ is the PSI in time period i+1 for a site that is ranked k in time period i as identified by the HSID method j.

2.3.4 Total Performance Difference Test (TPDT)

The SCT test requires the methods being evaluated to produce similar estimates of PSI. If one method yields higher PSI value in general, then it is expected to have higher SCT score under the test. To address the issue, Jiang et al. (2014) proposed the TPDT test which assumes that the hotspots identified by method j with all years of crash data are true hazardous sites. For the top k true hotspots, the absolute difference of PSI estimated in successive time periods is computed and summed. The smaller the difference, the better is the corresponding HSID method. TPDT is expressed as below:

$$TPDT_j = \sum_{k=1}^{n} \left| (PSI_{k(i),method=j,i+1} - PSI_{k(i),method=j,i}) \right| \tag{14}$$

Where i, j, k have the same definition as shown in Equation 11.

2.3.5 Total Score Test (TST)

The TST test was originally proposed by Montella (2010) and later modified by Jiang et al. (2014) for its application to PSI-based HSID methods. It is a weighted score of previous test criteria. The highest possible TST score is 100, which indicates the corresponding method performs best from every aspect. Specifically, TST is calculated as follows:

$$TST_j = \frac{100}{4} * \left[\frac{SCT_j}{maxSCT} + \left(1 - \frac{TPDT_j - minTPDT}{maxTPDT}\right) + \frac{MCT_j}{maxMCT} + (1 - \frac{TRDT_j - minTRDT}{maxTRDT})\right] \quad (15)$$

Finally, it is worth mentioning again the importance of the underlying homogeneity assumption for the above tests: the expected safety performance of cities remains virtually unaltered over successive periods. Therefore, it is highly recommended that the practitioner carefully check the cities and ensure they are in similar operational state across adjacent years. It is possible that some cities provided area-wide treatments and successfully reduced the number of crashes during the study period. It is advisable to exclude such cities when it is possible that some cities provided area-wide treatments and successfully reduced the number of crashes during the study period.

3. DATA PREPERATION

The data used in this study were collected from three sources: Statewide Integrated Traffic Records System (SWITRS), Highway Performance Monitoring System (HPMS) and California Department of Finance.

Collisions of various cities in California that occurred from 2008-2011 were obtained from SWITRS that contains 6 different categories of collisions (total fatal and injury collision, alcohol involved fatal and injury collision, pedestrian involved collision, bicycle involved fatal and injury collision, motorcycle involved collision, and property damage collision) and 5 categories of victim (vehicle driver, vehicle passenger, bicyclist, motorcyclist, and total victims count). The study focuses only on the total fatal and injury collisions of the cities. In addition, a main exposure-related factor of city safety performance, that is, Daily Vehicle Miles Travel (DVMT), was collected from Highway Performance Monitoring System (HPMS) for the same time periods. Furthermore, a main demographic factor, or, population, was gained from the California Department of Finance.

In order to improve the modeling accuracy, the collected data were then further separated into homogeneous groups based on Population and DVMT sizes. The results demonstrated in the paper are related with a group of 265 cities with small to medium size of population and DVMT.

Summary information for the various cities in terms of population, DVMT and total fatal and injury collision number is shown in Table 1.

Note: S.D. represents standard deviation.

4. DEVELOPMENT OF COLLISION PREDICTION MODELS

The development of collision prediction models (CPM, or, SPF under EB method) is described in this section. It is worth stressing that these models should not be judged on their ability to explain the causal factors related to collision occurrence. The main purpose for developing these functions is to provide the expected collision counts for specific cities that are required to apply the EB and FB methods—thus the focus is for crash prediction, not explanation. As shown in Table 1, originally both population and DVMT of the 265 cities were collected to serve as the explanatory variables. However, during the modeling diagnostic process, it was noticed that there is strong correlation between population and DVMT. In order to obliterate the potential bias caused by the collinearity between population and DVMT, the authors decided to drop the variable of Population from the CPM's. Therefore, only DVMT is included in the CPM's as an independent variable.

As mentioned in the Methodologies Section, due to overdispersion of collisions observed on various cities in CA, Negative Binomial regression models were generally fit in order to generate SPFs when using EB. Hereafter referred to as EBNB. In order to reduce the random effects of the collision data, the model was fitted with the mean collision frequency of each city in 4 years (2008–2011) as the target variable, and the logarithm of the mean DVMT as the predictor. Thus, the total number of observation for the EB NB model is 265.

Additionally, three models were fitted using FB approach which include the Negative Binomial (FBNB), Poisson Log-Normal (FBPLN), and Poisson temporal random effect (FBPTRE). These models were developed with the original panel data for 4 years separately. In other words, the total fatal and injury collision record for each city in each year was treated as one observation in the models. Hence, the total numbers of observations for the three models are 1060 (265*4).

When implementing the FB analysis in freeware WinBUGS version 1.4.3 package , uninformative priors were assumed with normal distribution (0, 1000) for all regression coefficients ($\boldsymbol{\beta}$), and with gamma distribution (0.001, 0.001) for hyper-parameters associated with the disturbance terms, i.e., 1/ a in FB NB model, τ^2 in FB PLN and FB PTRE models. In model calibration, two chains of 20000 iterations were set up for each model. After ensuring the convergence, first 5000 samples were discarded as adaptation and burn-in.

Modeling results of these models including, DIC, MSPE, the mean, standard deviation, and 2.5% and 97.5% quintiles of the posterior distribution, are presented in Table 2. Both the DIC and MSPE values indicate that the FBPTRE model is superior to the FBPLN and FBNB model in fitting the crash data, while the latter two share almost the same performance. FB PLN has relatively smaller MSPE, but its DIC is somewhat larger than that of FBNB. Note that Lord et al. (2008) previously found that PLN model could be a better alternative in case of low sample mean and small sample size. The same phenomenon is not revealed herein may be due to the large sample mean associated with the more aggregate level of city collision. It is also important to note that the DIC value of the EBNB model is not comparable to others in the sense that the number of observations is one fourth of other models. One more noteworthy point is that τ^2 in FBRTRE is larger than that in FBPLN. In comparison with Equations 6 & 7, it can be concluded that the variation among various sites represent the larger variance among collision counts.

HSID RESULTS

The five tests described previously (SCT, TPDT, MCT, TRDT, and TST) were used to assess the relative performance of the four HSID methods, EB, FBNB, FBPLN and FBPTRE. The evaluation experiment followed the following procedure:

1. The four year data (2008-2011) were evenly divided into 2 time periods, Period 1 (2008 2009) and Period 2 (2010-2011).
2. For each HSID method, cities are sorted in descending order of estimated PSI on the basis of data of Period 1.
3. Cities with the highest rankings are flagged for further investigation. Typically, a threshold is assigned based on law requirements or according to safety funds available for improvement, such as the top 5 % of cities. In this evaluation, both the top 10% and 5% cities are used as cutoff levels.
4. Estimated PSI's of both time periods were compared under each test to assess the performance of various HSID methods.

5.1 Site Consistency Test Results

Using SCT test it is shown in Table 3 that the FBPTRE method outperforms other HSID methods in identifying both the top 5% and 10% cities with highest sum of PSI's, 25,885.23 and 36,075.71, respectively, in Period 2 followed closely by the FBPLN method. The EB method performs the worst in both cases, with the identified cities experiencing the lowest PSI values, 9,617.61and 12,136.48 respectively. In comparison with the other two FB alternatives, the FBNB method performs relatively poorly, where the sites identified by this method in Period 1 produce smaller PSI values in Period 2.

However, as mentioned in previous test description, the limitation of SCT is this test requires the methods being evaluated to produce similar estimates of PSI because different methods have different ways to calculate the PSI. Therefore, the more insightful information might be obtained by comparing the two PSI values in two periods by the same method. As expected, the PSI in Period 1 would be larger than that of Period 2 as the HSID is conducted using the data of Period 1. The better method would yield the smaller PSI decrease in Period 2 relative to that in Period 1. For convenience of comparison, the relative difference was also calculated by dividing the difference of PSI in two periods by PSI in Period 1. The smaller relative difference, the more reliable the method in identifying the cities showing more PSI values in the future period. As shown in Table 3, the lowest relative difference percentages in both cases of identifying top 5% and 10% cities indicate the FBPTRE method has the best performance amongst the four HSID methods, with EB performing the worst.

5.2 The Method Consistency Test Results

Table 4 shows the number of similarly identified cities identified by alternate HSID methods over the two periods. The FBPTRE method is superior in this test by identifying the largest number of the same hot spots in both cases of top 5% and 10%, or,13 and 26 sites respectively. In other words, the FBPTRE method identified 26 sites in Period 1 that were also identified as hot spots in Period 2. The FBNB, which performs slightly better than the FBPLN method, places 2nd with identifying 13 consistent hot spots (in the case of top 5%) and 24 consistent hot spots (in the case of top 10%). The EB performed last with the lowest number of consistent hot spots identified in the two periods. Again, the FBPTRE method outperforms the other HSID methods.

Also shown in Table 4 are differences between percentages (shown in the parenthesis) of column 2 and column 3 for the four methods. There is a consistent drop in percentages as threshold values drop. The explanation is that the top sites suffer from greater random fluctuations in collisions and thus the higher the threshold the larger are the random fluctuations and the likelihood of not being identified in a prior year.

5.3 Total Rank Difference Test Results

Table 5 illustrates that the FBPTRE method is vastly superior using the Total Rank Difference Test. In both the cases of top 13 and 26 cities being identified, the FBPTRE method has significantly smaller summed ranked differences; by about 20% compared to the FBPLN and FBNB, and by more than 200% compared with the EB. FBPLN and FBNB in this test share very similar performance with almost the same rank difference in both cases. Again, the EB performs the worst by producing much larger rank differences. This result suggests that the FBPTRE method is the best HSID method (of the 4 methods evaluated here) for ranking cities consistently from period to period.

5.4 Total Performance Difference Test Results

Inspection of Table 6 reveals the results of the Total Performance Difference Test, which was proposed by Jiang et al. (13) to address the limitation of SCT test. In the case of top 5%, the FBPTRE appears to top other alternative methods by yielding lowest PSI difference. The FBPLN method trails closely behind the FBPTRE with slightly greater difference. In the case of top 10%, the FBPLN performs best by producing the smallest PSI difference. In both cases, the FBNB method remains in the 3rd place, followed by the EB, which generates the highest PSI differences in two periods under both conditions.

As mentioned previously, the underlying assumption is that the hotspots identified by HSID method with all years of crash data are considered as true hazardous sites. Hence, the results shown in this test should be interpreted with care due to the relatively small accident history (4 years). Maybe this is a potential reason that PBPTRE did not perform best in both cases as it did under other tests. Considering other test results, it is plausible to expect that the advantage of the FBPLN in this test relative to FBPTRE would diminish with longer collision histories, and would be surpassed by the FBPTRE method. Regardless, the FBPTRE and FBPLN method perform similarly using this test.

5.5 Total Score Test Results

Recall that TST, originally proposed by Montella (2010) and subsequently modified by Jiang et al. (2014), produces a synthetic index integrating all previous test results. The maximum possible score (100) indicate the corresponding method performs best in every test being used. Table 7 reveals that the FBPTRE outperforms others in both case studies (top 5% and 10% of the cities), obtaining values of 100 and 99.77, respectively. The FBPLN method was slightly worse than the FBPTRE, followed by the FBNB method, and lastly the EB method (which is significantly worse than its FB alternatives).

5.6 Discussion of Test Results

The test results at city level highlight that each of the FB approaches significantly outperforms the EB method under various tests. Overall, the study results are consistent with the results of the previous quantitative evaluations carried out by others (Huang et al., 2009; Jang et al., 2014; Miranda-Moreno and Fu, 2007; Miaou and Lord, 2003; Pawlovich et al., 2006). EB recently has enjoyed wide applications especially after it was made available through several safety design and evaluation tools, including the Interactive Highway Safety Design Model (IHSDM), SafetyAnalyst and Highway Safety Manual (HSM). It is highly recommended that more research be carried out to verify the study results because different HSID methods yield different sets of hazardous locations. It can be observed from Table 8 that the numbers of sites that were consistently identified as hotspots by the Empirical Bayesian and three full Bayesian methods in the top 5% and 10% levels are 9, 10, 10 and 16, 16, 12, respectively. The lowest common rate, 12 out of 26, occurred in the case of identifying top 10% cities by EB and FBPTRE. In other words, this means that 14 among the top 26 cities identified by these two methods do not match. On the contrary, the largest number of common sites, 23, happened in the case of identifying top 10% cities between FBPLN and FBPTRE.

Both Empirical Bayes and the full Bayes are under Bayes' theorem, which shrinks the observed collision number to the "real" mean with additional information borrowed from the reference population. The test results indicate that the FB approaches might integrate the borrowed information more reliably and smoothly than does the EB. Another classification of the alternative HSID methods being evaluated is the methods with and without the temporal random effect. Hence, the authors are also interested in investigating the impact of serial correlation of errors on the shrinkage of the two types of information. To this end, the authors randomly selected 10 cities whose PSI's range widely (from 27.4 to 906.1) and further compared the performance of FBPLN and FBPTRE on these cities. The PSI of the 10 cities under the two methods for two different periods is illustrated in Figure 1.

Inspection of Figure 1 shows that the PSI values (represented by the 2 solid lines) estimated by the PLN model change significantly from Period 1 to Period 2, while those estimated by the PTRE model (represented by the two dash lines) are more consistent across the time periods. This is consistent with our expectation as the PLN model assumes the same site in different years to be independent observations. In other words, under the PLN method, the information from other locations and the information from the same site of different time periods have a same weight on the estimation of PSI for each site. By contrast, the PTRE method includes a site specific error term, which shrinks the observed mean of each time period to the "real" mean over years. Hence, as demonstrated in Figure 1, given the assumption that no change was experienced on the sites, the model that takes into account the serial correlation is more reliable.

6. CONCLUSIONS

Numerous methods have been proposed in the past to conduct HSID at micro levels such as road segments and intersections. On the contrary, very few studies have been dedicated to the network screening of more aggregate levels which include the county level by Miaou and Song (2005) and zonal level by Lovegrove and Sayed (2007). To add more research centered on HSID of macrolevel to the current literature, the authors performed HSID by using the data from CA at city level. It is anticipated that the research results could facilitate improved decisions by city planners and engineers when evaluating the safety performance of city as a whole, and therefore allow the states to allocate appropriate proportion of funds to various cities. Additionally, the paper aims to investigate whether the previous HSID findings at microlevel can also be revealed at the city level, which has much larger sample mean and variance.

Four years of city data from the State of California were collected to compare alternate HSID methods, including the EB and three FB alternatives, FBNB, FBPLN and FBPTRE, for ranking the safety performance of cities. Five evaluation tests which contain the Site Consistency Test (SCT), the Method Consistency Test (MCT), the Total Rank Difference Test (TRDT), the Total Performance Difference Test (TPDT) and the Total Score Test (SCT) were applied to evaluate the performance of the four HSID methods. The intended use of these tests is akin to the selection of statistical models where multiple criteria are used to select the 'best' model, including adjusted R-square, F-ratio, t-statistics of model variables, signs, and magnitudes of coefficients, and mean square error. As in statistical modeling, a model will not be 'best' among all criteria, and the analyst must compare models against a set of criteria and subjectively choose the most appealing model. Potential for Safety Improvement (PSI) was adopted as a measure of the crash risk. Moreover, two cutoff levels, top5% and 10% cities, were employed for reliable results. After evaluating these four methods, the following conclusions are drawn:

- The FBPTRE method outperformed the other three HSID methods on the Site Consistency Test, followed very closely by the FBPLN method. That is, the FBPTRE and FBPLN methods identified cites in Period 1 that produced the highest PSI value in Period 2—demonstrating good consistency. The EB method performed the worst.
- The FBPTRE method is superior to other three methods in terms of the Method Consistency Test, That is, the FBPTRE method consistently identified a larger intersection of cites across observation periods. The FBNB and the FBPLN method

followed the FBPTRE method in 2nd and 3rd place, respectively, while the EB method performed worst.

- Compared with the Method Consistency Test, the Total Rank Difference Test revealed pronounced benefits associated with the FBPTRE method. The FBPTRE method outperformed all competing HSID methods on this criterion, showing great consistency in ranking cites across observation periods. The EB method performed the worst by a large margin.
- In the Total Performance Difference Test, the FBPTRE and FBPLN shared the best performance in the cases of top 5% and 10%, respectively. Again, the EB method performed the worst against this criterion.
- In the Total Score Test, the highest scores indicate the FBPTRE has the best performance under different tests combined. FBPLN and FBNB are ranked in 2nd and 3rd place with FBNB having slightly lower score values. The EB method performed significantly worse, with much lower score values in both situations.

Overall, our study results are consistent with the results of numerous previous quantitative evaluations focused on micro-level HSID. First, the three FB approaches significantly outperform the EB counterpart. However, it is important to note the performance difference might be due in part to the different number of observations for EB (265) and FB (1060). Second, FBPLN and FBNB have the similar HSID performance, with the former one slightly better than the latter one. Third, the method accounting for temporal random effect yields more reliable HSID results than do the ones without considering the serial correlation in collision counts. This result is somewhat alarming, as EB has recently been adopted in the Highway Safety Manual and used by many agencies to conduct HSID. Therefore, it is desirable to have further studies dedicated to improving the EB method by more efficiently combining the two types of safety clues, reference population and crash history of each specific site.

The results observed in this paper require some caveats. First, only one independent variable is included in the collision prediction models and as a result the functional forms might not be appropriate in some cases. Second, the advantages associated with the FBPTRE methods are obtained based on California collision data, and the relative performances of HSID methods may change when using other accident data (this result is possible but not expected). Third, only temporal random effect is analyzed in the paper. Incorporation of other random effects such as spatial correlation might change the benefits related with the method that accounts for various random effects. It is therefore highly recommended that future studies accounting for both temporal and spatial correlations be conducted for the city level to check the benefits of FB methods exhibited in this study. Finally, there are only two time periods employed for method performance assessment. It is desirable to have future studies with more time periods for more reliable results.

7. ACKNOWLEDGMENTS

This research received the grant from UCCONNECT summer research program.

8. REFERENCES

1. Abdel-Aty, M., C. Siddiqui, H. Huang, and X. Wang (2011). Integrating Trip and Roadway Characteristics to Manage Safety in Traffic Analysis Zones. In Transportation Research

Record: Journal of the Transportation Research Board, No. 2213, TRB, National Research Council, Washington, D.C., pp. 20–28.

2. Aguero-Valverde, J., Jovanis, P.P., 2009. Bayesian multivariate Poisson log-normal models for crash severity modeling and site ranking. In: Presented at the 88th Annual Meeting of the Transportation Research Board.
3. Aguero-Valverde, J. (2013) Multivariate spatial models of excess crash frequency at area level: Case of Costa Rica, Accident Analysis and Prevention 59, pp. 365-373.
4. Carlin, B.P., Louis, T.A., 2000. Bayes and Empirical Bayes Methods for Data Analysis, Monographs on Statistics and Applied Probability, 69. Chapman & Hall, London.
5. Cheng, W., and S. Washington. Experimental Evaluation of Hotspot Identification Methods. Accident Analysis and Prevention, Vol. 37, No. 5, 2005, pp. 870–881.
6. Cheng, W., and S. Washington. New Criteria for Evaluating Methods of Identifying Hot Spots. In Transportation Research Record 2083, TRB, National Research Council, Washington, D.C., 2008, pp. 76–85.
7. Davis, G.A., Yang, S.,Bayesian identification of high-risk intersections for older drivers via Gibbs sampling. Transportation Research Record 1746, 2001. pp. 84–89.
8. Deacon, J.A., C. V. Zegeer, and R.C. Deen. (1975). Identification of Hazardous Rural Highway Locations. In Transportation Research Record 543, TRB, National Research Council, Washington, D.C., pp. 16-33.
9. Elvik, R. State-of-the-art approaches to road accident black spot management and safety analysis of road networks. Report 883. Institute of Transport Economics, Oslo., 2007.
10. Greene, W.H (2010). Econometrics Analysis 7th Edition. Prentice Hall, Upper Saddle River, New Jersey.
11. Hadayeghi A., A. Shalaby, and B. Persaud (2007). Safety Prediction Models: Proactive Tool for Safety Evaluation in Urban Transportation Planning Applications. In Transportation Research Record: Journal of the Transportation Research Board, No. 2019, TRB, National Research Council, Washington, D.C., pp. 225–236.
12. Hakkert, A. S. and D. Mahalel (1978). Estimating the Number of Accidents at Intersections from a Known Traffic Flow on the Approaches. Accident Analysis and Prevention, Vol.10, No.1, pp. 69-79.
13. Hauer, E., and B. N. Persaud. Problem of Identifying Hazardous Locations Using Accident Data. In Transportation Research Record: Journal of the Transportation Research Board, No. 975,TRB, National Research Council, Washington, D.C., pp. 36-43.,1984.
14. Hauer, E. (1986). On the Estimation of the Expected Number of Accidents. Accident Analysis and Prevention, Vol.18, No.1, pp. 1-12.
15. Hauer, E. and B. N. Persaud (1987). How to Estimate the Safety of Rail-highway Grade Crossings and the Safety Effect of Warning Devices. In Transportation Research Record 1114, TRB, National Research Council, Washington, D.C., pp. 131-140.
16. Hauer, E., J.C.N. Ng, and J. Lovell (1988). Estimation of Safety at Signalized Intersections. In Transportation Research Record 1185, TRB, National Research Council, Washington, D.C., pp. 48-61.
17. Hauer, E., B. N. Persaud, A. Smiley, and D. Duncan (1991). Estimating the Accident Potential of an Ontario Driver. Accident Analysis and Prevention, Vol.23 No.2/3, pp. 133 152.

18. Hauer, E., D.W. Harwood, F.M. Council, and M. S. Griffith (2002). Estimating Safety by the Empirical Bayes Method: A Tutorial. In Transportation Research Record 1784, TRB, National Research Council, Washington, D.C., pp.126-131.
19. Higle, J. L., and M. B. Hecht. A Comparison of Techniques for the Identification of Hazardous Locations. In Transportation Research Record 1238, TRB, National Research Council, Washington, D.C., 1989, pp. 10–19.
20. Huang, H.L., Chin, H.C., Haque, M.M., 2009. Hotspot identification: a full Bayesian hierarchical modeling approach. Transportation and Traffic Theory, pp.441–462 http://link.springer.com/chapter/10.1007/978-1-4419-0820-9
21. Jiang, X.M., M. Abdel-Aty, and S. Alamili. 2014. Application of Poisson Random Effect Models for Highway Network Screening. Accident Analysis and Prevention 63, pp. 74-82.
22. Kim, K., M. Brunner, and E. Yamashita. (2006). Influence of Land Use, Population, Employment, and Economic Activity on Accidents. In Transportation Research Record 1953, TRB, National Research Council, Washington, D.C., pp. 56-64.
23. Kweon, Y. (2008). Examination of Macrolevel Annual Safety Performance Measures for Virginia. In Transportation Research Record: Journal of the Transportation Research Board, No. 2083, TRB, National Research Council, Washington, D.C., pp. 9-15.
24. Ladron de Guevara, F., S. P. Washington, and J. Oh. (2004). Forecasting Crashes at the Planning Level: Simultaneous Negative Binomial Crash Models Applied in Tucson, Arizona. In Transportation Research Record: Journal of the Transportation Research Board, No. 1897, TRB, National Research Council, Washington, D.C., pp. 191–199.
25. Lan, B., Persaud, B., Lyon, C., Bhim, R. Validation of a full Bayes methodology for observational before–after road safety studies and application to evaluation of rural signal conversions. Accident Analysis & Prevention 41 (3), 2009., pp. 574–580.
26. Lord, D., & 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, 751-770.
27. Lovegrove, G. and T. Sayed (2006). Using Macrolevel Prediction Models in Road Safety Planning Applications. In Transportation Research Record: Journal of the Transportation Research Board, No. 1950, TRB, National Research Council, Washington, D.C., pp. 73–82.
28. Lovegrove, G. and T. Sayed (2007) Macrolevel Collision Prediction Models to Enhance Traditional Reactive Road Safety Improvement Programs. In Transportation Research Record: Journal of the Transportation Research Board, No. 2019, TRB, National Research Council, Washington, D.C., pp. 65–73.
29. Maher, M. J., and L. J. Mountain. The Identification of Accident Blackspots: A Comparison of Current Methods. Accident Analysis and Prevention, Vol. 20, No. 2, 1988, pp. 143–151.
30. McGuigan, D. R. D.(1981). The use of Relationships between Road Accidents and Traffic Flow in 'Black-Spot' Identification. Traffic Engineering and Control, Aug.-Sept., pp.448-453.
31. McGuigan, D. R. D.(1982). Nonjunction Accident Rates and their Use in 'Black-Spot' Identification. Traffic Engineering and Control, Feb., pp.45-56.
32. Miaou, S., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes methods. Transportation research record. Journal of the Transportation Research Board 1840, pp.31–40.

33. Miaou, S-P., and J.J. Song. (2005). Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion, and spatial dependence. Accident Analysis and Prevention, Vol. 37 No. 4, pp 699–720.
34. Miranda-Moreno, L.F., and L., Fu. Traffic Safety Study: Empirical Bayes or Full Bayes? The 86th Annual Meeting of the Transportation Research Board, TRB 2007. Paper 07-1680.
35. Montella, A., 2010. A comparative analysis of hotspot identification methods. Accident Analysis and Prevention 42, 571–581.
36. NCHRP. Safety Management Systems, A Synthesis of Highway Practice. Transportation Research Board, National Academies, Washington, D.C., 2003.
37. Norden, M., J.Orlansky, and H.Jacobs (1956). Application of Statistical Quality-Control Techniques to Analysis of Highway-Accident Data. Bulletin 117, HRB, National Research Council, Washington, D.C., pp.17-31.
38. Pawlovich, M.D., Li,W., Carriquiry, A.,Welch, T. 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, D.C. 2006, pp.163–171.
39. Persaud, B.N. (1999). Empirical Bayes Procedure for Ranking Sites for Safety Investigation by Potential for Safety Improvement. In Transportation Research Record 1665, TRB, National Research Council, Washington, D.C., pp. 7-12.
40. Persaud, B., Lan, B., Lyon, C. and Bhim, R. Comparison of Empirical Bayes and Ful Bayes Approaches for Before-After Road Safety Evaluations. Accident Analysis & Prevention 42, 2010, pp. 38–43.
41. 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.
42. Stokes, R.W. and I. M. Mutabazi (1996). Rate-Quality Control Method of Identifying Hazardous Road Locations. In Transportation Research Record 1542, TRB, National Research Council, Washington, D.C., pp. 44-48.
43. Tamburri, T. N., and R.N. Smith (1970). The Safety Index: Method of Evaluating and Rating Safety Benefits. In Highway Research Record 332, HRB, National Research Council, Washington, D.C., pp.28-39.
44. Taylor, J.L., and H.T. Thompson (1977). Identification of Hazardous Locations. Report No. FHWA-RD-77-81. Federal Highway Administration, Washington, D.C.
45. Washington, S. and J. Oh. Bayesian Methodology Incorporating Expert Judgment for Ranking Countermeasures Effectiveness under Uncertainty: Example Applied to at Grade Railroad Crossings in Korea. Accident Analysis & Prevention 38, 2006., pp. 234–247.

TABLE 1 Descriptive Statistics of Collected Data of Various Cities

Variables	Description	Year	Min	Max	Mean	S.D.
Collision	Total Annual Fatal and Injury Collisions	2008	10	3,879	377.3	493.0
		2009	8	3,745	374.4	489.5
		2010	17	3,945	371.3	493.4
		2011	13	3,900	369.9	470.5
Pop	Population Number	2008	25,117	98,709	89,564.5	102,455.8
		2009	25,265	1,014,965	90,940.2	104,975.1
		2010	25,077	952,509	88,704.4	100,439.5
		2011	25,261	964,371	89,221.6	101,199.7
DVMT	Daily Vehicle Miles Traveled	2008	33,849	8,267,781	873,508.5	969,563.4
		2009	28,430	8,363,995	884,913.9	979,646.5
		2010	41,307	8,364,002	899,321.4	986,362.3
		2011	37,465	8,625,277	899,084.4	1,017,646.8

TABLE 2 Description of Results of Various Models

Variable	Mean	SD	2.5%	97.5%
EBNB				
Intercept	-7.86	0.42	-8.68	-7.04
logDVMT	1.02	0.03	0.96	1.08
Overdispersion φ=5.94; DIC= 3,195.3; MSPE=217,746.8				
FBNB				
Intercept	-0.45	0.41	-1.06	0.32
logDVMT	0.41	0.03	0.35	0.46
Alpha (α)	1.20	0.03	1.14	1.25
DIC=7233.5; MSPE=204,849.9				
FBPLN				
Intercept	0.35	0.99	-1.41	1.78
logDVMT	0.38	0.07	0.28	0.51
Tau (τ^2)	2.06	0.29	1.62	2.64
DIC=7364.1; MSPE=187,790.4				
FBPTRE				
Intercept	0.48	0.51	-0.29	1.30
logDVMT	0.29	0.05	0.22	0.38
Tau (τ^2)	3.10	0.33	2.58	3.68
DIC=6838.5; MSPE=151,036.7				

TABLE 3 Results of Site Consistency Test (SCT) of Various Methods

Method	*Top 5% (13 cities)*			*Top 10% (26 cities)*		
	PSI 2008-2009	PSI 2010-2011	Relative Difference	PSI 2008-2009	PSI 2010-2011	Relative Difference
EB	11,842.37	9,617.61	18.79%	14,937.13	12,136.48	18.75%
FBNB	21,062.05	20,187.85	4.15%	28,157.90	26,770.80	4.93%
FBPLN	25,563.00	24,948.50	2.40%	36,232.50	35,034.35	3.31%
FBPTRE	26,381.35	25,885.23	1.88%	37,451.95	36,705.71	1.99%

Notes: Relative difference is calculated as the PSI difference in two periods relative to PSI in Period 1.

TABLE 4 Results of Method Consistency Test (MCT) of Various Methods

Method	*Top 5% (13 cities)*	*Top 10% (26 cities)*
EB	11 (84.6%)	19 (73.1%)
FBNB	13 (100%)	24 (92.3%)
FBPLN	13 (100%)	23 (88.5%)
FBPTRE	13 (100%)	25 (96.1%)

Notes: The number represents locations identified by methods in both periods, the percent shown in parenthesis stands for the percentage of consistent hot spots, or the percentage of hot spots identified in Period 1 that were also identified in Period 2.

TABLE 5 Results of Total Rank Differences Test (TRDT) of Various Methods

Methods	***Top 5% (13 cities)***	***Top 10% (26 cities)***
EB	47	294
FBNB	20	63
FBPLN	18	64
FBPTRE	15	58

TABLE 6 Results of Total Performance Differences Test (TPDT) of Various Methods

Methods	***Top 5% (13 cities)***	***Top 10% (26 cities)***
EB	2594.85	3526.52
FBNB	1085.55	1654.86
FBPLN	690.23	1252.35
FBPTRE	645.56	1284.47

TABLE 7 Results of Total Score Test (TST) of Various Methods

Methods	*Top 5% (13 cities)*	*Top 10% (26 cities)*
EB	44.64	41.08
FBNB	87.60	88.96
FBPLN	97.07	96.35
FBPTRE	100.00	99.77

TABLE 8 Common Cities Identified by Various Methods

	Top 5% (13 cities)				Top 10% (26 cities)			
	EB	**FBNB**	**FBPLN**	**FBPTRE**	**EB**	**FBNB**	**FBPLN**	**FBPTRE**
EB	-	9	10	10	-	16	16	12
FBNB	9	-	11	10	16	-	21	20
FBPLN	10	11	-	12	16	21	-	23
FBPTRE	10	10	12	-	12	20	23	-

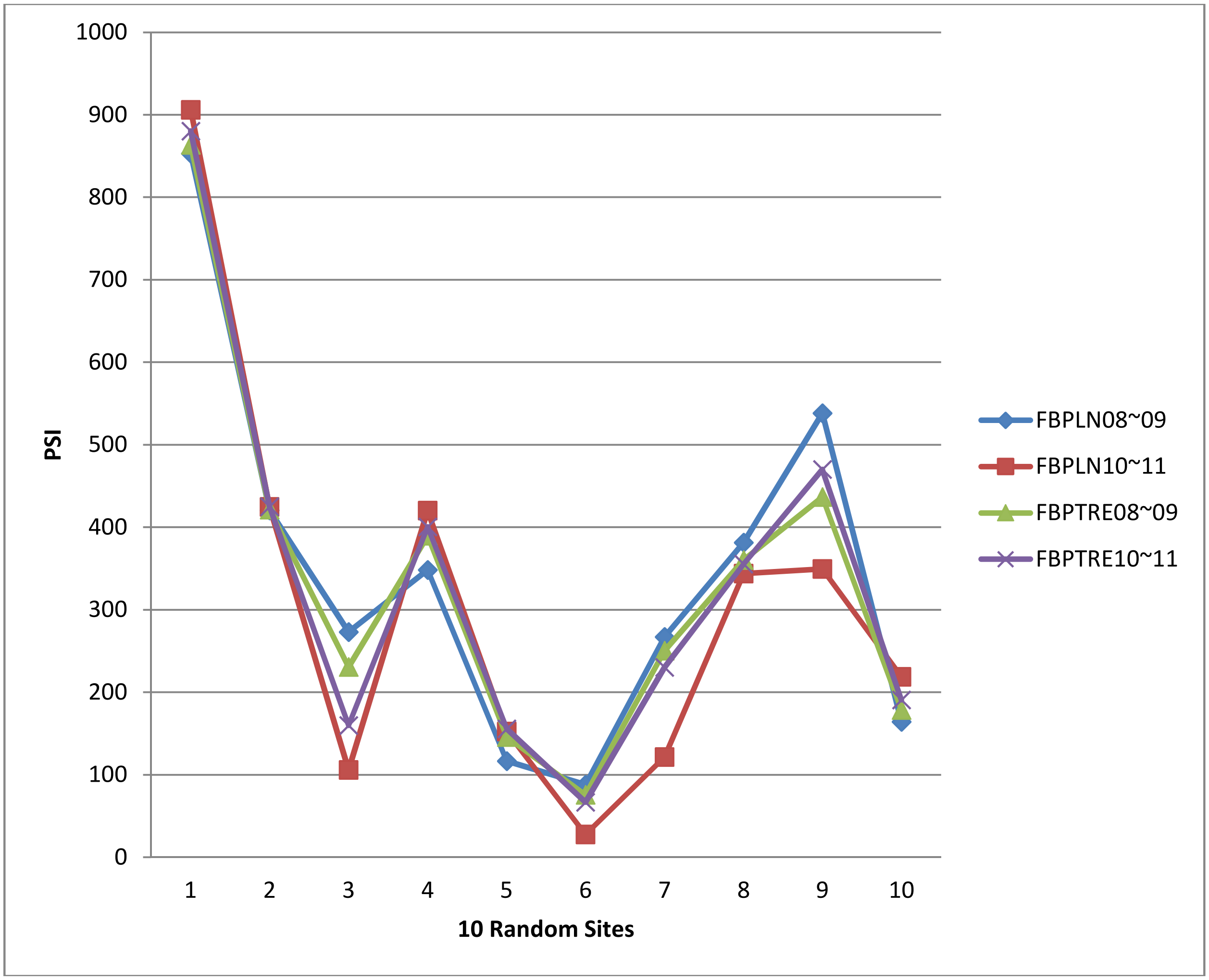


FIGURE 1 Comparison of PSI estimation under FBPLN and FBPTRE for 10 random sites

Notes: PSI represents the Potential for Safety Improvement

Paper #3: Evaluating the Influence of Neighboring Structures on Spatial Crash Frequency Modeling and Site Ranking Performance

Gurdiljot Singh Gill
Graduate Student, Department of Civil Engineering
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Fax: (909) 869-4342; Tel: (909) 869-4312; Email: gurdiljotg@cpp.edu

Wen Cheng, Ph.D. P.E. (corresponding author)
Associate Professor, Department of Civil Engineering
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Fax: (909) 869-4342; Tel: (909) 869-2957; Email:wcheng@cpp.edu

Meiquan Xie, Ph.D.
Visiting Scholar, Department of Civil Engineering,
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Fax: (909) 869-4342; Tel: (909)282-9142; Email: mxie@cpp.edu

Tom Vo
Modeler, Southern California Association of Governments
818 West 7th Street, 12th Floor, Los Angeles, CA 90017
Fax: (213) 236-1825; Telephone: (213) 236-1800; Email: Vo@scag.ca.gov

Xudong Jia, Ph.D. P.E.
Professor, Department of Civil Engineering
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Fax: (909) 869-4342; Tel: (909) 869-4312; Email:xjia@cpp.edu

Jiao Zhou
Graduate Student, Department of Civil Engineering
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Fax: (909) 869-4342; Tel: (909) 869-4312; Email:jiaozhou@cpp.edu

Word count: 5,147 words + 4 tables+4 figure =7,147 equivalent words

ABSTRACT

A large number of neighborhood weight matrices have been adopted for modeling crash spatial heterogeneity. However, there has been little evaluation of the influence of these different weight matrix structures on the crash prediction modeling performance. This study is focused on investigation of 17 different spatial-proximity matrices for development of spatial crash prediction models and site ranking using county-level data in California. Among the group of matrices being evaluated, traffic exposure-weighted and population-weighted distance-based matrices are first proposed in the traffic safety field. To address serial correlation of crashes in successive years, Bayesian spatial analysis was conducted with the combination of a first order autoregressive (AR-1) error process and time trend for crashes.

Two diagnostic measures were used for assessment of goodness-of-fit and complexity of models. In addition, seven evaluation criteria were employed to assess the benefits associated with better fitting models in site ranking. The results showed that modeling performance gets improved with the increase in number of neighbors being considered in the weight matrix. However, the larger number of neighbors also leads to larger variability of modeling performance. Specifically, Queen-2 and Decay-50 models proved to be superior among the adjacency and distance-based models, respectively. The significance of incorporating spatial correlations was highlighted by the consistently poor performance of the Base model which included only heterogeneity random effect. Finally, the model-fitting performance seems to be strongly correlated with the site ranking performance. The models with closer goodness-of-fit tend to yield more consistent ranking results.

Keywords: neighborhood weight matrix structures, first order autoregressive (AR-1) error process, Bayesian spatial crash prediction model, goodness-of-fit

INTRODUCTION

Previous research studies explored the spatial component of crashes as an advancement to the existing crash prediction models (*1-3*). Numerous spatial units have been taken into consideration to understand the implications of crash causing factors which operate at spatial scale (e.g. urban planning policy, census characteristics, highway classification, and so on). Depending on the purpose of study, the sites of interest could range from microscopic locations, such as block group (*4-5*), intersections (*6*), road segments (*7*), corridors (*8-9*), to macroscopic areas such as census tracts (*10*), health areas (*11*), traffic analysis zones (TAZs) (*12-15*), or counties (*16-20*). Comparatively speaking, the microscopic analysis is primarily centered on investigating the factors associated with geometric or traffic characteristics which influence the traffic safety on a network. Subsequently, engineering solutions are suggested for mitigation of risk. 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 policy-based countermeasures could be enactments of traffic rules, police enforcements, safety campaigns, and area-wide engineering treatments.

The literature review illustrates that a wide range of neighborhood weight matrix structures have been proposed to model crash spatial heterogeneity for both micro-level and macro-level analyses. Aguero-Valverde and Jovanis (*21*) 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 (*22*). Three adjacency-based weight matrices were developed for first, second and third order neighbors. The results demonstrated that the models with spatial correlations showed a significantly better fit than the Poisson lognormal model which considered only heterogeneity. Guo et al. (*9*) developed models to incorporate the spatial proximity at corridor level between intersections due to similarity in road design and environmental characteristics. The distance between intersections was adopted as the weight for CAR model. The modeling results demonstrated that the Poisson spatial model provided the best fit. Recently, Aguero-Valverde et al. (*23*) 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. Their results show that the model that considers both multivariate and spatial correlation has the best fit.

A wide array of geographical units and weight matrices have been explored in macro-level modeling as well. Best et al. (*24*) investigated the risk of leukemia in children at three different levels of data aggregation: Local Authority Districts (LADs), census wards and 1 km2 grid squares. They examined adjacency versus distance-based neighborhood spatial weights for each of analysis. Rhee et al. (*25*) 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 and an ordinary least squares baseline regression. Aguero-Valverde and Jovanis (*19*) 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. (*20*) 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.

Compared with the large amount of research dedicated to modeling spatial heterogeneity in crash counts using various weight matrix structures, there is little evaluation of the influence of these different weight matrices on the crash prediction modeling performance. Aguero-Valverde and Jovanis (*21*) evaluated the effects of different neighboring structures on the spatial correlation in crash frequency models using CAR model. The weight structures being investigated include exponential decay, adjacency-based, adjacency-route information, and distance order structures. Modeling results showed relatively inferior performance by exponential decay models. Also, the inclusion of spatial correlation substantially increased the random effects. Another study (*35*) presented an evaluation of crash prediction models at the TAZ levels with alternative types of spatial proximity structures containing 0–1 first-order adjacency, common-boundary length, and centroid distance-based models. The CAR model was also implemented. The results confirmed the extensive existence of cross-zonal spatial correlation in crash occurrence. The best predictive capability appeared to be associated with the model that used the common-boundary lengths. Moreover, full consideration of all possible spatial correlations for all zones significantly increased model complexity, which might lead to reduced predictive performance.

The first part of this paper compares alternative spatial-proximity structures and represents a natural continuation of the above two studies, with a number of important differences and unique contributions. First, more comprehensive weight matrices are evaluated which include 2 orders of Queen adjacency-based, 2 orders of Rook adjacency-based, common boundary length adjacency-based, 5 exponential decay, 5 pure distance order, population-weighted and traffic exposure-weighted distance order matrices. Amongst the 17 neighboring structures, the last two are first proposed in the traffic safety field. Moreover, the model without considering spatial heterogeneity is also developed to check the existence of cross-county spatial correlation in crash counts. Second, the serial correlation of the county-level crash count was taken into count via using the first order autoregressive (AR-1) error process and global time trend combined. Third, the relationship between the crash frequency modeling performance and the number of neighbors in the weight matrices is explored in greater detail.

In addition to evaluating the impact of weight structures on the crash frequency model complexity and fit, this paper also evaluates the effect of these structures on the site ranking performance. The ranking agreement of different weight matrices was assessed by seven different evaluation criteria: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), Cohen's kappa, total ranking difference (TRD) and mean absolute deviation (MAD).

The remainder of this paper first describes the methods employed for development of models with 17 different weight matrices, criteria for assessment of fit of those models, and the evaluation procedures implemented to analyze their performance. Then the source of data and its segregation is explained. Finally, the modeling and evaluation results are presented, followed by conclusions and recommendations for future research.

METHODOLOGY

Spatial autocorrelation occurs when events happening at different but nearby places are correlated. These were explored among 58 counties of California using a wide array of weight matrices ranging from simple to more sophisticated ones. This study analyzed 17 different neighborhood matrices and applied them to the county-level datasets of California. In addition, the model without accounting for spatial heterogeneity was also developed to assess the benefit of inclusion of spatial correlation in the crash count model. Hence, the process involved development of 18 different models using WinBUGS for estimation of crash rate, evaluation of their goodness-of-fit, and finally evaluation of relative site ranking performance of models. This study used the Full Bayesian (FB) hierarchical approach to account for the structural heterogeneities such as temporal and spatial ones. The model is of the form developed by Besag et al. (*36*):

$$y_{it} \sim Poisson\ (e_{it}\theta_{it}) \tag{1}$$

Where, y_{it} is the observed crash count at county i in time period t and θ_{it} is the mean expected crash rate for site i in time period t, and e_{it} is the exposure in county i of time period t. In this case, the exposure is the total daily vehicle-miles (DVMT) by county. The crash rate is modeled as shown in the following equation:

$$Log(\theta_{it}) = \beta_0 + \beta_1 t_k + \emptyset_i + u_i \tag{2}$$

Where, DVMT is the traffic exposure, β_0 is the intercept, β_1 is the fixed coefficient for the linear trend term t_k, $\emptyset_i$ is a spatially structured random effect, and u_i is a spatially unstructured random effect (heterogeneity). It should be noted that usually the model development incorporates some probable influential factors but the model in this study does not incorporate any covariates as the major focus of this study is to investigate the spatial correlations. This similar framework can be found in other studies (*26*, *27*). Likewise, the interaction between time and space was not included as it might potentially blur the comparison of different weight structures. In addition to the global time trend applying to all counties in California, this model also accounts for the autoregressive safety effect by specifying the distribution of u_i as a lag-1 dependence in errors. Lag-1 is the correlation one year apart in this study (hence k=1). AR-1 was chosen to capture the departure from tend and it was based on the assumption of stationarity restriction.

$$u_i \sim normal\left(0, \frac{\sigma_i^2}{(1-\gamma^2)}\right) \tag{3}$$

$$u_{it} \sim normal\left(\gamma \varepsilon_{i,t-1}, \sigma_{it}^2\right) \qquad \text{for } t>1 \tag{4}$$

Where, γ is the autocorrelation coefficient with the following range: $0 < \gamma < 1$. The combination of AR-1 and deterministic time trend for serial correlation can be found in the current practice. An example is the STEPAR (stepwise autoregressive method) specification provided in SAS/STAT 9.2 user's guide (*37*).

To accommodate the spatial correlation in the model, CAR prior was introduced for the spatial random effects. The formulation of the CAR model used in our analyses is shown below (*28*):

$$[\emptyset_i \mid \emptyset_j, i \neq j, \tau_\emptyset^2] \sim N(\emptyset_j, \tau_i^2) \tag{5}$$

The above equations show that the neighboring sites have an influence on the crash risk associated with an area. Subscripts i and j represent a county and its neighbor respectively, and $j \varepsilon N_i$ where N_i represents the neighbor set for region i. The weights are included as they also influence the risks, besides the neighbors. The weights for the adjacency and distance models are given by weights ij (w_{ij}) = 1 if i, j are adjacent, and 0 otherwise. Apart from the adjacency based models, different weights were used for other models which are explained in detail in the following subsections. For ease of illustration, all weight matrices being evaluated in the study are classified in Figure 1.

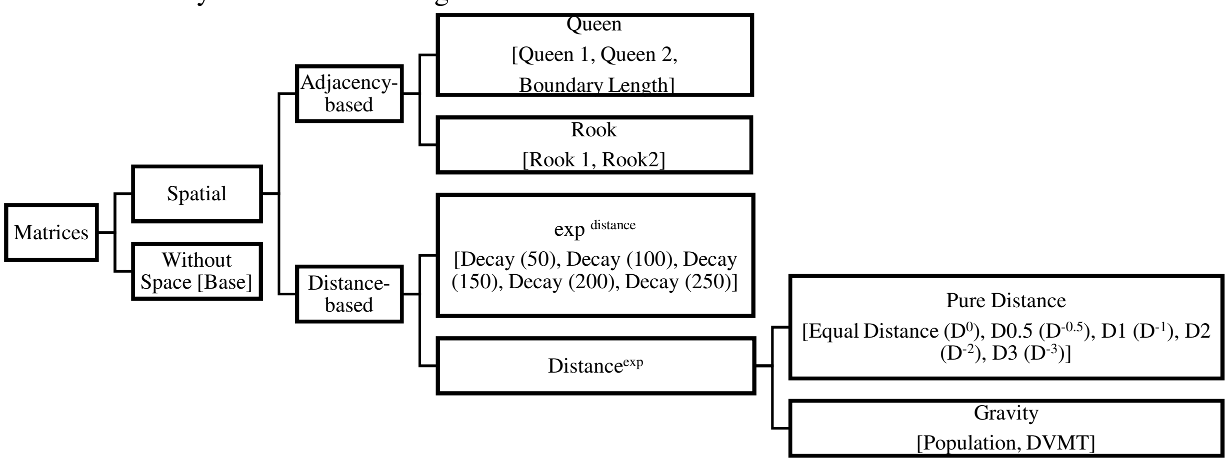


FIGURE 1 Types of weight matrices.

Adjacency-based Models

These models ignore the distance between sites of interest and focus only on neighboring structures based on proximity in space. Five different neighborhood adjacency-based weight matrices were developed, namely Queen-1, Queen-2, Rook-1, Rook-2, and Boundary Length (BL). The difference between Queen (Q) and Rook (R) is the criterion of assignment of neighbors. Queen uses the common boundaries as well as vertices to determine the adjacent neighbors while Rook considers only the common boundaries. For example, in top right corner of Figure 2, in case of Mariposa county: Rook neighbors are Tuolumne, Merced, and Madera counties, while Queen neighbors are Tuolumne, Stanislaus, Merced, Madera, and Mono counties. The numbers at the end reflect the order of contiguity, which means that the difference between Queen 1 and Queen 2 (and corresponding Rook 1 and Rook 2) is that Queen 1 includes the direct neighbors which share common points, while Queen 2 also includes the further neighbors of neighbors. It should be noted that for both cases, only the selection of neighbors is different, the first or second order neighbors contribute equal weights for the adjacency weight matrix. Finally, another weight matrix was developed for immediate neighbors based on the length of the boundary shared between counties. BL matrix placed more weightage among the neighboring counties which shared a longer boundary. For example, in the lower portion of Figure 2, it is known that there are five neighboring counties to San Bernardino County: Inyo, Kern, Los Angeles, Orange and Riverside. It is certain by mere visual inspection that Riverside has the longest boundary length while Orange has the smallest. Hence, the former would have the most weightage while the latter would have least weightage in BL matrix. The inclusion of this matrix is based on the hypothesis that a longer boundary increases the area for interaction of traffic between two counties and hence it may have a significant influence on the crash risks. It is important to note that BL is a special case of Queen 1 and they have the same number of neighbors. Clearly, adjacency-based weight matrices depict a binary or dichotomous situation where the wights have only two responses, zero and one. This approach requires relatively lesser data collection and computational efforts as compared to the other approach of distance-based weight matrices.

Distance-based Models

To account for a variety of scenarios, we also developed twelve models based on distance matrices. The simplest model, Equal Distance (ED), assigned equal weightage for weight matrix as it included all the counties as neighbors. The other models placed different weightages on neighbors. For Distance 0 (ED), 0.5, 1, 2, & 3, the following formulations were used: $w_{ij} = 1/dist^0_{ij}$, $w_{ij} = 1/dist^{0.5}_{ij}$, $w_{ij} = 1/dist_{ij}$, $w_{ij} = 1/dist^2_{ij}$ and $w_{ij} = 1/dist^3_{ij}$, respectively. These five models explored the different relationships between the weight and distance (e.g. linear, quadratic, cubic). These models only accounted for the relative distance between neighboring counties and placed more weight on counties that were closer together. Another similar set of models was developed ("Decay-50, 100, 150, 200 and 250") which was also based on the distance between neighbors. But these decay models were essentially different from regular distance models as there was a drastic reduction of weights as the distance between neighbors increased (*29*). The corresponding weight matrix for the "Decay" model was defined as:

$$w_{ij} = e^{\frac{-\mathrm{dist}_{ij}}{\delta_0}} \quad (6)$$

Where, w_{ij} = weight of the jth neighbor of the ith county, $dist_{ij}$ = geographic centroid distance between counties i and j, and δ_0 = a constant.

The decay was chosen based on the exploratory examination of correlogram of total collision count and average county distance of 250 miles. Five decays were chosen to incorporate different distances between the counties as the range of geometric centroid distances between counties was from 25-962 miles. In addition to the above two types of distance-based matrices, two Gravity models based on distance were also developed which borrows the idea of Gravity model from the standard "Four-Step" travel demand modeling process. The hypothesis is that sparsely populated neighbors and counties with less traffic exposure provide little information for spatial analysis. Population and DVMT were chosen in the study as they are commonly utilized by planners to assimilate the demographic changes for transportation modeling. The corresponding weight matrices were defined by $w_{ij} = p_i p_j/dist_{ij}$, where p_i and p_j are normalized populations and DVMTs of two counties under consideration, respectively. To the best knowledge of authors, such gravity models are first applied in the traffic safety field. The similar neighboring structures can be found in the public health field (*31*). It is noteworthy that the distance-based matrices contain more information, depending on their complexity, compared to the previously mentioned adjacency-based weight matrices. The calculated weightage offers more flexibility than the binary response of adjacency ones as it is calculated depending on the distance between target sites. The data collection and computational efforts significantly increase in this case due to the additional information.

Queen and Rook neighbors

FIGURE 2 California counties and associated centroids.

Goodness-of-Fit of Crash Frequency Models

This study used DIC (Deviance Information Criterion) to assess the complexity and goodness of fit of the models. DIC is a hierarchical modeling generalization of the AIC (Akaike Information Criterion) which was proposed by Spiegelhalter (*30*) to account for model fit and complexity. Specifically, DIC is defined as:

$$DIC = D(\bar{\theta}) + 2p_D = \bar{D} + p_D \tag{7}$$

Where, $D(\bar{\theta})$ is the deviance evaluated at the posterior means of estimated unknowns ($\bar{\theta}$), and posterior mean deviance $\bar{D}$ can be taken as a Bayesian measure of fit or "adequacy". p_D denotes the effective number of parameters in a model, as the difference between $D(\bar{\theta})$ and $\bar{D}$, i.e., mean deviance minus the deviance of the means. Generally, smaller values of DIC are preferred. As a general guideline by (*22*), a difference of 7+ points in the DIC is treated as significant for modeling performance.

To account for the random spatial effect explained by the model, we computed the fraction of total random variation as a ratio of the empirical variance of the spatial component against the total variance (*20*). The formulas are given below:

$$Var(u) = \sum_i \frac{(u_i - \bar{u})^2}{n-1} \tag{8}$$

$$Var(v) = \sum_i \frac{(v_i - \bar{v})^2}{n-1} \tag{9}$$

$$Fraction = \frac{Var(u)}{(Var(u) + Var(v))} \tag{10}$$

u_i is a spatially structured random effect and v_i is a spatially unstructured random effect, with i ranging from 1 to n = 58.

Site Ranking Evaluation

Naturally, one might be curious about the site ranking agreement of the models based on different weight matrices. To answer this question, the models were subjected to three levels of evaluation criteria (from basic to complex) which compared the posterior mean of crash rate resulting from the models.

Criteria based on Binary Partitions of Data

Under this type of criteria, the counties were divided into two groups based on estimated crash rates from various models with a certain threshold. In the present study, a cutline of 10% was used. It means that the top 10% counties were considered as high risk, while the remaining counties were treated as low risk. The common binary diagnostic criteria (*38*) including Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) were calculated with the following formulas:

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

$$PPV = \frac{TP}{TP + FP} \tag{13}$$

$$NPV = \frac{TN}{TN + FN} \tag{14}$$

Where, TP: number of truly high risk counties correctly identified as unsafe; FP: number of truly low risk counties identified as unsafe; TN: number of truly low risk counties identified as safe; and FN: number of truly high risk counties identified as safe. Since this study is only focused on the relative site ranking performance of various models, the model with best predictive performance (the lowest DIC) is chosen to establish the "truly" high risk or low risk counties. These criteria evaluate the ranking agreement from different perspectives. In specific, sensitivity is the ability of a model to correctly classify an individual as truly high risk. Specificity measures the ability of a model to correctly classify an individual as low risk. PPV represents the percentage of counties identified by one model as high risk which are truly high risk, while NPV indicates the percentage of counties identified by one model as low risk which are truly low risk. For all criteria, if the value is higher (as close to 100 as possible), then it suggests that the model of interest is doing as good as the model with the lowest DIC.

Criteria based on Multiple Partitions of Data

Aside from the above criteria based on the dichotomy of counties, the counties can also be divided into multiple groups to get more precise results. Cohen's Kappa (*39*) was chosen to determine the magnitude of county-ranking agreement based on three groups: High (top 10th percentile), Medium (11-89th percentile) and Low (90th percentile and lower). For each group, the number of common identified counties by two models was tallied and Kappa statistic was calculated with the following equation:

$$K = \frac{(p_O - p_e)}{(1 - p_e)} \tag{15}$$

Where p_o is the proportion of observations in agreement, and p_e represents the proportion in agreement due to chance. The Kappa statistic can ensure only truly high risk counties (excluding the random entries) were included for their respective groups. A relatively higher value of Kappa indicates a larger number of sites (agreement) which are commonly identified as unsafe between the crash prediction models.

Criteria based on Continuous Comparison of Data

Finally, in addition to categorizing counties into multiple groups, the relative site ranking performance of various models can be evaluated by comparing all counties in a continuous fashion. Two criteria of this category were implemented: the total ranking difference (TRD, *40*) and Mean Absolute Deviation (MAD, *41*). TRD is of the following form:

$$TRD = \sum_{i=1}^{n} \left| R_{i,j} - R_{i,k} \right| \tag{16}$$

Where, $R_{i,j}$ is the rank of county *i* by model *j*, and $R_{i,k}$ is the rank of county *i* by model *k*. The smaller TRD, the closer the two models in terms of site ranking.

Alternatively, the estimated crash rates by various models can be compared to determine MAD:

$$MAD = \frac{1}{n} \sum_{i=1}^{n} \left| CR_{i,j} - CR_{i,k} \right| \tag{17}$$

Where $CR_{i,j}$ is the estimated crash rate of county i by model j, and $CR_{i,k}$ is the estimated crash rate of county i by model k. This measure may be regarded as the safety component which represents the crash rate prediction performance of the models relative to the measure of "true safety". The degree of deviation from the expected crash rate would be reflected by the MAD value for a particular model: relatively greater value (compared to other competing models) would mean greater deviation from expected "true" crash rate, which reflects inferior prediction performance.

DATA PREPARATION

The data used in this study were collected from multiple sources: Statewide Integrated Traffic Records System (SWITRS), Highway Performance Monitoring System (HPMS), California Department of Finance, and Southern California Association of Governments (SCAG). Collisions of various counties in California that occurred from 2008-2013 were obtained from SWITRS that contains different severity levels. The study focuses only on the total fatal and injury collisions of the counties given the underreporting issue related with property damage only (PDO) collisions (*31*). In addition, a main exposure-related factor of county safety performance, that is, Daily Vehicle Miles Travel (DVMT) (*32*), was collected from HPMS for the same time periods. Furthermore, a main demographic factor, or, population, was gained from the California Department of Finance for the use in population-based gravity model which incorporates the population of two counties under consideration. In addition, the data for boundary length and the geometric centroid distances between all counties were obtained from SCAG. The boundary length data were used for the BL weight matrix. The weight in BL matrix was directly proportional to the length of shared boundary between neighboring counties. The data were incorporated indirectly to act as a measure for the distinction between two adjacency-based neighbors (Queen and Rook). All the distance-based models were based on the other dataset which had the distances among centroids of various counties. As the state of California has 58 counties, so the distance matrix had the size of 58x57 with a minimum value of 25 miles and maximum of 962 miles. Table 1 shows the characteristics of first and second order adjacency-based matrices, BL and distance-based matrices. Since the counties had very few cases of shared vertices, due to irregular shape, not much difference was observed between Queen and Rook count. However, a three-fold increase in the mean count of neighbors was noticed between first and second order neighbors.

TABLE 1 Characteristics of Neighborhood Weight Matrices

Neighbor matrices	Mean	Median	Min	Max	S.D.	Sum
Queen-1	4.91	5	2	8	1.32	285
Queen-2	13.1	14	4	21	3.54	762
Boundary Length	4.91	5	2	8	1.32	285
Rook-1	4.29	4	2	8	1.23	249
Rook-2	12.1	13	4	17	3.36	700
Distance-based	57	57	57	57	N/A	3306

Notes: S.D. represents standard deviation; boundary length is a special case of Queen-1.

Summary information for the various cities in terms of population, DVMT, boundary length, distance, and total fatal and injury collision number are shown in Table 2.

TABLE 2 Descriptive Statistics of Collected Data of Various Counties

Variables	Description	Year	Min	Max	Median	Mean	S.D.
Collision	Total Annual Fatal and Injury Collisions	2008	34	52,896	791	2,993	7,299
		2009	31	51,371	769	2,868	7,076
		2010	34	50,683	669	2,821	6,985
		2011	26	50,989	730	2,789	7,010
		2012	25	51,207	697	2,801	7,065
		2013	30	51,502	689	2,755	7,101
Pop	Population	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
DVMT	Daily Vehicle Miles Travelled (unit: miles)	2008	168,265	214,971,058	5,005,121	15,387,562	31,617,871
		2009	170,690	214,236,853	4,836,964	15,317,670	31,469,076
		2010	169,420	211,876,665	5,448,915	15,482,772	31,148,547
		2011	164,587	214,458,135	4,761,505	15,353,471	31,594,730
		2012	166,923	214,482,442	4,551,148	14,768,115	31,478,320
		2013	165,180	215,817,520	4,462,740	14,924,626	31,747,694
B.L.	Boundary Lengths among neighboring counties (unit: miles)	N/A	0.04	204	37	45	34

Distance	Distance among centroids of counties (unit: miles)	N/A	25	962	227	273	176

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

RESULTS

This study was aimed at investigation of 17 adjacency and distance-based weight matrices along with the base model in a variety of scenarios, which account for the spatial correlations between 58 counties of California for crash count models. These models were implemented in freeware WinBUGS package (*22*) using a MCMC algorithm. In the absence of strong prior information for factor effects and dispersion parameters, uninformative priors were assumed with normal distribution (0, 1,000) for all regression coefficients, and with gamma distribution (0.001, 0.001) for precision estimates. In model calibration, we discarded the first 5000 samples as burn-in and ran a further 25,000 iterations which were used in the calculation of the posterior estimates. Two chains were set up for each model starting from diverse initial values. Convergence was assessed using the Gelman-Rubin convergence statistic (*33*) and visual inspection of the trace plot. Plus, the sample MC errors of all parameters were less than 5% of the associated standard deviation. It should be noted that for the Gravity matrix, the neighbors were weighed according to the size of population of counties and Daily Vehicle Miles Traveled (DVMT), but they were not used as independent variables for development of models.

Modeling Results

Mean and significance

Table 3 shows the posterior mean and standard deviation of the coefficient estimates (Intercept, β_1, and γ) for all 18 models which indicate their quantitative proportionality on the crash rate by DVMT. β_1, the coefficient for yearly time trend, is statistically significant for all models except the base one. The Base model, which included only heterogeneity random effects, was observed to have four times larger standard deviation for β_0 and β_1 than rest of the spatial models. Apparently, the loss of precision due to greater variance was so high that β_1 was rendered to be insignificant considering the 95% confidence interval. Such significant loss of precision could be attributed to the exclusion of spatial correlations as all other spatial models had a remarkably higher precision. Moreover, the autocorrelation coefficient (γ) appeared to be statistically significant for all models which indicates the strong serial correlation among crashes of successive years. The relatively low magnitude of γ is expected after the linear time trend was applied to all counties.

TABLE 3 Parameter Estimates and Comparison of Model Fit

Models	β_0	β_1	γ	Goodness-of-fit			Average Predicted Crash Rate (units: 10^{-4})	Fraction
	Mean (sd)	Mean (sd)	Mean (sd)	$\overline{D}$	p_D	DIC		
Base model	-8.702 (0.058)	-0.022 (0.016)	0.238 (0.012)	2766.230	280.531	3046.760	1.66674	NA
Queen-1	-8.701 (0.015)	-0.022 (0.006)	0.062 (0.019)	2764.720	219.744	2984.460	1.66404	0.629
Queen-2	-8.699 (0.014)	-0.023 (0.006)	0.058 (0.021)	2764.420	168.707	2933.120	1.66426	0.629
Rook-1	-8.699 (0.013)	-0.023 (0.006)	0.058 (0.019)	2765.180	236.286	3001.470	1.66379	0.631

Rook-2	-8.697 (0.014)	-0.023 (0.006)	0.056 (0.021)	2764.420	178.442	2942.860	1.66413	0.629
Boundary Length	-8.699 (0.014)	-0.023 (0.006)	0.058 (0.023)	2755.210	207.302	2972.510	1.66381	0.629
D0 (Equal Distance)	-8.698 (0.014)	-0.023 (0.006)	0.051 (0.020)	2765.580	230.399	2995.980	1.66452	0.629
D0.5 (1/Distance$^{0.5}$)	-8.698 (0.014)	-0.023 (0.006)	0.053 (0.022)	2765.310	206.322	2971.630	1.66444	0.629
D1 (1/Distance)	-8.695 (0.014)	-0.024 (0.006)	0.053 (0.020)	2765.100	167.782	2932.880	1.66431	0.629
D2 (1/Distance^{2})	-8.701 (0.013)	-0.022 (0.006)	0.057 (0.022)	2765.250	198.287	2963.540	1.66416	0.631
D3 (1/Distance^{3})	-8.702 (0.013)	-0.022 (0.006)	0.056 (0.022)	2765.570	190.159	2955.730	1.66398	0.631
Decay (50)	-8.699 (0.013)	-0.023 (0.006)	0.057 (0.022)	2765.330	59.914	2825.250	1.66412	0.629
Decay (100)	-8.701 (0.014)	-0.022 (0.006)	0.054 (0.021)	2765.060	218.851	2983.910	1.66426	0.629
Decay (150)	-8.699 (0.014)	-0.023 (0.006)	0.054 (0.022)	2765.010	232.488	2997.500	1.66437	0.629
Decay (200)	-8.696 (0.014)	-0.024 (0.007)	0.056 (0.022)	2764.910	147.481	2912.390	1.66439	0.629
Decay (250)	-8.697 (0.014)	-0.023 (0.006)	0.052 (0.021)	2765.210	186.630	2951.840	1.66445	0.629
Gravity-population	-8.705 (0.014)	-0.021 (0.005)	0.077 (0.015)	2764.960	269.729	3034.690	1.66401	0.627
Gravity-DVMT	-8.706 (0.017)	-0.021 (0.007)	0.077 (0.015)	2764.620	265.822	3030.450	1.66419	0.626

Notes: 1.β_0 is the intercept coefficient; β_1 is the time trend coefficient; γ is the autocorrelation coefficient.

2. Nonsignificant variable is shown in the shaded cell.

3. Base model is the one without considering the spatial correlation among counties.

4. sd represents the standard deviation.

5. Refer to Figure 1 for the details of the weight matrices classification.

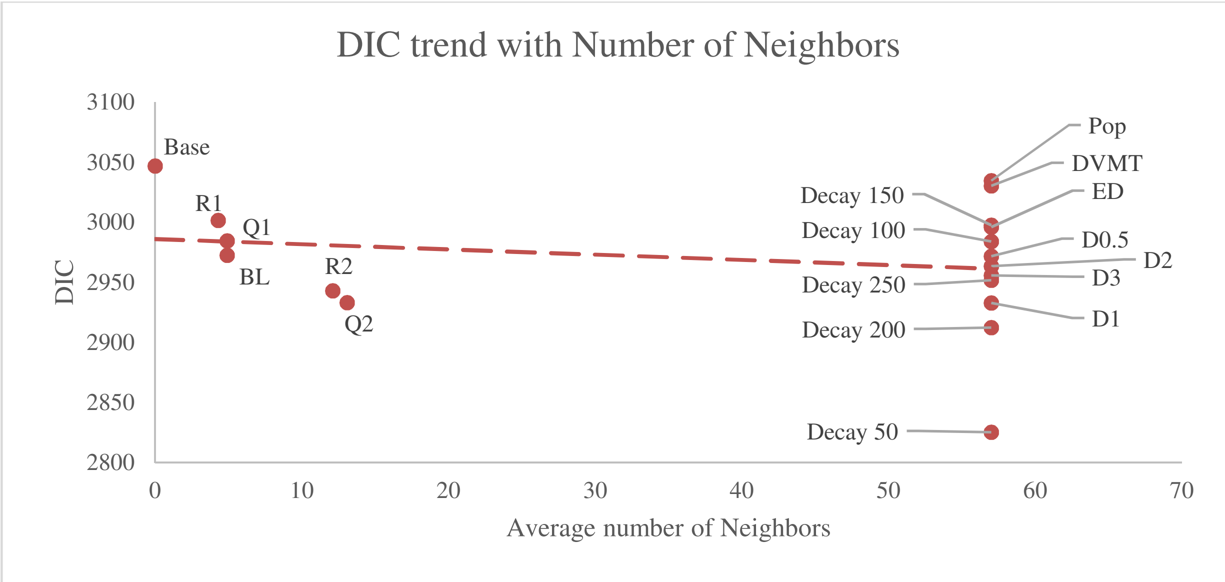


FIGURE 3 DIC trend with the corresponding number of neighbors for different matrices.

The lower value of DIC indicates a better fit of model. As shown in Table 3, the values for 18 matrices range from the lowest (2,825.25) for Decay 50 to the highest (3,046.76) for Base. It should be noted that the value of fit ($\overline{D}$) is mostly similar for all the matrices, but DIC varies greatly as it is governed mainly by the complexity (p_D), which has a wide range across the models with the lowest for Decay 50 (59.9) and highest for Base (280.5). The two Gravity models also exhibit complexities comparable to the Base model. The increase in complexity reflects more effective parameters in the model. Usually, such increase of complexity is compensated by better fitness to accomplish lower DIC value. But the $\overline{D}$ values of these three matrices are similar to the rest, which indicates that better fitness was achieved at the cost of much larger number of effective parameters. As this study employed both adjacency and distance based matrices, the authors investigated the possible influence of number of neighbors involved in a matrix on the goodness-of-fit. A graph was plotted between the DIC and the average number of neighbors for all matrices (Figure 3). It is worth mentioning that there was a significant difference between the number of neighbors for adjacency and distance-based matrices, and also within the adjacency matrices themselves (Table 1). This difference is clearly depicted in Figure 3 where similar matrices are aggregated. A drop in DIC is observed among the adjacency-based matrices, with Rook-2 and Queen-2 demonstrating better fitness of model with the real data. Both these matrices had twice the number of neighbors than their corresponding first order matrices. As for the Q1 and BL, BL has lower DIC even though they have the same number of neighbors. It is probably due to the more information embedded in the BL matrix, or, the length of common boundary length. An overall downward linear trend is observed with an increase in the number of neighbors, indicating that distance-based models had a better fitness collectively. However, it is clearly shown in Figure 3 that the distance models were scattered in a wide range on DIC scale. In order to better understand their behaviors, they were split into two parts: Pure Distance and Decay models and their DIC trend is shown in Figure 4. As shown in the graphs, both groups had a variable trend. In case of Pure Distance matrices, ED (Equal distance) had the worst fitting and then the DIC score steeply dropped until D1, which had a difference of 63 points from ED. The model fitness again drops for D2 and finally improves for D3. A similar highly variable trend was exhibited by Decay matrices with the decay constant of 50 revealing the best fit, which abruptly drops (DIC rises) for Decay 100 and then seems to oscillate till 250. Overall, the trend implies that careful consideration should be placed while dealing with distance-based weight matrices when the distance has a wide range and high variance. Selection of wrong orders for distance–based neighboring structure sometimes could lead to DIC even higher than adjacency-based matrices. It is therefore highly recommended that a sensitivity analysis be imperative for spatial models using distance-based matrices to ensure the appropriate distance-based weight structures are selected.

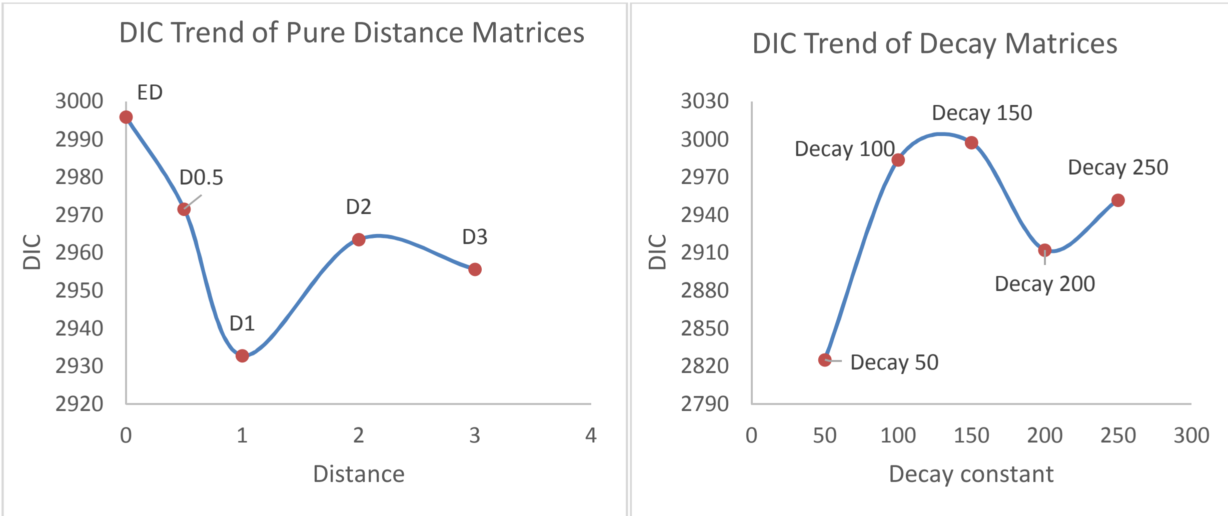


FIGURE 4 DIC trends of distance models.

Average Predicted Crash Rate

As shown in Table 3, the Base model significantly deviates from the rest in terms of model predicted crash rate. In case of distance-based models, a linear trend was observed. For decay models, the predicted crash rate reflected an increase for each increment of decay constant, as it went from 1.66412 for Decay 50 to an upward limit of 1.66445 for Decay 250. Similar linear trend was noted from models D0 to D3, where the average predicted crash rate decreased with an increase in the exponential power of distance function. This variation among the estimation results of models exhibits the tendency of prediction performance to be inclined on the type of distance weight matrices.

Fraction

In terms of fraction, which is a measure of spatial random effect explained by model, the Decay models had a marginal inferiority over Pure Distance as D2 and D3 were able to attribute relatively higher percentage of the variation to spatial effects. The Gravity models had the lowest scores. It seems that the advantage of distance-based modeling was diluted due to the inclusion of extra information of populations and traffic exposure of two counties under consideration. Such low fraction models might be preferable for studies focused on investigation of relationship between spatial covariates as the spatial structure is kept intact whereas high fraction models would serve better for exploring the spatial relationship between crash estimates.

Site Ranking Evaluation Results

The 18 models were subjected to the aforementioned criteria for assessment of complexity and fitness. But it is important to examine if the theoretical advantages of better fitting models are also carried over to the site ranking performance. For that purpose, evaluation tests were conducted to determine the relative performance of models in site ranking. It is worth mentioning here that in accordance with some previous studies (*34*), this study conducted the evaluation tests by comparing the site ranking of all models with Decay-50 since it displayed significantly better model fit and complexity. Table 4 exhibits the corresponding evaluation result.

TABLE 4 Site Ranking Evaluation Results of Alternate Weight Matrices Compared with Decay-50

	Sensitivity	Specificity	PPV	NPV	Cohen's Kappa	TRD	MAD
Base model	1	0.8182	0.9853	1	0.8365	517	2.2342
Queen-1	1	1	1	1	0.8380	130	0.5115
Queen-2	1	0.9545	0.9963	1	0.8384	116	0.4490
Rook-1	1	1	1	1	0.8380	136	0.4927
Rook-2	1	0.9545	0.9963	1	0.8384	124	0.4562
Boundary Length	1	1	1	1	0.8380	176	0.6213
D0 Equal Distance	1	0.9545	0.9963	1	0.8384	110	0.4708
D0.5 (1/Distance$^{0.5}$)	1	0.9545	0.9963	1	0.8384	90	0.4380
D1 (1/Distance)	1	0.9545	0.9963	1	0.8384	71	0.3579
D2 (1/Distance^2)	1	1	1	1	0.8380	52	0.3085
D3 (1/Distance^3)	1	1	1	1	0.8380	94	0.3934
Decay (100)	1	1	1	1	0.8380	57	0.3458
Decay (150)	1	1	1	1	0.8380	72	0.3259
Decay (200)	1	0.9545	0.9963	1	0.8384	81	0.3724
Decay (250)	1	0.9545	0.9963	1	0.8384	85	0.4137
Gravity-population	0.9963	1	1	0.9565	0.8379	170	0.9970
Gravity-DVMT	0.9963	1	1	0.9565	0.8379	170	0.8968
Correlation Coefficient between each criterion and DIC	**-0.55 (0.0224)**	-0.13 (0.6187)	-0.13 (0.624)	**-0.55 (0.0224)**	**-0.71 (0.0015)**	**0.60 (0.01)**	**0.68 (0.003)**

Notes: 1. Base model is the one without considering the spatial correlation among counties.

2. PPV-Positive Predictive Value; NPV-Negative Predictive Value; TRD-Total Ranking Difference; MAD-Mean Absolute Deviation.

3. Refer to Figure 1 for the details of the weight matrices classification.

4 .The statistical significances of the correlation coefficients are shown in parentheses.

5. The statistically significant correlation coefficients are shown in bold.

Sensitivity, Specificity, PPV, and NPV

In practice, the cut-off levels are significantly low due to budget constraints for the funds allocated towards the implementation of safety countermeasures. This study utilized a 10% threshold for filtering out the hazardous sites. As mentioned before, the top 10% counties identified by Decay-50 model are considered as “truly” high risk or hazardous ones. Table 4 demonstrates that, except for the Gravity models, all other models have values of 1 for both sensitivity and specificity. It means these models screen out the same top 10% high risk counties as did Decay 50 matrix, and the percentage of counties identified by these models as high risk which are truly high risk is 100%. For the criteria of Specificity and PPV, the Base model has the smallest values which are 0.8182 and 0.9853, respectively. These numbers suggest that about 18% of counties identified by the Base model as low risk are not truly safe, and the percentage of counties identified by the Base model as high risk which are truly high risk is 98.53%. Overall, the Base model and Gravity models have the lowest site ranking agreement with Decay-50 model in terms of the criteria based on the binary classification of all counties. It is interesting to note that the three models have the largest DIC difference with Decay-50. Such phenomenon indicates that the model with closer predictive capability tend to flag out more common high risk counties.

Cohen’s Kappa, TRD, and MAD

Kappa is another quantitative measure of agreement between Decay-50 and other models. The difference of this criterion lies in the fact that it is based on three categories of counties in the study: high risk, medium risk, and low risk. A higher score indicates more agreement. Since all the models revealed

83% agreement, further scrutiny was required for assessment of the relative site ranking agreement. The Base model had the lowest agreement, closely followed by the Gravity models. Again, such trend was also witnessed for model fitness assessment.

TRD and MAD are more sophisticated tools as they compare each county in a continuous fashion. Lower scores indicate greater agreement between the models and Decay-50. From Table 4, it is known that, under MAD, the Base model has significantly larger value of 2.2343, followed by Gravity models. The similar trend was found for TRD. Once more, these criteria show that the Base model has most ranking difference with Decay-50. Another noteworthy trend observed for TRD and MAD is the superior performance of most of the distance-based models with relatively lower values of TRD and MAD. For example, the lowest TRD score (52 for D2) belongs to a distance-based model which is three times lesser than the worst adjacency-based model (176 for Boundary Length). The same set of models exhibit similar performance for MAD too, where the deviation observed for Boundary Length is twice of D2. The potential rationale may be explained by the fact that distance-based weight matrices give a continuous output depending on the proximity between sites while the adjacency-based matrices are restricted to a binary output. Also, distance-based matrices incorporated more number of neighbors. This supplementary information seems to have a positive impact on the site ranking performance.

Correlation between Each Criterion and DIC

As discussed before, the above criteria seemed to have some correlation with the modeling goodness-of-fit measure: DIC. The authors were motived to run the Pearson's correlation between each site ranking criterion and DIC. The correlation coefficients and associated p-values are shown in the bottom line of Table 4. It is clearly shown that, except specificity and PPV, all other criteria (or, sensitivity, NPV, Kappa, TRD, MAD) demonstrated a statistically significant correlation with DIC. Plus, the associated coefficients were high in magnitude as well. This finding suggests that the models with closer fitness and predictive performance tend to have more similar site ranking performance.

CONCLUSIONS AND RECOMMENDATIONS

This study evaluated the impact of 17 different weight matrix structures on the spatial crash frequency modeling fitness and the relative site ranking performance as well. The serial correlations among crash counts of different years were addressed by the combination of linear time trend and a first order autoregressive (AR-1) error process. The Base model which includes only the heterogeneity random effect was also developed to evaluate the benefit of accounting for the spatial correlations among crash counts in the models. The results yielded the following major findings:

1. The highest values of $\overline{D}$, p_D, and DIC indicated that the inclusion of spatially structured random effect can fit the data better and reduce the effective number of parameters in the model, thereby leading to higher predictive capability.
2. The modeling performance appeared to be increased with the increase in number of neighbors in the weight matrices. For example, Queen adjacency performed better than Rook adjacency, the second order adjacency performed better than first order adjacency, and Decay-50 claimed the remarkably lower DIC than others.
3. The distance-based models had more neighbors being included in the weight matrices and tended to have better modeling performance than others. However, the distance-based models also demonstrated larger variability of DIC's compared with adjacency-based models. Such phenomenon recommends that a careful selection of the appropriate distance-based weight structures is important for the spatial models.
4. The newly proposed the Gravity models appeared to slightly better fit the data than other distance-based models based on the $\overline{D}$ value. However, such benefit was accompanied by much larger effective number of parameters, and therefore lead to a higher DIC.

5. Most of the site ranking evaluation criteria used in the study had a statistically significant correlation with DIC. It suggests that the models with similar predictive capabilities tend to yield similar site ranking performance.

Some recommendations for future research are shown below as well:

1. In our study, the different spatial-proximity matrices were compared by using the intrinsic Gaussian CAR prior. The other formulations of the CAR model such as proper Gaussian CAR, MCAR, mixture models are worthwhile to examine and see if the results reported here can be replicated in those models as well.
2. As the focus of this study was the comparison of different spatial proximity matrices, so no covariates were considered as influential factors for model development. But the authors acknowledge that the incorporation of significant covariates at county level may impact the performance of the spatial models.
3. A sensitivity analysis is recommended for the distance-based matrices as an interesting (though not straightforward) trend was obtained in this study with variation of constant in case of Decay models and exponent of distance in case of Pure Distance models.
4. Finally, such an investigation of spatial weight matrices should be conducted for different area levels (such as TAZs, census tracts, LADs) so that the results of this study may be verified, or the deviations from expected results examined in different spatial analysis units for better understanding of spatial correlations and the deemed benefits.

ACKNOWLEDGMENTS

Support for this study was provided by 2016 UCCONNECT Summer Research Program. The authors thank Southern California Association of Governments for providing common boundary length and geometric centroid distance data of counties used in this study. Finally, the authors would also like to thank anonymous reviewers for their constructive comments and suggestions on an earlier version of the paper.

REFERENCES

1. 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.
2. Guo, F.,Wang, X., Abdel-Aty, M. A., 2010. Modeling signalized intersection safety with corridor-level spatial correlations. Accid.Anal.Prev. 42(1), 84-92.
3. Jonathan, A. Wu, V., Donnell, K., F.K., Donnell, E.T., 2016. A multivariate spatial crash frequency model for identifying sites with promise based on crash types. Accid.Anal.Prev.87, 8-16.
4. 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.
5. Levine, N., K. E. Kim, and L. H. Nitz. Spatial Analysis of Honolulu Motor Vehicle Crashes: II. Zonal Generators. *Accident Analysis and Prevention,* Vol. 27, No. 5, 1995, pp. 675–685
6. Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis & Prevention 38 (6), 1137–1150.
7. Shankar, V., F. Mannering, and W. Barfield. Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies. *Accident Analysis and Prevention,* Vol. 27, 1995, pp. 371–389.
8. 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.
9. Guo, F., X. Wang, and M. A. Abdel-Aty. Corridor-Level Signalized Intersection Safety Analysis Using Bayesian Spatial Models. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
10. Noland, R.B., Quddus, M.A.A., 2004. Spatially disaggregate analysis of road casualties in England. Accident Analysis and Prevention 36, 973–984.

11. MacNab, Y.C., 2004. Bayesian spatial and ecological models for small-area crash and injury analysis. Accident Analysis & Prevention 36 (6), 1019–1028.
12. Abdel-Aty, M., Siddiqui, C., Huang, H., 2011. Integrating trip and roadway characteristics in managing safety at traffic analysis zones. Transportation Research Record 2213, 20–28.
13. De Guevara, F.L., Washington, S.P., Oh, J., 2004. Forecasting crashes at the planning level: simultaneous negative binomial crash model applied in Tucson, Arizona. Transportation Research Record 1897, 191–199.
14. Hadayeghi, Alireza, Amer S. Shalaby, and Bhagwant N. Persaud. "Development of planning level transportation safety tools using Geographically Weighted Poisson Regression." *Accident Analysis & Prevention* 42.2 (2010): 676-688.
15. Washington, S., Schalkwyk, I.V., You, D., Shin, K., Samuelson, J.P., 2010. PLANSAFE User Manual: Forecasting the Safety Impacts of Socio-Demographic Changes and Safety Countermeasures. NCHRP 8-44(2)
16. Amoros, E., Laumon, L.M., 2003. Comparison of road crashes incidence and severity between some French counties. Accident Analysis and Prevention 35, 537–547.
17. Miaou, S.-P., Song, J.J., Mallick, B.K., 2003. Roadway traffic crash mapping: a space-time modeling approach. J. Transp. Stat. 6 (1), 33–57.
18. Noland, R.B., Oh, L., 2004. The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. Accident Analysis and Prevention 36, 525–532.
19. Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accid. Anal. Prev. 38 (3), 618–625.
20. Huang, H., Darwiche, A.L., Abdel-Aty, M.A., 2010. County-level crash risk analysis in Florida: Bayesian spatial modeling. Transp. Res. Rec. 2148, 27–37.
21. Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of road crash frequency with spatial models. Transp. Res. Rec. 2061, 55–63.
22. 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*
23. Jonathan, Aguero-Valverde, Kun-Feng Ken Wu, and Eric T. Donnell. "A multivariate spatial crash frequency model for identifying sites with promise based on crash types." *Accident Analysis & Prevention* 87 (2016): 8-16.
24. Best N, Cockings S, Bennett J, Wakefield J, Elliott P: Ecological regression analysis of environmental benzene exposure and childhood leukaemia: sensitivity to data inaccuracies, geographical scale and ecological bias. *J R Statist Soc A* 2001, 164:155-174.
25. Rhee, K. A., Kim, J. K., Lee, Y. I., & Ulfarsson, G. F. (2016). Spatial regression analysis of traffic crashes in Seoul. *Accident Analysis & Prevention*, *91*, 190-199.
26. Earnest, A., Morgan, G., Mengersen, K., Ryan, L., Summerhayes, R., & Beard, J. (2007). Evaluating the effect of neighbourhood weight matrices on smoothing properties of Conditional Autoregressive (CAR) models. International journal of health geographics, 6(1), 1.
27. Best, N., Richardson, S., & Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical methods in medical research*,*14*(1), 35-59.
28. Lawson, A. B., Browne, W. J., & Rodeiro, C. L. V. (2003). *Disease mapping with WinBUGS and MLwiN* (Vol. 11). John Wiley & Sons.
29. Best NG, Arnold RA, Thomas A, Waller LA, Conlon EM: Bayesian models for spatially correlated disease and exposure data (with discussion). In *Proceedings of the 6th Valencia International Meetings on Bayesian Statistic. Bayesian Statistics 6* Edited by: Bernardo JM, Berger JO, David AP, Smith. Oxford: Oxford University Press; 1999:131-156.
30. Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society, Series B 64 (4), 583–616.

31. Ye, F., & Lord, D. (2011). Investigation of effects of underreporting crash data on three commonly used traffic crash severity models: multinomial logit, ordered probit, and mixed logit. *Transportation Research Record: Journal of the Transportation Research Board*, (2241), 51-58.
32. Miaou, S., Song, J.J., Mallick, B.K., 2003. Roadway traffic crash mapping: a space–time modeling approach. Journal of Transportation and Statistics 6 (1), 33–57.
33. Brooks, Stephen P., and Andrew Gelman. "General methods for monitoring convergence of iterative simulations." Journal of computational and graphical statistics 7.4 (1998): 434-455.
34. 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.
35. Dong, N., Huang, H., Xu, P., Ding, Z., & Wang, D. (2014). Evaluating Spatial-Proximity Structures in Crash Prediction Models at the Level of Traffic Analysis Zones. Transportation Research Record: Journal of the Transportation Research Board, (2432), 46-52.
36. Besag, J., York, J., Molli´e, A., 1991. Bayesian image restoration with two applications in spatial statistics. Ann. Inst. Stat. Math. 43, 1–59.
37. Jones, A., and E. Huddleston. "SAS/STAT 9.2 user’s guide." Cary, NC: SAS Institute Inc (2009).
38. Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. Indian journal of ophthalmology, 56(1), 45.
39. Byrt, Ted, Janet Bishop, and John B. Carlin. "Bias, prevalence and kappa."Journal of clinical epidemiology 46.5 (1993): 423-429.
40. Cheng, Wen, and Simon Washington. "New criteria for evaluating methods of identifying hot spots." Transportation Research Record: Journal of the Transportation Research Board 2083 (2008): 76-85.
41. Huang, Helai, Hoong Chin, and Mazharul Haque. "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 (2009): 32-41.

Paper #4: Alternative Multivariate Multimodal Crash Frequency Models

Wen Cheng, Ph.D., P.E. (corresponding author)
Associate Professor, Department of Civil Engineering,
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Tel: (909) 869-2957, Email: wcheng@cpp.edu

Gurdiljot Singh Gill
Graduate Research Assistant, Department of Civil Engineering,
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Tel: (909) 908-7084, Email: gurdiljotg@cpp.edu

Roya Falahati
Graduate Research Assistant, Department of Civil Engineering,
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Tel: (949) 231-7720, Email: rfalahati@cpp.edu

Xudong Jia, Ph.D., P.E.
Professor, Department of Civil Engineering,
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Tel: (909) 869-4312, Email:xjia@cpp.edu

Jiao Zhou
Graduate Student, Department of Civil Engineering,
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Tel: (714) 488-7965, Email:jiaozhou@cpp.edu

Tom Vo
Modeler, Southern California Association of Governments
818 West 7th Street, 12th Floor, Los Angeles, CA 90017
Telephone: (213) 236-1800, Email: Vo@scag.ca.gov

ABSTRACT

The central issue for successful implementation of multimodal approach is the development of appropriate crash frequency models which can jointly estimate the crash risk of different mode users. This study proposed two multivariate spatial-temporal models to analyze seven years of modal crash data from traffic analysis zones: one with fixed time trend applied to all modes, the other with mode-varying time trend coefficients. These models were compared with three other multivariate models from past studies. The major objective was to examine the benefits of the newly proposed models which have substantially increased computational cost, since both dimensions of time and space are considered, as well as their interactions. Moreover, the relative site-ranking performance among the alternative models was also evaluated with different evaluation criterion. The modeling results indicated that proposed multivariate space-time models had superior performance while the models without spatial and temporal correlations performed the worst. Site ranking evaluation revealed strong positive correlation between site ranking and modeling performance. Overall, the study is anticipated to enhance the understanding of safety impacts from the interaction of various modes.

Keywords: Multimodal approach; Multivariate Spatial-Temporal; Crash Frequency Modeling; Site Ranking; Traffic Analysis Zone

1. INTRODUCTION

Non-motorists are a vulnerable segment of the traveling public, which are defined as road users not in or upon a motor vehicle and consist of walking pedestrians, bicyclists, individuals in wheel chairs or motorized personal conveyances, skateboarders and others (NHTSA, 2012). Considering the potential implications for congestion, health, and the environment, safety for those engaging in active transportation remains a substantial issue. Encouraging individuals to indulge in active transportation, involving walking and bicycling, brings with it a societal obligation to protect commuters as they engage in these modes of travel. Despite the health and environmental benefits (Berrigan et al., 2006; Frank et al., 2010; Furie & Desai, 2012; Giles-Corti et al., 2010; Insall, 2013; Wanner et al., 2012), choosing walking or cycling as the desired mode exposes cyclists and pedestrians to safety risks due to the lack of a protective structure and difference in bodily mass between them and motor vehicles, which renders them prone to heightened injury susceptibility in case of a collision (Williams, 2013).

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 and a shift towards non-motorized modes of transportation, namely walking and cycling. Therefore, literature review illustrated fairly extensive research (Lee and Abdel-Aty, 2005; Moudon et al., 2011; Vivoda et al., 2008; Beck et al., 2007; Wardlaw, 2002; Cai et al., 2016) dedicated to the investigations into factors impacting non-motorist safety on roadways. The results generally concluded that, in addition to roadway design, driver and non-motorist inebriation, low-light conditions, and increased vehicle speed have detrimental effects on non-motorist safety. Nonetheless, while these previous studies are useful for identifying the factors contributing to cyclist, pedestrian and motor-vehicle injury occurrence, these have been modeled separately and few attempts have been made to combine these into a multimodal approach. As a matter of fact, more exhaustive safety portraits may be built by employing a multimodal approach, as it allows the flexibility to simultaneously determine the injury risk of different travel modes. For example, if intersection characteristics are found to increase crashes for cyclists and pedestrians, safety improvements can be carried out for both modes at once. A multimodal approach also may ease the task of selecting sites for safety improvement interventions as well as potentially provide a more economically viable solution, compared to separated analysis and interventions for pedestrians and cyclists.

A central issue to the successful implementation of multimodal approach is the development of multivariate crash frequency models which can jointly estimate the crash risk of different mode users. The simultaneous modeling is essential given the unobserved heterogeneity shared by various transportation modes. Ignorance of such correlation structures has been illustrated to reduce the efficiency of the model due to lesser precise parameters (Bijleveld, 2005; Park & Lord, 2007; Congdon, 2001). In comparison with the large number of univariate models dedicated to various mode users, very few studies have used the joint models to analyze the interaction between different modes. Recently, Convay et al. (2013) performed a bivariate correlation analysis to find the locations of conflict occurrence between bicycles and pedestrians, freight, passenger cars, and cabs in an urban area. The conflict was defined as the obstructions parked in or across the bicycle lane. The characteristics which influenced the conflicts for between these modes were also explored. This study recommended to develop a multivariate regression model for prediction of multimodal conflicts. In order to simultaneously analyze the injury and traffic flow outcomes for different modes, Strauss et al. (2014) subsequently employed Bayesian multivariate Poisson models for studying safety outcomes for motor-vehicle, cyclist and pedestrian flows at intersections. Safety performance functions were developed and crash contributing factors were identified for each mode. A comparative study based on injury risk determined motor-vehicles to be the main risk factor for cyclist and pedestrian injury occurrence at intersections.

One common limitation associated with above two studies lies in the lack of consideration for spatial or serial correlations within crash data. The significance of incorporating spatial correlations was highlighted by many studies (Guo et al., 2010; Abdel-Aty & Wang, 2006) with the consistent better performance of the spatial models over those accounting for heterogeneity random effect only. Likewise, addressing serial correlations has been found to enhance the model fitness and precision by numerous research (Andrey & Yagar, 1993; Hay & Pettitt, 2001; Wang et al., 2006; Wang et al., 2013). Even though multivariate spatial (Aguero-Valverde, 2013) and multivariate temporal models (El-Basyouny et al., 2014) have been applied to estimate various crash severities or outcomes, there is still no or very little research addressing traffic safety issues by employing the multivariate spatial-temporal modeling, whose application has been found in other areas such as environmental engineering and hydrology (Wheater et al., 1991; Choi et al., 2009; Li et al., 2016).

To fill this research gap, the authors proposed two multivariate spatial-temporal models to analyze the modal crash data: one with fixed time trend applied to all modes; the other with mode-varying time trend coefficients. These models were then compared with three types of multivariate models used in the past including multivariate without temporal and spatial random effects, multivariate spatial and multivariate temporal. The major objective is to examine the benefits of the newly proposed models which have substantially increased computational cost since both dimensions of time and space are considered, as well as their interactions. Moreover, the relative site ranking performance among the alternative models were also evaluated with different evaluation criteria with varying complexity. Overall, the study dedicated to the crash analysis of different modes (motor-vehicle only, pedestrian-involved, bicyclist-involved and motorcyclist-related) is anticipated to enhance the understanding of safety impacts from the interaction of various modes.

2. METHODOLOGY

This study analyzed four different transportation mode users-involved crashes occurring at the TAZs (Traffic Analysis Zone) of the City of Irvine in California. The process involved development of multivariate spatial-temporal models and compare their modeling and site ranking performance with three other competing multivariate models assuming a Poisson-Lognormal distribution for crash counts. All models in the study were developed using the Full Bayes approach. Similar to the Empirical Bayesian (EB) method, the Full Bayesian (FB) method has been widely used in safety analysis (Davis & Yang, 2001; Washington & Oh, 2006). Even though numerous studies have illustrated favorable results yielded by the EB method (Maher & Mountain, 1988; Higle & Hecht, 1989; Cheng & Washington, 2005), an FB was chosen due to some of its advantages over EB: its capability to seamlessly integrate prior information and all available data into a posterior distribution (rather than point estimates), its capability to provide more valid safety estimates in smaller data samples, and its capability to allow more complicated model specifications. In addition to the normal Poisson-Gamma distribution, the FB models are also capable of accommodating the Poisson-Lognormal distribution and various hierarchical Poisson distributions that can address the serial and spatial correlations among the sites (Pawlovich et al., 2006; Miranda-Moreno, 2006). The details of various models are presented as follows in the order of complexity.

2.1.1. Model 1: Multivariate Poisson-Lognormal Model (MVPLN)

This model assumes that crash count of certain modal crash **j** at a given location **i** in time **t** (in years), $\mathbf{y_{ijt}}$**,** obeys Poisson distribution, while the corresponding observation specific error term $\boldsymbol{\varepsilon}_{\mathbf{ijt}}$ follows a multivariate normal distribution:

$$y_{ijt}|\lambda_{ijt} \sim Poisson\ (\lambda_{ijt}) \tag{1}$$

$$\ln(\lambda_{ijt}) = X'_{ijt}\beta + \varepsilon_{ijt} \tag{2}$$

$$\varepsilon_{ijt} \sim Normal\ (0,\ \Sigma) \tag{3}$$

Where $$\boldsymbol{y_{ijt}} = \begin{pmatrix} y_{it}^1 \\ y_{it}^2 \\ y_{it}^3 \\ y_{it}^4 \end{pmatrix}, \quad \boldsymbol{\lambda_{ijt}} = \begin{pmatrix} \lambda_{it}^1 \\ \lambda_{it}^2 \\ \lambda_{it}^3 \\ \lambda_{it}^4 \end{pmatrix}, \quad \boldsymbol{\varepsilon_{ijt}} = \begin{pmatrix} \varepsilon_{it}^1 \\ \varepsilon_{it}^2 \\ \varepsilon_{it}^3 \\ \varepsilon_{it}^4 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{14} \\ \vdots & \ddots & \vdots \\ \sigma_{41} & \cdots & \sigma_{44} \end{pmatrix} \tag{4}$$

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

$$\Sigma^{-1} \sim Wishart(I, J) \tag{5}$$

Where I is the **J x J** identity matrix (Congdon, 2006), and **J** is the degree of freedom, J=4 herein representing 4 crash outcomes corresponding to four different modes.

2.1.2. Model 2: Multivariate Poisson-Lognormal with Time Trend (MVPLNT)

Under this model, a yearly trend term t is added to Equation 2 resulting in the new expression:

$$\ln(\lambda_{ijt}) = X'_{ijt}\beta + \varepsilon_{ij} + \gamma t_j * T \tag{6}$$

Where $\boldsymbol{\gamma t_j}$ is the trend coefficient vector for various crash types, and $\mathbf{T}$ is yearly trend. Various types of trend were explored in previous studies (Lawson et al., 2003). This study assumes a linear yearly trend for various crash types with a non-informative prior N (0, 100^2).

2.1.3. Model 3: Multivariate Poisson-Lognormal Spatial Model (MVPLNS)

In this model, a spatially structured error term $\mathbf{u_{ij}}$ is added to Equation 2 which leads to the following expression:

$$\ln\left(\lambda_{ijt}\right) = X'_{ijt}\beta + \varepsilon_{ijt} + u_{ij} \tag{7}$$

Where $\mathbf{u_{ij}}$ is fit by a zero-centered multivariate conditional auto-regressive model (Mardia, 1988) which has a conditional normal density shown as follows:

$$u_i|u_k, \ \ \textstyle\sum_i \ \sim N_j(\sum_{k\sim i} C_{ik}\, , u_k, \sum_i \ ) \tag{8}$$

Where each $\sum_{\boldsymbol{i}}$ is a positive definite matrix representing the conditional variance matrix, and the adjacency matrix $\mathbf{C_{ij}}$ is of the same dimension with $\sum_{\boldsymbol{i}}$ (Jonathan et al., 2016). The precision matrix $\sum^{-1}$ 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 $\mathbf{N_i}$ where $\mathbf{N_i}$ 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 (Wang & Abdel-Aty, 2006; Guo et al., 2010; Aguero-Valverde & Jovanis, 2006; Xu & Huang, 2015), weight structures such as various adjacency-based, distance-based models, and semi-parametric geographically weighted, have been explored. As the current study is focused on the evaluation of alternate model-based HSID methods, the distance-based structure was used as an example to explore the spatial correlations with the following formulation:

$$w_{ij} = \frac{1}{d_{ij}} \tag{9}$$

Where $\mathbf{w_{ij}}$ is the weight between intersection **i** and **j**, and $\mathbf{d_{ij}}$ is the distance between intersection **i** and **j**. With this weight structure, it is known that more weightage was assigned to intersections which are relatively closer.

2.1.4. Model 4: Multivariate Poisson-Lognormal Spatial-Temporal Model (MVPLNST) with Fixed Time Coefficient

This model represents the first multivariate space-time model with the assumption of fixed yearly trend for various crash types. The corresponding formula is shown as follows:

$$\ln(\lambda_{ijt}) = X'_{ijt}\beta + \varepsilon_{ij} + u_{ij} + (\gamma t + \delta_{ij}) * T \tag{10}$$

Where $\boldsymbol{\gamma}\mathbf{t}$ is the fixed yearly trend coefficient for all crash types, and $\boldsymbol{\delta}_{\mathbf{ij}}$ is an interaction random effect between space and time which allows different temporal trends in crash risk for different spatial locations. $\boldsymbol{\gamma}\mathbf{t}$ was assigned a non-information prior of $N(0, 100^2)$ and $\boldsymbol{\delta}_{\mathbf{ij}}$ was assumed to have the same prior with $\mathbf{u_{ij}}$.

2.1.5. Model 5: Multivariate Poisson-Lognormal Spatial-Temporal Model (MVPLNST) with Varying Time Coefficients for Crash Types

This model represents the second multivariate space-time model under the premise that the yearly trends for various crash types are different. The model for is of the following form:

$$\ln(\lambda_{ijt}) = X'_{ijt}\beta + \varepsilon_{ij} + u_{ij} + (\gamma t_j + \delta_{ij}) * T \tag{11}$$

Where $\boldsymbol{\gamma}\mathbf{t_j}$has the same definition and prior distribution as shown in Equation 6.

2.2. Goodness-of-Fit of the 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:

$$DIC = \overline{D} + p_D \tag{12}$$

Where $\overline{\mathbf{D}}$ is the sum of the posterior mean deviance which measures how well the model fits the data; the smaller the $\overline{\mathbf{D}}$, the better the fit. $\mathbf{p_D}$ represents the effective number of parameters. In general, $\overline{\mathbf{D}}$ will decrease as the number of parameters in a model increases. Therefore, the $\mathbf{p_D}$ term is mainly used to compensate for this effect by favoring models with a smaller number of parameters. This idea is analogous to other penalized fit criteria such as Akaike information criterion (Akaike, 2011) and Bayesian information criterion (Schwarz, 1978). Based on the model-selection decision criteria suggested by Best et al. (2005): the models with DIC values within 1 or 2 of the 'best' model are also strongly supported, values within 3 and 7, weakly supported, and models with a DIC greater than 7 have substantially inferior modeling performance.

2.3. Site Ranking Evaluation

One natural question is how the modeling performance would impact on the site ranking. Would the models with similar modeling performance yield close ranking results? To answer this question, the study employed the model with best performance as the base model and assumed it would yield the “true” safety estimates. Building on such premise, three types of criteria with the increasing group divisions were used for relative site ranking performance.

2.3.1. Sensitivity and Specificity

Under these criteria, the TAZs were divided into two groups based on estimated crash counts from various models with a certain threshold. In the present study, a cutline of 5% was used to replicate the general real-world practice. The true safe or unsafe locations can be identified with the specified cutline applied to estimated crash counts by the base model (Huang et al., 2009). The Sensitivity and Specificity can be calculated with the following equation:

$$Sensitivity = \frac{CP\ (correct\ positives)}{TP\ (true\ positives)} \tag{13}$$

$$Specificity = \frac{CN\ (correct\ negatives)}{TN\ (true\ negatives)} \tag{14}$$

Where, CP: number of true unsafe locations correctly identified by the comparative model; TP: total number of truly unsafe locations; CN: number of true safe locations correctly flagged out by the comparative model; and TN: total number of truly safe locations. The larger the values of sensitivity and specificity, higher the agreement between the comparative and base model in site ranking.

2.3.2. Cohen's Kappa

Compared with sensitivity and specificity, Kappa statistics has two unique features: 1. It allows to check the agreement of two models based on more than two groups; 2. It can exclude those true identifications by chance. Kappa can be calculated as follows:

$$K = \frac{(P_o - P_e)}{(1 - P_e)} \tag{15}$$

Where, $\mathbf{P_o}$ is the proportion of observations in agreement, and $\mathbf{P_e}$ represents the proportion in agreement due to chance. A relatively higher value of Kappa indicates a larger number of sites (agreement) which are commonly identified as unsafe between the crash frequency models.

To ensure more reliable comparison results among various models, the study calculated three Kappa statistics shown as below.

- Kappa (3): All TAZ's were divided into three groups based on estimated crash counts (top 5%, middle 90%, and bottom 5%).
- Kappa (4): the crash dataset of TAZ were equally divided into four groups (20% each).
- Kappa (5): the crash dataset of TAZ were equally divided into five groups (25% each).

2.3.3. Mean Absolute Deviation (MAD)

With this criterion, the relative site ranking performance of various models can be evaluated by comparing all TAZs in a continuous fashion. Technically, each TAZ is treated as an individual group. MAD aims to estimate the average difference of estimated crash counts by comparing models for each location, and it can be calculated with the following equation:

$$\mathrm{MAD} = \frac{1}{n}\sum_{i=1}^{n}\left|C_{i,j} - C_{i,k}\right| \tag{16}$$

Where $\mathbf{C_{i,j}}$ is the estimated crash count of TAZ **i** by model **j**, and $\mathbf{C_{i,k}}$ is the estimated crash count of TAZ **i** by model **k**. The smaller the MAD value, the more similarly the two models perform in site ranking.

3. DATA PREPARATION

This macro-level study analyzed the crashes of different modes which occurred in the City of Irvine in the period of 2006–2012. As demonstrated in previous research (Abdel-Aty et al., 2013), compared with other geographic units such as block groups and census tracts, TAZs have benefits of better homogeneity and easy integration into the transportation planning process. Various other studies have focused on the TAZs for planning-level analysis of crashes. With an aim to incorporate proactive safety measures in transportation planning, Siddiqui et al. (2012a) employed nonparametric statistical techniques for investigation of significant variables impacting severe crashes aggregated for TAZs. Subsequently, another study by Siddiqui et al. (2012b) examined the effect of spatial correlation for modeling bicycle and pedestrian crashes in the TAZs of two counties of Florida. Pulugurtha et al. (2013) developed crash estimation models at TAZ level based on the land use characteristics and revealed strong association of between such factors and crash occurrences. Observing a statistically significant role of such factors at estimation of crashes, this study recommended to utilize the modeling results for "safety conscious planning, land use decisions, long range transportation plans, and, to proactively apply safety treatments in high risk TAZs". Recently, a study by Cai et al. (2016) employed various traffic, roadway, and socio-demographic covariates from neighboring TAZs to explore the spatial spillover effects for analysis of pedestrian and bicycle crashes, and observed them to be significant for the concerned crashes. In view of the aforementioned studies, TAZs were selected as the base units for the current study, and the crash data were aggregated at the TAZ-level. Overall, there are 203 TAZs in the City. The map in Figure 1 displays the distribution of all TAZs and associated crash counts. Four different transportation mode-related crashes were collected from SWITRS (California Statewide Integrated Traffic Records System) which include pedestrian, bicyclist, motorcyclist and vehicle only crashes. Shape file of TAZ boundary and TAZ characteristics were provided by SCAG (Southern California Association of Governments).

Figure 1 about here

The variables used for model development and the associated descriptive statistics are shown in Table 1. The numbers of various transportation mode-involved crashes were used as the dependent variables. DVMT was utilized as the exposure variable. 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. In addition, the distance matrix containing distances among various TAZ centroids were also collected from SCAG for the estimation of distance-based spatial random effect. Since there are 203 TAZs in the city, the matrix includes 203x202 distances. Their descriptive statistics can be found in Table 1 as well.

Table 1 about here

4. RESULTS

4.1. Model Comparison

In general, the larger the effective number of parameters is, the easier it is for the model to fit the data. To obtain a parsimonious model and avoid risk of over-fitting, backward stepwise methods were employed in selecting covariates. Besides, a correlation matrix for the variables entered in the final models has been checked to avoid multi-collinearity issues. Results of parameter estimation and associated uncertainty estimates of significant variables in the final models are presented in Table 2. It is known that the same significant variables are identified for all five models across different crash types. The robustness of results indicates that the models yield no difference in selecting the influential factors of crashes. However, the goodness-of-fit measures reveal the different performance of models. The MVPLNT model leads to the lowest Dbar value, indicating the multivariate model with time trend only fits the data very well. However, such benefit is accompanied by the second largest p_D, showing the relative larger effective number of parameters. On the contrary, the MVPLNS model enjoys the second lowest value of p_D, while having the highest Dbar. Overall, MVPLNST models have relatively lower values of both Dbar and p_D, resulting in the lowest DIC value, followed by MVPLNS and MVPLNT, with MVPLN having the largest DIC. Since the DIC differences are more than 7 points among all models, it can be concluded that MVPLNST, especially the one with varying time coefficient for different transportation mode users, significantly improve the model-fitting performance by borrowing strength from neighbors as well as considering the time trend. On the other hand, the MVPLN model, which doesn't consider either temporal or spatial effect, has the inferior modeling performance. As for the MVPLNS and MVPLNT models, the former one has an overall better performance than the latter one.

The variable coefficients change little across the five multivariate models. Comparatively speaking, the change of coefficients for the smaller sample size outcomes (motorcyclists, backlists, and pedestrians) is larger than that of the larger one (vehicle-only). Regarding the variable significance, the constant and percent of arterials are significant in each crash outcome, with the coefficient of latter variable ranging from -6.43 to -0.43. Since the arterial streets generally have higher design standards and are more difficult to cross, it indicates that the improved design standards and fewer interactions among various transportation mode users could reduce the crash risks. The percent of population age 18-24, bike lane density and college enrollments appear to have a significant impact on all crash outcomes except motorcyclist-involved crashes, while the proportion of people age 65 and older exerts a significant influence on all types of crashes except pedestrian-related crashes. As for the yearly trend, if the fixed trend is assumed for all crash type, the coefficient is significant with a negative value. However, if the varying trends are expected for different crash outcomes, the coefficients are merely significant for vehicle-related (negative) and bicyclist-related crashes (positive). Since the fixed trend and the vehicle-related crash time trend has the same sign, it can be concluded that the larger sample size has more impact on the time trend than does the lower sample size.

Table 2 about here

4.2. Correlation among Crash Types

Similar to previous literature (Park & Lord, 2007; Aguero-Valverde & Jovanis, 2009; Aguero-Valverde et al., 2016), the variance estimates of all four crash types for all multivariate models are statistically significant at the 0.05 level of significance which indicate the presence of over-dispersion in all modal crashes. In addition, a correlation analysis of error terms for all models were conducted. The result for multivariate space time model with varying time trend (Model 5) is shown in Table 3 for the illustrative purpose. It exhibits that the correlations are statistically significant for heterogeneity error term, the spatial random effect, and the one interacted with time among various crash types, demonstrating that the occurrence of various crash types is highly correlated. The highest correlation is observed between MC and Bike crashes (0.75), which might suggest that the motorcyclists and bicyclists have closer behaviors than other mode users.

Table 3 about here

4.3. Performance Comparison in Site Ranking

As determined by the DIC, the estimated crashes by Model 5 exhibited the best fit with the actual crash data. Hence, it was used to establish the "true" safety estimates of each zone and served as the base for comparison of site ranking performance of other four models. This is in line with the same practice by other studies (Aguero-Valverde et al., 2016). The performance of models was evaluated by employing different criteria, ranging from binary (sensitivity and specificity), to multiple group divisions (Kappa 3, 4, and 5 levels) and finally a continuous approach where each site is considered to be an individual group (MAD).

Table 4 about here

4.3.1. Sensitivity and Specificity

To replicate the real-world practices of hotspot identification, the threshold was set at 5%. This approach is binary as for each site there are two possible outcomes in terms of detection: after the filtering at top 5%, either a site could be regarded as safe, or unsafe. As depicted from Table 4, Models 2 and 4 both had 100% sensitivity and specificity. This is the scenario where all the truly hazardous sites were detected as hazardous by the models. For Models 1 and 3, the sensitivity was lower than specificity. This implies that the models were "better" at correct identification of truly safe sites compared with the identification of truly unsafe sites. Overall, the criteria demonstrated that Models 2 and 4 had closer agreement with the base model, Model 5.

4.3.2. Kappa

This study calculated the agreement between estimated unsafe sites and assumed unsafe sites (from base model) using the kappa statistic. This approach is essentially a refinement of sensitivity, as random crossing of thresholds is prevented for sites which are not truly hazardous. Larger value of Kappa demonstrates a larger agreement. To conduct an expansive study over a wider spectrum, three types of Kappa were calculated which differed by the thresholds. This approach ensures the evaluation of agreement over a larger dataset and range. As shown in Table 4, Model 4 consistently claimed the higher Kappa values within all divisional groups demonstrating its closest ranking result to Model 5. On the other hand, Model 1 exhibits the lowest Kappa value which seems reasonable as no random effects were included to account for correlations, which is totally different than Model 5.

4.3.3. MAD

The previous two criteria divided the crash dataset into two groups and then a maximum of five groups. MAD is a more comprehensive approach since all the crash sites are considered to be individual group. This method takes advantage of the continuous nature of crash counts. A smaller value of MAD signifies that the crash estimates of a particular model show lesser deviance compared to the base model. From the MAD results shown in Table 4, it is evident that Method 4 again performs closet to Method 5, followed by Method 3 which incorporated only spatial correlations. It is noteworthy that the evaluation results consistently demonstrate the positive correlation between DIC and ranking performance. The closer DICs the models have, the more agreement in their site ranking.

5. CONCLUSIONS AND RECOMMENDATIONS

The traffic safety field has employed separate temporal and spatial correlations for simultaneous estimation of crash outcomes. However, this is no or little research considering both dimensions of time and space, as well as the associated interactions, for the multivariate models. To this end, this study proposed two multivariate spatial-temporal models. The proposed models were developed using Full Bayesian framework and incorporated the spatial-temporal random effects with fixed and mode-varying time coefficients for various modal crashes. This study was primarily focused on the comparison of the proposed models with the alternate multivariate models which either did not incorporate or incorporated only one correlation: spatial or temporal. The study area was selected to be TAZ level due to the benefit of easier integration into transportation planning.

The models were compared based on the fitness of estimated and observed crash data, and relative site ranking performance using three evaluation criteria. The model fitness results from DIC revealed that the proposed models significantly improved the model fitting by pooling strength from the neighbors, consideration of time trend, as well as their interactions. Among the two proposed models, the model with mode-varying time coefficients was observed to be better. The influential factors for all the models were the same. The relative site ranking performance using different evaluation criteria with increasing group divisions consistently showed the strong positive correlation between modeling and site ranking performances. In other words, the models with closer DIC's tend to yield more similar ranking results.

Although the study clearly demonstrated the advantages of the proposed models due to the capabilities of combination of spatial and temporal random effects, still there are some recommendations to further bolster the significance of these models. Firstly, this study was focused at the TAZ level with a set of influential variables. Somewhat different results may be expected for other geographic areas, like block level, county, or smaller entities like intersections. Secondly, the fitness of models was assessed by employing the DIC. Other techniques could be utilized for such assessment like MAPE, RMSE, MSPE, among others. Moreover, cross validation techniques would also help verify the expected advantages at crash prediction. Finally, this study used a linear time-space interaction for development of models. The fitness and performance of other time-space relationships could be explored and compared with the proposed models.

ACKNOWLEDGMENTS

Support for this study was provided by 2016 UCCONNECT Summer Research Program. The authors would also thank Southern California Association of Governments for providing covariates, shape files and geometric centroid distance data of TAZ used in this study.

REFERENCES

1. Abdel-Aty, M., & Wang, X. (2006). "Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors." *Transportation Research Record: Journal of the Transportation Research Board*, (1953), 98-111.

2. Abdel-Aty, M., Lee, J., Siddiqui, C., & Choi, K. (2013). "Geographical unit based analysis in the context of transportation safety planning." *Transportation Research Part A: Policy and Practice*, *49*, 62-75.

3. Akaike, H. (2011). "Akaike's Information Criterion." In *International Encyclopedia of Statistical Science* (pp. 25-25). Springer Berlin Heidelberg

4. 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.

5. Aguero-Valverde, J., & Jovanis, P. (2009). "Bayesian multivariate poisson lognormal models for crash severity modeling and site ranking." *Transportation Research Record*, *2136*(2136), 82-91.

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. Jonathan, A. V., Wu, K. F. K., & Donnell, E. T. (2016). "A multivariate spatial crash frequency model for identifying sites with promise based on crash types." *Accident Analysis & Prevention*, *87*, 8-16.

8. Andrey, J., & Yagar, S. (1993). "A temporal analysis of rain-related crash risk." *Accident Analysis & Prevention*, *25*(4), 465-472.

9. 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.

10. Berrigan, D., Troiano, R.P., McNee, T., Disogra, C., Ballard-Barbash, R., (2006). "Active transportation increases adherence to activity recommendations." *American journal of preventive medicine,* 31(3), 210-216.

11. Best N, Richardson S, Thomson A. (2005). "A comparison of Bayesian spatial models for disease mapping." *Statistical methods in medical research,* 14(1), 35-59.

12. Bijleveld, F. D. (2005). "The Covariance between the Number of Accidents and the Number of Victims in Multivariate Analysis of Accident Related Outcomes." *Accident Analysis and Prevention,* Vol. 37, No. 4, pp. 591–600.

13. 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.

14. Cheng, W., Washington, S.P., (2005). "Experimental evaluation of hotspot identification methods." *Accident Analysis & Prevention* 37 (5), 870–881.

15. Choi, J., Reich, B. J., Fuentes, M., & Davis, J. M. (2009). "Multivariate spatial-temporal modeling and prediction of speciated fine particles." *Journal of statistical theory and practice*, *3*(2), 407-418.

16. Congdon, P., (2001). "Bayesian Statistical Modeling." *John Wiley & Sons, West Sussex, United Kingdom*

17. Congdon, P., (2006). "Bayesian Statistical Modeling." *2nd ed. Wiley, New York.*

18. 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.

19. Davis, G.A., Yang, S., (2001). "Bayesian identification of high-risk intersections for older drivers via Gibbs sampling." *Transportation Research Record: Journal of the Transportation Research Board*,1746, 84–89

20. 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 & Prevention*, *73*, 91-99

21. Frank, L.D., Greenwald, M.J., Winkelman, S., Chapman, J., Kavage, S., (2010). "Carbonless footprints: promoting health and climate change stabilization through active transportation." *Preventive Medicine. 50, S99-S105.*

22. Furie, G. L., & Desai, M. M. (2012). "Active transportation and cardiovascular disease risk factors in US adults." *American journal of preventive medicine*, *43*(6), 621-628.

23. Giles-Corti, B., Foster, S., Shilton, T., Falconer, R., (2010). "The co-benefits for health of investing in active transportation." *N. S. W. Public Health Bull. 21* (5–6 (May–Jun)), 122–127

24. Guo, F.,Wang, X., Abdel-Aty,M. A., 2010. "Modeling signalized intersection safety with corridor-level spatial correlations." *Accid.Anal.Prev*. 42(1), 84-92.

25. 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.

26. Higle, J. L. and M. B. Hecht. (1989). “A Comparison of Techniques for the Identification of Hazardous Locations.” *Transportation Research Record 1238*, *TRB, National Research Council, Washington, DC*, pp. 10–19.

27. Huang, H., Chin, H., & 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), 32-41.

28. Insall, P., (2013). “Active travel: transport policy and practice for health.” *Nutrition Bulletin. 38* (1), 61–69,

29. Jonathan, Aguero-Valverde, Kun-Feng Ken Wu, and Eric T. Donnell (2016). "A multivariate spatial crash frequency model for identifying sites with promise based on crash types." *Accident Analysis & Prevention* 87: 8-16.

30. Lawson, A. B., Browne, W. J., & Rodeiro, C. L. V. (2003). *Disease mapping with WinBUGS and MLwiN* (Vol. 11). John Wiley & Sons.

31. Lee, C., Abdel-Aty, M., (2005). “Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida.” *Accid. Anal. Prev. 37* (4 (Jul)), 775–786.

32. Li, D., Liang, J., Di, Y., Gong, H., & Guo, X. (2016). “The spatial-temporal variations of water quality in controlling points of the main rivers flowing into the Miyun Reservoir from 1991 to 2011.” *Environmental monitoring and assessment*, *188*(1), 1-12.

33. Maher, M.J., Mountain, L.J., (1988). "The Identification of Accident Blackspots: A Comparison of Current Methods." *Accident Analysis and Prevention*, Vol. 20, No. 2, pp. 143–151.

34. Mardia, K.V., (1988). "Multi-dimensional multivariate Gaussian Markov random fields with application to image processing." *J. Multivar. Anal.* 24 (2), 265–268.

35. Miranda-Moreno, L.F., (2006). "Statistical Models and Methods for Identifying Haz-ardous Locations for Safety Improvements." *University of Waterloo.*

36. Moudon, A.V., Lin, L., Jiao, J., Hurvitz, P., Reeves, P., (2011). "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." *Accid. Anal. Prev. 43* (1 (Jan)), 11–24

37. National Highway Traffic Safety Administration, (2012). "Traffic Safety Facts: Bicyclists and Other Cyclists." *United States Department of Transportation* (DOT HS 811 624).

38. Park, E., & Lord, D. (2007). "Multivariate poisson-lognormal models for jointly modelling crash frequency by severity." *Transportation Research Record*, *2019*(2019), 1-6

39. 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.

40. Pulugurtha, S. S., Duddu, V. R., & Kotagiri, Y. (2013). "Traffic analysis zone level crash estimation models based on land use characteristics." *Accident Analysis & Prevention*, *50*, 678-687.

41. 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*

42. Schwarz, Gideon. (1978) "Estimating the dimension of a model." *The annals of statistics* 6.2: 461-464.

43. Siddiqui, C., Abdel-Aty, M., & Huang, H. (2012a). "Aggregate nonparametric safety analysis of traffic zones." *Accident Analysis & Prevention*, *45*, 317-325.

44. Siddiqui, C., Abdel-Aty, M., & Choi, K. (2012b). "Macroscopic spatial analysis of pedestrian and bicycle crashes." *Accident Analysis & Prevention*, *45*, 382-391.

45. Strauss, J., Miranda-Moreno, L. F., & Morency, P. (2014). "Multimodal injury risk analysis of road users at signalized and non-signalized intersections." *Accident Analysis & Prevention*, *71*, 201-209.

46. Vivoda, J.M., Eby, D.W., St Louis, R.M., Kostyniuk, L.P., (2008). "Cellular phone use while driving at night." *Traffic Inj. Prev. 9* (1 (Mar)), 37–41

47. Wang, X., Abdel-Aty, M., & Brady, P. (2006). "Crash estimation at signalized intersections: significant factors and temporal effect." *Transportation Research Record: Journal of the Transportation Research Board*, (1953), 10-20.

48. Wang, X. and Abdel-Aty, M. (2006). "Temporal and spatial analyses of rear-end crashes at signalized intersections." *Accident Analysis & Prevention* 38.6: 1137-1150.

49. Wang, C., Quddus, M., & Ison, S. (2013). "A spatio-temporal analysis of the impact of congestion on traffic safety on major roads in the UK." *Transportmetrica A: Transport Science, 9*(2), 124-148.

50. Wanner, M., Götschi, T., Martin-Diener, E., Kahlmeier, S., Martin, B.W., (2012). "Active transport, physical activity, and body weight in adults: a systematic review." *American journal of preventive medicine 42* (5 (May)), 493–502

51. Wardlaw, M., (2002). "Assessing the actual risks faced by cyclists." *Traffic Eng. Control* 352–356.

52. Washington, S., Oh, J., (2006). "Bayesian Methodology Incorporating Expert Judgment for Ranking Countermeasures Effectiveness under Uncertainty: Example Applied to at Grade Railroad Crossings in Korea." *Accid. Anal. Prev*. 38, 234–247.

53. Wheater, H. S., Butler, A. P., Stewart, E. J., & Hamilton, G. S. (1991). "A multivariate spatial-temporal model of rainfall in southwest Saudi Arabia. I. Spatial rainfall characteristics and model formulation." *Journal of Hydrology*,*125*(3), 175-199.

54. Williams, A.F., (2013). "Protecting Pedestrians and Bicyclists: Some Observations and Research Opportunities." *Insurance Institute for Highway Safety*.

55. Xu, P., Huang, H., (2015). "Modeling crash spatial heterogeneity: Random parameter versus geographically weighting." *Accid. Anal. Prev*. 75, 16-25.

Paper #5: Predicting Likelihood of Hit-and-run Crashes Using Real-time Loop Detector Data and Hierarchical Bayesian Binary Logit Model with Random Effects

Meiquan Xie, Ph.D.
Visiting Scholar, Department of Civil Engineering,
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Fax: (909) 869-4342; Tel: (909)282-9142; Email: mxie@cpp.edu
& Assistant Professor, School of Transportation and Logistics,
Central South University of Forestry and Technology,
Tianxin, Changsha, Hunan, China, 410000
Fax: (909) 869-4342; Tel: (909)282-9142; Email: mxie@cpp.edu

Wen Cheng, Ph.D. P.E. (corresponding author)
Associate Professor, Department of Civil Engineering
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Fax: (909) 869-4342; Tel: (909) 869-2957; Email: wcheng@cpp.edu

Gurdiljot Singh Gill
Graduate Student, Department of Civil Engineering
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Fax: (909) 869-4342; Tel: (909) 869-4312; Email: gurdiljotg@cpp.edu

Jiao Zhou
Graduate Student, Department of Civil Engineering
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Fax: (909) 869-4342; Tel: (909)869-4312; Email: jiaozhou@cpp.edu

Xudong Jia, Ph.D. P.E.
Professor, Department of Civil Engineering
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA 91768
Fax: (909) 869-4342; Tel: (909) 869-4312; Email: xjia@cpp.edu

Simon Choi, Ph.D., AICP CTP
Chief of Research and Forecasting
SOUTHERN CALIFORNIA ASSOCIATION OF GOVERNMENTS
818 West 7th Street, 12th Floor, Los Angeles, CA 90017
Fax: (213) 236-1963; Tel: (213) 236-1849; Email: choi@scag.ca.gov

ABSTRACT

Objective: Fairly extensive research studies have been dedicated to identifying the influential factors of hit-and-run (HR) crashes. Most of them utilized the typical Maximum Likelihood Estimation Binary Logit models, and none of them have employed the real-time traffic data. To fill this gap, the study focused on predicting the likelihood of HR crashes, as well the general ones.

Methods: This study employed the hierarchical Bayesian models with random effects within a sequential Logit structure. Two-year crash and real time loop detector data were collected from one freeway segment in Southern California. Along with the evaluation of impact of random effects on model fitness and complexity, k-fold cross validation technique was also used to examine the predictive capability of the proposed model. Stepwise incremental sensitivity and specificity were calculated for each training set. Finally, ROC (Receiver Operating Characteristic) curve is utilized to graphically illustrate the predictive performance of the model.

Results: The results indicated similar significant contributing factors to general crashes as in the previous research. As for the HR crashes, the factors of upstream vehicle speed, roadway segment length, and weekend were found to be positively correlated with the HR crash risk. K-fold cross validation technique resulting in ROC (Receiver Operating Characteristic) curves exhibited the satisfactory prediction accuracy of the developed models in the study.

Conclusions: The probable reason for higher upstream speed of vehicles to be an influential factor for HR crashes could be attributed to the theory that higher vehicle speeds usually indicate presence of relatively smaller number of vehicles, which may be perceived as a lower risk situation by the driver as it is easier to escape detection in absence of witnesses. The considerable impact of weekend on HR occurrence hints at the possible indulgence in alcohol and speeding. Moreover, relatively superior fitness of the proposed model (ROC graph) was observed which may be accredited to the inclusion of random effects into the binary logit model which addressed the unobserved heterogeneity. Overall, the research suggested that the real-time traffic data seems to be a promising area of interest for understanding the conditions of HR crashes.

Key Words: Hit-and-run Crashes, Hierarchical Bayesian Binary Logit Models with Random Effects; Sequential Logit Structure; Receiver Operating Characteristic; Real Time Loop Detector Data

INTRODUCTION

Road crashes are a leading cause of death and also put an economic burden, based on level of severity of a crash (Blincoe et al., 2002). Although the factors contributing to the frequency and severity of crashes are diverse and complex, one of the significant determinants of the fatality risk of a crash is hit-and-run (HR) behavior (Tay et al., 2008; Kim et al., 2008). Crash data from the National Highway Traffic Safety Administration show that the number of fatal HR crashes is trending upward, from 1,274 in 2009, to 1,393 in 2010, to 1,449 in 2011. Perhaps more significantly, the 13.7% increase in HR deaths over that three-year period occurred while traffic deaths overall were falling 4.5%, from 33,883 in 2009 to 32,367 in 2011. HR crashes are defined as collisions where the driver of the striking vehicle leaves the scene before offering information or aid to the victim, or reporting to the police. It is a punishable offence, along with being unethical, as it delays crash notification. Since about 35% of fatalities occur within 1–2 hours of crash occurrence (Roess et al., 2004), hence delaying emergency response and medical assistance significantly increases the fatality risks (Tay et al., 2008). Some victims who escape immediately death after the collision, later die as a result of poor trauma care (Mock et al., 1997; Peden et al., 2004). HR drivers also increase the victim's exposure to being struck again by a subsequent vehicle (MacLeaod et al., 2012). Most of these accident victims may not have died if they had been rushed to the hospital for medical treatment immediately following the accident.

Although the literature on hit-and-run crashes is fairly extensive, most of the studies concentrated on developing methods to identify the vehicles involved to aid the apprehension of the offenders who left the crash scene without reporting it (Baucom, 2006; Taylor et al., 1989; Locke et al., 1988). Some studies in medical field examined the type of injury sustained by victims in HR crashes to identify the types of vehicles involved (Teresinski and Madro, 2001; Karger et al., 2001). Limited research has been conducted to explore different characteristics that serve as a motivation for the driver to flee the crash scene. The study by Solnick and Hemenway (1995) was among the foremost attempts to investigate the contributing factors for HR pedestrian fatalities. Multiple logistic regression was used to examine the effect of significant variables on the probability of HR. The authors observed the victim's age had a significant impact on likelihood of HR as only 10% drivers ran after killing children or elderly. Also, the chances of escaping detection or being at fault for driving under influence (DUI) increased the propensity of HR. Kim et al., (2008) used rough set analysis (RSA), along with Binary logistic regression using maximum likelihood estimation method, for investigation of the critical attributes and determinants which contributed to the occurrence of 4,939 HR crashes of all severity levels. The authors found that the human and behavioral factors prevailed over roadway and environmental attributes (such as being a male, tourist, intoxicated, and driving a stolen vehicle). Tay et al. (2008) extended the previous studies by incorporating an array of different independent variables to investigate HR crashes in an urban setting of Singapore. The standard decision analysis framework was used to analyze the motivation behind the driver's decision to HR based on expected costs and benefits of the two choices: to stay and report the crash, or to flee the crash site. The results demonstrated that the perception of possible escape of detection and avoiding tough legal consequences greatly influence the decision of a driver to leave a crash site. Another study by Tay et al. (2009) focused only on fatal crash outcomes of HR. In addition to a multitude of factors, traffic characteristics were also investigated such as, traffic flow, speed limit, traffic control device, and so on. It is noteworthy that these covariates relied on the invariable traffic data (such as groups of speed limit or types of traffic control devices) rather than incorporating real-time traffic characteristics (such as occupancy or vehicular speed). Two macro-level studies employed binary logistic regression models on the national data to investigate the factors which influence HR crashes resulting in pedestrians fatalities: Macleod et al. (2012) in United States and Aidoo et al. (2013) in Ghana. The results from Macleod et al. (2012) indicated an increased risk of HR in the early morning, poor light conditions, and on the weekend. Aidoo et al. (2013) specified that unclear weather conditions, along with nighttime conditions, absence of medians and junctions, significantly increased the likelihood on occurrence of HR crashes. Building on the cost benefits approach used in previous study (Tay et al., 2008), Jiang et al. (2015) proposed the concept of subjective-responsibility-ratio (SRR) to explore different factors that motivate an offending driver to flee the scene when a crash occurs inside river-crossing tunnels of an urban area. This study tried to employ mixed logit model but finally opted for binary logit model as mixed model proved not to be cost-effective due to many failed attempts at convergence.

The aforementioned studies investigated different characteristics which influence HR crashes using typical invariable data, but not the real-time traffic data which has been widely explored in studies focused on prediction of crashes on freeways (Oh et al., 2001; Abdel-Aty et al., 2005; Golob et al. 2008; Zheng et al., 2010; Pande et al., 2011; Ahmed et al., 2012; Li et al., 2012). Lee et al. (2002) used the data from 38 loop detector stations to develop an aggregate log-linear model to link crash potential with various traffic flow characteristics observed upstream (crash precursors). Changes in speed and traffic density were found to be statistically significant predictors of crash frequency. Subsequently, Lee et al., (2003) addressed the limitations of previous study by suggesting rational methods for determination of crash precursors. It was concluded from the results that occurrence of a crash is accompanied by a significantly higher difference in the observed speed between the upstream and downstream detectors. Also, an estimate of the actual time of crash was found to be correlated with an abrupt drop in speed at upstream detector. Abdel-Aty et al. (2004) used matched case-control logistic regression approach where every crash and corresponding noncrash was considered to be case and control, respectively. It was observed that crash occurrence was significantly affected by the average occupancy observed at the upstream station 5 to 10 min before the crash, and coefficient of variation in speed at the downstream loop detector station. Utilizing the loop detector data, the study by Abdel-Aty and Abdalla (2004) found an increase in likelihood of a crash with high variability in speed upstream, presence of on-ramp, and low variability in volume. Pande et al. (2005) observed that traffic parameters were more significantly associated with crash likelihood when considered for a 5-min interval rather than 3-min. In terms of data collection, Hourdes et al. (2006) used a different approach by capturing real-time traffic data with video cameras. Large speed differences between adjacent lanes and abrupt changes in traffic flow due to compression waves were identified as the factors having significant impact on crash likelihood. Xu et al. (2013) expanded the previous research by developing a sequential logit model to predict crash likelihood for different severity levels. The results showed that different traffic characteristics influenced different severity levels: frequent lane changes and high variability in speed affected PDO, while high speed and large speed difference between adjacent lanes impacted the fatal and severe injury crashes. Similar study focusing on severity levels was conducted by Yu and Abdel-Aty (2014) using real-time traffic and weather data to develop hierarchical Bayesian binary probit models with random effects. Modeling results corroborated the previous studies by demonstrating high variability of speed prior to the crash occurrence to be associated with increase in the likelihood of severe crash, along with a substantial improvement in goodness-of-fit of model due to inclusion of unobserved heterogeneity. Similar correlation between speed and crash occurrence was observed in a recent study by Wang et al. (2015) which used real-time microwave vehicle detection system data, real-time weather data, and ramp geometric information for crash analysis of expressway ramps.

Given the benefits of real-time data, the primary purpose of this study is to contribute to current literature of HR crashes by incorporating the real-time traffic data from loop detectors (such as vehicle count, speed, occupancy, and so on) to predict the likelihood of general and HR crashes. Roadway geometric and other factors (weather, time, day of week, driver behavior) are also explored which have a probable influence on driver's decision to run from crash scene. Moreover, instead of the typical binary logit model used in previous HR studies, this research employs the random effect Bayesian binary logit model to account for the unobserved heterogeneity which is expected to provide better model fit. Further, in addition to the evaluation of impact of random effects on model fitness and complexity, k-fold cross validation approach is also used to examine the predictive capability of the proposed model. Stepwise incremental sensitivity and specificity are calculated for each training set. Finally, ROC (Receiver Operating Characteristic) curve is utilized to graphically illustrate the predictive performance of the model.

METHODOLOGY

This study investigated the influence of different factors (with emphasis on real-time traffic flow variables) on the likelihood of General crash occurrence and HR crashes on a 44-mile freeway segment in Southern California. Similar to the past research HR studies, this study also utilized the binary logistic regression model, as the response variable in our study was dichotomous (for General: a crash could occur or not; and for HR: offending driver could run or stay). Since the authors were also interested in General crashes, unlike the previous studies focused only on HR, hence a sequential model structure was employed with Crash and NonCrash at first level and HR and NHR at second level. This approach allowed us to estimate the evaluation of HR and General crash likelihood separately. A similar model structure can be found at one recent study focusing on crash severity analysis (Xu et al., 2013). Overall, there are two levels of modeling process shown as follows:

- Level 1: Crash occurs (binary response = 1) vs. NonCrash (binary response = 0).
- Level 2: HR (binary response = 1) vs. NHR (binary response = 0).

In a typical Binary Logit model, the probability P_i of the occurrence of a General or HR crash was estimated by the following equation:

$$Y = \text{logit}\ (P) = \beta X \tag{1}$$

Where β is the vector of model parameters to be estimated, X is a vector of independent variables. However, the advantage of Binary regression with random effects has been clearly illustrated in previous studies (e.g., Yu and Abdel-Aty, 2014). Therefore, random effects were also introduced in our hierarchical Bayesian Binary Logit using the error term for both levels to find the probability of occurrence of a crash and determine the statistically significant factors which influenced the outcome (dependent variable). The inclusion of random effects gave more flexibility to capture the different variations within the observed data. Eq. 1 was modified to take the following form:

$$Y = \text{logit}\ (P) = \beta X + \varepsilon \tag{2}$$

Where ε is the error term modeled as independent random effects to capture unobserved heterogeneity. The random effects were specified at the individual crash level by assuming ε to have the noninformative normal priors, in a effort to explore the inferences provided by the data (Wang et el., 2015):

$$\varepsilon \sim Normal\ (0, \sigma^2\ ) \tag{3}$$

Where σ^2 is the variance of the normal distribution for ε. The inverse of σ^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, representing high uncertainty or prior ignorance:

$$\sigma^2 \sim gamma(0.5, 0.0005) \tag{4}$$

The likelihood of crash occurrence was calculated at the two levels of the sequential model by using the estimates from the binary logit model at each level.

$$P\ (\text{Crash}) = P_{f1} \tag{5}$$

$$P\ (\text{HR}) = P\ (\text{Crash}) * P\ (\text{HR/Crash}) = P_{f1}\ (1\text{-}P_{f2}) \tag{6}$$

Where the *P (a)* represents the probability of occurrence of the concerned outcome *a* (Crash and HR); *P (a/b)* is the probability of occurrence of *a,* assuming that *b* has already occurred; and *Pfi* represents the probability estimated by the binary logit model, where *i* is the level of the sequential structure

The models were estimated at two levels with the Bayesian techniques using open source software WinBUGS. Two chains of 15,000 iterations were used to obtain the summary statistics of the posterior parameters, 5,000 were discarded as adaptation and burn-in. Model convergence was monitored by visual inspection of the MCMC trace plots for model parameters. Moreover, the convergence of multiple chains was assessed by ensuring the value of Brooks–Gelman–Rubin (BGR) statistic to be less than 1.2. The significant variables in both models were selected based on the 95% confidence interval.

For assessment of model complexity and fit, this study used the Deviance Information Criterion (DIC) developed by Spiegelhalter et al. (2003). DIC is a hierarchical modeling generalization of the AIC (Akaike information criterion). It is computed as the sum of the posterior mean deviance and estimated effective number of parameters:

$$DIC = D(\bar{\theta}) + 2p_D = \overline{D(\theta)} + p_D \tag{7}$$

Where $D(\bar{\theta})$ is the deviance evaluated at the posterior means of estimated unknowns ($\bar{\theta}$), posterior mean deviance $\overline{D(\theta)}$ may be regarded as a Bayesian measure of fit, and p_D denotes the complexity measure for the effective number of parameters in a model. Based on the model-selection decision criteria suggested by Best et al. (2005): the models with DIC values within 1 or 2 of the 'best' model are also strongly supported, values within 3 and 7, weakly supported, and models with a DIC greater than 7 have substantially inferior modeling performance.

To validate the prediction accuracy of the models, this study used the k-fold cross-validation method for both crash levels. This procedure was implemented with an aim to minimize the bias associated with the random sampling of the dataset used for model development and another dataset for validation (Olson and Delen, 2008). In this study, to ensure the enough data sample size for both training and validation groups, k was chosen as 5. In other words, the dataset was randomly divided into five mutually exclusive groups of approximately equal size. Among the five datasets, every group was used as the validation dataset, and the other four groups were combined to form a training dataset which was used for building the prediction models. Hence, five different combinations were used to train and test the prediction models.

The model evaluation was conducted using Sensitivity and Specificity, which served as diagnostic criteria for supplementary empirical evaluation of agreement between two datasets.

$$Sensitivity = \frac{CP\ (correct\ positives)}{TP\ (true\ positives)} \tag{8}$$

$$Specificity = \frac{CN\ (correct\ negatives)}{TN\ (true\ negatives)} \tag{9}$$

Where, CP: number of true Crashes (or HR) correctly predicted by model; TP: total number of observed true Crashes (or HR); CN: number of true NonCrashes (or NHR) correctly estimated by the model; and TN: total number of observed NonCrashes (or NHR)

Usually certain predefined thresholds are used for establishing the correct prediction. But a more subtle approach was implemented by employing a continuous set of thresholds (with increment of 5%) which performed the sensitivity/specificity analysis in steps. Further, ROC was used to graphically illustrate the agreement between the estimated likelihood of crashes by the model and observed crashes at different thresholds, starting with the minimum cut-off of 0% and progressing towards the maximum limit of consideration (100%).

DATA DESCRIPTION

As this study primarily explored the influence of real-time traffic conditions on the likelihood of General and HR crashes, the concerned traffic data were obtained from the Highway Performance Measurement System (PeMS) maintained by the California Department of Transportation (Caltrans). Same data source was used to obtain the certain roadway geometric factors which were expected to impact the crashes. The crash data were obtained from the Statewide Integrated Traffic Records System (SWITRS). The impact of weather was also incorporated. In case of weather conditions corresponding to crash observations, the data were readily available in the crash database. But due to the unavailability of such data for normal conditions (crash-free), the data were collected from National Climate Data Center (NCDC). A total of 7 weather stations were located within a radius of 5 miles along the I-5 freeway.

The area under focus for this study was 44-mile segment on the mainline part of I-5 freeway in the Orange County of California and the 200 loop detector stations in the northbound and southbound directions provided the real-time traffic data. Based on the crash time and milepost, suitable traffic data were extracted from the 30-s raw detector file. From time's perspective, the data within the time interval of 5 and 10 minutes prior to crash occurrence were extracted, as suggested by previous studies (Abdel-Aty et al. 2004). From the space's perspective, the corresponding station IDs for consecutive upstream and downstream detectors for every crash were first obtained, and then the data covered by the station IDs were extracted. Also, the vehicle count, occupancy, and average speed for each lane in 30-s aggregation intervals were collected.

For analysis of crash factors, “Hit & Run” (HR) was chosen as the base group, and “Non Hit & Run” (NHR) as the matched one. A total of 328 crashes, 82 for HR crash, and 246 for NHR one (such as run off the roadway, motorcyclist-involved, truck-involved, etc.), were identified and used in the study. 1200 non-crash (NC) data groups were considered for comparison. With due consideration to the previous studies where the data sample sizes range from 52 (Oh et al., 2001) to 1528 (Abdel-Aty and Pande, 2005) for crash, and from 201 (Hassan and Abdel-Aty, 2011) to 23,068 (Hossain and Muromachi, 2010) for non-crash, the sample size for this study was normal. Due to the small probability for HR crash occurrence, the data collected span over 2 years, from 2010~2011.

The different datasets (traffic, crash, and geometric) were aggregated by the integration of crash date, crash location, crash time, and milepost. Figure 1 illustrates the flow chart of extraction and filtering of raw data for obtaining the working dataset.

Figure 1 about here

Moreover, given the possible wrong traffic data might be obtained due to malfunctioning of loop detectors, the data were double checked and invalid or unusable data were excluded if they satisfied infeasible conditions such as the average occupancy was greater than 100%, the vehicle count was greater than 0 veh/30 s while the occupancy was 0%, the occupancy was greater than 0% while the vehicle count was equal to 0 veh/30 s, and so on. Finally, the 30-s raw detector readings from two consecutive upstream-downstream detector stations were aggregated into 5-min intervals and converted into the 20 traffic flow variables presented in Table 1.

Table 1 about here

RESULTS

One major concern about the large number of covariates is the potential risk of modeling overfitting. To obtain a parsimonious model and ensure the most appropriate subset of variables were entered for model development, various techniques were employed including multicollinearity analysis and backward stepwise selection. Correlation evaluation was also conducted to ensure no strong correlations for final model. The statistically significant independent variables at a significance level of 0.05 for the sequential models are shown in Table 2.

Table 2 about here

Model Fitness

Crash vs. non-crash Six variables were found to be significantly correlated with crash likelihood, four of which were traffic flow variables: average count of vehicles upstream, standard deviation of detector occupancies upstream, average speed downstream, and standard deviation of difference in upstream and downstream speeds; and two were roadway geometry variables: median width and distance between upstream and downstream stations. The results clearly demonstrate that except average speed of downstream vehicles and median width, increase in the quantity of all variables tend to increase the likelihood of crash occurrence. The rise in number of vehicles upstream means that more vehicles are joining the queue. Since the number of vehicles is increasing only for upstream, that means the road ahead would be possibly clearer and only a few vehicles (or vehicle) are slowing down the traffic. In such a scenario, the drivers of the vehicles stuck in queue may tend to get impatient and try changing the lanes. This behavior may be the reason behind increase in likelihood of crashes. The deviation in the occupancy of upstream detectors has the largest coefficient value, which indicates that this factor has a very high influence in crash occurrence. Greater difference between the occupancy times of detectors indicates that the traffic flow is erratic, rather than being a normal streamlined flow. Such flow largely increases the application of sudden brakes and since vehicle speed on freeway tends to be higher, so the drivers may not have much time to maneuver the vehicle while applying brakes. As the occupancy could be considered to reflect the traffic density (Xu et al., 2013), another interpretation of results could be that significant variations in the density of vehicles in a segment elevates the chances of crash downstream. Also, the variations in speed were noted to positively increase the likelihood of crashes as depicted by the two factors: speed of downstream vehicles and deviation of speed for upstream and downstream. As the speed on freeways is relatively higher, even small variations tend to heighten the risk of crashes. These findings are in line with previous research (Yu and Abdel-Aty, 2014). Among the geometric features, the distance between upstream and downstream detectors (which signifies segment length) and the width of median were observed to be influential factors. As noted by a previous research (Anastasopoulos and Mannering, 2009; Xu et al., 2013), comparatively longer road segments increase the likelihood of crashes as the vehicles get a longer stretch of road for interactions at higher speeds, which raises the risk of crash occurrence. Median width had a negative coefficient, which means that wider medians tend to alleviate the crash risk as the separation between the opposing vehicles significantly increases which aids in mitigating the head-on crashes and also the visibility hindrance caused due to headlight glare of opposing vehicles at night.

HR vs. non-HR The upstream vehicle speed was found to be positively correlated with driver's decision of running away from crash scene. Higher speed of vehicles is usually related to lesser traffic volume. In such scenarios, the offending driver perceives a lower risk of identification as there are better chances of escaping due to the possible absence of witnesses of the crash (Tay et al., 2009). Similar reasons could be attributed for the positive correlation between segment length and HR crashes as longer segments tend to have higher speeds without any mixing of traffic. This study also found that there was a greater likelihood of HR during the weekends. This finding is in line with some of the previous literature (Solnick and Hemenway, 1995; Tay et al., 2009; Christoforou et al., 2010; Quddus et al., 2010, El-Basyouny et al., 2014), where day-of-the week was statistically significantly related with various crash types. This may be attributed to the fact that during weekends, relatively more drivers tend to be under the influence of alcohol (Solnick and Hemenway, 1994; Kim et al., 2008; Macleod et al., 2012) and likelihood of severe crashes increases as the speed of vehicles is higher due to lesser traffic volume. Both these factors significantly increase the reporting cost (Tay at al., 2009) and hence the driver prefer to flee the scene instead of facing serious legal consequences. Similar to the observations of previous studies (Kim et al., 2008; Tay et al., 2008), the weather was found to be statistically insignificant.

Cross Validation

As previously stated, k-fold cross-validation approach was employed to assess the predictive capability of the developed models. To ensure enough sample size for training and validation sets, the data were divided into five groups for both General and HR datasets. The dependent variable for the sequential models was the probability of occurrence of Crash (or HR). The crash estimates from the models were compared with the observed outcomes (event = 1 and non-event = 0) to find the agreement between them. The criteria of sensitivity and specificity using Eqs. 8 and 9 were employed to measure the effectiveness of the model to estimate the probability of a crash. Overall, 20 thresholds with 5% increments of probability were used to calculate the associated sensitivity and specificity. Based on the 20 pairs of sensitivity and specificity values, ROC (Receiver Operating Characteristic) curves were prepared to show the overall predictive performance of both models, where the y-axis is sensitivity and x-axis is 1-specificity.

As exhibited in Figures 2 and 3, the areas under the ROC curves for both sequential logit models were found to be 0.823, and 0.765, respectively, which were considerably better than a random ROC which has an area of 0.5. In comparison with previous studies (Ahmed and Abdel-Aty, 2012, Xu et al., 2013) the predication accuracy of present models can be considered as satisfactory. The trade-off between the sensitivity and false alarm rate as displayed in ROC can be used to select the optimal threshold value in practice. Usually the "knees" of the curve (which represent the sharp turning points) are considered as a good choice, which lead to relatively larger sum of sensitivity and specificity. For example, a potential knee in Figure 2 indicates that stativity of 70% is roughly accompanied with the specificity of 90% (1-0.1). Likewise, one possible knee in Figure 3 shows the pair of (sensitivity=0.82, specificity=1-0.41). The thresholds correspond to the knees can yield somewhat larger values of both sensitivity and specificity. However, it is important to note that the threshold values selected must be subjected to the requirement of the practical implementation or the specific traffic agency policies. For instance, some agency might prefer having a larger sensitivity value, even though the associated specificity is far lower than the ones resulting from other thresholds.

Figure 2 and Figure 3 about here

CONCLUSIONS AND RECOMMENDATIONS

The primary objective of this study was to investigate the influence of real-time traffic factors on the occurrence of General and HR crashes. A sequential model structure was employed with Crash and NonCrash at first level and HR and NHR at second level. Bayesian random effects binary logit model was developed for both levels to find the probability of occurrence of a crash and determine the statistically significant factors which influenced the outcome (the probability of occurrence of a General crash/HR). The k-fold cross validation technique was used for estimation of prediction accuracy of models by randomly dividing the dataset into five groups of similar size. For evaluation of the performance of models, two complementary indicators were used: sensitivity and specificity. ROC curve was generated to depict the prediction performance of models at different thresholds.

In case of General crashes, most of the results of this study were consistent with the previous studies. Many traffic related factors were positively correlated with the tendency of occurrence of crashes such as vehicle count, density, and speed variations. The length of the segment and width of median were two roadway geometry factors which proved to be influential with a 95% confidence level. Since this study also explored the influence real-time traffic conditions on the likelihood of HR crashes, higher speed of upstream vehicles was observed to elevate the possibility of an erring driver to run away from a crash scene. The probable reason could be attributed to the theory that higher vehicle speeds usually indicate presence of relatively smaller number of vehicles, which may be perceived as a lower risk situation by the driver as it is easier to escape detection in absence of witnesses. Also, weekend had a considerable impact with more drivers committing HR during weekends due to possible indulgence in alcohol and speeding. The ROC indicated an area of 0.765 for HR model, which was better than the area of 0.725 observed by a recent study (Jiang et al., 2015). This superior fitness may be accredited to the inclusion of random effects into the binary logit model which addressed the unobserved heterogeneity.

This study employed the particular framework on the data from urban freeway. It is recommended to perform similar analysis for different environments (like mountainous or rural regions) to ensure that the empirical results hold true. Different procedures may be adopted for collection of real-time data (such as fixed video cameras or unmanned aerial vehicles) to compare the accuracy of performance. Moreover, the sample size for this study was of normal range. It is expected that a larger sample size would determine more traffic factors to have a statistically significant influence on the occurrence of crashes, or HR crashes in particular. This research indicated that the real-time traffic data seems to be a promising area of interest for understanding the conditions of HR and the traffic safety research could benefit with more investigation in this area.

ACKNOWLEDGMENTS

Financial support for this study was provided by 2016 UCCONNECT Summer Research Program.

REFERENCES

1. Abdel-Aty, M., Uddin, N., Abdalla, F., Pande, A., Hsia, L. Predicting freeway crashes based on loop detector data using matched case-control logistic regression. Trans Res Rec. 2004:1897, 88–95.
2. Abdel-Aty, M., and M. F. Abdalla. Linking Roadway Geometrics and Real-Time Traffic Characteristics to Model Daytime Freeway Crashes: Generalized Estimating Equations for Correlated Data. In Transportation Research Record: Journal of the Transportation Research Board, No. 1897, Transportation Research Board of the National Academies, 2004:Washington, D.C., pp. 106–115.
3. Abdel-aty M, Pande A. Identifying crash propensity using specific traffic speed conditions. J Safety Res. 2005;36(1):97-108.
4. Ahmed, M., Abdel-Aty, M., Yu, R. A Bayesian updating approach for real-time safety evaluation using AVI data. In: Presented at the 91th Annual Meeting of the Transportation Research Board, CD-ROM, 2012: Washington, D.C.
5. Ahmed, M., Abdel-Aty, M. The viability of using automatic vehicle identification data for real-time crash prediction. IEEE Trans on Intel Trans Sys. 2012:13 (2), 459–468.
6. Aidoo EN, Amoh-gyimah R, Ackaah W. The effect of road and environmental characteristics on pedestrian hit-and-run accidents in Ghana. Accid Anal Prev. 2013;53:23-7.
7. Anastasopoulos PCh, Mannering FL. A note on modeling vehicle accident frequencies with random-parameters count models. Accid Anal Prev. 2009;41(1):153-9.
8. Baucom B. Use of clinical forensic skills for a 9-year-old pedestrian who was struck by a motor vehicle in a hit and run crash. J Emerg Nurs. 2006;32(4):319-21.
9. Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. Stat Methods Med Res. 2005;14(1):35-59.

10. Blincoe, L., Seay, A., Zaloshnja, E., Miller, T., Romano, E., Luchter, S., & Spicer, R. The economic impact of motor vehicle crashes, 2000 (No. HS-809 446,). 2002:Washington, DC, National Highway Traffic Safety Administration.
11. Christoforou Z, Cohen S, Karlaftis MG. Identifying crash type propensity using real-time traffic data on freeways. J Safety Res. 2011;42(1):43-50.
12. El-basyouny K, Barua S, Islam MT. Investigation of time and weather effects on crash types using full Bayesian multivariate Poisson lognormal models. Accid Anal Prev. 2014;73:91-9.
13. Golob, T., Recker, W., Pavlis, Y. Probabilistic models of freeway safety performance using traffic flow data as predictors. Safety Science 2008:46 (9), 1306–1333.
14. Hassan, H., Abdel-Aty, M. Exploring visibility-related crashes on freeways based on real-time traffic flow data. In: Presented at 90th Annual Meeting of the Transportation Research Board, CD-ROM, 2011:Washington, D.C.
15. Hossain, M., Muromachi, Y. Evaluating location of placement and spacing of detectors for real-time crash prediction on urban expressways. In: Presented at 89th Annual Meeting of the Transportation Research Board, CD-ROM, 2010:Washington, D.C.
16. Hourdos, N., Garg, V., Michalopoulos, G., Davis, G. Real-time detection of crash-prone conditions at freeway high-crash locations. Trans Res Rec 2006:1968, 83–91.
17. Jiang, C.M., Lu, J.J., Lu, L.J. Analysis of single-vehicle crash injury severities in urban river-crossing road tunnels. Appl. Mech. Mater. 2015:743, 526–532.
18. Karger B, Teige K, Fuchs M, Brinkmann B. Was the pedestrian hit in an erect position before being run over?. Forensic Sci Int. 2001;119(2):217-20
19. Kim, K., Pant, P., Yamashita, E. Hit and run crashes: using rough set analysis with logistic regression to capture critical attributes and determinants. In: Transportation Research Board Annual Meeting, 2008: Washington, DC.
20. Lee, C., F. Saccomanno, and B. Hellinga. Analysis of Crash Precursors on Instrumented Freeways. In Transportation Research Record: Journal of the Transportation Research Board, No. 1784, TRB, National Research Council, 2002: Washington, D.C. pp. 1–8.
21. Lee, C., Saccomanno, F., Hellinga, B. Real-time crash prediction model for the application to crash prevention in freeway traffic. Trans Res Rec 2003:1840, 67–77.
22. Li, Z., Chung, K., Liu, P., Wang, W., Ragland, D. Surrogate safety measure for evaluating rear-end collision risk near recurrent bottlenecks. In: Presented at the 91th Annual Meeting of the Transportation Research Board, CD-ROM, 2012:Washington, DC.
23. Locke, J., Wilkinson, J., Hanford, T. A data collection of vehicle topcoat colors. 2. The measurement of color samples used in the vehicle refinishing industry. Forensic Sci. Int. 1988:37, 177–187.
24. Mock CN, Nii-amon-kotei D, Maier RV. Low utilization of formal medical services by injured persons in a developing nation: health service data underestimate the importance of trauma. J Trauma. 1997;42(3):504-11.
25. Macleod KE, Griswold JB, Arnold LS, Ragland DR. Factors associated with hit-and-run pedestrian fatalities and driver identification. Accid Anal Prev. 2012;45:366-72.

26. Oh, C., Oh, J., Ritchie, S. Real-time estimation of freeway accident likelihood. In: Presented at 80th Annual Meeting of the Transportation Research Board, CD-ROM, 2001:Washington, D.C.
27. Olson, D., Delen, D. Advanced Data Mining Techniques. 2008:Springer, Berlin, Germany.
28. Pande A, Abdel-aty M. Assessment of freeway traffic parameters leading to lane-change related collisions. Accid Anal Prev. 2006;38(5):936-48.
29. Pande, A., Dasand, A., Abdel-Aty, M., Hassan, H. Real-time crash risk estimation are all freeways created equal? Trans Res Rec 2011:2237, 60–66.
30. Peden, M. (2004). World report on road traffic injury prevention.
31. Quddus, M., Wang, C., Ison, S. Road traffic congestion and crash severity: econometric analysis using ordered response models. J Trans Eng 2010:136, 424–435.
32. Roess, R.P., Prassas, E.S., Mcshane, W.R. Traffic Engineering, Third ed. 2004:Pearson Education International.
33. Solnick SJ, Hemenway D. Hit the bottle and run: the role of alcohol in hit-and-run pedestrian fatalities. J Stud Alcohol. 1994;55(6):679-84.
34. Solnick SJ, Hemenway D. The hit-and-run in fatal pedestrian accidents: victims, circumstances and drivers. Accid Anal Prev. 1995;27(5):643-9.
35. Spiegelhalter, D., Thomas, A., Best, N., Lunn, D. Winbugs User Manual. MRC Biostatistics Unit, Cambridge 2003.
36. Taylor, M., Cousins, D., Holding, R., Locke, J., Wilkinson, J. A data collection of vehicle topcoat colours. 3. Practical considerations for using a National Database. Forensic Sci. Int. 1989:40, 131–141.
37. Tay R, Barua U, Kattan L. Factors contributing to hit-and-run in fatal crashes. Accid Anal Prev. 2009;41(2):227-33.
38. Tay R, Rifaat SM, Chin HC. A logistic model of the effects of roadway, environmental, vehicle, crash and driver characteristics on hit-and-run crashes. Accid Anal Prev. 2008;40(4):1330-6.
39. Teresinski, G., Madro, R. Knee joint injuries as a constructive factor in car-to pedestrian accidents. Forensic Sci. Int. 2001:124, 74–82.
40. Wang, L., Shi, Q., & Abdel-Aty, M. Predicting Crashes on Expressway Ramps with Real-Time Traffic and Weather Data. Trans Res Rec. 2015: (2514), 32-38.
41. Xu C, Tarko AP, Wang W, Liu P. Predicting crash likelihood and severity on freeways with real-time loop detector data. Accid Anal Prev. 2013;57:30-9.
42. Yu R, Abdel-aty M. Using hierarchical Bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data. Accid Anal Prev. 2014;62:161-7.
43. Zheng Z, Ahn S, Monsere CM. Impact of traffic oscillations on freeway crash occurrences. Accid Anal Prev. 2010;42(2):626-36.

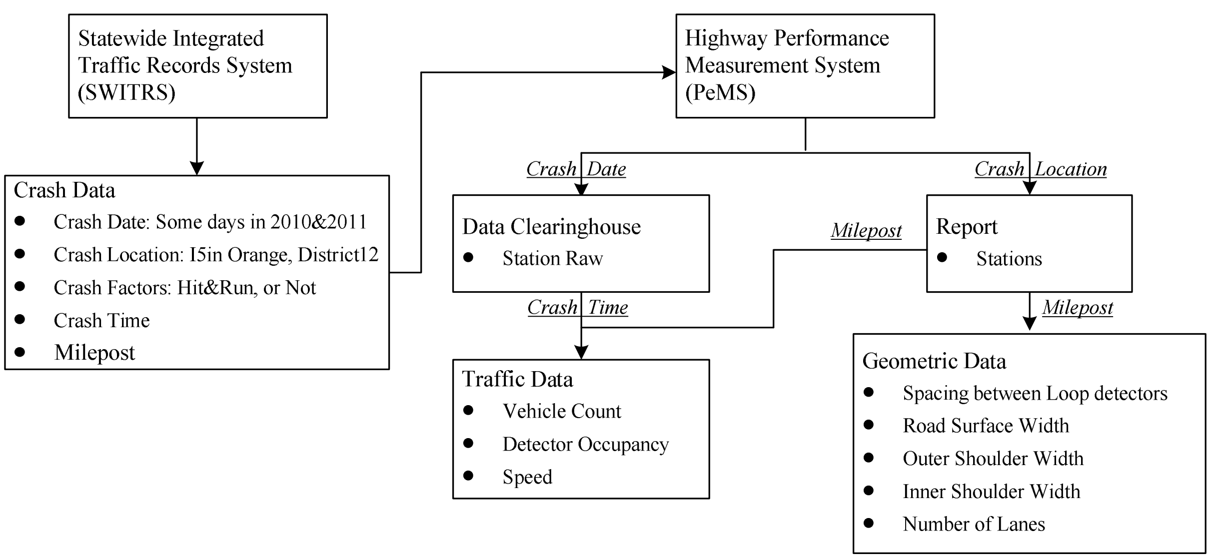


FIGURE 1 The flow chart of data integration from various sources.

TABLE 1 Variables Considered for the Models

Factor	Symbol	Variables
Traffic data	$VehCnt_u$	Average 30-s vehicle count at the upstream station (veh/30 s)
	$DetOcc_u$	Average 30-s detector occupancy at the upstream station (%)
	$AvgSpd_u$	Average 30-s speed at the upstream station (mile/h)
	$CntDev_u$	Std. dev. of 30-s vehicle count at the upstream station (%)
	$OccDev_u$	Std. dev. of 30-s detector occupancies at the upstream station (%)
	$SpdDev_u$	Std. dev. of 30-s mean speeds at the upstream station (mile/h)
	$CvSpd_u$	Coefficient of variation of 30-s mean speeds at the upstream station (mile/h)
	$VehCnt_d$	Average 30-s vehicle counts at the downstream station (veh/30 s)
	$DetOcc_d$	Average 30-s detector occupancy at the downstream station (%)
	$AvgSpd_d$	Average 30-s speed at the downstream station (mile/h)
	$CntDev_d$	Std. dev. of 30-s vehicle count at the Downstream station (%)
	$OccDev_d$	Std. dev. of 30-s detector occupancies at the downstream station (%)
	$SpdDev_d$	Std. dev. of 30-s mean speeds at the downstream station (mile/h)
	$CvSpd_d$	Coefficient of variation of 30-s mean speeds at the downstream station (mile/h)
	$AvgCnt_{u\text{-}d}$	Average absolute difference in vehicle counts between upstream and downstream

	$AvgOcc_{u-d}$	Average absolute difference in detector occupancies between upstream and downstream
	$AvgSpd_{u-d}$	Average absolute difference in speeds between upstream and downstream stations
	$DevCnt_{u-d}$	Std. dev. of absolute difference in vehicle counts between upstream and downstream
	$DevOcc_{u-d}$	Std. dev. of absolute difference in detector occupancies between upstream and
	$DevSpd_{u-d}$	Std. dev. of absolute difference in speeds between upstream and downstream stations
Geometry data	$DetDist_{u-d}$	Distance between upstream and downstream stations (mile)
	$Width_s$	Road surface width (ft)
	$Width_o$	1 = if outer shoulder width > 10 ft; 0 = otherwise
	$Width_i$	1 = if inner shoulder width > 10 ft; 0 = otherwise
	$Width_m$	Inner median width (ft)
	Lanes	Number of lanes
Others	Weather	1 = adverse weather conditions (rain or fog); 0 = otherwise
	Peak	1 = peak period (7-9 a.m.;4-7 p.m.); 0 = otherwise
	AlcoInv	1= Alcohol Involved; 0=otherwise
	Monday	1= Yes; 0=otherwise
	Tuesday	1= Yes; 0=otherwise
	Wednesday	1= Yes; 0=otherwise
	Thursday	1= Yes; 0=otherwise
	Friday	1= Yes; 0=otherwise
	Weekend	1= Yes; 0=otherwise
	Night	1= night period (8 p.m.-5 a.m.);0=otherwise
	Hit&Run	1=Hit &Run; 0=otherwise

TABLE 2 Model Fitness and Estimated Parameters

Goodness-of-fit of models

Model	**Dbar**	**Dhat**	**pD**	**DIC**
Crash vs. Non-Crash	1410.71	1403.59	7.111	1417.82
HR vs. Non-HR	354.400	350.341	4.059	358.458

Estimated model parameters

Model	**Parameters**	**mean**	**Sd**	**MC error**	**2.50%**	**median**	**97.50%**
Crash vs. Non-Crash	Intercept	-1.797	0.4952	0.02458	-2.79	-1.787	-0.848
	Avg Count-U	0.087	0.0255	9.581E-4	0.03709	0.0868	0.1363
	OcDev-U	15.0	4.917	0.1527	5.402	15.0	24.76
	Avg. Speed-D	-0.0234	0.0051	2.128E-4	-0.0335	-0.0234	-0.0134

	Dev-Speed U-D	0.1116	0.0214	4.357E-4	0.07	0.1115	0.154
	Det Dist U-D	1.035	0.281	0.005584	0.4827	1.036	1.581
	Widthm	-0.0276	0.0073	1.544E-4	-0.0424	-0.0274	-0.0137
HR vs. Non-HR	Intercept	-2.825	0.5441	0.01757	-3.899	-2.822	-1.782
	Avg Speed up.	0.0182	0.0084	2.583E-4	0.0021	0.0181	0.0348
	Det dist. U-D	1.064	0.5278	0.01085	0.004	1.059	2.102
	Weekend	0.6242	0.3037	0.00337	0.0289	0.622	1.215

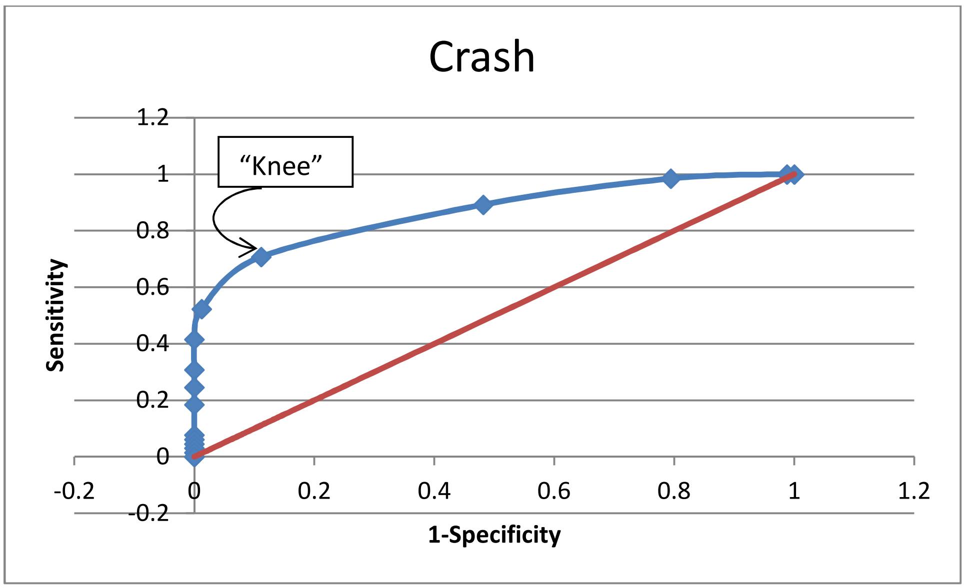


FIGURE 2 Receiver Operating Characteristic for Crash vs. Non-Crash Model.

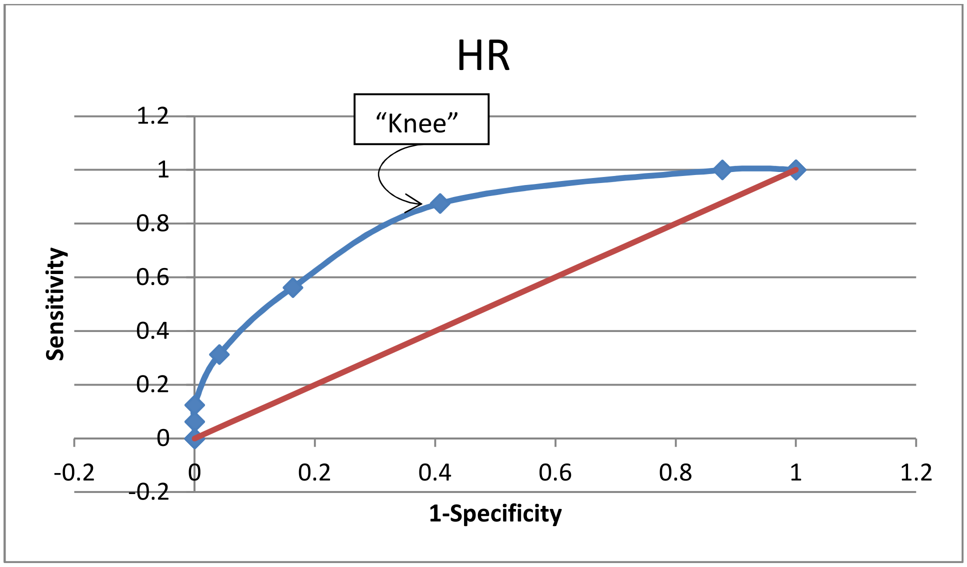


FIGURE 3 Receiver Operating Characteristic for HR vs. Non-HR Model.

Paper #6: A New Approach to Addressing Temporal Correlation in Crash Frequency Modeling: Combination of Time-varying Coefficients and Autoregressive Process

Wen Cheng[a,*], Gurdiljot Singh Gill[b], Simon Choi[c], Jiao Zhou[d], Xudong Jia[e], Meiquan Xie[f]

* Corresponding author, [a] Associate Professor, Department of Civil Engineering
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA, USA 91768
Fax: (909) 869-4342; Tel: (909) 869-2957; Email: wcheng@cpp.edu

[b] Graduate Student, Department of Civil Engineering
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA, USA 91768; Email: gurdiljotg@cpp.edu

[c] Chief of Research and Forecasting
SOUTHERN CALIFORNIA ASSOCIATION OF GOVERNMENTS
818 West 7th Street, 12th Floor, Los Angeles, CA, USA 90017
Fax: (213) 236-1963; Tel: (213) 236-1849; Email: choi@scag.ca.gov

[d] Graduate Student, Department of Civil Engineering
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA, USA 91768
Fax: (909) 869-4342; Tel: (909) 869-4312; Email: jiaozhou@cpp.edu

[e] Professor, Department of Civil Engineering
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA, USA 91768
Fax: (909) 869-4342; Tel: (909) 869-4312; Email: xjia@cpp.edu

[f] Visiting Scholar, Department of Civil Engineering,
California State Polytechnic University, Pomona
3801 W. Temple Ave., Pomona, CA, USA 91768
& Assistant Professor, School of Transportation and Logistics,
Central South University of Forestry and Technology,
Tianxin, Changsha, Hunan, China, 410000
Fax: (909) 869-4342; Tel: (909)282-9142; Email: mxie@cpp.edu

ABSTRACT

Compared with a large amount of research using various different ways of addressing serial correlations among crash data, there is relatively little research dedicated to the evaluation of the different temporal treatments on modeling performance. To add to the literature the much-needed research, this study proposed a new method which combines the strengths of time-varying coefficients and autoregressive process, and compared its performance with seven other temporal models used in the past. Ten years of crash data and other covariates associated with traffic analysis zones in the City of Irvine, California were used. Bayesian hierarchical approach was employed to account for the structural heterogeneities.

The comparisons were conducted for assessment of goodness-of-fit, the accuracy of crash estimation, and relative performance of site ranking. The modeling results indicated that the proposed model appeared to have the best fit with actual crash data and a relatively lower complexity than other competing models. Longitudinal and cross-sectional validations using RSS (Residual Sum of Squares) demonstrated that the proposed model had very significant superiority at crash prediction with an RSS score three times smaller than the worst performing model. The site ranking evaluation established that the models with similar DIC and RSS scores tend to have more similar ranking performance.

Keywords: Serial Correlations; Time-Varying Coefficients; Autoregressive Process; Bayesian Hierarchical Approach; Structural Heterogeneities

1. INTRODUCTION

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. Some studies developed univariate models by focusing on a particular crash outcome or total crashes (Ulfarsson and Shankar, 2003; Lord et al., 2005; Kim et al., 2006). These models operated on the assumption that crash outcomes are independent, while they were revealed to be multivariate due to sharing of unaccounted factors (Bijleveld, 2005; Park and Lord. 2007). Congdon (2001) observed that ignorance of such correlation structures may reduce the efficiency of the model due to lesser precise parameters.

Lately, more advanced models have been proposed to incorporate the correlation structures to account for the unobserved heterogeneity in crash data (MacNab, 2004; Miaou and Lord, 2003; Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Lord and Mannering, 2010). Significant correlations were observed by some research studies which jointly considered different crash severity levels (Tunaru, 2002; Ladron de Guevara et al., 2004; Miaou and Song, 2005; Song et al., 2006; Ma and Kockelman, 2006). Some research studies (Ma et al., 2008; Aguero-Valverde and Jovanis, 2009) noted that more precise estimates were obtained for model parameters with the inclusion of correlations in multivariate crash counts. Apart from the correlations of severity levels, some studies utilized spatial random effects to explore the spatial correlations between the crash sites at different area levels like intersections (Wang and Abdel-Aty, 2006; Mitra, 2009), segments (Aguero-Valverde and Jovanis, 2008), corridors (Abdel-Aty et al., 2006; Guo et al., 2009), Census tracts (Narayanamoorthy et al., 2013), Traffic Analysis Zones (Washington et al., 2010; Xu and Huang, 2015), counties (Miaou et al., 2003; Huang et al., 2010), and so on. The studies focused on simultaneously modeling crash types (Song et al., 2006; Jonathan et al., 2016) observed that inclusion of correlated spatial random effects in the model significantly increased the fitness of model with the crash data and superior site ranking performance.

Another dimension of studies investigated the inclusion of serial correlations to benefit from the time dependent factors which were not incorporated in the previous models (Andrey and Yagar, 1993; Hay and Pettitt, 2001; Wang et al., 2013). Wang et al. (2006) did a temporal analysis of rear-end collisions at intersections by utilizing generalized estimating equations with negative binomial link function. The three-year longitudinal data for 208 signalized intersections served for the development of four models with different correlation structures: independent, exchangeable, autoregressive (AR), and unstructured. The autoregressive structure was observed to have best goodness-of-fit and an estimated correlation of 0.4454 for each successive two years. Huang et al. (2009) employed the same AR model with a time step of one year (lag-1), along with five other models, for empirical evaluation of identification of hotspots by different approaches. The study conducted on intersection crash data revealed that the models based on Full Bayesian (FB) hierarchical approach significantly outperformed others in hotspot identification. Based on the same approach, the AR-1 model was observed to have the best fit with the crash rate-related parameters, as assessed by three goodness-of-fit criterion: DIC (Deviance Information Criterion), MAD (mean absolute deviance), and MSPE (mean-squared predictive error). Aguero-Valverde (2013) did a segment-level comparative study based on the precision of crash frequency estimates of different random effects models. Two of the models considered for comparison were Fixed-over-time and independent-over-time random effects. The results established that by fixing the random effects over time, the model parameters were able to 'pool strength' from the neighboring years as the model fit and precision of estimates significantly increased. The evaluation of models also revealed that the concerned model performed consistently better at site ranking due to the notable reduction in standard errors of estimates. Jiang et al. (2014) used site-specific fixed-over-time random effect to incorporate temporal correlations into a Poisson lognormal model for highway network screening. The comparison results with other competing models revealed that the inclusion of these correlations significantly improved the capability of the model to fit the crash data, which was in line with the observations made by a previous study (Aguero-Valverde, 2013). Moreover, the model exhibited a consistently superior performance for identification of hotspots. El-Basyouny and Kwon (2012) analyzed the yearly trend and random variation of parameters using yearly time period while investigating the effect of weather and time on crash types. Four multivariate Poisson lognormal models were developed using Full Bayesian framework: with and without linear time trend, yearly varying intercept, and yearly varying coefficients. The results confirmed the superiority of the model with varying coefficients to possess the best fit based on DIC. This study used day-of-week as a proxy variable due to the unavailability of traffic exposure, and it was recommended to investigate different datasets with traffic exposure with the temporal models to confirm the findings. Along with a linear time trend, an ecological study by Earnest et al. (2007) used a Conditional Autoregressive (CAR) model to evaluate different neighborhood weight matrices which incorporated quadratic time trend as well.

Compared with a large number of research studies using various temporal correlations among crash data, there is little research dedicated to the evaluation of the different temporal treatments on modeling performance. To add to the literature the much-needed research, the present study first compared seven alternate temporal models implemented in the past studies with varying complexity of random effects: (I) independent-over-time; (II) fixed-over-time; (III) linear time trend; (IV) quadratic time trend; (V) yearly varying intercept; (VI) yearly varying coefficients; and (VII) Autoregressive-1 (AR-1). Subsequently, based on the modeling results, this study proposed a new method which combines the strengths of two best models: the time varying coefficients and AR-1. To the best knowledge of authors, such combination is first proposed in the traffic safety field, even though the similar combination of AR and time trend can be found in the current practice, such as the STEPAR (stepwise autoregressive method) specification provided in SAS/STAT 9.2 software package (Jones and Huddleston, 2009). 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. This research chose AR-1, which has been used often in traffic safety research, as a representative of the large body of AR-involved models.

As this study is primarily focused on assessment of temporal treatments, hence a same distance-based weight matrix was incorporated in all models to account for the possible spatial correlations among crash sites to ascertain that the differences observed in results are mostly influenced by the temporal random effects. However, the time and space interactions were not included which might blur the comparison of alternative temporal treatments. All models under investigation were evaluated based on the goodness-of-fit, accuracy of crash estimation, and relative performance of safety ranking at Traffic Analysis Zones (TAZ) level. Firstly, DIC was used as a measure for assessment of the fitness of model estimates with the actual crash data. It is also utilized for determination of model complexity, or, number of effective parameters that are used for model development. Secondly, chi-square RSS (Residual Sum of Squares) was calculated to quantify the discrepancy in the crash estimation and observed crashes. This cross-validation is performed using two approaches: longitudinal and horizontal. The former is essentially a forward prediction evaluation, while the latter is cross-sectional prediction, which is performed by employing k-fold technique where the dataset were divided into k groups and each of them was used interchangeably as validation and training sets. Finally, to evaluate the relative site ranking performance among competing models, the model with best predictive capability was chosen as the base model which serves to yield the "truly" safety estimates. The ranking agreement was then evaluated using sensitivity, specificity and ROC (Receiver Operating Characteristic) based on a set of ordinal threshold lines.

The remainder of the paper first describes the development of eight models under the Full Bayesian framework, along with the convergence checks used. Then, the criteria for model fitness, validation, and empirical evaluation are described, followed by discussion of results. Conclusions and recommendations follow.

2. MATERIALS AND METHODS

2.1 Model Development

This study used the Full Bayesian (FB) hierarchical approach to account for the structural heterogeneities, such as temporal and spatial, for development of crash frequency models. Several recent studies have revealed the capability of hierarchical modeling technique to better fit the crash rate data by incorporating such heterogeneities (Song et al., 2006; MacNab, 2003; Huang et al., 2008a; Huang et al., 2008b).

The model formulation is presented in order of complexity, from the independent-over-time random effects to more sophisticated proposed model, which is a combination of AR-1 and time-varying coefficient. The models under investigations are specified as below:

Model 1: Independent-over-time random effects

At the first step of the hierarchical approach, the crash rate is modeled as the Poisson process:

$$y_{it} \sim poisson\ (e_{it}\theta_{it}) \tag{1}$$

Where, y_{it} is the observed crash count at zone i in time period t and θ_{it} is the mean expected crash rate for site i in time period t, e_{it} is the exposure in zone i of time period t. In our case, the exposure is the average daily vehicle-miles (DVMT) by TAZ.

Crash rate is modeled as a function of covariates and random effects, as shown in equation:

$$\log(\theta_{it}) = \beta_0 + \beta_k X_k + \emptyset_i + \varepsilon_{it} \tag{2}$$

where, β_0 is the vector of intercept, β_k is the vector of independent coefficients, X_k is the vector of independent covariates, $\emptyset_i$is the spatially structured random effect, and ε_{it} is the error term for TAZ i at time t. Since the main focus of this study was to compare the influence of time on crash rate, hence spatial random effects term ($\emptyset_i$) was separately introduced in the model (split from random effect) to avoid blurring the time comparison.

At the second step, the coefficients of covariates (β_k) are modeled using non-informative Normal priors (i.e. $\beta_k \sim$ Normal (0, 10^6)) while the error term and associated variance are modeled with normal and gamma distributions, respectively (Bernardinelli et al., 1995):

$$\varepsilon_{it} \sim normal\ (0, \sigma_{it}^2\ ) \tag{3}$$

$$\sigma_{it} \sim gamma\ (0.001{,}0.001) \tag{4}$$

Model 2: Fixed-over-time random effects

According to Equation 3, the error term ε_{it} varies across road sites and over time. However, it can be argued that the same zone shares identical unobserved features over years. Hence, an identical site-specific random effect across years is added taking the following form:

$$\log(\theta_{it}) = \beta_0 + \beta_k X_k + \emptyset_i + \varepsilon_i \quad (5)$$

$$\varepsilon_i \sim normal\ (0, \sigma_t^2\ ) \quad (6)$$

Model 3: Linear time trend

In this model, a linear time trend is assumed by employing time as a covariate

$$\log(\theta_{it}) = \beta_0 + \beta_k X_k + \beta_{k+1} t + \emptyset_i + \varepsilon_i \quad (7)$$

Where, β_{k+1} is the scalar parameter for linear yearly trend and assigned with non-informative Normal prior (i.e. β_k ~ Normal (0, 10^6))

Model 4: Linear as well as quadratic time trend

To explore a more complex trend of time, a quadratic temporal random effect term is also added

$$\log(\theta_{it}) = \beta_0 + \beta_k X_k + \beta_{k+1} t + \beta_{k+2} t^2 + \emptyset_i + \varepsilon_i \quad (8)$$

Where, β_{k+2} is the parameter for quadratic yearly trend and assigned with the same non-informative Normal prior as β_{k+1}.

Model 5: Varying intercept

This model has an intercept which varies with the yearly time period for every site

$$\log(\theta_{it}) = \beta_{0t} + \beta_k X_k + \emptyset_i + \varepsilon_i \quad (9)$$

Where, β_{0t} is the vector of yearly-varying intercept. Each intercept was assigned with a non-informative Normal prior.

Model 6: Varying coefficients

This model includes the yearly varying coefficients for the intercept as well as covariates. Essentially, it can be regarded as a "random parameter" model in terms of time periods.

$$\log(\theta_{it}) = \beta_{0t} + \beta_{kt} X_{kt} + \emptyset_i + \varepsilon_i \quad (10)$$

Where, β_{kt} is the matrix of time varying coefficients of independent covariates. Each element of the matrix was assigned with a non-informative Normal prior.

Model 7: 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. We chose AR-1 based on the assumptions of stationarity restriction.

$$\mathrm{Log}(\theta_{it}) = \beta_0 + \beta_k X_k + \emptyset_i + \varepsilon_{it} \quad (11)$$

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:

$$\varepsilon_{it} \sim normal\left(0, \sigma_{it}^2 / (1 - \gamma^2)\ \right) \quad (12)$$

$$\varepsilon_{it} \sim normal\left(\gamma \varepsilon_{i,t-1}, \sigma_{it}^2\right) \quad \text{for t >1} \quad (13)$$

Where γ is the autocorrelation coefficient with the range of $0 < \gamma < 1$.

Model 8: Combination of AR-1 and varying coefficients

Based on our modeling results, Model 6 and 7 appeared to be the best fitting models. This study proposes a new model which combines the strengths of the two models. As stated previously (Jones and Huddleston, 2009), the stepwise autoregressive method combines a time trend regression with an autoregressive model for departures from trend (Lunn et al., 2000). Building on that approach, this study combines the time-dependent random parameters with the autoregressive lag-1 model.

$$\log(\theta_{it}) = \beta_{0t} + \beta_{kt}X_{kt} + \emptyset_i + \varepsilon_{it} \tag{14}$$

$$\varepsilon_{it} \sim normal(\,0, \sigma_{it}^2/(1-\gamma^2)\,) \tag{15}$$

$$\varepsilon_{it} \sim normal\left(\gamma\varepsilon_{i,t-1}, \sigma_{it}^2\right) \qquad \text{for t >1} \tag{16}$$

2.2 Convergence Checking

These eight models were estimated with the Full Bayesian techniques using the open source software WinBUGS (Brooks and Gelman, 1998). While fitting these models to the crash data, summary statistics of the posterior inference of parameters were obtained via three chains with 15,000 iterations, initial 3000 of which were discarded as a burn-in sample. Convergence was monitored to ensure that posterior distribution has been found and sampling should be initiated (El-Basyouny and Kwon, 2012). Convergence was assessed by visual inspection of the Markov chains trace plots for the model parameters. Moreover, the number of iterations was selected such that the ratios of Monte Carlo error for each parameter in the model relative to standard deviations would be less than would be less than 0.05. Also, the convergence of multiple chains was assessed by ensuring the value of Brooks–Gelman–Rubin (BGR) statistic (Spiegelhalter et al., 2002) to be less than 1.2.

2.3 Model Comparison

This study used DIC (Deviance Information Criterion), for model comparison on the basis of the complexity and goodness of fit. DIC is a hierarchical Bayesian equivalent of Akaike's Information Criteria (AIC), which was proposed by Spiegelhalter et al. (2003) to account for model fit and complexity. Specifically, DIC is defined as an estimate of fit plus twice the effective number of parameters:

$$DIC = D(\bar{\theta}) + 2p_D = \bar{D} + p_D \tag{17}$$

Where, $D(\bar{\theta})$ is the deviance evaluated at the posterior means of the parameters of interest ($\bar{\theta}$), and posterior mean deviance $\bar{D}$ can be taken as a Bayesian measure of fit or "adequacy". p_D denotes the effective number of parameters in a model, which reflects the complexity of the model, as the difference between $D(\bar{\theta})$ and $\bar{D}$, i.e., mean deviance minus the deviance of the means. Generally, smaller values of DIC are preferred. As a general guideline by Spiegelhalter et al. (2003), a difference of 7+ points in the DIC is treated as significant for modeling performance.

2.4 Validation

The results from the models were further studied to ensure that the advantages of better goodness-of-fit were carried over to the estimation. Chi-squared RSS (Residual Sum of Squares) was employed to quantify the difference between the crash prediction results from the models and observed crash counts. Specifically, RSS is defined as:

$$RSS = \sum (O_i - \theta_i)^2 / \theta_i \tag{18}$$

with O_i and θ_i being the observed and estimated crash rate, respectively.

RSS was calculated by utilizing two different approaches: Longitudinal and Cross-sectional. In longitudinal method, the observed crash count of tenth year (O_i) and the prediction results from the models developed based on the first nine years were used to calculate RSS. It is noteworthy that for three models (Model 5, 6, and 8), this approach would not be feasible due to the presence of time varying coefficients. Hence, the authors also employed a cross-sectional approach which utilized the k-fold cross-validation method (Ye and Lord, 2011). The data were randomly divided into k mutually exclusive groups of approximately equal size. Among the k datasets, every group was used as the validation dataset, and the other groups were combined to form a training dataset which was used for building the prediction models. Overall, four groups were used as different combinations of training and validation sets to test the prediction models.

2.5 Evaluation of Relative Site Ranking Performance

To quantify the transferability of better model fitness and crash estimation to site ranking performance, the models were evaluated for relative site ranking performance. As demonstrated by a previous study (Aguero-Valverde, 2013), the comparative assessment was performed by assuming the 'best model' (based on DIC and RSS) as the ideal case, which served to establish the truly safe and unsafe sites. Sensitivity and Specificity served as diagnostic criteria for evaluation of models for their empirical performance (site ranking).

$$Sensitivity = CP\ (correct\ positives) / TP\ (true\ positives) \tag{19}$$

$$Specificity = CN\ (correct\ negatives) / TN\ (true\ negatives) \tag{20}$$

Where, CP: number of truly hazardous sites correctly identified as hazardous; TP: total number of truly hazardous sites based on 'best model'; CN: number of truly safe sites identified as safe; and TN: total number of truly safe sites based on 'best model'.

Under usual practice, one or two predefined thresholds (say top 10% of sites) are used for establishing the hazardous sites. Our study employed a continuous set of thresholds which performed the sensitivity/specificity analysis in steps, starting with the minimum number of sites and progressing towards the maximum limit of consideration of all 203 TAZs counties as hotspots. Further, ROC (Receiver Operating Characteristic) was used to graphically illustrate the ranking agreement of pairs of models at different threshold levels.

3. DATA DESCRIPTION

This study was conducted on the data of 203 TAZs from the city of Irvine, California. Various other studies have focused on the TAZs for planning-level analysis of crashes as they have benefits of better homogeneity and easy integration into the transportation planning process (Siddiqui et al., 2012; Abdel-Aty et al., 2013; Pulugurtha et al., 2013). As the objective of this study required development of crash prediction models, while accounting for spatial and temporal correlations, an array of variables were employed. In specific, 10-year crash data (2003-2012) used as independent variables were collected from Statewide Integrated Traffic Records System (SWITRS), and other covariates containing socioeconomic factors, transportation-related information and road environment factors covering the same time period were provided by SCAG (Southern California Association of Governments). The ten year time period was selected to ensure a clean serial trend in crashes since AR-1 model is reliant on the previous one year's crash count.

The crash dataset were comprised of fatal and injury collisions only, without the inclusion of property damage only (PDO) due to the underreporting issue related with those crashes (Ye and Lord, 2011). The map showing the TAZs and the distribution of hotspot and cold spots based on observed crash count is presented in Figure 1. Daily Vehicle Miles Travel (DVMT) was incorporated as the traffic exposure factor (Miaou et al., 2003). The models were developed based on data from 2003~2011. The crash dataset of 2012 was not used in model development, but rather employed for the longitudinal cross validation purpose. As the models also incorporated the spatial correlations, the distance between the geometric centroids of TAZs were also obtained from SCAG. As the city of Irvine has 203 TAZs, so the distance matrix had the size of 203x202 with a minimum value of 0.16 miles and maximum of 13.21 miles. Summary information for the various dependent and independent variables for the TAZs is shown in Table 1. The statistics reflect the respective measure (maximum, minimum, and so on) for a particular year while considering the whole dataset of 10 years (2003-2012). It is important to note that most of the variables have 'zero values' which may be attributed to the fact that some of the areas (TAZs) differed from the rest with respect to the 'area type'. For example, some areas may be completely vacant, farms, or only serve as office destinations without any permanent residents.

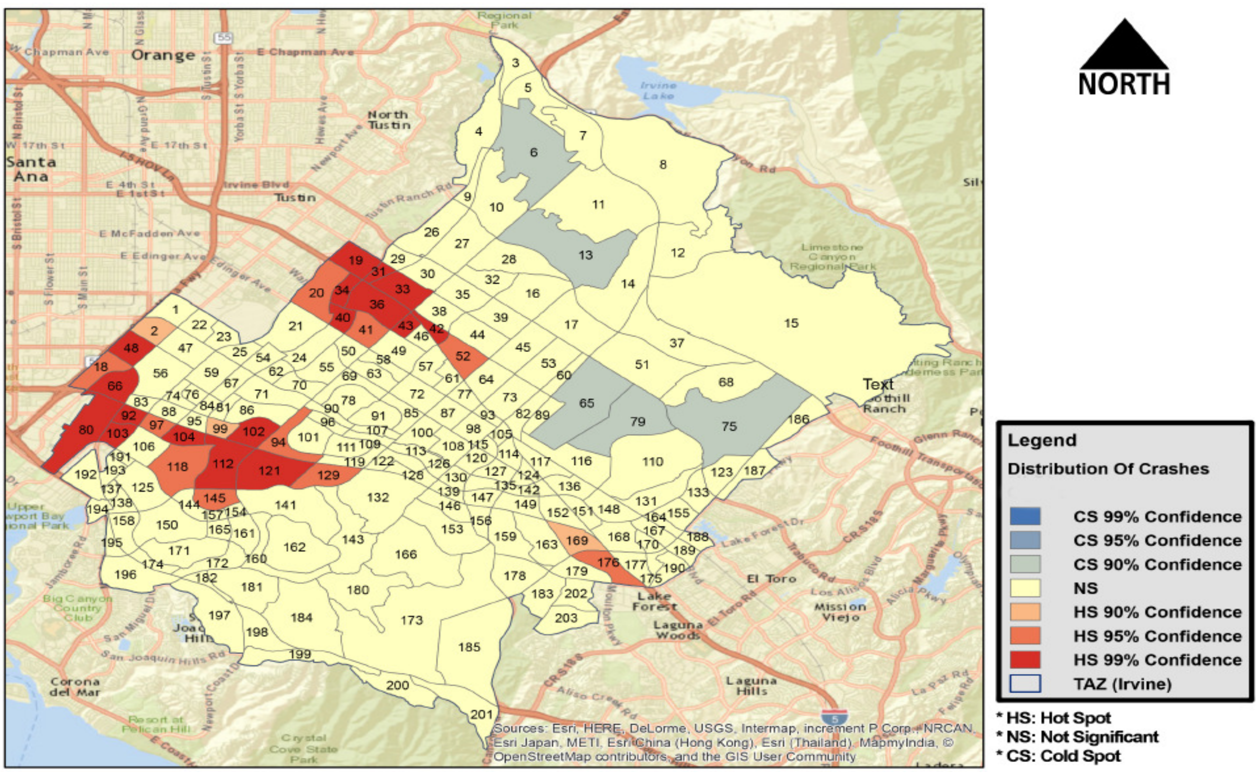


FIGURE 1 Crash distributions at TAZ level in the City of Irvine, California.

TABLE 1 Descriptive Statistics of Collected Data of Various TAZs

Variables	Description	Minimum	Maximum	Median	Mean	S.D.
Collision	Total Annual Fatal and Injury Collisions	0	56	3	5.06	6.6
VMT_per_day	Vehicle Miles Travelled per day	112.57	276079.92	34795.66	54262.44	56156.84

Acre	TAZ area in acre	0.69	5062.95	183.35	282.90	431.75
Med-inc	Median household income ($)	0.00	183347.00	41581.00	48440.78	50635.10
Pop_den	Population density (persons/acre)	0.00	32.40	0.79	6.19	7.96
HH_den	Household density (hh/acre)	0.00	13.62	0.10	2.34	3.15
Emp_den	Employment density (jobs/acre)	0.00	121.10	2.01	10.34	17.43
Ret_den	Rtail job density	0.00	17.45	0.08	0.79	2.02
RetSer_den	Retail+ Service (retail + FIRE + ArtsFood + Other Serv.) job density	0.00	50.60	0.44	2.99	6.29
Jobmix13	Job mix (13sectors); 1 = highest mix (jobs are equal for all sectors)	0.00	0.93	0.62	0.54	0.28
Int34_Den	Intersection density (3- and 4- legs)	0.00	0.62	0.08	0.12	0.12
BKlnAcc	Bike lane access (1=if a TAZ has bike lane)	0.00	1.00	1.00	0.92	0.28
Rail	1=at least one rail station in a TAZ	0.00	1.00	0.00	0.01	0.10
TTbus_D	Total Bus Stop Density	0.00	0.53	0.01	0.05	0.09
WalkAcc	Walk Accessibility	0.00	74.53	0.42	3.87	9.46
Pct_Art	Percent of main arterial (45-55 mph) of TAZ	0.00	0.80	0.00	0.11	0.17
Distance	Distance among centroids of TAZs (unit: miles)	0.16	13.21	4.24	4.42	2.23

4. RESULTS AND DISCUSSION

4.1 Model Fitness

For each model, preliminary multi-collinearity tests and stepwise selection methods were employed in selecting the best subset of predictors for final model development.

Table 2 shows the comparison results for model fitness and complexity. Substantial model-fitting improvement was observed in the models which accounted for serial correlations (Models 7 and 8). The models with independent and fixed-time random effects (Models 1 and 2) had a 40 point difference of DIC value compared to the two AR models. The AR models had the lowest DIC even though the complexity was highest due to the inclusion of more effective parameters, as reflected by their values of p_D. On closer scrutiny, it was revealed that the incorporation of γ reduced the deviance($\overline{D}$), which compensated for the increase in DIC due to the inclusion of more effective parameters. The worst model fitness was exhibited by Models 1 and 2, while Models 3, 4, 5, and 6 had almost similar DIC values, considering that the difference of 7 points in DIC reflects equivalent goodness-of-fit. It is noteworthy that the inclusion of more time-varying coefficients (Model 6) significantly reduced the deviance and improved the DIC, even though the complexity increased.

In comparison between the two best fitting models (AR models), Model 8 had a lower complexity than Model 7 (reflected by a difference of 25 points of p_D), though the number of effective parameters were same for both models. This could be attributed to the addition of varying coefficients to the basic AR model which lends support to refinement of goodness-of-fit, ultimately resulting in a lower DIC value. Overall, even though there was a modest DIC difference (3.1), but the proposed Model 8 had a lower DIC, along with the benefits of lower complexity.

Only the statistically significant variables were shown in Table 2, except for those shown in shaded cells for the purpose of comparison of time-varying coefficients. As evident from Table 2, the models had a varying number of variables that were identified to have a significant impact on crash risk. Models 1~4 had three significant factors (β_0, β_1, β_2) while Models 5~6 had one additional factor (β_0, β_1, β_2, β_3). This trend implies that the identification of contributing factors is highly influenced by the selection of model as some parameters may remain hidden if temporal random effects are not appropriately accounted. Both linear and quadratic time trends in Models 3~4 were significant. Both autocorrelation coefficients γ were statistically significant in Models 7 and 8. The magnitude of γ in model 8 was lower than that in Model 7. This is expected as the time-varying coefficients in Model 7 takes into account the serial correlation to some degree. Finally, for all models, the geographic area and percent of arterial streets were negatively related to crashes, while household density showed the positive relationship. Such phenomena indicate the more space and higher design standard of streets could lower the crash risk, while the more household would increase the likelihood of crash occurrence.

Table 2 Estimates of Variable Coefficients Obtained by various Models

Models	Variables and Parameters							Goodness-of-fit		
	β_0	β_1	β_2	β_3	β_{t1}	β_{t2}	γ	$\bar{D}$	p_D	DIC
1	-8.99 (0.11)	-9.11E-4 (2.75E-4)		-1.48 (0.52)				6856.2	195.6	7051.8
2	-9.00 (0.11)	-8.71E-4 (2.84E-4)		-1.50 (0.51)				6856.1	195.3	7051.4
3	-8.91 (0.10)	-8.84E-4 (2.55E-4)		-1.53 (0.52)	-0.02 (0.004)			6839.9	196.5	7036.4
4	-8.85 (0.11)	-8.88E-4 (2.61E-4)		-1.55 (0.52)	-0.05 (0.018)	0.003 (0.001)		6837.3	197.5	7034.8
5	-9.13 (0.14) -9.25 (0.14) -9.17 (0.14) -9.22 (0.14) -9.28 (0.14) -9.31 (0.14) -9.29 (0.14) -9.26 (0.14) -9.29 (0.14)	-9.23E-4 (2.64E-4)	0.09 (0.03)	-1.06 (0.51)				6833.3	203.3	7036.7

	-9.14 (0.16)	-8.80E-4 (3.02E-4)	0.08 (0.03)	-1.01 (0.57)						
	-9.34 (0.16)	-6.07E-4 (2.92 E-4)	0.09 (0.03)	-1.10 (0.56)						
	-9.16 (0.16)	-0.01E-3 (3.01E-4)	0.07 (0.03)	-0.83 (0.56)						
	-9.28 (0.16)	-7.56E-4 (2.97E-4)	0.08 (0.03)	-0.77 (0.56)						
6	-9.23 (0.16)	-9.66E-4 (3.11E-4)	0.07 (0.03)	-1.19 (0.56)				6805.9	226.6	7032.4
	-9.31 (0.16)	-9.27E-4 (3.11E-4)	0.11 (0.03)	-1.51 (0.57)						
	-9.33 (0.16)	-6.91E-4 (2.97E-4)	0.09 (0.03)	-1.30 (0.56)						
	-9.44 (0.16)	-5.18E-4 (2.87E-4)	0.11 (0.03)	-0.86 (0.56)						
	-9.42 (0.16)	-8.85E-4 (3.08E-4)	0.12 (0.03)	-0.82 (0.57)						
7	-9.34 (0.14)	-7.13E-4 (2.54E-4)	-0.09 (0.03)	-0.96 (0.49)			0.11 (0.02)	6692.5	327.1	7019.6
	-9.16 (0.16)	-8.67E-4 (3.01E-4)	0.08 (0.03)	-0.93 (0.57)						
	-9.36 (0.16)	-5.89E-4 (2.90E-4)	0.09 (0.03)	-1.05 (0.58)						
	-9.18 (0.16)	-9.88 E-4 (3.11E-4)	0.07 (0.03)	-0.75 (0.57)						
	-9.30 (0.16)	-7.42E-4 (2.97E-4)	0.08 (0.03)	-0.72 (0.57)						
8	-9.25 (0.16)	-9.50E-4 (3.08E-4)	0.08 (0.03)	-1.14 (0.58)			0.07 (0.03)	6713.9	302.6	7016.5
	-9.33 (0.16)	-9.06E-4 (3.10E-4)	0.11 (0.03)	-1.45 (0.58)						
	-9.36 (0.16)	-6.65E-4 (2.93E-4)	0.10 (0.03)	-1.23 (0.58)						
	-9.46 (0.16)	-4.96E-4 (2.84E-4)	0.11 (0.03)	-0.79 (0.57)						
	-9.44 (0.16)	-8.72E-4 (3.09E-4)	0.12 (0.03)	-0.74 (0.58)						

Notes: 1. β_0- Constant; β_1- Acre; β_2-HH_den; β_3- Pct_Art ; β_{t1}- linear time trend ; β_{t2}- quadratic time trend ; γ- autocorrelation coefficient (Refer to Table1 for detailed description of variables).
2. Numbers in parentheses represent uncertainty estimates, or, posterior standard deviations
3. The non-significant coefficients are shown in the shaded cells.

4.2 Cross Validation

As mentioned earlier, chi-square RSS was employed for cross validating the crash prediction results of the models. A relatively lower value of RSS indicates better capability to predict the crash count in an area. Longitudinal RSS was adopted for comparison between the observed crash count of last year and predicted crashes based on previous years. For the models with time varying coefficients (Models 5, 6, and 8), it was not feasible to conduct forward prediction. Hence, cross sectional prediction was conducted by utilizing 4-fold approach to divide the dataset randomly. The results are shown in Table 3.

In case of cross sectional prediction, the least average value of RSS was observed for Method 8, followed by Method 7 and 6. This reflects the accuracy of these methods to predict crashes with least deviation from the observed count. The largest value was observed for Method 3, which incorporated a linear time trend in the model. Probably the same global linear trend applicable to all data of successive years might not be appropriate. The comparable Method 4 exhibited better prediction capabilities, which could be attributed to the inclusion of quadratic time trend for providing a more subtle fitting to the data. Similar trends were revealed by the results from four groups. Group 4 demonstrates very significant differences among the eight methods, where the worst performing Method 3 has a roughly three times larger RSS score than the best Method 8. These differences were lesser pronounced in other groups. The longitudinal RSS corroborated the findings as similar trend was displayed in the cross-sectional one, with Method 3 performing the worst and Method 7 having the best prediction capabilities. As Method 7 bears the closest results to proposed Method 8, hence it may be inferred from the results that similar performance would have been exhibited by Method 8.

The results from longitudinal and cross-sectional validations further bolster the selection of Methods 6 and 7 for the development of proposed Method 8, as these two methods were best at both fronts, only inferior to the proposed Method 8.

TABLE 3 Cross Validation by RSS

Horizontal RSS (cross sectional prediction)								
Group	Method							
	1	2	3	4	5	6	7	8
1	9.753	9.755	9.7	9.72	10.296	10.556	11.124	10.744
2	9.71	9.742	9.819	9.815	8.713	8.681	8.773	9.06
3	18.632	18.724	18.661	18.735	12.811	13.006	14.18	12.986
4	10.58	16.449	19.403	11.814	11.147	10.326	6.57	6.365
Average	12.169	13.668	14.396	12.521	10.742	10.642	10.162	9.789
Longitudinal RSS (Forward Prediction)								
Group	Method							
NA	1	2	3	4	7	NA	NA	NA
	5.092	4.977	5.547	5.154	4.797			

Note: NA refers to not applicable

4.3 Relative Performance at Site Ranking

Based on the DIC and RSS, the newly proposed Model 8 was adjudged the best. Hence, the authors ordered the site ranking performance of all other models with reference to Model 8, where the sites identified by Model 8 at different thresholds were assumed to be "truly hazardous (unsafe)". As mentioned previously, different incremental thresholds were utilized for assessment of ranking performance, with every 5% increment roughly reflecting 10 sites.

4.3.1 Sensitivity and Specificity

A model with ideal agreement with Model 8 is supposed to have a 100% value of sensitivity and specificity. This would imply that the model successfully identified truly unsafe sites without any false alarms. Table 4 clearly shows the sensitivity and specificity for the seven models at different cut-off levels. Highly varying trends were observed up to the threshold of top 60% sites. Models 5~7 proved to be significantly better at the identification of truly safe and unsafe sites up to a threshold of almost top 50% sites. For top 10 unsafe sites, Models 1~4 had only 60% sensitivity while Model 7 had 100%. This practically means that out of 10 sites, only six were commonly identified with Model 8. For top 20 sites, this percentage improved to 70. These models slowly catch up with better models (Models 5-7) by reducing the gap for discrepancy of identified sites after the threshold of 60%. This trend indicates that only if the cut-off level for hotspots is more than 60% (obviously unpractical in real world), then most of the models would have similar performance.

Overall, Model 7 seemed to have largest values of sensitivity and specificity, followed by Model 6 and 5, in most cases, especially for thresholds less than 60%. Given that these three models are also closer to Model 8 in terms of DIC, it might be concluded that modeling performance is strongly correlated with site ranking performance. The models with similar modeling performance tend to identify more common hot spots.

TABLE 4 Sensitivity and Specificity of Different Ranking Approaches versus Model 8

Models	1		2		3		4		5		6		7	
# of hotspots	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
0	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
10	0.60	0.98	0.60	0.98	0.60	0.98	0.60	0.98	0.90	0.99	0.90	0.99	1.00	1.00
20	0.70	0.97	0.70	0.97	0.70	0.97	0.70	0.97	0.95	0.99	0.95	0.99	1.00	1.00
30	0.80	0.97	0.77	0.96	0.80	0.97	0.80	0.97	1.00	1.00	1.00	1.00	1.00	1.00
40	0.88	0.97	0.88	0.97	0.88	0.97	0.88	0.97	1.00	1.00	1.00	1.00	1.00	1.00
50	0.86	0.95	0.86	0.95	0.86	0.95	0.86	0.95	1.00	1.00	1.00	1.00	1.00	1.00
60	0.93	0.97	0.93	0.97	0.93	0.97	0.93	0.97	0.98	0.99	0.98	0.99	0.97	0.99
70	0.90	0.95	0.89	0.94	0.89	0.94	0.90	0.95	0.99	0.99	1.00	1.00	1.00	1.00
80	0.90	0.93	0.90	0.93	0.90	0.93	0.90	0.93	0.99	0.99	1.00	1.00	1.00	1.00
90	0.89	0.91	0.89	0.91	0.89	0.91	0.89	0.91	0.98	0.98	0.99	0.99	0.99	0.99
100	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.99	0.99	0.99	0.99	0.99	0.99
110	0.92	0.90	0.91	0.89	0.90	0.88	0.90	0.88	0.98	0.98	1.00	1.00	1.00	1.00
120	0.95	0.93	0.95	0.93	0.95	0.93	0.95	0.93	0.98	0.98	0.99	0.99	0.99	0.99
130	0.95	0.92	0.95	0.92	0.95	0.92	0.95	0.92	1.00	1.00	1.00	1.00	1.00	1.00
140	0.96	0.92	0.96	0.90	0.96	0.90	0.96	0.90	0.99	0.97	1.00	1.00	1.00	1.00
150	0.95	0.87	0.96	0.89	0.96	0.89	0.96	0.89	0.99	0.98	0.99	0.98	0.99	0.98
160	0.97	0.88	0.96	0.86	0.96	0.86	0.97	0.88	0.99	0.95	1.00	1.00	1.00	1.00
170	0.97	0.85	0.98	0.88	0.98	0.88	0.98	0.88	0.99	0.97	1.00	1.00	1.00	1.00
180	0.99	0.91	0.99	0.91	0.99	0.91	0.99	0.91	1.00	1.00	1.00	1.00	0.99	0.96
190	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
200	1.00	0.67	1.00	0.67	1.00	0.67	1.00	0.67	1.00	0.67	1.00	1.00	1.00	1.00
203	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00

Note: sens represents sensitivity; spec means specificity

4.3.2 ROC

The performance of all the models was assessed with reference to Model 8 by using ROC as well. Model 8 with the lowest DIC and RSS was considered to be the ideal case, where the area under the curve equals unity, which basically means that the particular method would have a 100% sensitivity and specificity (i.e. only truly hazardous sites are identified as hazardous). Figure 2 depicts ROCs of Models 1, 4, and 7 for the illustration purpose. The area-under-the-curve for

Models 1, 4, and 7 are 0.97, 0.98, and 0.99, respectively. Model 7 is clearly the closest to the ideal case. The area measures indicate that the overall ranking performance for all three models is very close to Model 8, significantly better than a random ROC of 0.5. But it is very important to note that there are still very big variations in top 5% and 10% sites for the Models 1 and 4. Usually in practice, such scenarios are common where the cut-off levels are in the range of top 5%, 10%, or 20% due to budget constraints for the funds allocated towards the implementation of safety countermeasures. In such cases, drastic differences would be observed between the hotspots identified by the two models which would result in significantly different empirical implications. These curves demonstrate that choosing a right model for site ranking proves to be a critical decision.

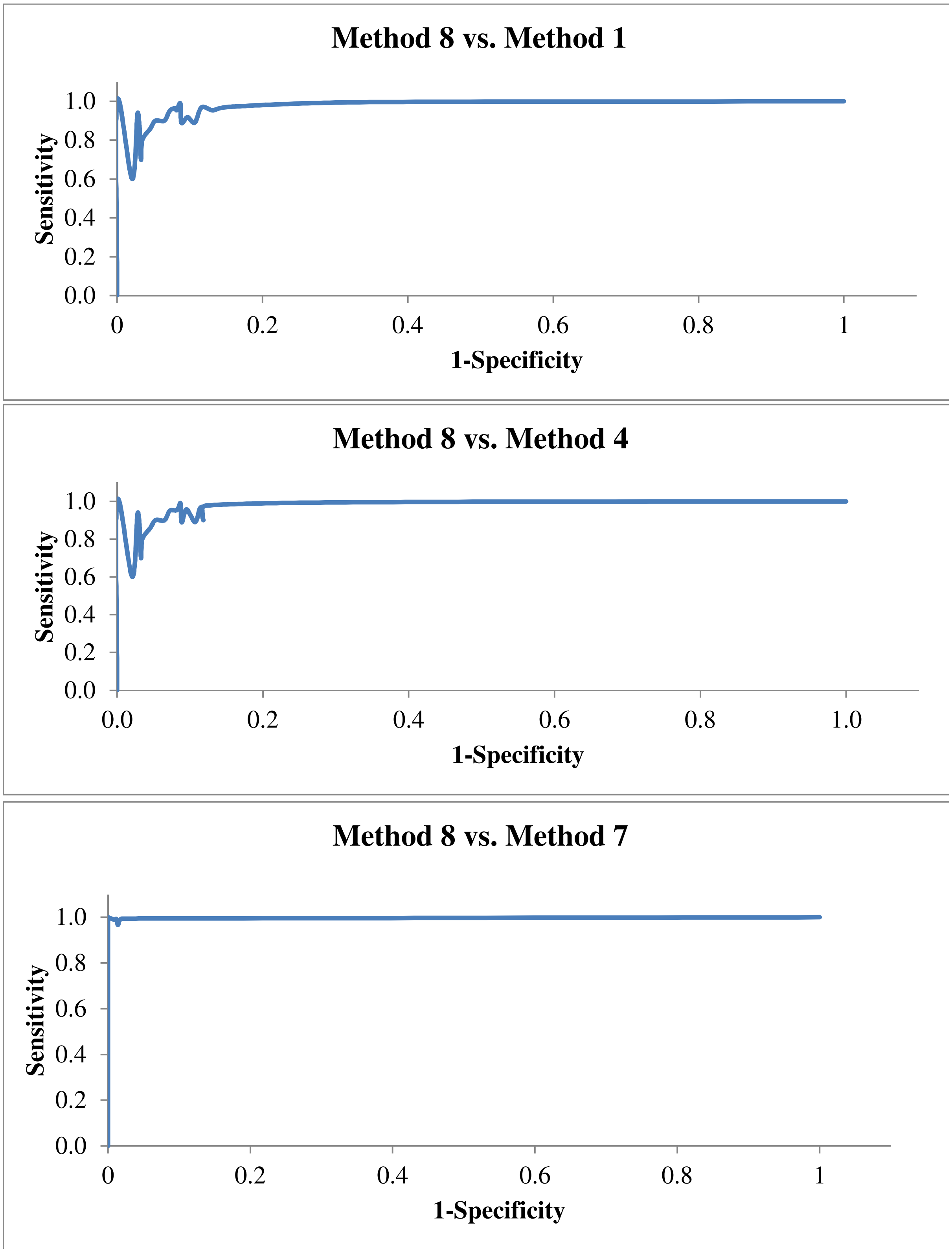


FIGURE 2 ROC of different ranking approaches.

5. CONCLUSIONS AND RECOMMENDATIONS

This study was mainly focused on the evaluation of alternative ways of addressing serial correlation in crash prediction models, along with the assessment of the newly proposed model's anticipated superiority. Ten-year traffic data of 203 TAZs from the city of Irvine were used to develop eight statistical models based on the Full Bayesian framework. The models varied in their complexity, with the simplest model being independent over time and the proposed model (most sophisticated) being a combination of time-varying coefficients and AR-1. It is noteworthy a distance-based spatial random effect term was intentionally introduced so as not to blur the probable advantages of temporal random effects. Comparisons were made among these models based on the fitness, crash prediction, and site ranking performance. Some important findings are shown as follows.

1. The results from DIC revealed significant modeling performance amongst the competing models. The models with independent and fixed-time random effects had 40 points of DIC higher than two AR models. The inclusion of varying coefficients was observed to result in significant reduction in the posterior deviance and improvement of model fitness. The proposed model exhibited best fitness, even though the complexity was highest due to the inclusion of more effective parameters.
2. The number of statistically significant variables being identified was different among the models. Generally speaking, the models with lower DIC tend to identify more significant variables. This indicates the importance of selection of model for identification of contributing factors, as some significant factors may remain hidden with an inferior model.
3. The results from both longitudinal and cross-sectional validations further bolstered the newly proposed Model 8 have the best predictive capabilities, followed by Model 7 and 6.
4. The relative site ranking performance evaluation showed that the site ranking performance is strongly correlated with the modeling performance.

Although the research here reflects an improved understanding of how various HSID perform, further work is still needed. Some of the recommendations are mentioned below:

1. The study used the zonal level data for evaluation of veracious temporal treatments of crash data. The studies based on different geospatial units are needed to verify the results obtained in this study.
2. Only AR-1 was chosen to represent the large body of models involving autoregressive process. Those models are also worthwhile to explore and check whether there are better models than the proposed one of this study.
3. Chi-squared RSS was employed to cross validate the predictive capabilities of various models. More criteria might be used to confirm the results obtained herein.
4. Full Bayesian hierarchical approach was employed for incorporating heterogeneities in crash data. Other frameworks may also be explored.

ACKNOWLEDGMENTS

The authors thank Southern California Association of Governments for providing covariates and geometric centroid distance data of TAZ used in this study.

FUNDING

This research received the grant from UCCONNECT summer research program.

REFERENCES

1. Abdel-Aty, M., Wang, X., (2006). Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors. *Transportation Research Record: Journal of the Transportation Research Board, No.* 1953. *Transportation Research Board of the National Academies, Washington, DC, pp*. 98–111.
2. Abdel-Aty, M., Lee, J., Siddiqui, C., & Choi, K. (2013). Geographical unit based analysis in the context of transportation safety planning. *Transportation Research Part A: Policy and Practice*, *49*, 62-75.
3. 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.
4. Aguero-Valverde, J., Jovanis, P.P., (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis and Prevention* 38 (3), 618–625.
5. Aguero-Valverde, J., Jovanis, P.P., (2008). Analysis of road crash frequency with spatial models. *Transp. Res. Rec.* 2061, 55–63.
6. Aguero-Valverde, J., & Jovanis, P. (2009). Bayesian multivariate poisson lognormal models for crash severity modeling and site ranking. *Transportation Research Record*, *2136*(2136), 82-91.
7. Andrey, J., & Yagar, S. (1993). A temporal analysis of rain-related crash risk. *Accident Analysis & Prevention*, *25*(4), 465-472.
8. Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., & Songini, M. (1995). Bayesian analysis of space—time variation in disease risk. *Statistics in medicine*, *14*(21-22), 2433-2443.
9. Bijleveld, F. D. (2005). The Covariance between the Number of Accidents and the Number of Victims in Multivariate Analysis of Accident Related Outcomes. *Accident Analysis and Prevention,* Vol. 37, No. 4, pp. 591–600.
10. Brooks, S.P., Gelman, A., (1998). Alternative methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. 7, 434–455.
11. Castro, M., Paleti, R., & Bhat, C. R. (2012). A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. *Transportation research part B: methodological*, *46*(1), 253 272.
12. Congdon, P., (2001). Bayesian Statistical Modeling. *John Wiley & Sons, West Sussex, United Kingdom*
13. Earnest, A., Morgan, G., Mengersen, K., Ryan, L., Summerhayes, R., & Beard, J. (2007). Evaluating the effect of neighborhood weight matrices on smoothing properties of Conditional Autoregressive (CAR) models. *International journal of health geographics*, *6*(1), 1.
14. El-Basyouny, K., & Kwon, D. W. (2012). Assessing time and weather effects on collision frequency by severity in Edmonton using multivariate safety performance functions. In *Transportation Research Board 91st Annual Meeting* (No. 12-0494).
15. Guo, F., X. Wang, and M. A. Abdel-Aty. (2009). Corridor-Level Signalized Intersection Safety Analysis Using Bayesian Spatial Models. *Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C.*

16. 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.
17. Huang, H., Darwiche, A.L., Abdel-Aty, M.A., (2010). County-level crash risk analysis in Florida: Bayesian spatial modeling. *Transportation Research Record* 2148, 27–37.
18. Huang, H., Chin, H., 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, 32–41
19. Huang, H. L., H. C. Chin, and M. M. Haque. (2008) Severity of Driver Injury and Vehicle Damage in Traffic Crashes at Intersections: A Bayesian Hierarchical Analysis. *Accident Analysis and Prevention,* Vol. 40, pp. 45–54.
20. Huang, H., H. C. Chin, and M. M. Haque. (2008) Bayesian Hierarchical Analysis of Crash Prediction Models. *Presented at 87th Annual Meeting of the Transportation Research Board*, Washington, D.C.
21. Jiang, X., Abdel-Aty, M., & Alamili, S. (2014). Application of Poisson random effect models for highway network screening. *Accident Analysis & Prevention*, *63*, 74-82.
22. Jonathan, A. V., Wu, K. F. K., & Donnell, E. T. (2016). A multivariate spatial crash frequency model for identifying sites with promise based on crash types. *Accident Analysis & Prevention*, *87*, 8-16.
23. Jones, A., & Huddleston, E. (2009). SAS/STAT 9.2 user's guide. *Cary, NC: SAS Institute Inc*. Madsen, H. (2007). *Time series analysis*. CRC Press.
24. Kim, D. G., Washington, S., & Oh, J. (2006). Modeling crash types: New insights into the effects of covariates on crashes at rural intersections. *Journal of Transportation Engineering*, *132*(4), 282-292.
25. Ladron de Guevara, F., Washington, S., & Oh, J. (2004). Forecasting crashes at the planning level: simultaneous negative binomial crash model applied in Tucson, Arizona. *Transportation Research Record: Journal of the Transportation Research Board*, (1897), 191-199.
26. Lord, D., Mannering, F., (2010). The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A* 44, 291–305.
27. Lord, D., Washington, S.P., Ivan, J.N., (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention* 37 (1), 35–46.
28. Lunn, D., Thomas, A., Best, N., Spiegelhalter, D., (2000). WinBUGS – a Bayesian modeling framework: concepts, structure, and extensibility. *Stat. Comput*. 10, 325–337.
29. Ma, J., Kockelman, K.M., (2006). Bayesian multivariate Poisson regression for models of injury count by severity. *Transportation Research Record* 1950, 24–34.
30. Ma, J., Kockelman, K., & Damien, P. (2008). A multivariate poisson-lognormal regression model for prediction of crash counts by severity, using bayesian methods. *Accident Analysis and Prevention*, *40*(3), 964-975
31. MacNab, Y.C., (2004). Bayesian spatial and ecological models for small-area crash and injury analysis. *Accident Analysis and Prevention* 36 (6), 1019–1028.
32. MacNab, Y. C. A Bayesian Hierarchical Model for Accident and Injury Surveillance. *Accident Analysis and Prevention,* Vol. 35, 2003, pp. 91–102.

33. Miaou, S.-P., Lord, D., (2003). Modeling traffic crash–flow relationships for intersections: dispersion parameter, functional form and Bayes versus empirical Bayes. *Transportation Research Record* 1840, 31–40.
34. Miaou, S. P., Song, J. J., & Mallick, B. K. (2003). Roadway traffic crash mapping: a space-time modeling approach. *Journal of Transportation and Statistics*, *6*, 33-58.
35. Miaou, S. P., & Song, J. J. (2005). Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. *Accident Analysis & Prevention*, *37*(4), 699-720.
36. Mitra, S. (2009). Spatial autocorrelation and Bayesian spatial statistical method for analyzing intersections prone to injury crashes. *Transportation Research Record: Journal of the Transportation Research Board*, (2136), 92-100.
37. Narayanamoorthy, S., Paleti, R., & Bhat, C. R. (2013). On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. *Transportation research part B: methodological*, *55*, 245-264.
38. Park, E., & Lord, D. (2007). Multivariate poisson-lognormal models for jointly modelling crash frequency by severity. *Transportation Research Record*, *2019*(2019), 1-6
39. Pulugurtha, S. S., Duddu, V. R., & Kotagiri, Y. (2013). Traffic analysis zone level crash estimation models based on land use characteristics. *Accident Analysis & Prevention*, *50*, 678-687.
40. Quddus, M.A., (2008). Time series count data models: empirical application to traffic accidents. *Accident Analysis and Prevention* 40 (5), 1732–1741.
41. Shankar, V., F. Mannering, and W. Barfield (1995). Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies. *Accident Analysis and Prevention,* Vol. 27, 1995, pp. 371–389.
42. Siddiqui, C., Abdel-Aty, M., & Choi, K. (2012b). Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis & Prevention*, *45*, 382-391.
43. Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van der Linde, A., (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 64 (4), 583–616.
44. 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*
45. Song, J. J., M. Ghosh, S. Miaou, and B. Mallick (2006). Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping. *Journal of Multivariate Analysis,* Vol. 97, 2006, pp. 246–273.
46. Tunaru, R. (2002). Hierarchical Bayesian Models for Multiple Count Data. *Austrian Journal of Statistics,* Vol. 31, Nos. 2 and 3, pp. 221–229
47. Ulfarsson, G.F., Shankar, V.N., (2003). An accident count model based on multi-year cross-sectional roadway data with serial correlation. *Transportation Research Record* 1840, 193–197.
48. Wang, X., Abdel-Aty, M., & Brady, P. (2006). Crash estimation at signalized intersections: significant factors and temporal effect. *Transportation Research Record: Journal of the Transportation Research Board*, (1953), 10-20.
49. Washington, S., Schalkwyk, I.V., You, D., Shin, K., Samuelson, J.P., (2010). PLANSAFE User Manual: Forecasting the Safety Impacts of Socio-Demographic Changes and Safety Countermeasures. *NCHRP* 8-44(2)

50. Wang, C., Quddus, M., & Ison, S. (2013). A spatio-temporal analysis of the impact of congestion on traffic safety on major roads in the UK.*Transportmetrica A: Transport Science*, *9*(2), 124-148.
51. Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis & Prevention 38 (6), 1137–1150.
52. Xu, P., Huang, H., (2015). Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. *Accident Analysis and Prevention*. 75, 16-25.
53. Ye, F., & Lord, D. (2011). Investigation of effects of underreporting crash data on three commonly used traffic crash severity models: multinomial logit, ordered probit, and mixed logit. *Transportation Research Record: Journal of the Transportation Research Board*, (2241), 51-58.