Journal of Safety Research xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Safety Research

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

Comparative analysis of zonal systems for macro-level crash modeling

Qing Cai, * Mohamed Abdel-Aty, Jaeyoung Lee, Naveen Eluru

Department of Civil, Environment and Construction Engineering, University of Central Florida, Orlando, FL 32816, United States

ARTICLE INFO

Article history:
Received 30 August 2016
Received in revised form 28 November 2016
Accepted 27 February 2017
Available online xxxx

Keywords:
Macro-level crash modeling
Census tracts
Traffic analysis zones
Traffic analysis districts
Poisson lognormal

ABSTRACT

Macro-level traffic safety analysis has been undertaken at different spatial configurations. However, clear guidelines for the appropriate zonal system selection for safety analysis are unavailable. In this study, a comparative analysis was conducted to determine the optimal zonal system for macroscopic crash modeling considering census tracts (CTs), state-wide traffic analysis zones (STAZs), and a newly developed traffic-related zone system labeled traffic analysis districts (TADs). Poisson lognormal models for three crash types (i.e., total, severe, and non-motorized mode crashes) are developed based on the three zonal systems without and with consideration of spatial autocorrelation. The study proposes a method to compare the modeling performance of the three types of geographic units at different spatial configurations through a grid based framework. Specifically, the study region is partitioned to grids of various sizes and the model prediction accuracy of the various macro models is considered within these grids of various sizes. These model comparison results for all crash types indicated that the models based on TADs consistently offer a better performance compared to the others. Besides, the models considering spatial autocorrelation outperform the ones that do not consider it. Finally, based on the modeling results and motivation for developing the different zonal systems, it is recommended using CTs for socio-demographic data collection, employing TAZs for transportation demand forecasting, and adopting TADs for transportation safety planning.

1. Introduction

Safety and mobility are two fundamental requirements of transportation services. Unfortunately, a recent study revealed that the total cost of traffic crashes is almost two times greater than the overall cost of traffic congestion (Meyer, Systematics, C., & Association, A.A., 2008). Hence, it is very important to devote efforts to enhance road safety and thus reduce the social burden. Towards this end, a common approach is the application of macroscopic level crash modeling, which can integrate safety into long-range transportation planning at zonal level.

In the past decade, several studies have been conducted for crash modeling at a macro-level (see (Yasmin & Eluru, 2016) for a detailed review). Across these studies, various zonal systems have been explored including: block groups (Levine, Kim, & Nitz, 1995), census tracts (LaScala, Gerber, & Gruenewald, 2000), traffic analysis zones or TAZs (Abdel-Aty, Siddiqui, & Huang, 2011; Cai, Lee, Eluru, & Abdel-Aty, 2016; Hadayeghi, Shalaby, & Persaud, 2003; Hadayeghi, Shalaby, & Persaud, 2010; Ladrón de Guevara, Washington, & Oh, 2004; Lee, Abdel-Aty, Choi, & Siddiqui, 2013; Yasmin & Eluru, 2016), counties (Aguero-Valverde & Jovanis, 2006; Huang, Abdel-Aty, & Darwiche, 2010), and ZIP code areas (Lee, Abdel-Aty, Choi, & Huang, 2015; Lee et al., 2013). Most of these zonal systems were developed for different specific usages. For example, the block groups and census tracts are developed by census bureau for the presentation of statistical data while TAZs are delineated for the long-term transportation plan. Meanwhile, the area of census tracts and TAZs are greater than the block groups (Abdel-Aty, Lee, Siddiqui, & Choi, 2013). As a result, within the study area, the number of units, aggregation levels and zoning configuration can vary substantially across different zonal systems. Regarding this, Kim, Brunner, and Yamashita (2006) developed a uniform 0.1 mile^2 grid structure to explore the impact of socio-demographic characteristics such as land use, population size, and employment by sector on crashes. Compared with other existing geographic units, the grid structure is uniformly sized and shaped which can eliminate the artifact effects. However, considering the availability and use of the various zonal systems for other transportation purposes creating a uniform grid structure would not be feasible from the perspective of state and regional agencies. Hence, as part of our study, we investigate the performance of safety models developed at various zonal configurations to offer insights on what zonal systems are appropriate for crash analysis and long term transportation safety planning.

* Corresponding author.
E-mail addresses: qingcai@knights.ucf.edu (Q. Cai), M.Aty@ucf.edu (M. Abdel-Aty), jaeyoung@knights.ucf.edu (J. Lee), Naveen.Eluru@ucf.edu (N. Eluru).

http://dx.doi.org/10.1016/j.jsr.2017.02.018
0022-4375/© 2017 National Safety Council and Elsevier Ltd. All rights reserved.

Recently, several research studies have been conducted to compare different geographic units. Abdel-Aty et al. (2013) conducted modeling analysis for three types of crashes (total, severe, and pedestrian crashes) with three different types of geographic entities (block groups, TAZs, and census tracts). Inconsistent significant variables were observed for the same dependent variables, validating the existence of zonal variation. However, no comparison of modeling performance was conducted in this research. Lee, Abdel-Aty, and Jiang (2014) aggregated TAZs into traffic safety analysis zones (TSAZs) based on crash counts. Four different goodness-of-fit measures (i.e., mean absolute deviation, root mean squared errors, sum of absolute deviation, and percent mean absolute deviation) were employed to compare crash model performance based on TSAZs and TAZs. The results indicated that the model based on the new zone system can provide better performance. Instead of determining the best zone system, Xu, Huang, Dong, and Abdel-Aty (2014) created different zoning schemes by aggregating TAZs with a dynamical method. Models for total/severe crashes were estimated to explore variations across zonal schemes with different aggregation levels. Meanwhile, deviance information criterion, mean absolute deviation, and mean squared predictive error were calculated to compare different models. However, the employed measures for the comparison can be largely influenced by the number of observations and the observed values. Thus, the comparison results might be limited in the two studies (Lee et al., 2014; Xu et al., 2014) since the measures were calculated based on zonal systems with different number of zones. Ignoring such limitation may result in inaccurate crash prediction results and inappropriate transportation safety plans.

To address the limitation, one possible solution is to compute the measures based on a third-party zonal system so that the calculation would have the same observations. Towards this end, a grid structure that uniformly delineates the study region is suggested as a viable option. Specifically, the crash models developed for the various zonal systems will be tested on the same grid structure. To ensure that the result is not an artifact of the grid size, several grid sizes ranging from 1 to 100 mile^2 will be considered.

The current paper aims to conduct comparative analysis of different geographic units for macroscopic crash modeling analysis and provide guidance for transportation safety planning. Towards this end, both aspatial model (i.e., Poisson lognormal (PLN)) and spatial model (i.e., PLN conditional autoregressive (PLN-CAR)) are developed for three types of crashes (i.e., total, severe, and non-motorized mode crashes) based on census tracts, traffic analysis zones, and a newly developed zone system — traffic analysis districts (see the following section for detailed information). Then, a comparison method is proposed to compare the modeling performance with the same sample sizes by using grids of different dimensions. By using different goodness-of-fit measures, superior geographic units for crash modeling and transportation safety planning are identified.

2. Configuration of geographic units

In this study, crash models were developed based on three different geographic units, which are discussed in the following subsections.

2.1. Introduction of geographic units

2.1.1. Census tracts

According to the U.S. Census Bureau, census tracts (CTs) are small, relatively permanent subdivisions of a county or equivalent entity to present statistical data such as poverty rates and income levels. On average, a CT has about 4,000 inhabitants. CTs are designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions.

2.1.2. Traffic analysis zones

Traffic analysis zones (TAZs) are geographic entities delineated by state or local transportation officials to tabulate traffic-related data such as journey-to-work and place-of-work statistics (23). TAZs are defined by grouping together census blocks, block groups, or census tracts. A TAZ usually covers a contiguous area with a 600 minimum population and the land use within each TAZ is relatively homogeneous (Abdel-Aty et al., 2013).

2.1.3. Traffic analysis districts

Traffic analysis districts (TADs) are new, higher-level geographic entities for traffic analysis (FHWA and Census Transportation Planning Products (CTPP), 2011). TADs are built by aggregating TAZs, block groups or census tracts. In almost every case, the TADs are delineated to adhere to a 20,000 minimum population criteria and more likely to have mixed land use.

2.2. Comparison of geographic units

In Florida, the average area of CTs, TAZs, and TADs is 15.497, 6.472, and 103.314 mile^2, respectively. Across the three geographic units, which are shown in Fig. 1, a TAD is considerably larger than a CT and TAZ while a TAZ is most likely to have the smallest size.

CT boundaries are generally delineated by visible and identifiable features, with the intention of being maintained over a long time. On the other hand, both TAZs and TADs are developed for transportation planning and are always divided by physical boundaries, mostly arterial roadways. Usually, CTs and TAZs nest within counties while TADs may cross county boundaries, but they must nest within metropolitan planning organizations (MPOs) (FHWA and Census Transportation Planning Products (CTPP), 2011).

3. Data preparation

Multiple geographic units were obtained from the U.S. Census Bureau and Florida Department of Transportation (FDOT). The state of Florida has 4,245 CTs, 8,518 TAZs, and 594 TADs. Crashes that occurred in Florida in 2010–2012 were collected for this study. A total of 901,235 crashes were recorded in Florida among which 50,039 (5.6%) were severe crashes and 31,547 (3.5%) were non-motorized mode crashes. In this study, severe crashes were defined as the combination of all fatal and incapacitating injury crashes while non-motorized mode crashes were the sum of pedestrian and bicyclist involved crashes. On average, TADs have highest number of crashes since they are the largest zonal configuration. Given the large number of crashes in the Florida data, units with zero count are observed for CTs and TAZs. However, within a TAD no zero count units exist for the time period of our analysis.

A host of explanatory variables are considered for the analysis and are grouped into three categories: traffic measures, roadway characteristics, and socio-demographic characteristics. For the three zonal systems, these data are collected from the Geographic Information System (GIS) archived data from Florida Department of Transportation (FDOT) and U.S. Census Bureau (USCB).

The traffic measures include VMT (Vehicle-Miles-Traveled), proportion of heavy vehicle in VMT. Regarding the roadway variables, roadway density (i.e., total roadway length per square mile), proportion of length roadways by functional classifications (freeways, arterials, collector, and local roads), signalized intersection density (i.e., number of signalized intersection per total roadway mileage), length of bike lanes, and length of sidewalks were selected as the explanatory variables. Concerning the socio-demographic data, the distance to the nearest urban area, population density (defined as population divided by the area), proportion of population between 15 and 24 years old, proportion of population equal to or older than 65 years old, total employment density (defined as the total employment per square mile), proportion of unemployment, median household income, total commuters density (i.e., the

Please cite this article as: Cai, Q., et al., Comparative analysis of zonal systems for macro-level crash modeling, *Journal of Safety Research* (2017), http://dx.doi.org/10.1016/j.jsr.2017.02.018