Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership

Xiaolei Ma[a,b], Jiyu Zhang[a], Chuan Ding[a,b,*], Yunpeng Wang[a,b]

[a] School of Transportation Science and Engineering, Beijing Key Laboratory for Cooperative Vehicle Infrastructure System and Safety Control, Beihang University, Beijing 100191, China

[b] Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100191, China

ARTICLE INFO

Keywords:
Transit ridership
Built environment
Spatio-temporal analysis
Traffic analysis zone

ABSTRACT

Understanding the influence of the built environment on transit ridership can provide transit authorities with insightful information for operation management and policy making, and ultimately, increase the attractiveness of public transportation. Existing studies have resorted to either traditional ordinary least squares (OLS) regression or geographically weighted regression (GWR) to unravel the complex relationship between ridership and the built environment. Time is a critical dimension that traditional GWR cannot recognize well when performing spatiotemporal analysis on transit ridership. This study addressed this issue by introducing temporal variation into traditional GWR and leveraging geographically and temporally weighted regression (GTWR) to explore the spatiotemporal influence of the built environment on transit ridership. An empirical study conducted in Beijing using one-month transit smart card and point-of-interest data at the traffic analysis zone (TAZ) level demonstrated the effectiveness of GTWR. Compared with those of the traditional OLS and GWR models, a significantly better goodness-of-fit was observed for GTWR. Moreover, the spatiotemporal pattern of coefficients was further analyzed in several TAZs with typical land use types, thereby highlighting the importance of temporal features in spatiotemporal data. Transit authorities can develop transit planning and traffic demand management policies with improved accuracy by utilizing the enhanced precision and spatiotemporal modeling of GTWR to alleviate urban traffic problems.

1. Introduction

The rapid increase in private car ownership aggravates metropolitan traffic congestion, thereby causing a series of issues, such as air pollution, high energy consumption, and accidents (Ding, Wang, Liu, Zhang, & Jiawen Yang, 2017). A possible countermeasure to alleviate the aforementioned negative impacts is to prioritize public transportation (Chakour & Eluru, 2016). Transit authorities aim to optimize public transportation planning and improve service quality to achieve the goal of promoting public transit systems, and ultimately, increase the attractiveness of public transit. In particular, identifying the key determinants that affect transit ridership and analyzing the spatial and temporal evolution of influences is crucial (Taylor & Fink, 2003). A thorough understanding of the factors that influence transit ridership can enable transit authorities to efficiently allocate the limited resources for the deployment of transit service and to develop additional targeted policies for pricing and investments.

In the past decades, studies on the influential factors of transit ridership can be divided into two categories. The first category comprises descriptive research using traveler attitudes and perceptions (Brown, Hess, & Shoup, 2001; Dueker, Strathman, & Bianco, 1998; Mineta, 2002). Most descriptive analyses aid transit operators and emphasize internal factors, such as fare innovation, marketing change, and other strategies or programs. Descriptive analyses generally use survey and interview data. Essential data are highly subjective and dependent on the perceptions and assumptions of respondents regarding internal and external factors related to ridership (Dueker et al., 1998; Stanley, 1998). The data are likely biased given the limited or incorrect information. Therefore, inherent data deficiency leads to reduced usage of descriptive study on transit ridership (Taylor & Fink, 2003). The second category consists of causal analyses to examine the influential factors of transit ridership (Hartgen & Mather, 1994; Kohn, 2000; Syed & Khan, 2000). Causal analyses have been conducted in complex empirical studies because such studies commonly use more objective and heterogeneous data sources than descriptive studies (Taylor, Miller, Iseki, & Fink, 2009). The factors that are hypothesized

* Corresponding author at: School of Transportation Science and Engineering, Beijing Key Laboratory for Cooperative Vehicle Infrastructure System and Safety Control, Beihang University, Beijing 100191, China.

E-mail addresses: xiaolei@buaa.edu.cn (X. Ma), jiyu1613108@buaa.edu.cn (J. Zhang), cding@buaa.edu.cn (C. Ding), ypwang@buaa.edu.cn (Y. Wang).

https://doi.org/10.1016/j.compenvurbsys.2018.03.001
Received 13 December 2017; Received in revised form 2 March 2018; Accepted 2 March 2018
0198-9715/

to influence ridership and the variables that are operationalized in causal analyses are more diverse than those in descriptive research (Holmgren, 2007). These models explain ridership based on internal and external variables. External factors have a greater impact on ridership than internal factors (Chung, 1997; Gomez-Ibanez, 1996). Furthermore, the built environment, as a key component of external factors, actively changes the travel behavior of people (Ding, Chen, & Jiao, 2018), thereby resulting in the fluctuation of transit ridership (Cao, Mokhtarian, & Handy, 2007; Cao, Mokhtarian, & Handy, 2009; Wang, Chai, & Li, 2011). A few existing studies have acknowledged the significant contribution of land use attributes, including density (Cervero, 2002; Lee & Moudon, 2006), land use mixture (Cervero & Kockelman, 1997), and other relevant factors, to public transit attractiveness.

Two representative transit ridership levels, namely, station and region, are considered dependent variables in studies on the influential factors of transit ridership. Research on station-level ridership counts average daily boarding or alighting passengers at each bus stop or subway station, whereas that on region-level ridership is more macroscopic, with a summary of passenger count in an administrative region or a traffic analysis zone (TAZ). Chakour and Eluru (2016) quantified the influences of the attributes of the built environment on bus stop-level boarding and alighting ridership in Montreal. They concluded that improving public transport service and accessibility would be the most effective measure to attract bus passengers. Thompson, Brown, and Bhattacharya (2012) built a ridership estimation model with the built environment and transit travel price. The results suggested that traffic operators should focus on service in decentralized employment centers. A large number of variables are used to explain ridership. Studies on the influence of the built environment on transit ridership are summarized in Table 1, including the explanatory variables used in these studies.

Ordinary least squares (OLS) regression is the most representative and widely used approach among statistical methods for unraveling the complex relationship between the built environment and transit ridership (Sohn and Shim, 2010; Sung and Oh, 2011; Guerra, Cervero, & Tischler, 2011; Zhao et al., 2013). In the OLS model, the basic assumption is that station- or region-level ridership data are independent and stationary in space. However, ridership data from a particular place do not conform to the independence hypothesis because of the local interaction and spatial nonstationarity among its adjacent stations and zones. Thus, the applicability of the OLS approach to transit ridership modeling has been criticized for neglecting spatial variation (Jun et al., 2015). The spatial nonstationarity is that the ridership is sensitive to urban form and location while the stations will impact each other (Qian & Ukkusuri, 2015). Spatial nonstationarity among different ridership units will cause the estimation coefficients of the explanatory variables to vary spatially across different observations (Clark, 2007). Lloyd (2010) emphasized the necessity of integrating spatial nonstationarity into the regression model for spatial heterogeneity of transit ridership.

Several OLS-like extended models have been proposed to consider spatial heterogeneity in parameter estimation and to overcome the drawback of neglecting the spatial autocorrelation effect in the traditional OLS method. Typical examples include distance-decay weighted regression (Gutiérrez, Cardozo, & García-Palomares, 2011), two-stage least squares (2SLS) regression (Estupiñán & Rodríguez, 2008; Taylor et al., 2003), the passion model (Chu, 2004), and the geographically weighted regression (GWR) model (Jun et al., 2015; Qian & Ukkusuri, 2015). Among these models, GWR is specifically designed to deal with spatial data regression (Brunsdon, Fotheringham, & Charlton, 1996; Fotheringham, Charlton, & Brunsdon, 1998). The essence of the GWR model is to design a weight matrix for each observation; the matrix depends on the distance between the locations of observations. This feature allows the GWR model to capture the spatial pattern of data effectively via spatial varying coefficients (Cardozo et al., 2012), and thus, the GWR model has been widely used in transportation planning. Zhao et al. (2013) developed a GWR model to investigate spatial variations in the relationship between ridership and potential contributing factors and then demonstrated the fluctuation of this relationship over the entire study area. Cardozo et al. (2012) found that the GWR model yielded better fitting result than the OLS model in station-level transit ridership forecasting. Qian and Ukkusuri (2015) related taxi ridership to two categories of explanatory variables using GWR. The results provided valuable insights into taxi demand prediction. Jun et al. (2015) investigated the influence of a mixed GWR on station-level ridership. The aforementioned examples demonstrate the superior capability of the GWR model in determining the spatial dependencies of transit ridership at station or TAZ scale.

However, when modeling spatiotemporal data (e.g., ridership, traffic count, and house price) using GWR, the input (i.e., dependent variable) requires being aggregated or averaged by a certain period, such as average annual traffic data and daily boarding passengers. Time is another critical dimension that cannot be adequately learned by traditional GWR models. This condition is particularly true when modeling transit ridership in a TAZ where morning and evening peak passenger flows are over 50% of the daily ridership. It indicates that ridership is also temporal nonstationarity. Similar to spatial nonstationarity, temporal nonstationarity represents that the ridership sensitive to the time and will be influenced by the historical ridership (Lin & Shin, 2008). The spatiotemporal variations of ridership should be considered (Goodchild, 2013). That is Spatiotemporal analysis has always been a popular topic in transport studies (Lockwood, Srinivasan, and Bhat, 2005; Hanson & Huff, 1988). Over the past decade, the

Table 1
Summary of the literature review on the impact of the built environment on transit ridership.

Author	Dependent variable	Model	Key explanatory variables
Qian and Ukkusuri (2015)	Taxi ridership in the zip code tabulation areas	Geographically weighted regression	Road density, bike lane density, parking spaces, subway and bus accessibility
Taylor et al. (2009)	Transit ridership for each of the 265 urbanized areas	Two-stage simultaneous equation regression	Regional geography, metropolitan economy, population characteristics, auto/highway system, transit system characteristics
Estupiñán and Rodríguez (2008)	BRT station boarding passengers	Two-equation simultaneous model	Station characteristics, physical attributes, perceived characteristics, neighborhood attributes
Jun, Choi, Jeong, Kwon, and Kim (2015)	Station-level ridership of the pedestrian catchment areas	Mixed geographically weighted regression	Population and employment densities, mixed land use, intersection density, road density, number of bus stops
Cardozo, Garca-Palomares, and Gutirrez (2012)	Stop-level boarding passengers	Geographically weighted regression	Land-use mix, street density, number of metro lines, number of urban bus lines, number of suburban bus lines
Thompson et al. (2012)	TAZ-level commuting transit ridership	Negative binomial regression	Total population, total employment, walkability, parking fees, local in-vehicle travel time, size of destination zone
Zhao, Deng, Song, and Zhu (2013)	Ridership within the pedestrian catchment area of metro stations	Ordinary least squares regression	Area of residential, office, and other-use buildings; number of educational institutions, hotels, restaurants, entertainment venues, shopping centers, and hospitals; road length; number of feeder bus lines
Taylor, Miller, Iseki, and Fink (2003)	Total transit ridership/transit ridership per capita	Two-stage least squares regression	Service supply, service attributes, regional and urban characteristics