Accident Analysis and Prevention 70 (2014) 110–120

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Sensitivity analysis in the context of regional safety modeling: Identifying and assessing the modifiable areal unit problem

Pengpeng Xu[a], Helai Huang[b,*], Ni Dong[a], Mohamed Abdel-Aty[c]

[a] Urban Transport Research Center, School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan 410075, PR China
[b] School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan 410075, PR China
[c] Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, FL, USA

ARTICLE INFO

Article history:
Received 26 July 2013
Received in revised form 31 January 2014
Accepted 19 February 2014
Available online 10 April 2014

Keywords:
Modifiable areal unit problem
Sensitivity analysis
Regional safety prediction model
REDCAP

ABSTRACT

A wide array of spatial units has been explored in current regional safety analysis. Since traffic crashes exhibit extreme spatiotemporal heterogeneity which has rarely been a consideration in partitioning these zoning systems, research based on these areal units may be subjected to the modifiable areal unit problem (MAUP).

This study attempted to conduct a sensitivity analysis to quantitatively investigate the MAUP effect in the context of regional safety modeling. The emerging regionalization method-RECDAP (regionalization with dynamically constrained agglomerative clustering and partitioning) was employed to aggregate 738 traffic analysis zones in the county of Hillsborough to 14 zoning schemes at an incremental step-size of 50 zones based on spatial homogeneity of crash risk. At each level of aggregation, a Bayesian Poisson lognormal model and a Bayesian spatial model were calibrated to explain observed variations in total/severe crash counts given a number of zone-level factors.

Results revealed that as the number of zones increases, the spatial autocorrelation of crash data increases. The Bayesian spatial model outperforms the Bayesian Poisson-lognormal model in accurately accounting for spatial autocorrelation effects, unbiased parameter estimates, and model performance, especially in cases with higher disaggregated levels. Zoning schemes with higher number of zones tend to have increasing number of significant variables, more stable coefficient estimation, smaller standard error, whereas worse model performance. The variables of population density and median household income show consistently significant effects on crash risk and are robust to variation in data aggregation. The MAUP effects may be significantly reduced if we just maintain at about 50% of the original number of zones (350 or larger). The present study highlights MAUP that is generally ignored by transportation safety analysts, and provides insights into the nature of parameter sensitivity to data aggregation in the context of regional safety modeling.

1. Introduction

Regional crash prediction model to explain observed cross-sectional variations in crash counts using macro-structural covariates becomes a fairly routine component in traffic safety research. Three essential incentives ensure its rapid development. First, many factors affecting crashes operate at a spatial scale (e.g., trip distribution and generation, land use pattern, and various demographical characteristics). It is desired to relate safety with zonal level factors. Then, there is a need for state agencies to regularly monitor region level safety and provide incentives to reduce the number of traffic casualties in a region's safety program. Therefore, a reliable assessment of safety is indispensable by estimating the aggregating crash potentials associated with the target road network on different spatial scales. Furthermore, road safety is increasingly considered as a necessary component in transportation planning (FHWA, 2005). Regional safety prediction models have been suggested as means of incorporating safety considerations into long term transportation planning (Washington et al., 2006).

To response these needs, the last decade witnessed fast growing scope of scientific research in the context of regional safety analysis. Various spatial units are utilized in previous research. However, most of the employed zoning schemes have specific usage and may not be appropriate for regional safety analysis, as traffic crashes

* Corresponding author. Tel.: +86 731 82656631.
E-mail addresses: huanghelai@csu.edu.cn, huanghelai@hotmail.com (H. Huang), mabdel@mail.ucf.edu (M. Abdel-Aty).

http://dx.doi.org/10.1016/j.aap.2014.02.012
0001-4575/

exhibit extreme spatiotemporal heterogeneity which has rarely been a consideration in partitioning zonal systems (Huang et al., 2013). Therefore, research based on these modifiable areal units is probably subject to problems related to the scale and zonal effects.

Up till now, there is a little comprehensive investigation assessing how scale and zonal effects influence statistical results of safety modeling. The assessment of MAUP effects becomes a pressing issue as it may lead to unreliable and inaccurate safety estimation (Wang et al., 2012a; Huang et al., 2013).

The objective of this study is to quantitatively investigate the impacts of specific aggregation configurations on the variations of regional safety modeling. The emerging regionalization method-REDCAP (regionalization with dynamically constrained agglomerative clustering and partitioning, Guo, 2008; Guo and Wang, 2011) is employed to produce a new set of aggregated areal units based on spatial homogeneity of safety related factors. Through this sensitivity analysis, it is possible to explore the scope of statistical results that can be produced for various zoning systems, to determine which safety factors are sensitive to the variation in scale or areal unit definition, to what extent, and to identify which level of data aggregation minimizes the impact of the MAUP. It is expected to highlight the MAUP that is generally ignored by transportation safety analysts, and provide insights into the nature of parameter sensitivity to data aggregation in the context of multivariate safety modeling analysis.

2. Literature review

2.1. Regional crash prediction model

Regional crash prediction models have been widely applied to explain observed variations in crash counts associated with area-wide factors. Different area-wide characteristics were considered, including road characteristics such as intersections density (e.g., Huang et al., 2010), road length with different speed limit (e.g., Abdel-Aty et al., 2011; Siddiqui et al., 2012), road length with different functional classifications (e.g., Quddus, 2008; Hadayeghi et al., 2010; Huang et al., 2010), junctions and roundabouts (e.g., Noland and Quddus, 2004; Quddus, 2008); traffic patterns such as traffic flow and traffic speed (Noland and Quddus, 2005; Quddus, 2008; Hadayeghi et al., 2010); trip generation and distribution (Abdel-Aty et al., 2011); environment conditions such as total precipitation/snowfall, and number of rainy/snowy days per year (Aguero-Valverde and Jovanis, 2006); land use (e.g., Hadayeghi et al., 2010; Siddiqui et al., 2012; Pulugurha et al., 2013); and socioeconomic factors such as population density (e.g., Hadayeghi et al., 2003; Huang et al., 2010; Siddiqui et al., 2012), age cohorts (e.g., Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Hadayeghi et al., 2010; Huang et al., 2010), household incomes (e.g., Huang et al., 2010; Siddiqui et al., 2012) and employment (Quddus, 2008; Hadayeghi et al., 2010; Huang et al., 2010).

As crash data are typically collected with reference to location dimension, two problems arise (LeSage, 1999): (1) spatial correlation exists among the observations, and (2) spatial heterogeneity occurs in the relationships that are modeled. Traditional crash prediction models such as Poisson lognormal model and negative binomial model (i.e., Poisson gamma model) have largely ignored the issue of spatial correlation (spatial dependency) in traffic crash data, which would be misleading as this cannot reflect the true underlying data generating processes. Recently, Bayesian hierarchical models have been widely employed to address the issue of unmeasured spatial correlation in regional safety analysis (Miaou et al., 2003; Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Huang et al., 2010; Siddiqui et al., 2012).

A wide array of spatial units have been explored in previous regional safety studies, such as regions (Washington et al., 1999), counties (Miaou et al., 2003; Aguero-Valverde and Jovanis, 2006; Huang et al., 2010), districts (i.e., a level of sub-national division used for the purposes of local government; Haynes et al., 2007), traffic analysis zones (TAZs, Hadayeghi et al., 2003; Guevara et al., 2004; Hadayeghi et al., 2006, 2010; Abdel-Aty et al., 2011; Siddiqui et al., 2012; Pulugurha et al., 2013), wards (i.e., the primary unit of electoral geography for civil parishes; Noland and Quddus, 2004; Quddus, 2008), enumeration districts (a geographic area assigned to an individual census taker, or enumerator, usually representing a specific portion of a city or county; Noland and Quddus, 2005), and grid-based structure (MacNab, 2004). However, most of above established zoning schemes have specific usage and may not be the optimal zone systems for macro-level safety analysis. As a result, road safety research based on these modifiable areal units is probably subject to modifiable areal unit problem (MAUP).

2.2. Sensitivity analysis for modifiable areal unit problem

MAUP refers to situations that when the boundary of zones used in a spatial analysis changes, the statistical inference and interpretation derived from the zones are also different (Openshaw, 1984). Typically, it can 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 that are aggregated at different levels, zoning effect refers to the variability introduced by different zoning configurations.

Although the MAUP research has a substantive history, the solutions to MAUP are not straightforward as different geographical phenomena subjects to varied spatial and temporal distributions. Fotheringham et al. (2000) suggested that a sensitivity analysis which examines the parameters estimates using various spatial and scale configurations is desired before presenting analysis results to decision makers. This suggestion of sensitivity analysis is usually implemented by aggregating the smallest available base units into large regions and alternate zones. Based on the sensitivity analysis, the impacts of specific configurations on the variations of statistical analysis results can then be quantitatively investigated.

By considering the MAUP on a line, Thomas (1996) empirically investigated the effect of systematic variation of road segment length on the statistical description of crash counts. Three distinct groups of roadway segments were found: (1) for small segments about 100 m, crash counts followed almost Poisson distribution, (2) for medium size segments (e.g., 300–2000 m), crash counts had an intermediate empirical distribution, and (3) for lager segments (e.g., more than 2000 m), crash counts were almost normally distributed. The empirical work of Thomas (1996) demonstrated that the definition of road segments may also affect the statistical results due to MAUP. However, the study was limited to a univariate exploratory sensitivity analysis and the author did not further investigate how this influence could be inserted into safety modeling.

Recently, Abdel-Aty et al. (2013) investigated the effects of zonal variation (e.g., TAZs, block groups, and census tracts) on the performance of Bayesian Poisson lognormal models with three different dependent variables (e.g., total crashes, severe crash and pedestrian crashes, respectively). Inconsistent parameter estimates were found across the three geographic units, which demonstrated the existence of MAUP effects in regional safety analysis. However, the research was limited to three regions, probably too few to fully assess the impacts of spatial aggregation.

In a more extended research, Fotheringham and Wong (1991) performed linear regression to comprehensively investigate the impact of scale and zoning effects on the stability of parameter