Journal of Transport Geography 69 (2018) 221–233

Contents lists available at ScienceDirect

Journal of Transport Geography

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

Examining spatial relationships between crashes and the built environment: A geographically weighted regression approach

Yuan Huang[a], Xiaoguang Wang[b,*], David Patton[c]

[a] Department of Geography, University of South Carolina, 318 Callcott Building, Columbia, SC 29208, USA
[b] Department of Geography, Central Michigan University, 287 Dow Science Complex, Mount Pleasant, MI 48859, USA
[c] Department of Geography, Central Michigan University, ET Building 228C, Mount Pleasant, MI 48859, USA

ARTICLE INFO

Keywords:
Vehicle crashes
Built environment
Geographically weighted regression
Spatial non-stationarity
Detroit region

ABSTRACT

A better understanding of the relationships between vehicle crashes and the built environment is an important step in improving crash prediction and providing sound policy recommendations that could reduce the occurrence or severity of crashes. Global statistical models are widely used to explore the relationships between vehicle crashes and the built environment, but these models do not incorporate a spatial component and are unable to deal with the issues of spatial autocorrelation and spatial non-stationarity. Our research utilizes a geographically weighted regression (GWR) model to explore the relationships between crashes and the built environment in the context of the Detroit region in Michigan. We find that the relationships between the built environment and crashes are spatially non-stationary: both the strength and the direction of their relationships differ over space. Our study also identifies several built environment variables, such as commercial use percentage, local road mileage percentage, and intersection density, that have relatively stable relationships with crashes. Our research demonstrates the feasibility and value of using spatial models in traffic, transportation, and land use research.

1. Introduction

1.1. The influence of the built environment on vehicle crashes

Vehicle crashes are among the top five causes of death in the United States. The 2011 death and mortality statistics listed vehicle accidents as the number one cause of death for people under the age of 44 (Miniño and Murphy, 2013; Pirdavani et al., 2014). Approximately 2.22 million people were injured and 32,367 people were killed due to vehicle accidents in the United States in 2011 (NHTSA, 2013). Researchers continue to seek ways to understand the key factors that contribute to crashes, with the aim of improving crash prediction and providing policy recommendations that could reduce crash occurrence or severity. Previous research has shown that vehicle crashes result from the interaction of five major factors: driver, traffic, road, vehicle, and environment (Miaou, 1996; Song et al., 2006). While numerous studies have examined the impacts of driver characteristics (McGwin and Brown, 1999; Ulak et al., 2017), traffic volume (Ewing and Dumbaugh, 2009), and driving speed (Aarts and Van Schagen, 2006), studies examining the relationship between the built environment and crashes have been scarce (Dumbaugh and Li, 2010).

The built environment, as described by land use patterns and road configurations, may be crucially linked to travel safety because it directly influences traffic volume and speed. For example, land use patterns along roads create travel demand, which generates traffic (Kim and Yamashita, 2002). Road systems, an integral part of the built environment, determine the maximum traffic flow obtainable on a given roadway, as well as regulating and influencing traffic speed, which is a key determinant of crash severity (Kmet and Macarthur, 2006). Land access to roads, such as curb cuts and intersections, influences the level of conflicts between vehicles, between vehicles and pedestrians, and between vehicles and the environment, which subsequently impacts the type, frequency, and severity of crashes (Aarts and Van Schagen, 2006).

Existing studies have shown that certain land use types are more likely to cause accidents than others. Kim and Yamashita (2002) examined the relationships between motor vehicle crashes and land use using overlay analysis and descriptive statistics. Their findings showed that most crashes happen in built-up urban areas with mixed residential and commercial land use. Dumbaugh and Li (2010) examined the relationships between the built environment and vehicle crashes and found strip commercial uses and big box stores to be major risk factors for vehicle accidents. Ukkusuri et al. (2012) found that there was a

* Corresponding author at: 287 Dow Science Complex, Department of Geography, Central Michigan University, Mount Pleasant, MI 48859, USA.
E-mail addresses: yhuang@email.sc.edu (Y. Huang), wang9x@cmich.edu (X. Wang), patto1d@cmich.edu (D. Patton).

https://doi.org/10.1016/j.jtrangeo.2018.04.027
Received 4 October 2017; Received in revised form 24 April 2018; Accepted 28 April 2018
0966-6923/

greater likelihood of vehicle–pedestrian crashes in census tracts with a greater proportion of industrial, commercial, and open land and schools than in tracts with a greater proportion of residential land use. Ha and Thill (2011) applied kernel density estimation and spatial modelling to identify systematic variations of collision hazard intensities across neighborhoods. Their findings showed that vehicle–pedestrian crashes were more common in business districts and in densely settled residential neighborhoods with large and less affluent populations.

In addition to land use types, road configuration and roadway design have also been shown to influence the likelihood and severity of crashes. Noland (2003) found infrastructure elements such as road categories and number of lanes to be highly correlated with crash frequency and severity. They determined that network characteristics such as intersection density, speed limit, and transit access were good predictors for both vehicle-only crashes and vehicle–pedestrian crashes. Ladron de Guevara et al. (2004) found intersection density and road types were significantly related to crashes resulting in injury or property damage. Pulugurtha and Sambhara (2011) showed that the number of lanes, speed limit, and pedestrian and vehicular volume were good predictors of vehicle–pedestrian crashes. Clifton et al. (2009) found that good transit access and pedestrian connectivity were significantly and negatively associated with injury severity in vehicle–-pedestrian crashes. Dai (2012) found that suburban corridors with high levels of activity had significantly increased injury risks for vehicle–-pedestrian crashes than other areas.

1.2. Methodological challenges

Most existing traffic safety studies have relied on global statistical models to disentangle the relationships between crashes and various characteristics of the built environment. Global statistical models assume that the relationships between explanatory variables (crash-inducing factors) and response variable (s) (crash-related variables) are consistent over space. Global relationships among variables are typically modelled and estimated based on all available observations in a study area. Commonly used global models in the existing traffic safety literature include Poisson and negative binomial regression models (Dumbaugh and Li, 2010), generalized ordered probit models (Clifton et al., 2009), logistic regression models (Ossenbruggen et al., 2001), multivariate models (Clifton and Kreamer-Fults, 2007), and hierarchical Bayesian models (Li et al., 2007).

Although global models are powerful in discerning global relationships between crashes and crash-inducing factors, one major limitation of such a modelling approach is that the models are non-spatial: they ignore the special characteristics of spatial data, such as spatial autocorrelation and spatial non-stationarity (Anselin, 1994). Most global statistical models require observations to be independent; however, spatial phenomena, including crash occurrences, are usually spatially correlated (Anselin, 1993): accident "black spots" places where traffic accidents are historically and disproportionately concentrated, are found in many communities (Geurts et al., 2004; Mandloi and Gupta, 2003; Shao et al., 2008). Global statistical models also assume that the modelled relationships between explanatory variables and a response variable are consistent over space (spatial stationarity) (Miller and Hanham, 2011). However, spatial relationships at one location may be different from those in other locations (spatial non-stationarity). The "global" parameters estimated by using all available observations may not be adequate to reveal local relationships in different parts of the study area, because they are unable to incorporate spatial autocorrelation and spatial non-stationarity. Global models could create unstable parameter estimates and yield unreliable significance test results (Nam and Song, 2008).

Researchers are calling for more spatially explicit modelling techniques that are capable of revealing spatial processes and solving spatial issues (Atkinson, 2005). Spatial lag, spatial error, Bayesian hierarchical models, and geographically weighted regression (GWR) models are common spatial models. Spatial lag model, spatial error model, and GWR model are commonly used in the field of geography. Spatial lag model is used when there is spatial correlation in the dependent variable and spatial error model is used when there is spatial correlation in the residuals (Anselin et al., 1996). However, neither spatial lag model nor spatial error model can model spatial non-stationarity - the spatial variations in the relationships between the independent variable and dependent variables. Among the various spatial models, GWR is unique in that it estimates local parameters for a regression model by allowing the relationship between variables to vary over space. Given that locality is important and measuring local relationships is vital to understanding spatial processes, GWR was designed to capture local spatial relationships by generating a separate regression for every location in a dataset based on the unique spatial relationships derived from its neighbors (Brunsdon et al., 1996). The coefficients, which include signs and significant level, of the independent variables derived from GWR can yield intuitive results. GWR has been used in the fields of ecology (Wang et al., 2005), sociology (Calvo and Escolar, 2003; Wang and Chen, 2017), transportation (Cardozo et al., 2012; Selby and Kockelman, 2013), geography (Qian and Ukkusuri, 2015; Su et al., 2012), and other fields related to social science. Existing crash studies that have utilized GWR models suggest that there exist spatial dependency and spatial heterogeneity in the occurrence and severity of traffic accidents (Li et al., 2013; Park et al., 2010; Pirdavani et al., 2014) and that GWR may provide a better statistical fit than traditional ordinary least squares (OLS) models when modelling the severity of accidents (Zheng et al., 2011). However, studies analyzing the relationship between crashes and the built environment have been limited until now, as most crash studies utilizing GWR have emphasized examining the influence of networks and sociodemographic variables on traffic crashes (Li et al., 2013; Pirdavani et al., 2014).

This paper presents the results of our research utilizing a GWR model to understand the spatial relationship between crashes and the built environment in Detroit, Michigan, where 527,749 vehicle-related crashes occurred from 2007 to 2011 (MDSP, 2013). For this study, we define the Detroit region as the area within the boundaries of Wayne, Oakland, and Macomb Counties. These three counties encompass all of the urbanized areas of Detroit as defined by the U.S. 2010 Census. Our research aims to address two main questions: Does the relationship between the built environment and crashes vary across space? If so, what are the local relationships? Our GWR model aims to uncover detailed spatial relationships between the built environment and vehicle crashes. We hope our findings will help local transportation planners, traffic engineers, and land use planners to better understand the interactions between the built environment and crashes within their communities. Our research also demonstrates the feasibility and value of using spatial models in traffic, transportation, and land use studies.

2. Research method: a GWR approach

2.1. Research framework

Our research framework is presented in Fig. 1. As a first step, we conduct an OLS regression in which the dependent variable is modelled as a linear function of multiple predictors using least square approach (Brunsdon et al., 1996). Moran's I is then used to examine the presence of spatial autocorrelation in the residuals (Moran, 1950). A statistically significant *p* value of Moran's I indicates that the OLS model violates the assumption of linear regression that the errors should be uncorrelated between observations. To solve the problem of spatial autocorrelation, which may largely result from the issue of spatial heterogeneity, we perform GWR. We use different bandwidths (distances and neighbors) to find the optimal scale for GWR analysis.

The GWR model is a spatial model that extends traditional linear