Journal of Transport Geography 69 (2018) 45–57

Contents lists available at ScienceDirect

Journal of Transport Geography

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

Spatial variations in urban public ridership derived from GPS trajectories and smart card data

Wei Tu[a,b,*], Rui Cao[d], Yang Yue[a,b], Baoding Zhou[a,b], Qiuping Li[e], Qingquan Li[a,b,c]

[a] Shenzhen Key Laboratory of Spatial Information Smart Sensing and Services, Research Institute for Smart Cities, School of Architecture and Urban Planning, Shenzhen University, Shenzhen, PR China

[b] Key Laboratory for Geo-Environmental Monitoring of Coastal Zone of the National Administration of Surveying, Mapping and GeoInformation, Shenzhen University, Shenzhen, PR China

[c] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, PR China

[d] International Doctoral Innovation Centre, University of Nottingham, Ningbo, PR China

[e] Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, PR China

ARTICLE INFO

Keywords:
Ridership
Big data
Trajectory
Smart card data
Geographically weighted regression
Transit
Taxi

ABSTRACT

Understanding urban public ridership is essential for promoting public transportation. However, limited efforts have been made to reveal the spatial variations of multi-modal public ridership (such as buses, metro systems, and taxis) and the underlying controlling factors. This study explores multi-modal public ridership and compares the similarities and differences of the associated factors. Daily bus, metro, and taxi ridership patterns are first extracted from multiple sources of big transportation data, including vehicle (bus and taxi) GPS trajectories and smart card data. Multivariate regression analysis and geographically weighted regression analysis are used to reveal the associations between these data and demographic, land use, and transportation factors. An empirical study in Shenzhen, China, suggests that employment, mixed land use, and road density have significant effects on the ridership of each mode; however, some effects vary from negative to positive across the city. The results also indicate that road density, income, and metro accessibility do not have significant effects on metro, transit or bus ridership. These findings suggest that the effects of the associated factors vary depending on the mode of travel being considered and that the city should carefully consider which factors to emphasize in formulating future transport policy.

1. Introduction

Faced with the increasing social and environmental challenges posed by overpopulation, traffic congestion, climate change, and energy consumption, many governments worldwide have prioritized public transportation and focused on improving public transportation services (Vuchic, 2007). Recently, China began the Thirteenth Five-Year Program, which is intended to increase the public transport usage rate to more than 50%, thus alleviating traffic congestion and reducing transportation-related emissions in the cities (MOT China, 2017). However, public transportation continues to be used relatively infrequently in many countries (APTA, 2016). For example, in the United States, only 5.3% of trips are made using general public transportation system (including buses, light rail, subways, metro systems, and taxis) in many cities (PTFT, 2016). Understanding urban public transportation ridership is essential for promoting public transportation.

Numerical studies have quantified urban ridership. The most widely used models are the aggregated four-step model (McNally, 2007) and the disaggregated activity-based model (Bowman and Ben-Akiva, 2001). Both of these models offer a basis for quantifying urban ridership; however, they have weaknesses, including their high cost, low accuracy, and lack of elasticity (Gutiérrez et al., 2011; Zhao et al., 2014). An alternative approach that integrates geographic information systems (GIS) and multivariate analysis has been developed to explore urban ridership. The direct association between the ridership of one mode public transportation (i.e., buses, metro systems, light rail, taxis) and potential demographic, land use, and transportation factors has been revealed (Taylor and Fink, 2003; Kuby et al., 2004; Chow et al., 2006; Yao, 2007; Zhang and Wang, 2014; Kepaptsoglou et al., 2017). For example, Gutiérrez et al. (2011) developed a distance-decay weighted regression function to model the metro ridership in Madrid. Chakour and Eluru (2016) examined the association of bus ridership and stop-level infrastructure and the built environment in Montreal using a composite marginal likelihood-based ordered response probit

* Corresponding author at: Room 1402, Science and Technology Building, Shenzhen University, No. 3688 Nanhai Avenue, Nanshan District, Shenzhen, Guangdong 518060, PR China.
E-mail address: tuwei@szu.edu.cn (W. Tu).

https://doi.org/10.1016/j.jtrangeo.2018.04.013
Received 18 June 2017; Received in revised form 2 March 2018; Accepted 11 April 2018

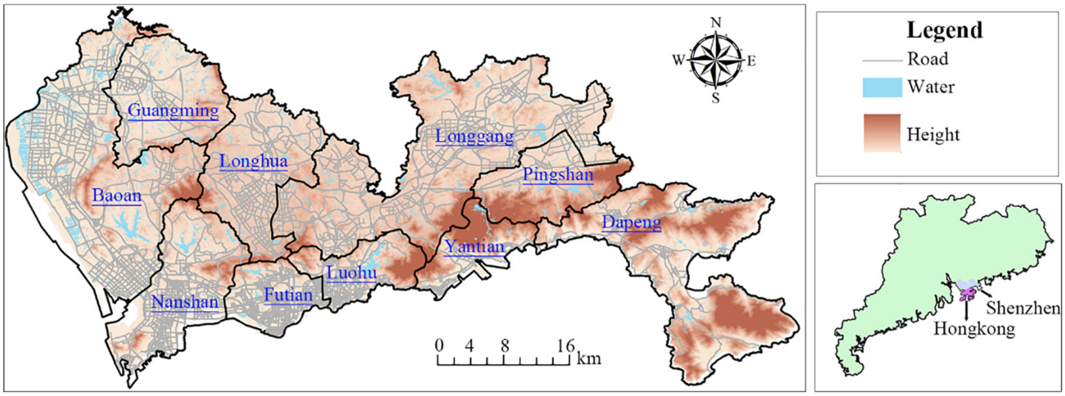

Fig. 1. Study area.

model. Qian and Ukkusuri (2015) uncovered the relationship between taxi ridership and socio-demographic and built-environment variables in New York City. Focusing on subway operational factors, Ding et al. (2016) devised a machine learning method known as gradient boosting decision trees (GBDT) to estimate short-term station-level subway ridership. Empirical studies in many cities worldwide suggest that the significant factors for urban ridership include population (Cardozo et al., 2012; Zhao et al., 2014), employment (Gutiérrez et al., 2011; Cardozo et al., 2012), land use (Chakraborty and Mishra, 2013; Hu et al., 2016), the presence of residential areas (Chiou et al., 2015), road density (Chiou et al., 2015), the distance to metro stations (Qian and Ukkusuri, 2015) and the presence of points-of-interest (Hu et al., 2016) Few studies have attempted to untangle the relationships between multi-modal public ridership and the potential controlling factors and compare them, thus providing a comprehensive understanding for decision makers to promote the sustainable development of urban public transportation. Furthermore, a case study of an Asian city is rare.

Exploring citywide public ridership confronts great challenges. In general, public transportation systems (including buses, light rail, metro systems, and taxis) are very complex. From a global view, buses, light rail, metro systems and taxis cooperate to fulfill daily travel demands and compete with one other (Cao et al., 2016; Li et al., 2017a; Zhang et al., 2018). Therefore, clarifying the influences on multi-modal public transport systems is a non-trivial task. Collecting multi-modal travel ridership data is an essential requirement. In addition, because public travels are driven by various human needs, and the level of urban functionality varies from place to place, the ridership distribution in urban areas depends strongly on the underlying urban form and functions (Pulugurtha and Aguila, 2012). Therefore, it is necessary to explicitly consider spatial variations when exploring urban ridership (Pulugurtha and Aguila, 2012; Qian and Ukkusuri, 2015). Geographically weighted regression (GWR) analysis considers local effects explicitly, thus offering a reasonable explanation for spatial phenomena (Brunsdon et al., 1996) and a promising method to reveal the underlying factors of citywide ridership.

Recently, information and communication technology (ICT) has provided us with massive amounts of useful transportation-related data, such as smart card data (SCD) (Zhong et al., 2014; Zhou et al., 2017; Li et al., 2017b), bus GPS trajectories (Ehmke et al., 2012; Tao et al., 2014) and taxi GPS trajectories (Fang et al., 2012; Liu et al., 2012; Zhou et al., 2015; Tu et al., 2016; Yang et al., 2018). Thus, we have an unprecedented opportunity to assess and characterize citywide human travel (Alsger et al., 2016; Ma et al., 2013) and fixed activities (Yue et al., 2014; Tu et al., 2017). Using SCD, bus and taxi GPS trajectories, and related GIS data, this article endeavors to reveal the spatial variations of bus, metro, transit (the sum of buses and metro systems), and taxi ridership in a city and the driving forces. The spatially resolved effects of a set of demographic, land use and transportation factors on the ridership of each form of transit are explored using GWR analysis. The results of the model indicate that bus, metro, transit (equal to total bus and metro), and taxi systems exhibit common controlling factors, including employment, land use mix, and road density. The modes nonetheless exhibit non-negligible differences, such as bus accessibility, metro accessibility, and income. Such factors significantly affect taxi ridership, but not bus, metro or transit ridership. The spatial variations of the effects are also visualized. These findings can help policy makers to attain a much higher public transportation usage.

The rest of this paper is organized as follows. The next section introduces the study area and the dataset. Section 3 describes the methodology, including the dependent and independent variables, the global regression model, and the GWR model. Section 4 reports the results and compares them. Section 5 compares the results with those of worldwide cities. The final section concludes the paper and suggests directions for future work.

2. Study area and dataset

This study uses SCD, vehicle GPS trajectories, and related GIS data collected in Shenzhen, China, to investigate the spatial variations of multi-modal urban ridership. In this section, we outline the geography of Shenzhen, as well as the public transportation systems.

2.1. Study area

Shenzhen is located in the Pearl River Delta, adjacent to Hong Kong. Since its establishment in 1979 as a Special Economic Zone (SEZ) of China, Shenzhen has become one of the largest and most innovative cities in China. Currently, Shenzhen has a residential population of over 10 million people and a mobile population of 6 million people (Shenzhen Statistics Department, 2015). It covers 1995 km^2, 800 of which are built-up areas. As shown in Fig. 1, the city includes 6 administrative districts (Luohu, Futian, Nanshan, Yantian, Baoan, and Longgang) and 4 functional districts (Guangming, Longhua, Pingshan, and Dapeng). Luohu, Futian, Nanshan and Yantian belong to the old SEZ, which were well planned and built. Futian and Luohu are recognized as the city center, and they contain high buildings and are characterized by dense employment. Nanshan is the high technology district with many innovative companies and factories. Yantian is a port area for international trade. Baoan, Guangming, Longhua, Longgang, and Pingshan are suburbs that developed independently at the beginning of the SEZ; they possess compact neighborhoods with residential communities and factories, water bodies, mountains, and other features characteristic of open spaces. Dapeng is a tourist destination, and it contains mountains and beaches.

Since its founding in 1983, Shenzhen has experienced rapid growth in the ownership of privately owned vehicles and thus a dramatic increase in the use of privately owned vehicles for daily travel. However, for most urban residents, public transport remains the backbone of