Proc. of the 6th Intl. Symposium on Reliability Engineering and Risk Management (6ISRERM)
31 May – 1 June 2018, Singapore
Editor(s) Xudong Qian, Sze Dai Pang, Ghim Ping Raymond Ong, Kok-Kwang Phoon
Copyright © 2018 Author(s). All rights reserved.

SPATIAL EFFECTS IN TRAFFIC SAFETY: A REVIEW AND ASSESSMENT OF METHODOLOGICAL ALTERNATIVES

Helai HUANG[1], Ya-ru GU[1], Xiaoqi ZHAI[1]

[1]*Urban Transport Research Center, School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan, 410075 P.R. China.*

E-mail: huanghelai@csu.edu.cn

Since crash occurrence are typically aggregated as clusters in space and crash data are always collected with certain spatial scale, intrinsic spatial effects exist extensively in road safety analysis. It is a common issue for crash prediction models and has gained lots of focus in the past decades toward different aspects. Therefore, having a comprehensive understanding of spatially distributed crash data is indispensable for safety inspection, and incorporating spatial effects in both micro-level and macro-level crash prediction modeling is expected to represent the true underlying data generating processes. This paper provides a detailed review of the spatial data characteristics in road safety analysis, including three key issues, i.e., multilevel data structure, spatial dependence and heterogeneity, as well as methodological approaches that have been used to solve these problems. Whereas these spatial analysis technique substantially improved the accuracy and robustness of crash prediction, zonal level practice has been highlighted in this study to specify the use of these methods.

Keywords: spatial effects, road safety analysis, methodological approaches, zonal level practice

1 Introduction

With the enormous losses to social economy and lives resulting from vehicle crashes, road safety has been attracting more and more attentions from government to citizens in the past decades. Benefiting from continually research efforts, safety analysis has been carried out toward many aspects, e.g., micro-level crash prediction models for road entities, macro-level models for regional units, multivariate analysis considering the correlations between different road users or crash severities, et al. (see Lord and Mannering 2010). Although different modeling technique is suited for each research direction, spatial effects exist extensively in influencing these models' accuracy, since crashes occurrence are typically aggregated as clusters in space and crash data are always collected with certain spatial scale. There are two main benefits to include spatial dependence in safety research, for one thing, spatial effect makes it possible for site estimation to pool strength from neighbors (Aguero and Jovanis 2008); for another, spatially correlated random effect can serve as a part of structured disturbance. The absence of spatial effects can lead to biased and inconsistent parameter estimation. There are two main objectives of this paper, the first one is to provide a review of contemporary thinking in the spatial characteristics of crash data and show a steady advancement of research in methodological approaches to address these data-relate problems. With the similar topic, Wang et al. (2012) did a review focusing on four practical issues (i.e., modifiable areal unit problem, ecological fallacy, spatial dependence and matching individual observations to the correct spatial units) of spatial models in transport, while this paper is mainly about the crash data characteristics and methodological alternatives in spatial analysis. The second objective of this paper is to specify the use of these approaches with several previous studies of zonal level safety analysis, which has been suggested as means of incorporating safety considerations into long term transportation planning (Washington et al. 2006).

2 Spatial Characteristics of Crash Data

Considering the crash generation process, each crash could be analyzed as a Bernoulli trial. The most natural crash analysis unit would be in the individual-crash level (e.g., crash sites, drivers, pedestrians) (Jonathan 2013). However, because of the tremendous data demands for analysis, researchers often aggregate crash data into spatial scales ranging from road segments or intersections to zonal level. As a result, multilevel data structure exists in the traffic safety analysis regarding to spatial dimension. Huang and Abdel-Aty (2010) proposed a five-level hierarchy structure to represent various traffic entities with spatial distribution including from macroscopic to the microscopic levels, i.e., Geographic region level – Traffic site level – Traffic crash level – Driver-vehicle unit level – Occupant level.

There are two major problems related to the location data: spatial dependence and spatial heterogeneity (LeSage 1998). For spatial dependence, it refers to the phenomenon that the crash data of a particular area may be intrinsically correlated with the neighboring attributes. Specifically, spatially observable factors such as road geometry features or traffic characteristics of neighboring areas may have a 'spillover' effect on the interested study unit from the nature of continuity; and spatially unobservable factors such as zonal regulations at proximity locations may cause a spatial error correlation effects. There are two main benefits to include spatial dependence in safety research, for one thing, spatial effect makes it possible for site estimation to pool strength from neighbors (Aguero and Jovanis 2008); for another, spatially correlated random effect can serve as a part of structured disturbance. Additionally, the relationship between crashes and risk factors may not be constant over space, and the cross-area variations could be referred to as spatial heterogeneity. The spatial heterogeneity comes from both area-specific variables and unobserved factors, what's more, these unobserved factors may be collected with observed ones (Xu et al. 2017). Without considering the spatial dependence and spatial heterogeneity, may lead to biased parameter estimation and serious errors (Anselin and Griffith 1988).

3 Methodology toward Spatial Analysis

3.1 *Methodology for multilevel data structure*

The hierarchical modeling is one of the most pervasive methods in the literature dealing with the multilevel structure of crash data. Hierarchical models provide an avenue for including different level data information into a model framework whether in macro or micro crash prediction modeling, for example, Wang and Huang (2016) used the Bayesian hierarchical joint model composed of a road-entity-level model and a TAZ-level model to identify risk factors at both micro level and macro level; In Aguero and Jovanis (2010), crash counts by road segment are modeled at the first level of the hierarchy while the segments are aggregated by road functional class at the second level.

Lately, with the development of computer science, artificial intelligent models (AI) have been widely used with multilevel data structure such as neural networks (Zeng et al. 2016) and support vector machine (Dong et al. 2015). But this method has been criticized for being block boxes incapable of generating explicit functional relationships and statistically interpretable results.

3.2 *Methodology for spatial dependence*

Generally, there are two main methods in spatial econometrics to specify the spatial dependence: (1) traditional econometric methods suitable for cross-sectional continuous data, and (2) Bayesian hierarchical methods suitable for non-negative random count data (Quddus 2008). As discussed above, spatial dependence includes the spatial spillover effect and the spatial error correlation effect, and two specifications have been used to accommodate them in the conventional spatial econometric models: (1) the spatial lag specification can identify spatial spillover effects as well as spatial error correlation effects by adding an explanatory variable in the form of a spatially lagged dependent variable, while (2) the spatial error specification only considers the spatial error correlation effects by adding spatially lagged error structure (Anselin 1988). A specification of these two methods could be referred to Quddus (2008), where simultaneous autoregressive (SAR) model and spatial error model (SEM) were presented. However, because of these methods do not show the true underlying data generation process, the inferences derived from the traditional spatial models could be misleading (Bhati 2005). Also, as the analysis units are less aggregated, leading the numbers of units with zero count increasing, the distribution of crash counts will become highly skewed to the right.

A more flexible and widely used method to model spatial dependence is to incorporate a spatial random effect term, which is commonly given a conditional autoregressive prior (CAR, e.g., Quddus 2008, Huang et al. 2010, Siddiqui et al. 2012) or joint prior (e.g., Mitra 2009, Macnab 2004). Such models are primarily based on a Bayesian framework, using the Markov Chain Monte Carlo (MCMC) method, which has widely been proved advantageous over traditional likelihood estimation (Dong et al. 2016). These approaches always assume the crash data follows a Poisson or Negative Binomial (NB) distribution, so they are more appropriate for the count data analysis. Unfortunately, these approaches only consider spatial error correlation effects without spatial spillover effects, and the models estimation can be difficult with the number of spatial units increases (Narayanamoorthy et al. 2013).

Another potential method that can be used to solve the spatial correlation effect is the generalized estimating equations (GEEs) (Lord and Persaud 2000, Wang and Abdel-Aty 2006). GEEs is an extension of generalized linear models to the analysis of spatially (or temporally) correlated crash frequencies at the intersection level, but it may not be able to capture spatial dependence at spatial units (Quddus 2008).

A common challenge in most of the spatial correlation models, such as CAR and SAR, is the specification of spatial neighboring structure. Different correlation structures could be obtained by giving various weight matrix to each neighbor. The majority of previous studies used the 0-1first order neighboring structure (e.g., Miaou et al. 2003), that is, the adjacent index equals one when two zones are adjacent and zero otherwise. While most previous studies assumed equal weights for adjacent zones, several literatures defined varying spatial-proximity structures both in road entity level and zonal level and studied their effects on model performance. Aguero and Jovanis (2010) and Flask and Schneider (2013) both compared different neighboring structures, and found the consistent result that first order neighbors are effective in the analysis, suggesting that spatial correlation is a more significant factor in regions that directly neighbor the unit of interest. Dong et al. (2014) presented an evaluation of crash prediction models at the level of TAZs with four types of weight structures: 0-1 first order adjacency, common-boundary length, geometry-centroid distance, and crash-weighted centroid distance. The results showed that model with 0-1 first order neighboring structure underperformed compared with the others, and model considering proximity of neighboring zones by weighting their common-boundary lengths performed the best. Therefore, the selection of proper neighboring structures is important to the spatial correlation assessment, and the frequently used 0-1 first order method in previous studies may be not very effective for some case analysis, more comprehensive investigation of different neighboring structures are needed.

3.3 *Methodology for spatial heterogeneity*

Models accommodating for spatial heterogeneity could be classified into three types: global models, semi-local models and local models (Hadayeghi et al. 2010). Specifically, global models assume that the relationship between the dependent variables and each explanatory variable does not vary across geographic areas, as a result, fixed coefficients for the entire study area are estimated. The extensively used Generalized Linear Modeling (GLM) procedure with Negative Binomial or Poisson distribution are cases of global models. But this approach treats variables dependently across areas, which may hide some important spatial factors affecting the occurrence of crashes and produce biased results. To address this problem, spatial correlation methods (demonstrated in the previous section of this paper) are developed to recognize the local nature of spatial data by relaxing the assumption that the error terms for each observation are independent. Although spatial relationships are incorporated into the modeling framework through the covariance of the error terms, such models are not thought of local models, but semi-local models instead, in that these models generate fixed parameters as global models. However, some explanatory variables may have various predicting force of different areas, it might be stronger at certain locations while weaker at others. As a consequence, the constant parameters cannot present the variability. In this way, local models, such as random parameters, geographically weighted Poisson regression (GWPR) and Bayesian spatially varying coefficients (BSVC) models are promising to evaluate the spatial heterogeneity.

With respect to random parameter model, the parameters are drawn from some univariate distributions, and are assumed to vary randomly from case to case. Considerable studies have used this approach to accommodate unobserved factors at road segment or intersection level (e.g., Milton et al. 2008, Wu et al. 2013), while only a few applied for the regional crash modeling (e.g., Xu and Huang 2015, Bhat et al. 2017). Despite the fact that the random parameter models outperform the traditional fixed parameters models, it was criticized that the parameters are estimated independently without considering the locations to which the parameters refer, which may be inappropriate when unobserved heterogeneity are correlated over space (Xu et al. 2017). GWPR and BSVC are two competing approaches can be used to demonstrate this potential spatial correlation in varying parameters. GWPR for count data was developed by Fotheringham et al. (2002) to calibrate multivariate regression models of locally non-stationary processes, through the explicit treatment of spatial coordinates. The output of GWPR is a set of local spatial parameters where the weight are linked to the distance between the observation and the location where independent variables are measured (Hadayeghi et al. 2010). In the past decade, GWPR was extensively used in the field of epidemiology, spatial economy and ecology analysis, with only a few efforts directly towards road safety (e.g., Hadayeghi et al. 2010, Li et al. 2013). A recent study of Xu et al. (2017) introduced an original BSVC model to interpret the spatially non-constant parameters under the Bayesian framework, which is the first practice of this method in the traffic safety analysis. In the BSVC approach, the spatially varying coefficients are modeled as a multivariate spatial process, instead of fitting spatially local regression models as in the GWPR approach. It is worth noting that, although GWPR and BSVC models are proved better prediction performance than the fixed effect models, the validity of inferences derived from the GWPR and SVCP models is needed further assessment (Wheeler and Calder 2007).

4 Zonal Level Spatial Analysis

An increasing research effort has been made on the zonal level safety analysis incorporating the aforementioned spatial issues, which is supposed to provide viable approaches supporting the transportation safety planning techniques. In zonal models, area-wide covariates including socio-demographics (e.g., Wang et al. 2007), road characteristics (e.g., Siddiqui et al. 2012), weather conditions (e.g., Aguero and Jovanis 2006), and commute characteristics (e.g., Abdel-Aty et al. 2013) are related to road crashes, which are aggregated with specific spatial scale, ranging from states (e.g., Noland 2003), counties (e.g., Aguero and Jovanis 2006, Huang et al. 2010), traffic analysis zones (e.g., Hadayeghi et al. 2010, Dong et al. 2015), block groups (e.g., Abdel-Aty et al. 2013), wards (e.g., Quddus 2008). Different zonal levels could be contained in a hierarchical model, for example, Flask and Schneider (2013) proposed a hierarchical negative binomial model with mixed effects including two zonal level (county level and township level) to evaluate single vehicle motorcycle crashes. Within-group correlation and heterogeneity are the two most prevalent topics in the previous research. Huang et al. (2010) used Bayesian spatial model with CAR prior to relate total crashes and severe crashes with socio-demographics factors (e.g., median household income, unemployment rate), road characteristics (e.g., freeway density, principal arterial density, intersection density), and traffic characteristics (e.g., traffic intensity, truck AADT,).

5 Summary and Conclusion

As the preceding discussion indicates, several spatial characteristics of crash data are supposed to be considered in the crash prediction models, including multilevel data structure, spatial dependence and spatial heterogeneity. With the progress in methodological technique, a number of innovative approaches have been proposed to deal with these data-related problems. In the past few years in particular, advanced models, such as hierarchical model for multilevel data, Bayesian spatial model with CAR prior, random parameter models, GWPR as well as BSVC have been widely introduced into site-level, zonal level and even recently emerging meso-level crash prediction process. Incorporating these approaches into crash prediction models helps to reflect the true relationship between crash occurrence and explanatory variables, especially in the macro analysis framework, it is supposed to provide viable approaches supporting the transportation safety planning techniques. Although numerous achievements have been gained, the exploration of spatially distributed crash data is still in its growing period. In addition, there are several problems related to the spatial analysis, such as the modifiable area unit problem, the process of border crashes, need to be further resolved.

References

Abdel-Aty, M., Lee, J., Siddiqui, C. and Choi, K., Geographical Unit Based Analysis in the Context of Transportation Safety Planning, *Transp. Res. Pt. A-Policy Pract.,* 49(49), 62-75, Mar, 2013.

Aguero, V. J. and Jovanis, P., Spatial Analysis of Fatal Injury Crashes in Pennsylvania, *Accid. Anal. Prev.,* 38(3), 618-625, May, 2006.

Aguero, V. J. and Jovanis, P., Analysis of Road Crash Frequency with Spatial Models, *Transport. Res. Rec.,* 2061(2061), 55-63, Dec, 2008.

Aguero, V. J. and Jovanis, P., Spatial Correlation in Multilevel Crash Frequency Models: Effects of Different Neighboring Structures, *Transport. Res. Rec.,* 2165(3), 21-32, Dec, 2010.

Aguero, V. J., Multivariate Spatial Models of Excess Crash Frequency at Area Level: Case of Costa Rica, *Accid. Anal. Prev.,* 59(4), 365-373, Oct, 2013.

Anselin, L., *Spatial Econometrics: Methods and Models, (Vol. 4),* Kluwer Academic Press, Netherlands, 1988.

Anselin, L. and Griffith, D., Do Spatial Effects Really Matter in Regression Analysis?, *Pap. Reg. Sci.,* 65(1), 11-34, Jan, 1988.

Bhat, C. R., Astroza, S. and Lavieri, P. S., A New Spatial and Flexible Multivariate Random-Coefficients Model for the Analysis of Pedestrian Injury Counts by Severity Level, *Anal. Methods Accident Res.,* 16, 1-22, Dec, 2017.

Bhati, A.S., *Modeling Count Outcomes with Spatial Structures: An Information Theoretic Approach*, Justice Policy Center, The Urban Institute, Washington DC., 2005.

Dong, N., Huang, H., Xu, P. and Wang, D., Evaluating Spatial-Proximity Structures in Crash Prediction Models at the Level of Traffic Analysis Zones, *Transport. Res. Rec.*, 2432(6), 46-52, Dec, 2014.

Dong, N., Huang, H. and Zheng, L., Support Vector Machine in Crash Prediction at the Level of Traffic Analysis Zones: Assessing the Spatial Proximity Effects, *Accid. Anal. Prev.,* 82, 192-198, Sep, 2015.

Dong, N., Huang, H., Lee, J., Gao, M. and Abdel-Aty, M., Macroscopic Hotspots identification: A Bayesian Spatio-Temporal Interaction Approach, *Accid. Anal. Prev.,* 92, 256-264, Jul, 2016.

Flask, T. and Schneider, IV. W., A Bayesian Analysis of Multi-Level Spatial Correlation in Single Vehicle Motorcycle Crashes in Ohio, *Saf. Sci*, 53(2), 1-10, Mar, 2013.

Fotheringham, A.S., Brunsdon, C. and Charlton, M., *Geographically Weighted Regression the analysis of spatially varying relationships,* John Wiley and Sons, 2003.

Hadayeghi, A., Shalaby, A. S. and Persaud, B. N., Development of Planning Level Transportation Safety Tools Using Geographically Weighted Poisson Regression, *Accid. Anal. Prev.,* 42(2), 676-688, Mar, 2010.

Huang, H. and Abdel-Aty, M., Multilevel Data and Bayesian Analysis in Traffic Safety, *Accid. Anal. Prev.,* 42(6), 1556-1565, Nov, 2010.

Huang, H., Abdel-Aty, M. and Darwiche, A. L., County-Level Crash Risk Analysis in Florida, *Transport. Res. Rec.,* 2148(4), 27-37, Dec, 2010.

LeSage, J.P., *Spatial Econometric,* Department of Economics, University of Toledo, 1998.

Li, Z., Wang, W., Liu, P., Bigham, J.M. and Ragland, D.R., Using Geographically Weighted Poisson Regression for County-Level Crash Modeling in California, *Saf. Sci.,* 58(10), 89-97, Oct, 2013.

Lord, D. and Mannering, F., The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives, *Transp. Res. Pt. A-Policy Pract.,* 44(5), 291-305, Jun, 2010.

Macnab, Y. C., Bayesian Spatial and Ecological Models for Small-Area Accident and Injury Analysis, *Accid. Anal. Prev.,* 36(6), 1019-1028, Nov, 2004.

Miaou, S. P., Song, J. J. and Mallick, B. K., Roadway Traffic Crash Mapping: A Space-Time Modeling Approach, *J. Transport. Stat.*, 6(1), 33-57, Nov, 2003.

Milton, J., Shankar, V. and Mannering, F.L., Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Aanalysis, *Accid. Anal. Prev.,* 40(1), 260-266, Jan, 2008.

Mitra, S., Spatial Autocorrelation and Bayesian Spatial Statistical Method for Analyzing Intersections Prone to Injury Crashes, *Transport. Res. Rec.,* 2136(11), 92-100, Dec, 2009.

Nakaya, T., Local Spatial Interaction Modelling Based on the Geographically Weighted Regression Approach, *Geo J.,* 53(4), 347-358, Apr, 2001.

Narayanamoorthy, S., Paleti, R. and Bhat, C. R., On Accommodating Spatial Dependence in Bicycle and Pedestrian Injury Counts by Severity Level, *Transp. Res. Pt. B-Methodol.,* 55(3), 245-264, Sep, 2013.

Noland, R. B., Traffic Fatalities and Injuries: The Effect of Changes in Infrastructure and Other Trends, *Accid. Anal. Prev.,* 35(4), 599-611, Jul, 2003.

Noland, R. B., Klein, N. J. and Tulach, N. K., Do Lower-Income Areas Have More Pedestrian Casualties?, *Transport. Res. Board 92nd Annual Meeting,* 59(10), 337-345, Jan, 2013.

Quddus, M. A., Modelling Area-Wide Count Outcomes with Spatial Correlation and Heterogeneity: An Analysis of London Crash Data, *Accid. Anal. Prev.,* 40(4), 1486-1497, Jul, 2008.

Siddiqui, C., Abdel-Aty, M. and Choi, K., Macroscopic Spatial Analysis of Pedestrian and Bicycle Crashes, *Accid. Anal. Prev.,* 45(3), 382-391, Mar, 2012.

Wang, C., Quddus, M., Ryley, T. and Lisa, D., Spatial Models in Transport: A Review and Assessment of Methodological Issues, *Transport. Res. Board Annual Meeting*, Jan, 2012.

Wang, X. and Abdel-Aty, M., Temporal and Spatial Analyses of Rear-End Crashes at Signalized Intersections, *Accid. Anal. Prev.,* 38(6), 1137-50, Nov, 2006.

Wang, J. and Huang, H., Road Network Safety Evaluation Using Bayesian Hierarchical Joint Model, *Accid. Anal. Prev.,* 90, 152-158, Mar, 2016.

Washington, S.P., Van, Schalkwyk, I., Mitra, S., Meyer, M., Dumbaugh, E. and Zoll, M., *Incorporating Safety into Long-Range Transportation Planning.* In: NCHRP Report 546, Transport. Res. Board, Washington, DC., 2006.

Wheeler, D. C. and Calder, C. A., An Assessment of Coefficient Accuracy in Linear Regression Models with Spatially Varying Coefficients, *J. Geogr. Syst.,* 9(2), 145-166, Feb, 2007.

Wu, Z., Sharma, A., Mannering, F.L. and Wang, S., Safety Impacts of Signal-Warning Flashers and Speed Control at High-Speed Signalized Intersections, *Accid. Anal. Prev.,* 54(5), 90-98, May, 2013.

Xu, P. and Huang, H., Modeling Crash Apatial Heterogeneity: Random Parameter versus Geographically Weighting, *Accid. Anal. Prev.,* 75, 16-25, Feb, 2015.

Xu, P., Huang, H., Dong, N., and Wong, S. C., Revisiting Crash Spatial Heterogeneity: A Bayesian Spatially Varying Coefficients Approach, *Accid. Anal. Prev.,* 98, 330-337, Jan, 2017.

Zeng, Q., Huang, H., Pei, X. and Wong, S.C., Modeling Nonlinear Relationship Between Crash Frequency by Severity and Contributing Factors by Neural Networks, *Anal. Methods Accident Res.,* 10, 12-25, Jun, 2016.

Lord, D. and Persaud, B., Accident Prediction Models with and without Trend: Application of the Generalized Estimating Equations (GEE) Procedure, *Transport. Res. Rec,* 1717(1), 102-108, Jan, 2000.

Wang, C., Quddus, M. and Ison, S., The Effects of Area-Wide Road Speed and Curvature on Traffic Casualties in England, *J. Transp. Geogr.,* 17(5), 385-395, Jun, 2009.