

INVESTIGATION OF ROADWAY GEOMETRIC AND TRAFFIC FLOW FACTORS FOR VEHICLE CRASHES USING SPATIOTEMPORAL INTERACTION

G. Gill ^a, T. Sakrani ^a, W. Cheng ^{a,*}, J. Zhou ^a

^a California State Polytechnic University-Pomona, California, USA (gurdiljotg, tsakrani, wcheng, jiaozhou)@cpp.edu

KEY WORDS: spatiotemporal; interaction; random parameters; time-varying coefficients; roadway segments; Bayesian

ABSTRACT:

Traffic safety is a major concern in the transportation industry due to immense monetary and emotional burden caused by crashes of various severity levels, especially the injury and fatality ones. To reduce such crashes on all public roads, the safety management processes are commonly implemented which include network screening, problem diagnosis, countermeasure identification, and project prioritization. The selection of countermeasures for potential mitigation of crashes is governed by the influential factors which impact roadway crashes. Crash prediction model is the tool widely adopted by safety practitioners or researchers to link various influential factors to crash occurrences. Many different approaches have been used in the past studies to develop better fitting models which also exhibit prediction accuracy. In this study, a crash prediction model is developed to investigate the vehicular crashes occurring at roadway segments. The spatial and temporal nature of crash data is exploited to form a spatiotemporal model which accounts for the different types of heterogeneities among crash data and geometric or traffic flow variables. This study utilizes the Poisson lognormal model with random effects, which can accommodate the yearly variations in explanatory variables and the spatial correlations among segments. The dependency of different factors linked with roadway geometric, traffic flow, and road surface type on vehicular crashes occurring at segments was established as the width of lanes, posted speed limit, nature of pavement, and AADT were found to be correlated with vehicle crashes.

* Corresponding author

1. INTRODUCTION

Traffic safety is a major concern in the transportation industry due to immense monetary and emotional burden caused by crashes of various severity levels, especially the injury and fatality ones. During the year 2014, roadway crashes were the top cause of deaths among youngsters 16-24 years of age. The direct economic cost for crashes of all severities totaled to \$242 billion for the year 2010, while the comprehensive cost summed up to \$836 billion (NHTSA, 2017). To reduce such crashes on all public roads, the safety management processes are commonly implemented which include network screening, problem diagnosis, countermeasure identification, and project prioritization. The selection of countermeasures for potential mitigation of crashes is governed by the influential factors which impact roadway crashes. One issue central to crash analysis is the understanding of contributing factors to crash occurrence. Crash prediction model is the tool widely adopted by safety practitioners or researchers to link various influential factors to crash occurrences (Gill et al., 2017b). Many different approaches have been used in the past studies to develop better fitting models which also exhibit prediction accuracy. There are a wide range of models including the conventional univariate Poisson and negative binomial (NB) models which estimate the crash of different severity levels or outcomes separately (Ulfarsson and Shankar, 2003), the multivariate Poisson and /or Poisson-lognormal specifications which estimate the crashes of various severity levels or outcomes simultaneously (Ma and Kockelman, 2006), the zero-inflated Poisson (Shankar et al., 1995), among others. The aforementioned Poisson lognormal models allow more flexibility to incorporate the overdispersion of crash data by specifying the error term. Many studies have incorporated different specifications to account for the unobserved heterogeneity among crash data which is not addressed by the explanatory variables considered while development of models. The random effects were introduced in the traditional overdispersed Poisson lognormal models to account for spatially structured and unstructured correlations. The spatial correlations among crash entities have been explored to understand the implications of crash causing factors and draw less biased inferences at different levels such as block group, intersections, road segments, corridors, census tracts, health areas, traffic analysis zones (TAZs), or counties. The literature review illustrates a wide range of neighborhood weight matrix structures that have been proposed to model crash spatial heterogeneity such as adjacency-based weight matrices for first, second, and third order neighbors for segments, Rook and Queen adjacency-based weight matrices of first and second order, Decay and pure-distance based weight matrices of varying orders (Gill et al., 2017a). The aforementioned studies observed significant improvement of model fit and superiority at crash prediction due to the inclusion of spatial correlations from different perspectives. In case of spatially unstructured random effects, several studies incorporated the temporal correlations to address the serial changes for the traffic environment and also lend the continuity for the covariates included during model development.

In this study, a crash prediction model is developed to investigate the vehicular crashes occurring at roadway segments. The spatial and temporal nature of crash data is exploited to form a spatiotemporal model which accounts for the different types of heterogeneities among crash data and geometric or traffic flow variables. As this study incorporates

the spatial correlation, time-varying coefficients, and spatiotemporal interaction, hence the model needs to be flexible for accommodate such complexity. This study utilizes the Poisson lognormal model with random effects, which can accommodate the yearly variations in explanatory variables and the spatial correlations among segments.

2. DATA AND METHODOLOGY

2.1 Data Description

The five-year data used for this study were provided by HSIS (Highway Safety Information System) which collected the data in form of different raw files from California TASAS (Traffic Accident Surveillance and Analysis System) and NHTSA (National Highway Traffic Safety Administration). This data contained records which particularly focused on the freeway sections of California. Each year had four files pertaining to different types of data linked with Road, Occupancy, Vehicle, and Crash characteristics. The information extracted from these four files had crash number along with other factors like geometric (lane width, number of lanes, median type), traffic (Average Annual Daily Traffic), design speed, and so on. The crash data for 279 segments in a rural roadway section of California was considered for model development over a period of five years (2007-2011). This relatively long time period was required to fully incorporate the temporal trends which influenced the explanatory variables of crashes.

2.2 Model Development

As the literature review highlighted that many studies employed the Full Bayesian framework to develop Poisson lognormal models for overdispersed data, this study also adopted the same approach of generalized linear model as the foundation. The necessity of inclusion of different types of correlation structures stem from the superiority attained by the model to address the variability. This micro-level study of roadway segments introduces a spatially structured random effects term in the basic model formulation which addresses the spatial similarity of neighboring segments. This spatial correlation is transferred from one segment to another by the vehicles implying on these segments. Also, the nature of traffic flow is similar for continuous roadway entities which needs to be included through spatial correlation based on proximity as such impact is not fully addressed by the variables. Secondly, the random effects also account for the temporal trends which are expected to improve the model estimation as they capture the serial changes happening in the traffic environment and try to render continuity to the mostly discreet explanatory variables which helps make the data more robust. Finally, this study also accommodated the interaction of space and time. The univariate spatiotemporal Poisson lognormal model developed in this study was the following formulation:

$$y_{ijt} | \lambda_{ijt} \sim \text{Poisson}(\lambda_{ijt} e_{it}) \quad (1)$$

$$\ln(\lambda_{ijt}) = X'_{ijt} \beta_t + \varepsilon_{ijt} + (\delta_{ij}) * t \quad (2)$$

$$\varepsilon_{ijt} \sim \text{Normal}(0, \tau^2) \quad (3)$$

Where i is the site index, j is the total crash count, t is the time period index, y is the recorded crash number, λ is the expected crash number, e_{it} is the segment length which is used as an offset for segment i for time period t , X is the matrix of risk factors, β_t is the vector of model parameters varying with time t , ε_{ijt} is the independent random effects, δ_{ij} is the spatial error term which is fit by the conditional auto-regressive model (CAR) (Besag, 1974), and τ^2 is the variance of the normal distribution for ε_{ijt} . The varying parameters allow for the accommodation of temporal changes experienced by the influential factors, which may not be explained by the variables if they are rendered fixed as the aggregation of such parameters tends to ignore the potentially important variations among the parameters which may lead to biased inferences drawn from the erroneous estimates. The inclusion of time varying parameters allows the flexibility to account for variations which are not reported by the data due to lack of robustness but in reality impact the crash frequency. The spatial error term directly interacts with the linear trend and this approach is expected to provide a subtler fit to the crash data.

$$\tau^2 \sim \text{Gamma}(0.01, 0.01) \quad (4)$$

The CAR model has the capability to accommodate the spatial correlations based on different approaches. Various weight structures have been explored in previous studies (Guo et al., 2009; Xu and Huang, 2015) but for spatial analysis in present study, the weight matrix was developed by considering the adjacency of segments as the governing point to establish the neighbors. The outcome of the weight was binary as the only two possibilities were: either two segments are immediate neighbors (weight is equal to one in this case), or the segments are not immediate neighbors (weight is zero in this case). This class of weight matrices belong to adjacency-based matrix structures of first order (as only the immediate neighbors were considered). Such a matrix structure was selected to represent the traditional approach employed while accounting for the spatial correlation among micro-level study of segments. The CAR formulation allows for the incorporation of number of neighboring segments (which is two for all segments, except first and last segment which have only one immediate neighboring segment), the identity of neighboring segment for each concerned segment, and the assignment of weight (discussed earlier).

3. RESULTS

Variable Coefficients	Time in years				
	1	2	3	4	5
Intercept	-5.96	-7.75	-3.85	-6.21	-5.33
Lane width 9, 10 feet	0.18	0.16	-1.41	-0.27	-0.16
Lane width 11 feet	-0.08	0.84	-1.99	-0.62	-0.56
Lane width 12 feet	0.223	1.6	-0.23	1.18	1.32
Speed limit 25-40 miles	0.33	1.24	1.20	0.46	-0.31
Speed limit 45 miles	0.42	1.46	0.92	0.34	0.32

AADT	0.27	0.36	0.03	0.12	0.16
Bridge Deck Asphalt concrete <7 inches thick	3.92	3.25	3.20	5.1	2.6
	3.34	3.25	3.31	4.48	3.29

Table 2. Modeling results

The model was developed in a freeware statistical software WinBUGS which employing the Markov Chain Monte Carlo (MCMC) algorithms and Gibbs sampling techniques for model estimation. For the model calibration, two chains of 45,000 iterations were set up. Convergence was ensured by visual inspection of chains and the observing the desired threshold condition of MC errors to be lower than 5% of the standard deviation of the parameters. After ensuring the convergence, first 5,000 samples were discarded as adaptation and burn-in and the rest of the samples were used to draw parameter estimates. It is important to note that indicator variables (categorical) were used for lane width, speed limit, and surface type, with the baseline or reference as lane width more than 12 feet, speed limit 50-65 mph, and Asphalt Concrete (Oiled Earth-Gravel), respectively. The variables selected for model development were checked for multi-collinearity issues and only non-collinear ones were incorporated. As shown in Table 1, the modeling results demonstrated the significant correlations among the geometric and traffic flow variables with the crash data. The increase in lane width generally decreased the risk of indulging in a crash. This was observed for different dummy variables corresponding to different lane widths: such as 9, 10, 11, 12 and greater than 12 feet wide. This result seems logical as the direct exposure of vehicles decrease when the gap between adjacent lane vehicles increases due to increased lane width. However, for lane width 12 feet, the trend became opposite. The rationale may be the perceived maneuver safety due to more space which motivates the drivers to drive at higher speeds and less cautiously. The speed limit on the segment was found to be positively correlated with increased crash risk. The higher vehicle speed mean that the driver has less reaction time to handle the vehicle in case of urgency and hence the risk is greater. Another strong correlation was observed for the type of road surface and crash risk. The chance of getting into a crash heightened to a great extent while driving on a bridge deck. Similar trend was observed for asphalt concrete pavement with thickness less than seven inches. The average annual daily traffic (AADT) was weakly correlated with crashes which may be attributed to the lesser vehicular volume on the rural road under analysis.

4. CONCLUSIONS

Traffic safety agencies have a constant job to dedicate funds for safety interventions to mitigate roadway crashes. However, the identification of influential factors from a plethora of potential explanatory variables renders the fund allocation to be a complicated process. This study presented the dependency of different factors linked with roadway geometric, traffic flow, and road surface type on vehicular crashes occurring at segments. The width of lanes, posted speed limit, nature of pavement, and AADT were found to be correlated with vehicle crashes at segments. The variation of parameter estimates for every year illustrate the importance of considering the random parameter models to uncover the underlying trends which may escape the traditional models.

Moreover, the complex model formulation employed in this study will also benefit at hotspot identification as the estimation of crash counts at segments improves with such correlations which account for the variability and generate more precise estimates. The precision and accuracy of the estimates highly govern the allocation of funds towards safety countermeasures to regulate the influential factors which eventually improve the safety performance of entities due to mitigation of crashes. The future study may explore different types of explanatory variables and incorporate different spatial levels. Additionally, the crashes may be separated based on severity in an effort to study different or potentially common factors responsible for different severities.

REFERENCES

- Besag, Julian. "Spatial interaction and the statistical analysis of lattice systems." *Journal of the Royal Statistical Society. Series B (Methodological)* (1974): 192-236.
- Gill, G., W. Cheng, M. Xie, T. Vo, X. Jia, J. Zhou. "Evaluating the Influence of Neighboring Structures on Spatial Crash Frequency Modeling and Site Ranking Performance", *Transportation Research Record: Journal of the Transportation Research Board*, No. 2659 (2017a) (DOI: 10.3141/2659-13)
- Gill, G., R. Falahati, and W. Cheng. "Influential Factors of Right-Turn Violation during Concurrent Traffic/Pedestrian Signal Timing". (2017b) *WDSI Annual Conference Proceedings*, Vancouver, Canada
- Guo, Feng, Xuesong Wang, and Mohamed A. Abdel-Aty. "Corridor-Level Signalized Intersection Safety Analysis Using Bayesian Spatial Models." *Transportation Research Board 88th Annual Meeting*. No. 09-1163. 2009.
- Ma, Jianming, and Kara Kockelman. "Bayesian multivariate Poisson regression for models of injury count, by severity." *Transportation Research Record: Journal of the Transportation Research Board* 1950 (2006): 24-34.
- National Highway Traffic Safety Administration (NHTSA). *Traffic Safety Facts 2015 Data*. Washington, DC, 2017
- Shankar, Venkataraman, Fred Mannering, and Woodrow Barfield. "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies." *Accident Analysis & Prevention* 27.3 (1995): pp. 371-389.
- Ulfarsson, Gudmundur, and Venkataraman Shankar. "Accident count model based on multiyear cross-sectional roadway data with serial correlation." *Transportation Research Record: Journal of the Transportation Research Board* 1840 (2003): 193-197.
- Xu, Pengpeng, and Helai Huang. "Modeling crash spatial heterogeneity: random parameter versus geographically weighting." *Accident Analysis & Prevention* 75 (2015): 16-25.