The Influence of Zonal Configurations on Macro-level Crash Modeling

Xiaoqi ZHAI[a], Helai HUANG[a]*, Pengpeng XU[b], N.N.SZE[c]

[a]*School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan, China*

[b] *Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China*

[c] *Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong, China*

**Correspondence: huanghelai@csu.edu.cn*

Word count 7414 words + 2 figures + 4 tables

The Influence of Zonal Configurations on Macro-Level Crash Modeling

Abstract: Traffic safety has increasingly become an important concern in developing long-term transportation planning strategies. Since transportation planning steps always involve some kinds of geographic entity, predicting crashes for those entities is not only a mere avenue of analytic methods in safety research, but also influential to practical application in road infrastructure design and management. However, the selection of the geographic entity may be subjected to the modifiable areal unit problem (MAUP), which refers to the issue of inconsistent statistical results when dealing with geographic data of different aggregation configurations. This study investigated the impacts of zonal configurations on macro-level traffic safety analysis for crashes of different severity levels. Bayesian multivariate Poisson-lognormal models with multivariate conditional auto-regressive priors were developed to account for the spatial autocorrelation between adjacent geographical units and correlations among crash types of four ordinal severity levels, i.e. fatality, severe injury, slight injury and no injury. For the purpose of evaluating the effects of zonal configurations on macro-level traffic safety analysis, the proposed model was calibrated using crash data of four types of geographical units, i.e. block group (BG), traffic analysis zone (TAZ), census tract (CT) and ZIP code tabulation area (ZCTA), in Hillsborough County of Florida. Results indicated that there were remarkable variations, associated with zoning configuration, in parameter estimation in the terms of sign and magnitude of coefficients and statistical significance. Results of BG and CT based models were consistent in the term of parameter estimation due to both BG and CT were defined for processing the census data. Also, the spatial entities with smaller unit sizes tended to be highly correlated. The results of identification of high-crash locations also had significant difference for different zonal configurations. The study empirically revealed the extensive presence and the significance of MAUP in macro-level safety analysis based on the existing zonal configurations.

Keywords: Modifiable Areal Unit Problem, Macro-level traffic safety analysis, Geographical configuration, Crash severity

1. Introduction

Safety has increasingly become an important concern in developing long-term transportation planning strategies (Washington, 2006). Since transportation planning steps always involve some kinds of geographic entity, predicting safety levels for those entities is not only an avenue of safety analysis, but also a demanded application in the practice of transportation planning. For this purpose, several types of zonal configurations have been extensively examined, including traffic analysis zone (TAZ) (Abdel-Aty et al., 2011; Dong et al., 2014; Dong et al., 2015; Pulugurtha et al., 2013; Siddiqui et al., 2012; Xu and Huang, 2015, Xu et al., 2017; Lee et al., 2018; Guo et al., 2018), traffic analysis districts (TADs) (Cai et al., 2017), ZIP code tabulation area (ZCTA) (Lee et al., 2014; Lee and Abdel-Aty, 2017), census tract (CT) (Ukkusuri et al., 2011; Ukkusuri et al., 2012; Wang and Kockelman, 2013), block group (BG) (Levine et al., 1995), English wards (Quddus, 2008), and enumeration districts (Noland and Quddus, 2005).

In applying different zoning schemes, the macro-level safety evaluations have been conducted for a variety of zonal factors in association with socioeconomic and demographic status, transportation network, road facilities and traffic flow condition. The results could be significantly beneficial for regional safety inspection, countermeasures development and safety-oriented planning decision making. However, significant variations in effect of zonal factors have been found in analytic context between different zonal configurations. The problem is clearly in need of effort to address, and referred as modifiable areal unit problem (MAUP) in geographic science.

MAUP refers to the issue of inconsistent statistical results when dealing with geographic data with different aggregation configurations (Fotheringham and Wong, 1991; Openshaw, 1984). Typically, it could be decomposed into two components: the scale and the zoning effects (Openshaw, 1984; Fotheringham and Wong, 1991). While the scale effect describes the occurrence of statistical result variations using data aggregated at different levels, the zoning effect refers to the variability introduced by different zoning configurations at the same aggregate level. Although the MAUP has been addressed by geographers in various topics for decades (since Gehlke and Biehl, 1934), it has received little attention in the traffic safety domain. Xu et al. (2014) conducted a sensitivity analysis based on specific aggregation configurations of TAZs to investigate the MAUP effect in macro-level safety modeling and highlighted the general neglect of MAUP in transportation safety analysis. To compensate for the MAUP effects in traffic safety analysis, Lee et al. (2014) created optimal zone systems for macroscopic safety modeling. Current TAZs with homogenous crash rates were combined into new single zones. It was found that the zone system with about 1:2 aggregation have better model performance and can minimize boundary crashes. Moreover, several recent studies, though not directly tackling the MAUP, have given due consideration to different zonal configurations comparison in terms of zonal safety estimates. Abdel-Aty et al. (2013) investigated the effects of zonal variations (i.e. TAZ, BG and CT) on the prediction performance for three different dependent variables, namely overall crash, severe injury crash and pedestrian crash, using the Bayesian Poisson-lognormal model. Results indicated that the TAZ-based models revealed more roadway/traffic related significant factors while BG-based models generated more commute related variables respectively. Therefore, for macroscopic safety analysis, it is worthwhile though challenging to have a deeper understanding on the effects of zonal configurations in terms of the changes in the patterns of parameter estimation.

In a more recent study, Amoh-Gyimah et al. (2017) examined the effect of six different spatial aggregation units on unobserved heterogeneity. This study applied random parameter models to address the unobserved heterogeneity by allowing the parameter to vary across observations. Results revealed that the variation in spatial units had an impact on unobserved heterogeneity. Unobserved heterogeneity refers to the complex unobserved effects due to the limited data. The omitted variables might interact with the observed ones and result in biased parameter estimates and incorrect inferences. As reviewed by Mannering et al. (2016), various statistical methods have been developed to address the issue of unobserved heterogeneity such as random effect (Ma et al., 2017), random parameter (Xu & Huang, 2015), latent class approaches (Yu et al., 2017), as well as multivariate models (Huang et al., 2017) and spatial correlation models (including the geographically weighted Poisson regression models (GWPR; Fotheringham et al., 2002; Xu & Huang, 2015). Most recent studies of macro-level safety analysis have confirmed that both correlation across different crash severity levels and spatial autocorrelation should be taken into account in estimating the unobserved effects. First, there may be extensive correlation between crash counts of different severity levels due to heterogeneous effects of unobserved factors (El-Basyouny et al., 2014; Huang et al. 2017; Bhat et al. 2017). The contributing factors might interact with crash of each severity level and generate distinct effects. Second, spatial dependency of zonal crash count had been explicitly identified (Huang and Abdel-Aty, 2010; Quddus, 2008; Siddiqui et al., 2012; Xu et al., 2014; Huang et al., 2016; Saha et al., 2018). Common unobserved location factors might result in spatial correlation effects in the error terms of the injury-risk propensity at proximally located crash locations (Mannering et al., 2014). Multivariate Bayesian spatial model has been the state of the art in modeling both types of correlations to proxy the unobserved heterogeneity (Aguero-Valverde 2013).

This study quantified the impacts of zonal configurations on macro-level crash prediction based on four conventional zonal configurations including BG, TAZ, CT, and ZCTA using an identical dataset. Bayesian multivariate Poison-lognormal models with multivariate conditional auto-regressive (CAR) priors were developed. To demonstrate the impacts of spatial aggregation on crash prediction performance, results of identification of high-crash locations under the four aforementioned zonal configurations were compared. This paper intended to reveal the extensive existence of variations in parameter estimation from both geographical and transportation-related aspects, and prudential application of macro-level safety analytic methods was recommended.

2. Data

This study employed the data of Hillsborough County of Florida, United States. The descriptive statistics of the collected data, with respect to four different zonal configurations, were summarized in Table 1.

GIS maps for each of the four zonal configuration methods were extracted from U.S. Census Bureau website. Of the four zonal configurations under investigation, BG represents the smallest geographic entity, with median area of 239.5 acre (Range: 7.3-5018.2 acre). Overall, there are 795 BGs in Hillsborough County. TAZ is essentially spatial aggregation of more than one census block, and partially a function of population (Peters and MacDonald, 2009). TAZ has higher level of homogeneity than census block because TAZ is specially delineated by the State government and concerned bureaus for land use and transportation planning purposes (Census, 1994). In Hillsborough county, there are 738 TAZs in total (Median area: 306.6 acre, Range: 3.9 – 22605.1 acre). CT is a relatively larger spatial unit. There are 249 CTs (Median area: 1086.3 acre, Range: 104.4- 57096.0 acre) in the study area. CTs are delineated to achieve homogeneous demographics, socio-economic and housing development conditions. ZCTA entails both residential and non-residential land uses and is the largest spatial unit among the four candidate geographic units. Hillsborough county has 52 ZCTAs (Median area: 6761.5 acre, Range: 579.3 – 103816.1 acre).

Crash data in Hillsborough County during the period 2005-2007 was extracted from the Florida Department of Transportation (FDOT) Crash Analysis Reporting System. Crash counts were stratified by injury severity level. In particular, there were five injury severity categories: Property Damage Only (No injury), Possible injury (Slight injury), Non-Incapacitating evident injury, Incapacitating injury, and Fatal injury. To simplify the analysis, Non-Incapacitating evident injury and Incapacitating injury were combined into one class, defined as "Severe injury" in this study.

A number of exploratory variables were collected from FDOT's Roadway Characteristics Inventory. The exploratory variables were geographically aggregated into the four zonal configurations under investigation by the PLANSAFE GIS tool and PLANSAFE Census Tool. The PLANSAFE GIS tool allows the user to aggregate point and line features by the given polygon units. We use this tool to aggregate the crash, intersection and roadway length/ polygon, DVMT/polygon at different levels. The PLANSAFE Census Tool was used to collect and aggregate census information by the given polygon units. In particular, Daily-Vehicle-Miles-Traveled (DVMT) was used to proxy the crash exposure. Variables related to road network include proportions of road segment with speed limits of 25 mph, 35 mph, 45 mph and 55 mph or higher respectively, and intersection density. Demographics and socioeconomic variables in consideration consisted of population density, proportion of female population, proportion of population of age under 15, proportion of population of age 15 to 64, proportion of population of age above 64, and median annual household income.

Prior to the establishment of crash prediction models, the multi-collinearity test was employed. Favorably, all the seven explanatory variables in consideration had low values of variance inflation factor (TAZ: 2.313; BG: 2.838; CT: 3.199 and ZCTA: 4.405, respectively) and were therefore retained in the models.

3. Methodology

3.1 Model development

Poisson lognormal model is comparable with negative binomial model while provides more flexibility, thus has been suggested as a substitute for the crash count data (Lord and Mannering, 2010). To account for the effects of spatial autocorrelation and correlations among crash counts of different severity levels, multivariate Poisson lognormal model with multivariate conditional auto-regressive prior (MVPLN-MCAR model) were applied. The models can be efficiently estimated in the freeware WinBUGS package (Lunn et al., 2000).

Y_{ik} denoted the observed number of crash of areal unit i of severity level k, where k = 1, 2, 3 and 4 denotes No injury, Slight injury, Severe injury, and Fatal respectively. λ_{ik} was the Poisson parameter. X_{ikl} was the l th explanatory variable for areal unit i of severity level k, and E_{vik} was the DVMT of areal unit i. The MVPLN-MCAR model (Aguero-Valverde, 2013) could be specified as:

$$Y_{ik} : \ Poisson(\lambda_{ik}) \tag{1}$$

$$\ln(\lambda_{ik}) = \beta_{0k} + \beta_{1k}\ln(E_{vik}) + \beta_{ikl}\sum_{l=2}^{L} X_{ikl} + \theta_{ik} + \varphi_{ik} \tag{2}$$

where, β_{0k} was the intercept for severity level k, β_{lk} was the coefficient for l th covariate of severity k, and θ_{ik} was assumed to be multivariate normal prior distributed as $\theta_i : \ N(0,\Sigma)$, where

$$\theta_i = \begin{pmatrix} \theta_{i1} \\ \theta_{i2} \\ \theta_{i3} \\ \theta_{i4} \end{pmatrix}, \sum = \begin{pmatrix} \sigma_{11}\sigma_{12}\sigma_{13}\sigma_{14} \\ \sigma_{21}\sigma_{22}\sigma_{23}\sigma_{24} \\ \sigma_{31}\sigma_{32}\sigma_{33}\sigma_{34} \\ \sigma_{41}\sigma_{42}\sigma_{43}\sigma_{44} \end{pmatrix} \tag{3}$$

The diagonal element σ_{11} of the covariance matrix represented the variance of θ_{i1} , and the off-diagonal element σ_{12} represented the covariance of θ_{i1} and θ_{i2}, so far and so on. Besides, $\sum$ was the variance-covariance matrix for heterogeneous effects with a hyper-prior defined by $\Sigma^{-1} : \ Wishart(R,k)$, where R was the $k \times k$ identity matrix (Aguero-Valverde, 2013).

Spatial autocorrelation between adjacent zones was realized by specifying a conditional autoregressive prior model (CAR model) to the residual term of the link function in an ordinary Poisson regression (Banerjee et al., 2014).

$$\phi_i \sim N(\bar{\phi}_i, \frac{1}{\tau_i}) \text{ with } \bar{\phi}_i = \frac{\sum_{i \neq j} \phi_j \omega_{ij}}{\sum_{i \neq j} \omega_{ij}} \text{ and } \tau_i = \frac{\tau_c}{\sum_{i \neq j} \omega_{ij}} \quad (4)$$

in which ω_{ij} was binary entries of proximity matrix as described above. And if i and j were adjacent (i.e., a shared border), ω_{ij} = 1, otherwise, ω_{ij} = 0. τ_c was the precision parameter in the CAR prior.

For the multivariate version of the spatial residual term, the MCAR (Huang et al., 2017) was used as the following expression:

$$\phi_{ik} \mid (\phi_{i1}, \ldots, \phi_{ik}) \sim MN\left(\overline{\phi_{ik}}, \frac{\Omega}{n_i}\right) \quad (5)$$

where $(\phi_{i1}, \ldots, \phi_{ik})$ denoted the intersections of the $k \times n$ matrix ϕ_{ik}, excluding the ith unit. Ω was the variance-covariance matrix for spatial heterogeneity. The diagonal element of the variance-covariance matrix represented the spatial variance. The off-diagonal elements represented the spatial covariance between different crash severities. A non-informative Wishart distribution was used as a hyperprior for the precision matrix, as defined above.

3.2 Performance assessment

Three common measures, Mean Absolute Deviance (MAD), Mean Squared Predictive Error (MSPE) and Sum of Absolute Deviance (SAD) were used to assess the goodness of fit of models. The MAD, MSPE and SAD were expressed as:

$$\text{MAD} = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{Y}_{ik} - Y_{ik} \right| \quad (6)$$

$$\text{MSPE} = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{Y}_{ik} - Y_{ik} \right)^2 \quad (7)$$

$$\text{SAD} = \sum_{i=1}^{n} \left| \hat{Y}_{ik} - Y_{ik} \right| \tag{8}$$

where n = number of observations, and Y_{ik} and $\hat{Y}_{ik}$ are observed and predicted values for i th zone and k th severity, respectively.

Model having lower value of MAD, MSPE and SAD implies a better fit.

3.3 Standard difference-in-means test

To evaluate the sensitivity of parameter estimate to the variation of the level of spatial aggregation, Difference-in-mean of parameter estimates were calculated (Fotheringham and Wong, 1991). The standard Difference-in-Mean test is a useful tool to evaluate the consistency of parameter estimates across diverse levels of geographical aggregation and the formulation is given as follows,

$$t = \frac{\hat{\beta}_{lkj_1} - \hat{\beta}_{lkj_2}}{SE\left| \hat{\beta}_{lkj_1} - \hat{\beta}_{lkj_2} \right|} \tag{9}$$

where $\hat{\beta}_{lkj_1}$ and $\hat{\beta}_{lkj_2}$ were estimated coefficients of independent variable l in severity k with zonal configurations j_1 and j_2 respectively. SE was the standard error.

3.4 Identification of high-crash locations

One potential application of zonal CPMs is to allow identification of areas with promise to support decision making and long term transportation planning. The potential for safety improvement (PSI) or excess crash frequency is a general approach to identify the high-crash locations, which is defined as the difference between crash frequency expectation at the zone and the expectation at the similar zones (Huang et al., 2009; Aguero-Valverde and Jovanis, 2010; Aguero-Valeverde, 2013; Lee, 2014; Lee et al., 2015). The measure could effectively identify whether a zone suffers more number of crashes compared with other zones with similar characteristics. When the potential for improvement value is greater than zero, a site experiences more crashes than expected. When the potential for improvement value is less than zero, a site experiences fewer crashes than expected. To identify the high-crash locations, PDO equivalency factor f was computed (Washington et al., 2014). Equivalent Property Damage Only (EPDO) crash count is the weighted sum of no injury, slight injury, severe injury, and fatal crashes of an entity. For instances, one fatal injury crash is equivalent to 1330 PDO crashes, one severe injury crash is equivalent to 949 PDO crashes, and one slight injury crash is equivalent to 11 PDO crashes respectively (Washington et al. 2014, adapted from Blincoe et al., 2002).

$$P_{ik} = \lambda_{ik} - \exp(\beta_{0k} + \beta_{1k}\ln(\mathrm{DVMT}_i) + \sum_{l=2}^{L} X_{ilk}\beta_{lk}) \tag{10}$$

$$P_i = \sum_{k=1}^{4} f \times P_{ik} \tag{11}$$

where P_i was the Potential for Safety Improvement for unit i.

4. Results and discussions

Four separated MPLVN-MCAR prior models were estimated for the four concerned zonal configurations respectively. Impacts of the variations in zonal configuration on the macro-level crash predictions were evaluated from three different perspectives, namely model performance, parameter estimate, and identification of high-crash locations.

4.1 Model performance comparison

The MAD and MSPE were used to compare the prediction precision, while the SAD was used to compare models performance from different zone systems. MAD and MSPE are widely used to evaluate the prediction precision because they are not affected by sample size. As shown in Tables 4, MAD and MSPE are smallest in BG based models compare to those in TAZ, CT and

ZCTA based models, which indicate the BG based model has the highest prediction precision. However in our study, we are more interested in comparing prediction model performance based on different zonal schemes with different zone numbers. Having an appropriate zone size is also a crucial criterion to evaluate whether the zonal system is appropriate for traffic safety analysis. Thus, SAD was used to compare models from different zone systems. As shown in Tables 4, values of SAD were lower for the models with larger zones (i.e. CT and ZCTA based models). It implied that the prediction performances of the models with larger zones were superior. It was consistent to the findings of previous studies (Fotheringham and Wong, 1991 ; Zhang and Kukadia, 2005; Wang et al., 2012; Xu et al., 2014; Lee et al., 2017). For TAZ and BG based models in which their zone sizes were comparable, prediction performance of TAZ-based model was superior to that of BG-based model, in particular for No injury, Slight injury and Fatal crashes. It could be due to the demographics and socioeconomics data were collected based on BG. Bias might have been induced when the census data was aggregated in another way.

Table 2 presented the results of multivariate correlations of crash counts across different severity levels. Results indicated that correlations of crash counts across different severity levels were high in general (Range: 0.419-0.977). It was consistent to the findings of previous study (Aguero-Valverde and Jovanis 2009). Also, the crash counts of consecutive severity levels were highly correlated (for example, 0.977 for correlation between no injury and slight injury crashes for BG-based model). This could be attributed to the effects of unobserved common factors on crash occurrence. For example, land use, trip generation and geometric design can simultaneously affect the crash likelihood regardless of severity level. This justified the use of multivariate models, as it could improve the model efficiency and avoid potential biases in parameter estimates, especially for TAZ and BG based models that had higher correlation between crashes of different severity levels (BG and TAZ: 0.908-0.946, CT: 0.419-0.449).

As shown in Table 3, there were high spatial correlations of crash count of same severity leve since the crashes were often in close spatial proximity. It was consistent to the findings of previous studies (Quddus, 2008; Huang et al., 2010; Siddiqui et al., 2012 and Xu et al. 2014). These unobserved factors such as the weather and climate effects are correlated over space, and ignoring the spatial correlation of data will inevitably result in inefficient and possibly inconsistent parameter estimates. Besides, crashes of different severity levels could also be subject to remarkable spatial correlation, especially for crashes of consecutive severity levels. For example, in TAZ based model, the results indicated that the slight injury crash at specific unit had high spatial correlation (0.917) with the no injury crash in adjacent units. This finding demonstrated the multivariate spatial model had satisfactory performance in accurately accounting for spatial autocorrelation effects and unbiased parameter estimates, supporting its widely usage in existing safety analysis (El-Basyouny et al., 2014; El-Basyouny and Sayed, 2009; Lee et al., 2015; Ma et al., 2008; Wang and Kockelman, 2013). It was also observed that the spatial correlation was lower for models with larger zone size. Spatial correlation was the strongest for BG-based model, and the weakest for ZCTA-based model respectively. If we denote spatial correlation of 0.6 or higher as strong, the numbers of pair of crash count with strong spatial correlation were BG, 8; TAZ, 5; CT, 6; ZCTA, 4, respectively.

Using 200 ft as a buffer for zone boundary (Ivan et al., 2006), proportion of boundary crash was computed for each zonal configuration. As shown in Table 1, proportion of boundary crash was the highest for TAZ-based model (82.30%), and the lowest for ZCTA-based model (22.68%). Unsurprisingly, TAZ generated longer boundary lines due to smaller zone size, thus larger total buffer area and higher likelihood for crash to be located on the buffer zones. Bias in parameter estimate could be attributed to the prevalence of boundary crash and thus the zonal configuration.

4.2 Parameters estimates variation across spatial units

Table 4 presented the results of MVPLN-MCAR models for the four zonal configurations. As shown in Table 4, there were obvious variations in parameter estimates for different zonal configurations across crashes of different severity levels, in the terms of sign, magnitude and statistical significance of parameter estimates. It confirmed the extensive presence and significance of MAUP in safety analysis.

DVMT was the only common statistically significant variable for all models regardless of crash severity level and zonal configuration. DVMT was found positively correlated to crash frequency. The increment in DVMT indicated more trip production, which enlarged the crash likelihood in a great extent.

Intersections have been recognized as hazardous locations on roads for a long time (Wang et al., 2009). As shown in Table 4, crash rates tended to increase with the increment of intersection density regardless of crash severity, in BG and CT based models. However, no significant association between intersection density and crash occurrence could be established for ZCTA-based model. The safety effect of intersection density, which was relatively localized, might be diluted when zone size increased. Results also showed that the effects of intersection density diminished for the frequencies of crash with higher severity level. It could be attributed to the lower vehicular speed and higher level of drivers' alertness near the intersections, and therefore, the lower collision impacts (Ladron de Guevara et al, 2004).

Appropriate speed limit should be set in accordance to the land use, landscape, road class, traffic pattern and their interactions (Xu and Huang, 2015). As shown in Table 4, except for fatal crash in CT based model, higher proportion of road segment with speed limit of 25 mph was associated with lower crash frequencies, regardless of crash severity, in BG and CT based models. Higher proportion of road segment with speed limit of 45 mph significantly associated with the increases in crash frequencies in TAZ models, and higher proportion of road segment with speed limit of 55 mph or above significantly associated with the reduction of crash frequencies in ZCTA models.

Crash frequencies were found significantly increased with population density (Huang and Abdel-Aty, 2010). Results of parameters estimation for population density were not consistent across the four concerned models. Higher population density significantly associated with the increment in the prevalence of No injury, Slight injury and Severe injury crashes in TAZ-based model, and No injury and Slight injury crashes in ZCTA-based models. However, population density was found significantly associated with the reduction in the crash frequencies regardless of severity level, for BG and CT based models. The favorable safety effect of population density could be attributed to higher expenditure on highway infrastructure improvement in the areas with higher level of residential and economic developments.

Median house income was found significantly associated with the reduction in crash frequencies, regardless of zonal configuration and crash severity. Such finding was consistent to that of previous research (Xu and Huang, 2015; Xu et al., 2014). It was because wealthier areas could afford to develop and implement more effective risk mitigation and avoidance measures, and thus resulted in favorable safety effects. Besides, higher income households could afford to buy more expensive vehicles, which were equipped with more advanced safety features.

Fig. 1 presented the results of different-in-mean test for each individual risk factor, with respect to crash severity and zonal configuration. It was obvious that the results of parameter estimates fluctuated across zonal configurations for all variables, except median annual household income. However, results of parameter estimates were generally consistent across crash severity.

4.3 Identification of high-crash locations

Fig. 2 presented the results of identification of high-crash locations, which were based on the estimates of Potential for Safety Improvement. The results of identification of high-crash locations portrayed a quite perplexing picture for different zonal configurations. For ZCTA-based model, the high-crash locations appeared to concentrate in central zones. For other three zonal configurations, the high-crash locations were quite dispersed. Besides, discrepancies in the results of identification of high-crash locations were substantial between models of different zonal configurations. For example, an area, as circled in the Figure 2, that was recognized as high-crash locations for CT-based model was identified as 'safe' for TAZ, BG and ZCTA models. It could be because of the effects of information scarcity for the models with larger zone size. Again, the discrepancies could impact on the design and implementation of cost-effective crash reduction plan, e.g. targeted treatments for problematic sites, corridors or areas.

5. Conclusions

This study evaluated the effects of zonal configuration on the performance of macro-level crash prediction. Bayesian multivariate Poisson-lognormal models with MCAR priors were established based on the historical crash data in Hillsborough County, Florida. First, we systematically revealed the extensive presence and the significance of MAUP in macro-level safety analysis from three aspects: parameter estimates, model assessment and high-crash locations identification. Results indicated that there were remarkable variations in model estimates, in the terms of sign, magnitude and statistical significance of parameter, between different zonal configurations. Zonal configurations with larger zone size, e.g. CT and ZCTA, had superior predictive performance. The results of identification of high-crash locations portrayed a quite perplexing picture for different zonal configurations. Together, these results have important implications for how traffic safety could be viewed in different zonal schemes. Second, we examine the spatial correlation and correlations across different severities for different zonal systems. There were high correlations between crashes of different severity level for zonal configurations with smaller zone size. Third, we brought up the possible causes of the MAUP are as follows: (1) Safety effect estimation of risk factors heavily depended on the zonal configuration; (2) Prevalence of boundary crashes may induce bias to model estimation; (3) There were possible interactions for spatial correlation and correlation between crashes of different severity levels. It is shown that the MAUP has serious implications on the spatial and statistical representations of transportation safety. Policy initiatives must determine if the MAUP is an important problem that needs to be taken into account before statistics grouped into areas, can be reliably reported.

In the future research, it is worth exploring the effect of MAUP on the transferability of the results of parameter estimation among different zonal configurations. Optimal zone size and spatial data resolution for the best crash prediction performance should be established for different purposes of risk mitigation. However, several enhancements may be pursued. Only limited variables were used in this study, the land use, trip generation and road network information should be further explored. Besides, it is necessary to explore the potential of the random parameter approach to model the multivariate outcomes, in particular taking into account the unobserved heterogeneity.

Acknowledgements

This work was jointly supported by the Joint Research Scheme of National Natural Science Foundation of China/Research Grants Council of Hong Kong (Project No. 71561167001 and N_HKU707/15), the Natural Science Foundation of China (Project No. 71371192), and the Research Committee of the Hong Kong Polytechnic University (Project No. 1-ZE5V). We would like to thank Dr. Mohamed Abdel-Aty at the University of Central Florida and the Florida Department of Transportation for providing the data.

References

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.

Abdel-Aty, M., Siddiqui, C., Huang, H., Wang, X., 2011. Integrating trip and roadway characteristics to manage safety in traffic analysis zones. Transportation Research Record: Journal of the Transportation Research Board, 20-28.

Aguero-Valverde, J., 2013. Multivariate spatial models of excess crash frequency at area level: Case of Costa Rica. Accident Analysis and Prevention 59, 365-373.

Amoh-Gyimah, Richard., Saberi, Meead., Sarvi, Majid., 2017. The effect of variations in spatial units on unobserved heterogeneity in macroscopic crash models. Analytic Methods in Accident Research, 13, 28-51.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2014. Hierarchical modeling and analysis for spatial data. Crc Press.

Bhat, C. R., Astroza, S., and Lavieri, P. S., 2017. A new spatial and flexible multivariate random-coefficients model for the analysis of pedestrian injury counts by severity level. Analytic Methods in Accident Research, 16, 1-22.

Blincoe, L., Seay, A., Zaloshnja, E., Miller, T., Romano, E., Luchter, S., Spicer, R., 2002. The economic impact of motor vehicle crashes, 2000. Citeseer.

Cai, Q., Abdel-Aty, M., & Lee, J., 2017. Macro-level vulnerable road users crash analysis: a Bayesian joint modeling approach of frequency and proportion. Accident Analysis & Prevention, 107, 11-19.

Census, U.S.B.o.t., 1994. Geographic areas reference manual. US Dept. of Commerce, Economics and Statistics Administration, Bureau of the Census.

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, 46-52.

Dong, N., Huang, H., Zheng, L., 2015. Support vector machine in crash prediction at the level of traffic analysis zones: assessing the spatial proximity effects. Accident Analysis and Prevention 82, 192-198.

El-Basyouny, K., Barua, S., Islam, M.T., 2014. Investigation of time and weather effects on crash types using full Bayesian multivariate Poisson lognormal models. Accident Analysis and Prevention 73, 91-99.

El-Basyouny, K., Sayed, T., 2009. Collision prediction models using multivariate Poisson-lognormal regression. Accident Analysis and Prevention 41, 820-828.

Fotheringham, A.S., Wong, D.W., 1991. The modifiable areal unit problem in multivariate statistical analysis. Environment and planning A 23, 1025-1044.

Fotheringham, A.S., Brunsdon, C., Charlton, M.E., 2002. Geographically WeightedRegression: The Analysis of Spatially Varying Relationship. Wiley, Chichester.

Gehlke, C.E., Biehl, K., 1934. Certain effects of grouping upon the size of the correlation coefficient in census tract material. Journal of the American Statistical Association 29, 169-170.

Guo, Y., Osama, A., & Sayed, T., 2018. A cross-comparison of different techniques for modeling macro-level cyclist crashes. Accident analysis and prevention, 113, 38-46.

Huang, H., Abdel-Aty, M., 2010. Multilevel data and Bayesian analysis in traffic safety. Accident Analysis and Prevention 42, 1556-1565.

Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., & Abdel-Aty, M., 2016. Macro and micro models for zonal crash prediction with application in hot zones identification. Journal of Transport Geography, 54(54), 248-256.

Huang H., Zhou H., Wang J., Chang F., 2017. A multivariate spatial model of crash frequency by transportation modes for urban intersections, Analytic Methods in Accident Research, 14,10-21.

Ivan, J.N., Deng, Z., Jonsson, T., 2006. Procedure for allocating zonal attributes to link network in gis environment, Transportation Research Board 85th Annual Meeting.

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, 191-199.

Lee, J., Abdel-Aty, M., Choi, K., 2014. Analysis of residence characteristics of at-fault drivers in traffic crashes. Safety science 68, 6-13.

Lee, J., Abdel-Aty, M., & Jiang, X., 2014. Development of zone system for macro-level traffic safety analysis. Journal of Transport Geography, 38, 13-21.

Lee, J., Abdel-Aty, M., Jiang, X., 2015. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. Accident Analysis and Prevention 78, 146-154.

Lee, J., & Abdel-Aty, M., 2017. Macro-level analysis of bicycle safety: Focusing on the characteristics of both crash location and residence. International Journal of Sustainable Transportation, 1-8.

Lee, J., Abdel-Aty, M., Cai Q., 2017. Intersection crash prediction modeling with macro-level data from various geographic units. Accident analysis and prevention 102, 213-226.

Lee, J., Yasmin, S., Eluru, N., Abdel-Aty, M., & Cai, Q., 2018. Analysis of crash proportion by vehicle type at traffic analysis zone level: a mixed fractional split multinomial logit modeling approach with spatial effects. Accident Analysis & Prevention, 111, 12-22.

Levine, N., Kim, K.E., Nitz, L.H., 1995. Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. Accident Analysis and Prevention 27, 675-685.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice 44, 291-305.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and computing 10, 325-337.

Ma, J., Kockelman, K.M., Damien, P., 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accident Analysis and Prevention 40, 964-975.

Mannering, F. L., and Bhat, C. R., 2014. Analytic methods in accident research: methodological frontier and future directions. Analytic Methods in Accident Research, 1, 1-22.

Mannering, F. L., Shankar, V., and Bhat, C. R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research, 11, 1-16.

Ma, Z., Zhang, H., Chien, S. I., Wang, J., & Dong, C., 2017. Predicting expressway crash frequency using a random effect negative binomial model: a case study in china. Accident Analysis & Prevention, 98, 214-222.

Noland, R.B., Quddus, M.A., 2005. Congestion and safety: A spatial analysis of London. Transportation Research Part A: Policy and Practice 39, 737-754.

Openshaw, S., Openshaw, S., 1984. The modifiable areal unit problem. Geo Abstracts University of East Anglia.

Peters, A., MacDonald, H., 2009. Unlocking the census with GIS. Esri Press.

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

Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accident Analysis and Prevention 40, 1486-1497.

Saha, D., Alluri, P., Gan, A., & Wu, W., 2018. Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models. Accident Analysis & Prevention.

Sheppard, E., McMaster, R.B., 2008. Scale and geographic inquiry: Nature, society, and method. John Wiley and Sons.

Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accident Analysis and Prevention 45, 382-391.

Ukkusuri, S., Hasan, S., Aziz, H., 2011. Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency. Transportation Research Record: Journal of the Transportation Research Board, 98-106.

Ukkusuri, S., Miranda-Moreno, L.F., Ramadurai, G., Isa-Tavarez, J., 2012. The role of built environment on pedestrian crash frequency. Safety science 50, 1141-1151.

Wang, C., Quddus, M.A., Ison, S.G., 2009. Impact of traffic congestion on road accidents: a spatial analysis of the M25 motorway in England. Accident Analysis and Prevention 41, 798-808.

Wang, C., Quddus, M.A., Ryley, T., Enoch, M., Davison, L., 2012. Spatial models in transport: a review and assessment of methodological issues, Proceedings of the 91st Annual Meeting of the Transportation Research Board, Washington, DC, USA.

Wang, Y., Kockelman, K.M., 2013. A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. Accident Analysis and Prevention 60, 71-84.

Washington, S., 2006. Incorporating safety into long-range transportation planning. Transportation Research Board.

Washington, S., Haque, M.M., Oh, J., Lee, D., 2014. Applying quantile regression for modeling equivalent property damage only crashes to identify accident blackspots. Accident Analysis and Prevention 66, 136-146.

Xu, P., Huang, H., Dong, N., Abdel-Aty, M., 2014. Sensitivity analysis in the context of regional safety modeling: Identifying and assessing the modifiable areal unit problem. Accident Analysis and Prevention 70, 110-120.

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

Xu P., Huang H., Dong N., Wong SC., 2017. Revisiting crash spatial heterogeneity: a Bayesian spatially varying coefficients approach. Accident Analysis and Prevention 98, 330-337.

Yu, R., Wang, X., & Abdel-Aty, M., 2017. A hybrid latent class analysis modeling approach to analyze urban expressway crash risk. Accid Anal Prev, 101, 37-43.

Zhang, M., Kukadia, N., 2005. Metrics of urban form and the modifiable areal unit problem. Transportation Research Record: Journal of the Transportation Research Board, 71-79.

Table 1 Summary of variables and descriptive statistics.

	Block group				Traffic analysis zone				Census tract				ZIP code tabulation area			
Variables	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Dependent variables																
Fatal crash	0.69	1.31	0	10	0.76	1.25	0	11	8.06	2.30	6	17	10.73	8.59	0	37
Sever injury crash	7.59	9.46	0	72	8.29	8.77	0	58	24.22	19.86	0	101	116.42	89.72	0	352
Slight injury crash	36.13	43.63	0	369	39.54	36.23	0	215	115.34	94.88	3	505	555.13	492.47	1	1846
No injury crash	59.03	80.50	0	632	64.53	65.97	0	426	188.43	175.96	4	1036	906.15	813.70	0	2992
Boundary crash (% of total crash)	66326 (76.13%)				69955 (82.3%)				55381 (63.57%)				19759 (22.68%)			
Road Environment/Traffic variables																
DVMT (Daily vehicle miles traveled*10^4)	9.28	12.78	0.03	94.79	9.58	11.02	0.01	78.88	29.41	31.07	0.04	220.83	185.15	115.43	0.60	442.75
Intersection density (Num of interception per mile)	1.81	1.81	0.00	15.89	3.17	5.61	1.00	66.12	1.56	1.32	0.00	8.73	3.44	4.71	0.00	25.22
P_seglen25(%of segment length with 25 mph PSL)	0.79	0.15	0.01	1.00	0.72	0.21	0.00	1.00	0.77	0.12	0.41	1.00	0.70	0.10	0.44	0.88
P_seglen35(%of segment length with 35 mph PSL)	0.14	0.10	0.00	0.59	0.18	0.15	0.00	1.00	0.15	0.07	0.00	0.40	0.18	0.06	0.07	0.37
P_seglen45(%of segment length with 45 mph PSL)	0.01	0.03	0.00	0.38	0.02	0.05	0.00	0.44	0.01	0.03	0.00	0.15	0.02	0.03	0.07	0.37
P_seglen55_65 (% of segment length with 55 mph or higher PSL)	0.03	0.07	0.00	0.55	0.05	0.10	0.00	0.83	0.04	0.06	0.00	0.27	0.06	0.07	0.00	0.36
Demographics and Socioeconomics variables																
Population density (Num of Population per acre)	5.54	5.96	0.00	101.29	3.76	3.38	0.00	19.01	4.45	3.41	0.04	16.07	2.85	2.27	0.07	7.58
P_female (% of female population）	0.51	0.06	0.00	1.00	0.49	0.49	0.00	1.00	0.51	0.03	0.35	0.61	0.51	0.02	0.47	0.61
POP_15 (% of people under 15 years old)	0.22	0.07	0.00	0.52	0.21	0.08	0.00	0.43	0.23	0.06	0.00	0.42	0.23	0.06	0.00	0.38
POP_15_65 (% of people from 15 to 65 years old)	0.65	0.11	0.00	1.00	0.65	0.13	0.00	1.00	0.65	0.08	0.15	1.00	0.66	0.09	0.17	1.00
POP_65 (% of people more than 65 years old)	0.13	0.12	0.00	0.90	0.14	0.12	0.00	1.00	0.12	0.10	0.00	0.85	0.12	0.11	0.00	0.83
MHINC (Median Annual Household Income*10^3)	43.57	21.39	0.00	159.95	40.14	20.24	0.00	115.66	42.86	16.81	9.46	96.59	46.51	14.22	19.46	78.77

Table 2 Results estimates of multivariate correlation

	No injury			Slight			Severe			Fatal		
BG	Mean	5%	95%	Mean	5%	95%	Mean	5%	95%	Mean	5%	95%
No injury				**0.977**	0.965	0.986	**0.969**	0.950	0.983	**0.943**	0.879	0.977
Slight							**0.964**	0.944	0.981	**0.946**	0.890	0.977
Severe										**0.944**	0.891	0.977
TAZ												
No injury				**0.974**	0.963	0.984	**0.966**	0.945	0.982	**0.908**	0.794	0.971
Slight							**0.970**	0.950	0.984	**0.921**	0.817	0.973
Severe										**0.928**	0.846	0.975
CT												
No injury				**0.926**	0.866	0.963	**0.906**	0.821	0.959	**0.422**	0.020	0.718
Slight							**0.892**	0.798	0.952	**0.419**	0.018	0.716
Severe										**0.449**	0.060	0.733
ZCTA												
No injury				**0.941**	0.878	0.984	**0.840**	0.659	0.956	**0.732**	0.410	0.932
Slight							**0.881**	0.753	0.965	**0.781**	0.514	0.943
Severe										**0.804**	0.565	0.946

Note: The bold numbers mean statistical significance at 95% significance level.

Table 3 Results estimates of spatial correlation

	No injury			Slight			Severe			Fatal		
BG	Mean	5%	95%	Mean	5%	95%	Mean	5%	95%	Mean	5%	95%
No injury	**0.782**	0.635	0.961	**0.913**	0.871	0.948	**0.720**	0.584	0.832	0.312	-0.060	0.572
Slight				**0.798**	0.666	0.952	**0.844**	0.759	0.908	**0.492**	0.138	0.712
Severe							**0.801**	0.647	0.984	**0.643**	0.380	0.834
Fatal										**0.681**	0.452	0.927
TAZ												
No injury	**0.524**	0.417	0.625	**0.848**	0.759	0.909	**0.577**	0.401	0.716	0.219	-0.165	0.554
Slight				**0.515**	0.387	0.628	**0.748**	0.623	0.847	0.408	-0.009	0.720
Severe							**0.672**	0.562	0.788	**0.673**	0.408	0.859
Fatal										**0.620**	0.399	0.886
CT												
No injury	**0.677**	0.500	0.865	**0.917**	0.854	0.958	**0.679**	0.466	0.826	0.382	-0.147	0.765
Slight				**0.673**	0.522	0.823	**0.821**	0.697	0.907	0.424	-0.109	0.788
Severe							**0.675**	0.498	0.849	0.417	-0.109	0.773
Fatal										**0.136**	0.094	0.188
ZCTA												
No injury	**0.292**	0.143	0.515	0.540	-0.095	0.919	-0.144	-0.756	0.639	0.073	-0.671	0.770
Slight				**0.210**	0.113	0.381	0.067	-0.611	0.748	0.088	-0.621	0.754
Severe							**0.277**	0.151	0.446	0.017	-0.681	0.732
Fatal										**0.239**	0.128	0.410

Note: The bold numbers mean statistical significance at 95% significance level.

Table 4 Results of model estimates

Variables	Block group				Traffic analysis zone				Census tract				ZIP code tabulation area			
	Mean	S.D.	BCI		Mean	S.D.	BCI		Mean	S.D.	BCI		Mean	S.D.	BCI	
			5%	95%			5%	95%			5%	95%			5%	95%
No injury																
Intercept	**3.331**	0.032	3.281	3.388	**3.677**	0.035	3.622	3.736	**4.838**	0.051	4.755	4.924	**6.326**	0.263	5.844	6.752
LnDVMT	**0.739**	0.038	0.680	0.811	**0.746**	0.033	0.693	0.798	**0.411**	0.041	0.344	0.477	**1.047**	0.172	0.767	1.344
Inter_density	**0.218**	0.035	0.158	0.276	**0.086**	0.036	0.026	0.145	**0.247**	0.054	0.158	0.338	0.184	0.165	-0.074	0.433
P_seglen25	**-0.257**	0.055	-0.348	-0.161	-0.025	0.038	-0.089	0.035	**-0.214**	0.059	-0.311	-0.117	-0.189	0.160	-0.458	0.076
P_seglen45	0.031	0.033	-0.018	0.089	**0.095**	0.028	0.049	0.141	0.010	0.052	-0.073	0.095	-0.061	0.147	-0.279	0.202
P_seglen55_65	-0.032	0.040	-0.099	0.032	-0.021	0.032	-0.072	0.031	0.089	0.053	-0.002	0.174	**-0.235**	0.132	-0.447	-0.023
POP_density	**-0.116**	0.041	-0.187	-0.049	**0.200**	0.038	0.139	0.261	-0.084	0.053	-0.171	0.002	**0.239**	0.144	0.022	0.492
MHINC	**-0.113**	0.034	-0.168	-0.055	**-0.093**	0.038	-0.157	-0.029	**-0.101**	0.048	-0.181	-0.021	-0.106	0.110	-0.276	0.079
MAD	2.234				2.475				2.937				3.199			
MSPE	8.489				10.268				14.059				16.951			
SAD	1776.064				1826.577				731.381				166.361			
Slight injury																
Intercept	**2.943**	0.029	2.896	2.994	**3.265**	0.032	3.214	3.321	**4.411**	0.040	4.346	4.478	**5.861**	0.211	5.476	6.204
LnDVMT	**0.668**	0.036	0.609	0.733	**0.656**	0.032	0.605	0.708	**0.410**	0.036	0.350	0.469	**1.022**	0.147	0.785	1.272
Inter_density	**0.247**	0.034	0.190	0.303	**0.071**	0.036	0.013	0.131	**0.257**	0.049	0.177	0.338	0.146	0.142	-0.079	0.358
P_seglen25	**-0.236**	0.053	-0.329	-0.144	-0.045	0.037	-0.108	0.013	**-0.212**	0.054	-0.300	-0.124	-0.175	0.137	-0.406	0.049
P_seglen45	0.023	0.031	-0.025	0.078	**0.057**	0.028	0.012	0.104	-0.013	0.046	-0.088	0.065	-0.063	0.128	-0.253	0.163
P_seglen55_65	**-0.090**	0.040	-0.157	-0.025	**-0.081**	0.032	-0.133	-0.029	0.028	0.047	-0.051	0.104	**-0.204**	0.113	-0.385	-0.024
POP_density	**-0.173**	0.041	-0.242	-0.106	**0.200**	0.036	0.141	0.258	-0.059	0.047	-0.139	0.019	**0.268**	0.122	0.083	0.481
MHINC	**-0.150**	0.035	-0.207	-0.092	**-0.116**	0.037	-0.179	-0.053	**-0.115**	0.045	-0.189	-0.041	-0.103	0.092	-0.247	0.050
MAD	2.398				2.643				3.532				3.651			
MSPE	9.813				11.588				19.456				20.853			

SAD	1906.767				1950.776				879.398				189.851			
Sever injury																
Intercept	**1.372**	0.037	1.311	1.434	**1.623**	0.036	1.566	1.683	**2.867**	0.047	2.790	2.945	**4.396**	0.176	4.096	4.676
LnDVMT	**0.678**	0.041	0.614	0.749	**0.655**	0.036	0.598	0.716	**0.395**	0.041	0.328	0.463	**1.004**	0.143	0.776	1.249
Inter_density	**0.124**	0.040	0.060	0.189	0.019	0.054	-0.071	0.109	**0.108**	0.056	0.017	0.200	0.034	0.141	-0.192	0.264
P_seglen25	**-0.215**	0.053	-0.306	-0.124	-0.028	0.045	-0.102	0.044	**-0.237**	0.061	-0.337	-0.137	**-0.235**	0.134	-0.459	-0.016
P_seglen45	-0.004	0.033	-0.055	0.054	**0.070**	0.032	0.018	0.122	-0.049	0.051	-0.134	0.036	-0.156	0.126	-0.357	0.052
P_seglen55_65	-0.034	0.042	-0.104	0.034	-0.012	0.038	-0.075	0.050	0.075	0.053	-0.013	0.160	**-0.232**	0.109	-0.412	-0.056
POP_density	**-0.199**	0.049	-0.281	-0.118	**0.174**	0.041	0.106	0.241	**-0.090**	0.055	-0.181	0.000	0.152	0.119	-0.038	0.352
MHINC	**-0.174**	0.040	-0.237	-0.106	**-0.100**	0.044	-0.172	-0.025	**-0.162**	0.052	-0.248	-0.076	**-0.150**	0.089	-0.293	-0.005
MAD	1.351				1.400				2.075				2.501			
MSPE	3.375				3.412				6.941				10.173			
SAD	1074.227				1033.384				516.647				130.054			
Fatal																
Intercept	**-1.182**	0.087	-1.326	-1.039	**-0.813**	0.074	-0.937	-0.691	**2.061**	0.025	2.019	2.101	**2.033**	0.149	1.781	2.268
LnDVMT	**0.639**	0.078	0.510	0.767	**0.541**	0.064	0.436	0.648	**0.058**	0.026	0.015	0.100	**0.956**	0.161	0.697	1.228
Inter_density	0.106	0.079	-0.022	0.239	-0.225	0.149	-0.484	0.009	-0.036	0.034	-0.092	0.019	0.066	0.155	-0.185	0.322
P_seglen25	**-0.273**	0.086	-0.417	-0.133	-0.006	0.085	-0.147	0.134	-0.054	0.040	-0.119	0.011	0.004	0.154	-0.249	0.255
P_seglen45	0.035	0.051	-0.050	0.120	**0.135**	0.049	0.053	0.215	-0.002	0.032	-0.055	0.050	-0.069	0.138	-0.295	0.158
P_seglen55_65	-0.046	0.060	-0.144	0.053	0.012	0.061	-0.089	0.112	**0.057**	0.033	0.003	0.111	**-0.213**	0.132	-0.433	0.000
POP_density	**-0.426**	0.118	-0.622	-0.234	-0.036	0.079	-0.166	0.095	**-0.061**	0.035	-0.118	-0.004	**-0.285**	0.134	-0.504	-0.063
MHINC	**-0.141**	0.076	-0.267	-0.018	-0.118	0.076	-0.242	0.007	-0.046	0.031	-0.097	0.005	**-0.367**	0.100	-0.529	-0.201
MAD	0.485				0.577				1.084				1.780			
MSPE	0.527				0.589				2.070				5.128			
SAD	385.401				425.669				269.960				92.579			

Note: The bold numbers mean statistical significance at 95% significance level

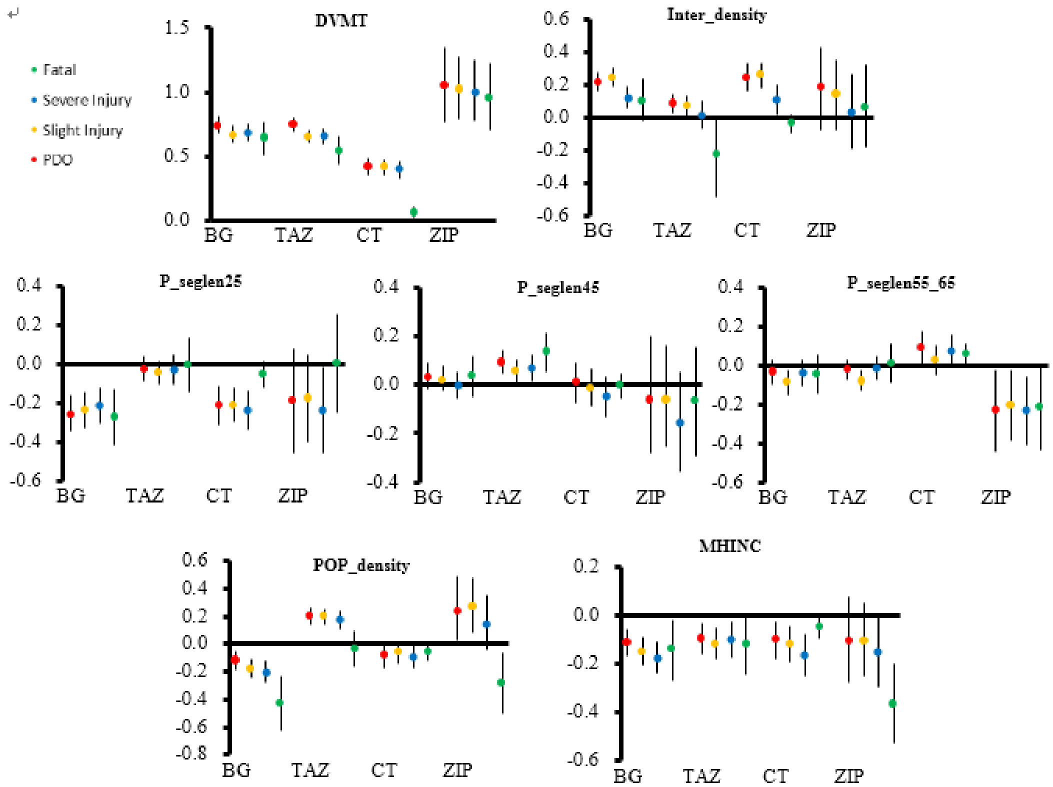

Fig. 1 Results of standard difference in means test for model parameter estimates

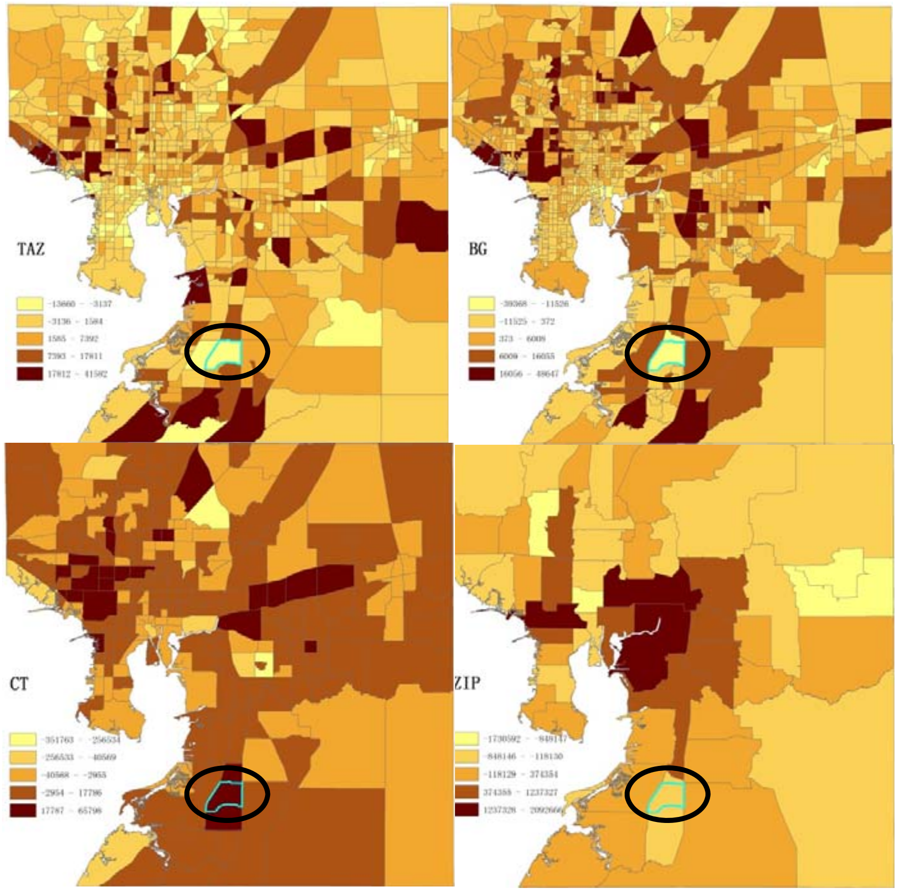

Fig. 2 The identification of high-crash location