ARTICLE IN PRESS

Accident Analysis and Prevention xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models

Dibakar Saha[a,*], Priyanka Alluri[b], Albert Gan[b], Wanyang Wu[b]

[a] *Collaborative Sciences Center for Road Safety, School of Urban and Regional Planning, Florida Atlantic University, 777 Glades Road, SO 376, Boca Raton, 33431, FL, United States*

[b] *Department of Civil and Environmental Engineering, Florida International University, 10555 West Flagler Street, EC 3680, Miami, 33174, FL, United States*

ARTICLE INFO

Keywords:
Bicycle safety
Conditional autoregressive model
Besag's model
Leroux's model
STRAVA bicycle ridership data

ABSTRACT

The objective of this study was to investigate the relationship between bicycle crash frequency and their contributing factors at the census block group level in Florida, USA. Crashes aggregated over the census block groups tend to be clustered (i.e., spatially dependent) rather than randomly distributed. To account for the effect of spatial dependence across the census block groups, the class of conditional autoregressive (CAR) models were employed within the hierarchical Bayesian framework. Based on four years (2011-2014) of crash data, total and fatal-and-severe injury bicycle crash frequencies were modeled as a function of a large number of variables representing demographic and socio-economic characteristics, roadway infrastructure and traffic characteristics, and bicycle activity characteristics. This study explored and compared the performance of two CAR models, namely the Besag's model and the Leroux's model, in crash prediction. The Besag's models, which differ from the Leroux's models by the structure of how spatial autocorrelation are specified in the models, were found to fit the data better. A 95% Bayesian credible interval was selected to identify the variables that had credible impact on bicycle crashes. A total of 21 variables were found to be credible in the total crash model, while 18 variables were found to be credible in the fatal-and-severe injury crash model. Population, daily vehicle miles traveled, age cohorts, household automobile ownership, density of urban roads by functional class, bicycle trip miles, and bicycle trip intensity had positive effects in both the total and fatal-and-severe crash models. Educational attainment variables, truck percentage, and density of rural roads by functional class were found to be negatively associated with both total and fatal-and-severe bicycle crash frequencies.

1. Introduction

According to the National Highway Traffic Safety Administration (NHTSA), bicycle commuting in the U.S. has increased by more than 60% since 2000 (NHTSA, 2016a). Yet, the safety risks to bicyclists are evident and are on an upward trend in recent years. During the period 2010–2015, the number of bicycle fatalities has increased by 31% across all states in the U.S. Florida, the southernmost state of the U.S. mainland, has witnessed 1.8 times as many bicycle fatalities in 2015 (150 fatalities) as in 2010 (83 fatalities). Overall, Florida leads the nation with 755 bicycle fatalities during 2010-2015. The bicyclist fatality rate in Florida in 2015 is 7.4 per million population, which is approximately three times higher than the nation's average of 2.5 per million population (NHTSA, 2016b). These grim statistics underscore the need for a thorough investigation of bicycle crashes in Florida. An essential step is therefore to develop crash prediction models that explain the relationship between bicycle crashes and their contributing factors.

Crash prediction models are broadly divided into two categories: micro-level or disaggregate models, and macro-level or aggregate models. In micro-level modeling, crashes are analyzed based on small, homogeneous road entities, such as roadway segments, ramps, and intersections, with one of the following objectives: identify geometric design features and traffic attributes contributing to crashes, select high crash locations, and suggest countermeasures to reduce crashes (e.g., Hadi et al., 1995; Abdel-Aty and Radwan, 2000; Lu et al., 2013; Chen and Persaud, 2014). On the other hand, in macro-level analysis, crashes are aggregated over some geographic areas to investigate the influence of socio-economic, demographic, land use, and infrastructure-related factors on crash occurrence for long-term planning and policy implications in improving traffic safety in neighborhoods (e.g., Oppe, 1989; Hadayeghi et al., 2003; Wier et al., 2009). Bicycle crashes are typically analyzed at macro-level due to the following characteristics: (a) bicycle crashes are often severe compared to other motor-vehicle

* Corresponding author.
E-mail address: sahad@fau.edu (D. Saha).

https://doi.org/10.1016/j.aap.2018.02.014
Received 12 July 2017; Received in revised form 14 February 2018; Accepted 14 February 2018
0001-4575/ Published by Elsevier Ltd.

crashes (e.g., bicyclists are more likely to be severely injured); (b) bicycle exposure is different from vehicle exposure and is difficult to quantify; and (c) bicycle crash trends are quite distinctive as evident from the fact that bicycle riding depends on demographics, built environment, and roadside infrastructure.

Noland and Quddus (2004), Wedagama et al. (2006), and Wei and Lovegrove (2013) developed bicycle crash frequency models at different areal units using the negative binomial (NB) distribution. While the NB regression is most commonly used to model crash data, particularly at micro-level, it is not always the best approach for modeling crash data at macro-level (Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Siddiqui et al., 2012). This is because crash aggregation over geographical areas introduces spatial dependency, also called spatial autocorrelation, which violates the assumption of full independence among observations in the NB model. Spatial autocorrelation is said to be present when observations in adjacent areas or neighborhoods are clustered, i.e., have similar data values (Dale and Fortin, 2002). Ignoring the possible effect of spatial autocorrelation among neighborhoods in macro-level data might lead to inaccurate parameter estimates.

Levine et al. (1995a; 1995b) demonstrated the presence and pattern of spatial dependence in crash data using a series of spatial statistical functions, and then developed spatial lag models to quantify the impact of zonal-level predictors (e.g., population, employment, and road characteristics) on motor vehicle crashes aggregated at the census block group level in Honolulu, Hawaii. Crashes were found to be more clustered by census block groups instead of a random distribution. One caveat of this study is that spatial lag models are appropriate for continuous data (Haining, 2003), which is not in agreement with the non-negative discrete properties of crash data. Moran's test is another measure to detect spatial autocorrelation in data; however, the test could be used as an exploratory measure to assess spatial autocorrelation, and is not intended for macro-level modeling (Banerjee et al., 2004).

The aforementioned issues pertaining to macro-level crash modeling could be addressed by adopting a hierarchical Bayesian modeling approach (MacNab, 2004; Aguero-valverde and Jovanis, 2006; Quddus, 2008; Huang et al., 2010). Spatial autocorrelation is accounted for in Bayesian models as a random coefficient, and its effect is measured by specifying conditional autoregressive (CAR) prior probability distributions using the hierarchical framework (Held and Rue et al., 2010). The class of Bayesian models with CAR prior distributions, simply referred to as CAR models, have several model specifications as suggested by Besag (1974); Besag et al. (1991); Cressie (1993); Leroux et al. (1999); Dean et al. (2001); and Simpson et al. (2017). Among them, the model proposed by Besag et al. (1991), herein referred to as the Besag's model, is the most popular and has been extensively used in macro-level crash predictions (Huang et al., 2010; Siddiqui et al., 2012; Aguero-Valverde, 2013; Barua et al., 2014). Among others, the model proposed by Leroux et al. (1999), herein referred to as the Leroux's model, was implemented in a few traffic safety studies to predict motor vehicle crashes (MacNab, 2004; Xu et al., 2017). However, no studies were found to have examined the relative performance of these two models in macro-level crash predictions.

Another important issue with bicycle safety research is with the unavailability of accurate and complete bicycle volume data. Researchers often address this issue by using surrogate measures of bicycle exposure such as vehicular traffic and population or employment data (Siddiqui et al., 2012; Lee et al., 2015; Cai et al., 2017). However, it is obvious that bicycle volume data is a more accurate measure of bicycle exposure. Because agencies do not usually collect bicycle volume data, smartphone applications intended for bicyclists to track their routes using device's GPS can be used as an alternative source of bicycle ridership data. *Strava* is one of the most popular GPS-based smartphone applications, which facilitates bicyclists to save and keep track of their trip information, including trip routes, trip miles, trip durations, etc. (Figliozzi and Blanc, 2015; Jestico et al., 2016). Florida Department of Transportation (FDOT) has acquired the *Strava* data from Strava Metro to learn about bicycle ridership in Florida. Investigating the *Strava* data as a possible source of bicycle exposure may provide valuable insights for safety modeling.

The objective of this study was to develop macro-level bicycle crash frequency models at census block groups in Florida. The hierarchical Bayesian approach was implemented to identify the contributing factors for total bicycle crash frequency and fatal-and-severe bicycle crash frequency. Two CAR model specifications, the Besag's model and the Leroux's model, were examined and their performance was evaluated. An array of variables comprising census block group-level demographic and socio-economic characteristics, roadway infrastructure, traffic characteristics, and bicyclists' trip characteristics obtained from the *Strava* smartphone application were investigated in this study.

2. Literature review

The majority of previous studies in traffic safety literature that employed CAR models have focused on motor vehicle crashes (Miaou et al., 2003; MacNab, 2004; Aguero-Valverde and Jovanis, 2006; Song et al., 2006; Quddus, 2008; Huang et al., 2010; Wang et al., 2012; Abdel-Aty et al., 2011; Aguero-Valverde, 2013; Wang and Kockelman, 2013; Lee et al., 2014; Barua et al., 2015; Dong et al., 2016). Recently, Abdel-Aty et al. (2011), Siddiqui et al. (2012), Chen (2015), Kaplan and Prato (2015), Lee et al. (2015), Amoh-Gyimah et al. (2016), Osama and Sayed (2016), and Cai et al. (2017) estimated CAR models to investigate the associations between bicycle crashes and area-level predictors.

Of these studies, Siddiqui et al. (2012) and Chen (2015) analyzed bicycle crash frequency data at traffic analysis zones (TAZs) and found that the models that accounted for spatial autocorrelation had performed better compared to those that did not consider the effect of spatial autocorrelation. Spatial autocorrelation in both studies was accounted for over 50% of total random effects. In another effort, Cai et al. (2017) analyzed non-motorized crash data (i.e., bicycle and pedestrian crashes) in the entire state of Florida using three zonal systems: census tracts, traffic analysis districts (TADs), and TAZs. The effect of spatial autocorrelation was found statistically significant in the models for TADs and TAZs. It is therefore essential to consider the effect of spatial autocorrelation in macro-level bicycle crash frequency models.

Several types of variables, including demographic and socio-economic characteristics, roadway characteristics, traffic characteristics, land use, and travel demand related variables were considered in previous studies concerning bicycle safety. Abdel-Aty et al. (2011) examined the effects of road characteristics and various types of trip productions and attractions on combined bicycle and pedestrian crashes at TAZs of four counties in Florida: Citrus, Hernando, Pasco, and Hillsborough. Trip productions and attractions, however, were found to have no significant impact on bicycle and pedestrian crash casualties. Siddiqui et al. (2012) estimated the following variables to have significant impact on bicycle crashes in Hillsborough and Pinellas Counties in Florida: road length with different speed limits, number of intersections, number of dwelling units, population density, household car ownership, and employment, where population density and employed were regarded as surrogate measures of bicycle exposure. Neither studies did consider traffic related variables and traffic exposure in model fit.

Chen (2015) investigated built environment factors to analyze bicycle crashes in Seattle, Washington. The study used the following travel demand variables as bicycle exposure: number of bicycle trips, number of total trips, and bike modal share. Of these exposure variables, the number of total trips was the only credible variable. Other credible variables included three-way intersections density, bike lane density, parking signs density, traffic signals density, entropy of mixed land use, and zonal mean of driving speed limits.