Journal of Transport Geography 64 (2017) 184–194

Contents lists available at ScienceDirect

Journal of Transport Geography

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

Spatial variations in active mode trip volume at intersections: a local analysis utilizing geographically weighted regression

Hongtai Yang[a], Xiaozhao Lu[a,*], Christopher Cherry[b], Xiaohan Liu[a], Yanlai Li[a]

[a] School of Transportation and Logistics, Southwest Jiaotong University, 01117, No.111, First Section, North of Second Ring Road, Chengdu, Sichuan 610031, China
[b] Department of Civil and Environmental Engineering, The University of Tennessee at Knoxville, 321 John D. Tickle Building, Knoxville, Tennessee, United States

ARTICLE INFO

Keywords:
Spatial variations
Active mode trip volume
Intersections
Geographically weighted regression

ABSTRACT

Geographically weighted regression (GWR) models have been employed in previous studies regarding vehicular travel demands, but few studies have locally modeled walking travel demands at intersections to address the issue of spatially varying relationships. Harnessing a comprehensive collection of walking and bicycling traffic counts over 10 years in Chittenden County, Vermont, US, along with socioeconomic characteristics, transit accessibility indices, land use attributes and characteristics of intersections and roadway networks, this study utilizes GWR models to identify whether there are spatially varying relationships between active mode travel demands and ambient built-environment attributes. One Ordinary Least Square (OLS) model and two GWR models were parametrically calibrated: a full GWR model of all local variables and a mixed GWR model of both global and local variables. K-fold cross-validation method is used to select variables that significantly influence the volume of active travel modes in the OLS model. The uniform set of variables is investigated in two GWR models. Only residuals of the mixed GWR model exhibit spatial independence. The prediction accuracy of the three models is respectively compared by means of the k-fold cross-validation method. Results show that the mixed GWR model has higher prediction accuracy, while the other two models have roughly the same level of performance. We find that not all independent variables possess a spatially varying relationship with active mode volumes. The flexibility of the mixed GWR model that allows some independent variables to be global strengthens its prediction power. With these findings, transportation planners can dynamically estimate bicycle and pedestrian volumes at widespread intersections, and this geographical realism would facilitate local transportation planning, facility design, safety enhancement and operation analysis, as well as instilling new insights into interdisciplinary spatial research domain.

1. Introduction

New Urbanism, among the most influential movements in contemporary urban design and planning field, arose in the United States during the early 1980s, which is still widely reshaping urban planning practice, municipal land-use strategy and real-estate development policy. New Urbanism promotes the creation of livable places (from a single building to a residential community), the reform of built-environment design and improvements in residents' living standards and quality of life. As a result of the New Urbanism movement, a critical objective in transportation studies has been to understand how planners can encourage people to travel by active modes such as walking and cycling (Katz et al., 1994). Active travel modes play a key role in not only promoting public health, but also reducing environmental impact, diminishing energy consumption and improving travel safety. Additionally, active mode trips reinforce the ridership of transit systems, and spur the demand for building pedestrian and bicycling facilities which serve for mass transit systems.

A New Urbanist principle aims to design more compact, diverse, mixed-use and pedestrian-friendly neighborhoods that encourage people to cut down on motorized trips, reduce vehicle miles traveled (VMT) and reshape how Americans take routine trips (Cervero and Kockelman, 1997; Cervero and Radisch, 1996). However, in order to allocate limited resources for building transportation facilities that will encourage walking and bicycling trips, accurate travel demand measurement (Pulugurtha and Repaka, 2008) is a necessity. In particular, it is important to understand how, as active mode travel demand is generated across a spatial domain, the strength of the relationship between travel demand and ambient land-use features (e.g., building density, mixed uses, etc.) varies locally across that domain: at some sites the

* Corresponding author.
E-mail addresses: yanghongtai@home.swjtu.edu.cn (H. Yang), lxz@swjtu.cn (X. Lu), cherry@utk.edu (C. Cherry), liuxiaohan1@my.swjtu.edu.cn (X. Liu), yanlaili@home.swjtu.edu.cn (Y. Li).

http://dx.doi.org/10.1016/j.jtrangeo.2017.09.007
Received 2 October 2016; Received in revised form 4 September 2017; Accepted 13 September 2017
Available online 26 September 2017

relationship will be stronger and at others weaker. If there are, in fact, such variations, it would be helpful to quantify and analyze them with an intention to obtaining effective demand measurements, sustainable land-use policies and well-informed investment strategies.

This study aims at investigating spatial variations in the nexus between pedestrian counts at intersections and ambient built-environment attributes with regards to land use features, socioeconomic factors, street connectivity and accessibility indices. Regression methods are commonly used to develop the relationship between the factor of research interest (response variable) and other possible influential factors (predictors). Many assumptions underlie basic linear regression methods, one of which is spatial homogeneity: the dependent variable and independent variables follow the same relationship at different locations. However, this assumption is always questioned when dealing with spatial data because the data-generation processes may be distinctive across space. There have been quite a few models already developed to address this issue, such as Spatial Expansion method (Casetti, 1972) and Geographically Weighted Regression (GWR) (Fotheringham et al., 2003). GWR has been successfully used to estimate multimodal travel demand, including daily vehicle counts (Selby and Kockelman, 2013; Zhao and Park, 2004), transport accessibility (Du and Mulley, 2006), transit use (Chow et al., 2006), and rail demand (Blainey, 2010).

This study attempts to establish local models capable of calibrating spatial variation and visually exhibiting the varied effects in a geographic information setting using GWR, specifically aimed at bicycle and pedestrian counts. Such findings could help planners and researchers better discern what infrastructure and surrounding built-environment attributes significantly contribute to active mode travels and how their effects on pedestrian demands vary across a spatial coverage. With a deepened understanding, it is possible to fill the voids in sparse collections of pedestrian counts and determine the true fraction of local pedestrian activities. To that end, the remainder of the paper consists of the following sections. Section 2 surveys the historical literature relevant to the research theme. Section 3 describes the data collection activity plus a multifaceted integration process. Section 4 states a spatial analysis methodology for parametrically estimating local models. Section 5 presents the modeling results. Sections 6 concludes with some remarks.

2. Literature review

Many studies have been performed to analyze the relationship between pedestrian, bicycle count at an intersection and ambient environment. The total number of pedestrians and bicycles crossing any leg of an intersection during a specified time period is often modeled. Four types of variables are commonly used in modeling: population density, land use information, characteristics of the intersection, and transit accessibility. The selection of variables and coefficients of these variables are different across models. There are three possible explanations: (1) there exist unobserved variables that are not unique to each community; (2) the selection of potential predictors (population density with 0.5 mile buffer vs. population density with 0.2 mile buffer) are different across studies; (3) some of the potential predictors are highly correlated (such as population density and density of people under age 18) so that a variable could easily be replaced by its highly correlated variable in a model. Despite these differences, previous studies reveal a similar pattern: pedestrian count has a positive relationship with population density, land use mix, number of buildings, proximity to downtown and existence of sidewalk, and a negative relationship with road slope and household automobile availability. A summary of previous bibliographies is presented in Table 1, which is a synoptic update of intersection-based pedestrian count studies performed by Schneider et al. (2012).

Mathematical models and geographical science have also been used to investigate behaviors of people and other entities, including modeling movements of people and goods as well as spatial locations of people (Walker, 2006). It is noteworthy that only global models are applied to model intersection-based pedestrian and bike counts (Pulugurtha and Repaka, 2008; Schneider et al., 2009; Schneider et al., 2012) in previous studies. With the evolution of methodological approaches, an awareness of the importance of geography has increased. Varying relationships between transportation related metrics and ambient environment have been observed by allowing the model parameters to vary over space. Entities at nearby locations often share more similarities than do entities that are far apart, and this notion is usually termed “Tobler's first law of geography” (Tobler, 1970). Accordingly, a GWR approach has been developed to study the varying relationship between metrics of interest and ambient environment by putting more weight on points close to the study point when estimating model parameters (Brunsdon et al., 1996; Fotheringham et al., 2003; et al., 2004).

Until now, GWR methods have been used in multimodal travel demand studies, including daily vehicle counts (Selby and Kockelman, 2013; Zhao and Park, 2004), transport accessibility (Du and Mulley, 2006), transit use (Chow et al., 2006), travel behavior (Srinivasan, 2010), zonal population (Vichiensan et al., 2006), and rail demand (Blainey, 2010). These studies have shown that local models validated spatial variations and improved prediction accuracy over global models. However, to the best knowledge of the authors, no previous studies have *locally* modeled walking and biking travel demands at widespread intersections, nor have any ones explored the inherent spatially varying relationship between active mode demands and infrastructure and land use variables. The question of whether GWR can perform better than global models in this regard remains to be explored. To begin to respond to this question, pedestrian and bike counts, along with certain attributes of street networks, characteristics of intersections, demographics and land use features were collected at several hundred intersections in Chittenden County, Vermont, U.S. in order to develop local GWR models. The GWR models are used to identify how spatial factors influence travel demand estimation locally.

3. Data preparation

This study combined multiple comprehensive sets of field data collected in Chittenden County. The first set is an assembly of walking and biking traffic counts that were garnered at hundreds of intersections. Next, shape files containing the point location information for intersections and surrounding infrastructure that can be used in geographic information systems (GIS) are prepared.

Out of 14 counties in Vermont, Chittenden County is the most densely populated, which also possesses unique characteristic from spatial modeling perspectives. First, the county is a relatively isolated 3-hour drive from the nearest major American and Canadian cities, which means it can be modeled as a “closed economic system” (a frequently held but often violated assumption in land-use related modeling) (Voigt et al., 2009). Second, the county is an excellent place to study spatial patterns of walking and biking traffic demands because it has a fairly extensive network of dedicated infrastructure for pedestrians and bicyclists. Prior to this study, from 2000 to 2009, the activities of pedestrians (and much fewer bicyclists) had been systematically counted at 428 intersections by the Chittenden County Metropolitan Transportation Organization (CCMPO) via manual traffic counts. The counts were conducted in different months, days and hours in one year or multiple years of the 10-year period. Almost all counting sites spanned the 4:00–6:00PM peak period, and many of them included up to 12 hours of counts (on average 8 hours per site). The counts at all intersections were compiled as hourly totals. The initial dataset consisted of 3,541 hourly records. In total, five records, with hourly counts larger than 500 pedestrians, were regarded as exceptional outliers and disregarded, which reduced the dataset to 3,536 records.

It is critical to understand seasonal, weekly, daily and hourly