Travel Behaviour and Society 28 (2022) 38–49

Contents lists available at ScienceDirect

Travel Behaviour and Society

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

Association between built environment characteristics and metro usage at station level with a big data approach

Long Chen [a], Yi Lu [a,b,*], Yanfang Liu [c], Linchuan Yang [d], Mingjun Peng [e], Yaolin Liu [c]

[a] Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong SAR
[b] City University of Hong Kong Shenzhen Research Institute, Shenzhen, China
[c] School of Resource and Environmental Sciences, Wuhan University, Wuhan, China
[d] Department of Urban and Rural Planning, Southwest Jiaotong University, Chengdu, China
[e] Wuhan Natural Resources and Planning Bureau, Wuhan, China

ARTICLE INFO

Keywords:
Transit-oriented development
Built environment
TOD outcomes
Mode share
Commuting
Smart card data

ABSTRACT

Transit-oriented development (TOD) planning strategy has been widely implemented worldwide to formulate dense, mixed-use built environment in the past three decades. The primary goal of TOD is to promote public transit usage including both transit mode share and ridership. Research supports that built environment characteristics around metro stations affect residents' travel behaviors and metro usage. However, the evidence remains inconsistent in different urban contexts. Furthermore, research focusing on mode share such as commuting trips at station level is still scarce. In this study, a rule-based model was used to identify commuting trips using metro service with smart card data (SCD), covering more than 90 percent of all metro passengers in Wuhan, China. Built environment characteristics around metro stations were measured with a 3Ds framework (density, diversity, and design). Results suggest that population density is negatively associated with metro commuting mode share, while street intersection shows a positive relationship. Office-oriented urban function and street intersection are positively correlated with metro ridership. Hence, exploring the fine-grained relationship of metro usage and built environment factors around transit stations in different urban and social contexts warrants further research attention.

1. Introduction

Many global cities have witnessed rapid urbanization and increased private car use recently, which lead to far-reaching negative consequences such as traffic congestion, urban sprawl, environmental pollution and physical inactivity (Bertolini et al., 2012; Cervero et al., 2004; Frank et al., 2019). Urban planning interventions to reduce car use and its associated urban problems have been an important research focus in recent decades. Since the early 1990s, transit-oriented development (TOD) systematically elaborated by Calthorpe (Calthorpe, 1993) has gained extensive attention as an effective and promising approach to address these concerns.

TOD generally has the following principles: including mixed land use, dense residential population, accessible transits, pedestrain-friendly envrionment, compact urban form around transit stops (e.g., metro station, light rail station or bus stop) (Calthorpe, 1993; Cervero et al., 2004; Singh et al., 2017). TOD planning principles are widely adopted to shape urban development in North America, Europe, Australia, and some Asia governments. For example, TOD strategy is primarily implemented to limit urban sprawl and reduce car dependency in American cities (Guerra, 2014). In Europe, TOD is generally introduced in urban renewal programs (Bertolini et al., 2012).

China has been experiencing a rapid urbanization process accompanied with urban population expansion since the 1990s (Zhao et al., 2018). Many large cities in China have built or plan to build urban rail transit system as a sustainable development mode to cope with rapid urbanization and population aggregation, alleviate traffic congestion and living environment deteriorating, and boost their economy (Cervero and Day, 2008; Huang et al., 2017; Wu and Hong, 2017; Zhao et al., 2018; Zhou, 2016). Since 2011, the central government has planned a transit-oriented metropolis program to fund many large cities around the nation to implement TOD projects (Zhou, 2016).

* Corresponding author at: Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong SAR.
E-mail addresses: lochen6-c@my.cityu.edu.hk (L. Chen), yilu24@cityu.edu.hk (Y. Lu), yfliu610@163.com (Y. Liu), yanglc0125@swjtu.edu.cn (L. Yang), pmj@zrzyhgh.wuhan.gov.cn (M. Peng), yaolin610@163.com (Y. Liu).

https://doi.org/10.1016/j.tbs.2022.02.007
Received 1 August 2020; Received in revised form 15 December 2021; Accepted 22 February 2022
Available online 1 March 2022
2214-367X/

A substantial body of literature has evaluated TOD impacts on various performance and outcomes (Arrington and Cervero, 2008), such as on property values (Duncan, 2010; Pan and Zhang, 2008), relocation of job and housing (Pagliara and Papa, 2011), travel behavior (Huang et al., 2019; Wu and Hong, 2017), and station area accessibility (Lyu et al., 2020; Papa and Bertolini, 2015). Among all performance and outcomes, the primary one is to increase transit ridership and reduce car dependency (Cervero et al., 2004; Singh et al., 2017).

The built environment characteristics around transit stops can exert an influence on transit usage (Ding et al., 2019; Ewing and Cervero, 2010; Loo et al., 2010). Some studies suggest that higher density in station service area can induce travel demand and increase transit ridership (Arrington and Cervero, 2008; Nasri and Zhang, 2014). However, other studies report inconsistent findings. For instance, land use mix is positively correlated with metro ridership in Seoul (Sung and Oh, 2011) and Nanjing (Zhao et al., 2013), but not in New York (Loo et al., 2010). The inconsistent findings may be due to different urban contexts, as well as inherent data limitations associated with widely-used travel survey data (Nasri and Zhang, 2014; Park et al., 2018). For example, the sample size and study areas of travel survey is limited and may be unrepresentative to the whole population.

Recently, urban big data (e.g., smart sard data, SCD) bring new opportunities to assess travel behaviors. Big data can accurately monitor travel behaviors in more fine-grained spatiotemporal resolution for all urban residents using smart card system. Hence, SCD can make up for shortcomings of traditional travel survey data with a larger and more representative sample of passengers' travel behaviors (Huang et al., 2018; Sung and Oh, 2011; Zhou et al., 2019). Most studies using SCD focused on boarding and alighting ridership at station level (Jun et al., 2015; Loo et al., 2010; Zhao et al., 2013). Studies also investigated travel mode choice in TOD areas versus non-TOD areas by travel survey (Nasri and Zhang, 2014; Shen et al., 2016). To date, research focusing on mode share such as commuting trips at station level is still scarce. In this study, we identified commuting trips using metro service with SCD, covering more than 90 percent of all metro passengers in Wuhan, China. Metro usage for commuting mode share and metro ridership were measured to investigate association with built environment characteristics at station level.

2. Literature review

2.1. TOD around the world and in China

The potential benefits of TOD to integrate urban development with rail networks make it a popular practice in urban planning worldwide. Many American cities have seen a rebound in rail transit program to increase development density and diversity, centralize suburban sprawl, and create pedestrian-friendly urban design near transit nodes (Guerra, 2014). In Australia, for instance, Brisbane has actively carried out a strong TOD approach following the principles including diverse land use patterns and higher employment density to foster a vibrant community, higher residential density to support frequent public transport service, well-connected street network to increase street movement and pedestrian activity (Kamruzzaman et al., 2016). In Europe, TOD was able to shape suburban development into satellite suburbs along transit corridors (Singh et al., 2017), and put in practice in urban redevelopment programs (Bertolini et al., 2012).

China is a newcomer in TOD implementation. However, TOD and metro construction projects have dramatically increased in the past decade. In 2016, more than 3000 km of metro lines has been constructed and another 3000 km has been planned with an annual investment of 300 billion Yuan (Zhao and Li, 2018). The China central government encourages metro construction to integrate land use with transportation development, and shape sustainable mobility in the future. Nonetheless, only a handful of studies focused on metro ridership and mode choice with TOD in China such as Shanghai and Nanjing (An et al., 2019; Shen et al., 2016; Zhao et al., 2013). The relationship between TOD outcomes and the built environment characteristics in station area remains unclear under Chinese urban context.

2.2. TOD's performances and outcomes

Many studies suggest that TOD strategy can help to achieve a host of social, economic, and environmental goals with intense and mixed land use, healthy and equitable mobility. Compelling evidence, from San Diego, Seoul, Beijing and Shanghai, confirms that TOD can improve transit accessibility and increase urban density (Duncan, 2010; Pan and Zhang, 2008; Sung and Oh, 2011; Zhang and Wang, 2013). An appropriate TOD planning achieves more efficient land use patterns, which is able to attract residential and commercial activities, and prevent urban sprawl in the long run (Ratner and Goetz, 2013; Singh et al., 2017). Additionally, if combined with favorable urban design, TOD can reduce car dependency and promote more sustainable travel modes, e.g. walking, cycling, and public transit (Huang et al., 2017; Loo and du Verle, 2017; Nasri and Zhang, 2014). In this sense, TOD is in line with active living planning goals that can reinforce residents' active travel behaviors in shaping more sustainable mobility (Frank et al., 2019; Zhao and Li, 2018).

Numerous studies have developed methods and tools to evaluate TOD based on the built environment characteristics around transit stops. The most widely used is the node-place model, which evaluate the balance between land use-driven demand and transportation-driven supply (Kamruzzaman et al., 2014; Li et al., 2019; Lyu et al., 2016; Singh et al., 2017). Based on this model, all station areas should strive a balance between transport and land use. Hence, stressed transit node (where demand $>$ supply) needs to increase transit service; stressed place (where demand $<$ supply) needs to promote density and diversity in station service area.

Several studies assessed the impact of TOD on travel behaviors and transit ridership (Nasri and Zhang, 2014; Pan et al., 2017; Park et al., 2018; Wu and Hong, 2017; Zhao et al., 2014). For instance, direct ridership model is widely used to investigate factors affecting transit ridership (Cardozo et al., 2012; Durning and Townsend, 2015; Jun et al., 2015; Zhao et al., 2013). Arrington and Cervero (Arrington and Cervero, 2008) analyzed 17 TOD projects and found that people living in TOD areas use transit for commuting trips more frequently compared to those living in non-TOD areas. Using household survey data collected surrounding eight rail stations, Noland and DiPetrillo (Noland and DiPetrillo, 2015) found that residents living in TOD use public transit more frequently and drive less frequently than those living farther out. Nasri and Zhang (Nasri and Zhang, 2014) suggest that people living in TOD areas tend to reduce vehicles miles traveled compared to non-TOD areas even with similar land use patterns.

2.3. Built environment factors and transit usage

In addition to proximity of transit stations (i.e., TOD and non-TOD areas) (Nasri and Zhang, 2014; Noland and DiPetrillo, 2015), built environment factors around transit stations also shape residents' travel behaviors (Ewing and Cervero, 2010; Sung and Oh, 2011; Vergel-Tovar and Rodriguez, 2018). These built environment factors may play more important role than transit service (i.e., frequency, speed of rail operation) and passengers' socioeconomic status in determining transit ridership (Durning and Townsend, 2015; Loo et al., 2010; Sung and Oh, 2011; Walton and Sunseri, 2010). In a synthesis review (Ewing and Cervero, 2010), a 3Ds framework, including Density, Diversity and Design, have been identified as major built environment factors shaping various travel mode choice, e.g., transit ridership.

Density is considered the most influential factor. High urban density attracts the concentration of population, employment and destinations to support public transit use (Durning and Townsend, 2015; Nasri and Zhang, 2014; Vergel-Tovar and Rodriguez, 2018). However, researchers

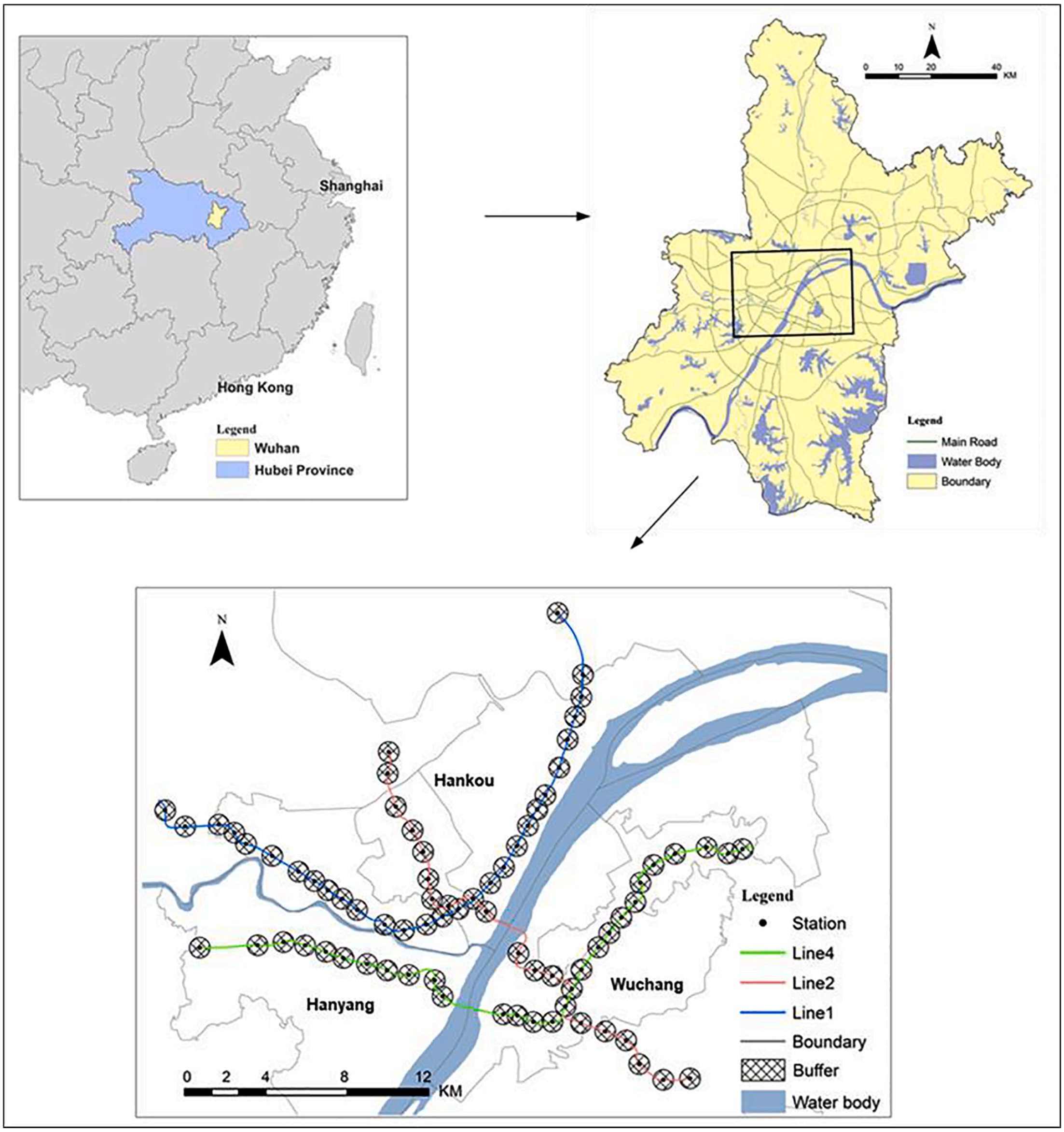

Fig. 1. The study area in Wuhan and 500-m radius station served buffers.

disagree with which types of density is more influential. Using household travel survey data, Park and colleagues (Park et al., 2018) found that density plays weaker role on transit use than street network design and land use diversity. Nasri and Zhang (Nasri and Zhang, 2014) suggested that higher employment density is linked with higher transit use based on travel survey. Loo and colleagues (Loo et al., 2010) indicated that population density and commercial-related land use explain rail ridership in New York and Hong Kong using smart card data. Similarly, positive association of population, employment, residential and commercial density with metro ridership was found in Seoul (Jun et al., 2015; Sung and Oh, 2011).

Diversity also shapes public transit use behavior. Mixed land use increase the propensity of walking to and from transit stops for different purposes (Ewing and Cervero, 2010). It has been confirmed in some studies (Jun et al., 2015; Loo et al., 2010; Park et al., 2018; Sung and Oh, 2011; Vergel-Tovar and Rodriguez, 2018). For example, Park and colleagues (Park et al., 2018) found that transit use is strongly correlated with land use diversity in station areas of eight U.S. metropolitan areas. Similar finding was revealed that higher land use mix associated with less vehicle miles traveled and more walking, cycling and transit use (Nasri and Zhang, 2014).

Design, such as street network design, also influences public transit usage. Design of street network enhances accessibility to various destinations as well as transit stations, and thereby increase transit ridership (Park et al., 2018). In particular, well-connected street networks around transit stations can shorten walking distance and provide more route options for transit use (Ewing and Cervero, 2010; Nasri and Zhang, 2014). Evidence from New York, Hong Kong and Seoul show that rail transit stations located in areas with more street intersections associated with more passengers (Loo et al., 2010; Sung and Oh, 2011).

2.4. The research gaps

Even though many studies claim that built environment characteristics of TOD areas could impact transit use, there are three major research gaps. First, the associations between built environment factors and transit usage are still inconclusive. For instance, some studies suggest that land use mix is strongly corelated with transit use and ridership (Jun et al., 2015; Park et al., 2018), while such association is not found in some other studies (Lin and Shin, 2008; Loo et al., 2010). The inconsistent finding is partially due to different urban contexts with distinctive built environment characteristics. Existing studies usually

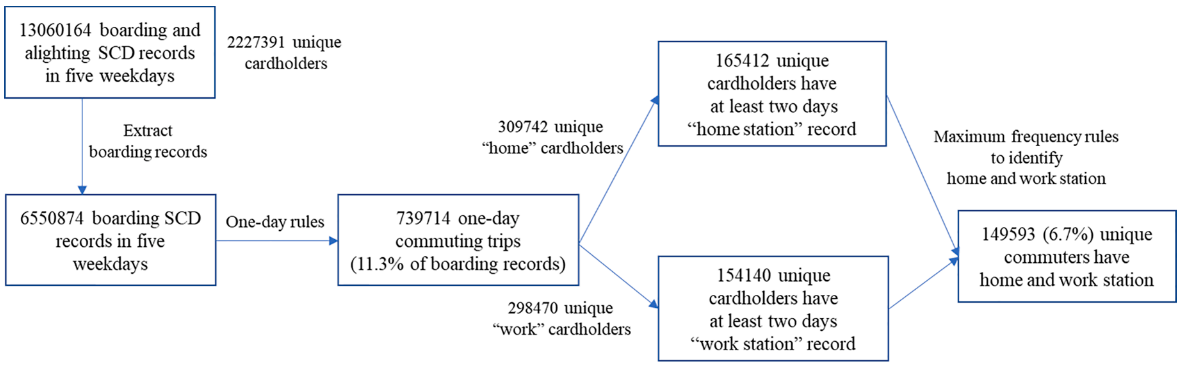

Fig. 2. The process to identify metro commuters.

focused on cities in developed countries, whereas transit usage and TOD in China has not been fully investigated.

Second, traditional travel survey has inherent data limitations such small sample size which could cause potential bias (Nasri and Zhang, 2014; Park et al., 2018). Recently, the emerging of big data such as transit smart card data (SCD) provides new opportunities to assess transit usage. The new data sources have overcome such bias (Huang et al., 2018). SCD has large data volume with abundant spatiotemporal travel information of cardholders compared with traditional small travel survey data. Large datasets generated by SCD provide accurate records of cardholders' trip information to better represent travel behaviors (Huang et al., 2018; Pelletier et al., 2011). Although such data sources lacking specific travel purposes, many aspects of passenger travel behaviors can be inferred from smart card data when travel patterns are regular. There are many studies have been able to infer origin and destination from smart card data (Huang et al., 2018; Pelletier et al., 2011; Zhou and Long, 2014). For instance, Hasan and colleagues (Hasan et al., 2013) proposed a simple frequency-based mobility model predicting that the most frequently visited places were home locations for individuals, whereas the second-most-visited paces were considered as the work locations. Studies also developed rule-based models (e.g., time threshold and frequency) for fulltime worker to identify home and work locations (Huang et al., 2019; Zhou and Long, 2014).

Third, it remains unclear what built environment factors at station level could impact metro usage, especially for commuting trips. Even though we already know that the residents in TOD areas tend to use public transit to commute compared with non-TOD areas by survey (Arrington and Cervero, 2008; Shen et al., 2016; Wu and Hong, 2017). Studies also developed regression models for ridership characteristics with built environment characteristics at station level (An et al., 2019; Sung and Oh, 2011; Zhao et al., 2013) and station-to-station level (Gan et al., 2020; Zhao et al., 2014). However, research focusing on metro mode share such as commuting trips at station level is still scarce.

In this study, we investigated the association between built environment characteristics and metro usage (i.e., metro mode share and ridership for commuting trips, and overall metro ridership) at station level with big data from SCD in Wuhan, China. More specifically, we developed a rule-based model using SCD to identify commuting trips with metro service. And metro mode share for commuting trips was measured. This study could contribute TOD research on mode choice at station level. The findings from this study would generate insights under Chinese urban contexts and help decision-makers improve metro usage of commuting trips through built environment interventions around transit stations.

3. Method

3.1. Data and study area

Wuhan is the capital city of Hubei province and the economic, political and cultural center in central China. As of 2015, Wuhan accommodated 10.3 million residents and most of the dwellers live in the main urban area. For example, Hankou district, the most representative urban area, had a resident population of 2.5 million within an area of 148.6 km^2. Its population density reaches a level of 16,824 person/km^2 in this district, making it a densely populated region.

To investigate the relationship between built environment factors and transit-oriented development (TOD) outcomes (i.e., metro commuting mode share, commuting ridership and metro ridership), we focused on metro station service area within a 500-meter radius buffer which is an appropriate scale to measure station catchment area of the stations. In 2015, there are three metro lines including 75 metro stations in Wuhan. The metro station service area is within urbanized region that is Hankou, Wuchang, and Hanyang district, which is regarded as the main urban area. Fig. 1 illustrates our study area, urban metro lines and the 500-m radius station served buffers in this study.

3.2. Outcomes of metro usage

We identified commuting trips and overall ridership with metro from one-week metro smart card data (SCD) from March 23 to 29 in 2015. Among the three metro lines in 2015, smart cards have been swiped around 2.5 million times per day, covering more than 90% of metro trips in Wuhan. The smart card data used in this study has a set of cardholders' travel information including unique card ID number (which may represent unique cardholders because most people used only one card during the study period), boarding or alighting stations, boarding or alighting time.

The overall metro ridership for each station was directly counted with SCD and further classified by different periods of a day (peak-hour 7AM-9AM vs. non-peak-hour 9AM-12AM), different days of the week (weekday vs. weekend) (Sung and Oh, 2011). We used Tuesday and Saturday SCD record to represent weekday and weekend respectively. There are 2,565,858 boarding and alighting records for weekday, and 3,168,024 for weekend.

We further extracted metro commuting trips with a rule-based model using the five-day weekday SCD records of one week. Two-day weekend SCD records were excluded to avoid bias because we assume that most commuting trips occur in weekdays. SCD in weekdays contain 13,060,164 records, and 2,227,391 unique cardholders in total. Firstly, we retrieve daily metro travel diary for all cardholders by combining their metro trips in each day. Then, we developed a one-day job-housing commuting model to identify the home and work location (represented by metro station). For cardholder with smart card number N, daily boarding travel dairy (Eq. (1)) can be summarized as:

$$N : [t_1 : s_1, t_2 : s_2, \ldots, t_k : s_k] \tag{1}$$

In this travel dairy, t represents time ($t_i < t_j$), s stands for station, and k donates number of metro boarding records. Following this, the boarding time, boarding station on each day can be reconstructed for

Fig. 3. Community and the built environment factors in station served buffer.

Table 1
Mid-categories in Gaode Map of three POI categories.

Categories	Micro-level categories in Gaode Map	Count	Percentage
Residential POI	Residential building	2321	9.1%
Commercial POI	Clothing store, sports store, Chinese food restaurant, fast food restaurant, hotel, recreation center	15,396	60.3%
Office POI	Company, enterprises, bank, finance & insurance service institution, governmental organization	7797	30.6%

each cardholder using this data model. The home (H) and work (W) location are identified using the boarding travel diary following the one-day rules (Eq. (2)) below.

$$H = s_1 \& t_1 \leqslant 10:00$$

$$W = s_p \& t_p \geqslant 16:30 \& t_{p-1} < 16:30 \ (2 \leqslant p \leqslant k) \tag{2}$$

The model rules can be interpreted that, for a boarding travel diary of a commuter who has a full-time job, the home location is supposed to be the first departure station before 10:00 am; the work location should be the first boarding station after 16:30. Meanwhile, the minimum time interval between morning and afternoon boarding time is 6.5 h which is a suitable range for a full-time job. The home and workplace are assumed to be within walking distance of a metro station. In this way, if the travel diary meets the model rules, we assumed that this cardholder has one-day commuting trips.

A total of 739,714 commuting trips were identified from SCD in weekdays. The cardholders with at least two days' commuting trips were treated commuters, which were assessed based on the card ID number. Because the home and work station identified from a same cardholder may differ by different days, we proposed a maximum-frequency rule to solve potential conflicts. For example, one cardholder may be identified to use place A as home in three days, and to use place B as home in the other two days. Then, the place with maximum frequency (place A) will be used as home. In case, place A and B have equal frequency (say two days each), the place exhibits the greatest residential potential is regarded as home. The concept of "residential potential" was based on the land use pattern in station service area. Similarly, the maximum-frequency rule was used to identify work station. The identification of home and work station was independent. Finally, 149,593 unique commuters were identified with home (origin) and work (destination) station of commuting trips which account for 6.7% of total cardholders in weekdays. The process to identify metro commuters is illustrated in Fig. 2.

Furthermore, metro mode share for commuting trips was calculated by the ratio between metro commuting origin trips and all residential population aged 19–59 within the station buffer. All residents aged 19–59 could largely represent overall commuters in the buffer, because the unemployment rate for this population group is <5% in Wuhan (Bureau Statistics, 2016). The equation for mode share is as follows (Eq. (3)):

$$ModeShare = \frac{NT_o}{\sum_i^k (A_{ib}/A_{ic}) \times P_i} \tag{3}$$

where NT_o is the number of commuting origin trips in the target station; A_{ib}, A_{ic} are the area of community i in the station buffer zone, and the total area of community i respectively; P_i is the population size aged 19–59 of community i; k is the number of communities that intersected in the station buffer zone. The community is a spatial administrative unit in Wuhan. The 500-m station buffer generally intersected with around 10 communities (Fig. 3). Yet, three metro stations in suburban areas can only intersect with one or two communities because these suburban communities are larger but less densely populated. Hence, these three stations have mode share value greater than 1. We tentatively increase population size aged 19–59 in catchment area by 50% of the community population to avoid potential outliers of mode share value.

Number of origin trips and destination trips of the metro commuters in each station were also considered as TOD outcomes. Weekday and weekend metro ridership were measured using Tuesday and Saturday SCD record respectively. Daily ridership on weekday and weekend were added up by daily boarding ridership. Boarding and alighting ridership in peak-hour (7AM-9AM), and alighting ridership in non-peak-hour (9AM-12AM) were considered both on weekday and weekend. Table 3 illustrates the dependent variables of metro usage at station level and summary statistics.

3.3. Built environment factors

The built environment factors are assessed according to the 3Ds framework (Ewing and Cervero, 2010; Lu et al., 2017) including density, diversity and design. The built environment variables were obtained and measured using geographic census data and point of interest (POI) data in ArcGIS 10.5. As a new form of urban geographic data retrieved from Gaode Map (https://lbs.amap.com/api/webservice), POI data provide more comprehensive information for spatial analysis in urban space compared with land use data. Therefore, the study developed residential, commercial, office POI density variables and POI mix for three (residential, commercial, and office) to make a more inclusive framework of built environment factors. It should be noted that, POI data in Gaode map were divided into two category levels that are macro- and micro-levels. There are 23 macro-level categories and more than 200 micro-level categories in original Gaode POI dataset. Different categories of POI may have different impact on ridership. For instance, office building may have stronger impact than a convenient shop. To avoid

Table 2
Summary statistics of dependent variables of metro usage and independent variables of built environment. (N = 75 stations).

Variables (Unit)		Mean	SD	Min	Max	VIF1	VIF2
Dependent variables							
	Commuting mode share (0–1)	0.250	0.221	0.026	0.822		
	Commuting origin trips (N)	1994.573	1650.107	29	9029		
	Commuting destination trips (N)	2004.481	2139.229	36	9533		
	Weekday daily boarding (N)	17161.187	15532.857	1232	89,618		
	Weekday peak boarding (N)	3262.773	2593.237	118	15,890		
	Weekday peak alighting (N)	3043.693	2905.324	108	14,088		
	Weekday non-peak alighting (N)	2927.187	3022.763	155	15,517		
	Weekend daily boarding (N)	21201.547	22729.779	1199	145,855		
	Weekend peak boarding (N)	2622.627	2269.347	114	15,748		
	Weekend peak alighting (N)	2341.427	2344.569	101	13,008		
	Weekend non-peak alighting (N)	4040.4	4595.525	184	26,087		
Independent variables							
Density	Population density (10000 N)	1.677	1.120	0.079	4.486	6.885	3.901
	Building floor area (km^2)	0.914	0.491	0.012	1.988	7.635	—
	Residential land ratio (0–1)	0.340	0.148	0.059	0.668	7.733	3.123
	Commercial land ratio (0–1)	0.116	0.092	0.002	0.457	2.732	2.512
	Office land ratio (0–1)	0.127	0.114	0	0.603	3.164	2.955
	Residential POI density (10 N)	3.227	2.709	0.1	12.1	9.090	—
	Commercial POI density (10 N)	22.072	30.069	0	155.8	4.184	3.688
	Office POI density (10 N)	11.117	12.59	0	55.9	5.905	4.815
Diversity	Land use mix for four (0–1)	0.798	0.144	0.292	0.997	5.389	3.566
	Land use mix of two (0–1)	0.851	0.178	0.323	0.998	8.012	—
	POI mix for three (0–1)	0.766	0.163	0.283	0.989	1.894	1.640
Design	Number of metro exit (N)	4.160	1.480	2	10	1.672	1.602
	Number of bus station (N)	9.333	5.006	1	35	1.959	1.930
	Number of bus line (N)	40.227	28.556	1	118	3.235	2.789
	Street length (km)	14.206	4.27	2.764	24.73	4.886	4.141
	Number of street intersection (N)	44.773	20.411	8	92	4.252	3.545

Note: SD = Standard Deviation; Min = Minimum; Max = Maximum; N = Number. VIF1 and VIF2 indicate variance inflation factor of independent variables before and after multicollinearity test.

heterogeneity of POI categories, our study sampled a small part representative micro-level categories in Gaode Map as residential, commercial, and office POIs based on their functionality and quantity. Table 1 shows the details of classification of POI dataset.

- Density. Population density, building floor area, residential land ratio, commercial land ratio, office land ratio, and residential POI density, commercial POI density, and office POI density are selected as density factors of the built environment. Population density is measured by population size within 500-m radius buffer using community-level population data, as it is measured in metro commuting mode share. Building density is assessed by total building floor area of all buildings contained in a station buffer. Land-use ratios of residential, commercial, and office land are measured by different land use area divided by station buffer area. Three land uses were re-classified by land block functionality in raw land use dataset based on national standard (Standard, 2014). For instance, commercial land usually includes land use for business and entertainment facilities; office land usually includes land use for administration, office affairs, hospital, and school; other land use generally refers to land use for industrial, transportation, green space and barren land. And residential, commercial and office POI density are assessed by number of POIs of each category in a station buffer.
- Diversity. Land use mix for four (residential, commercial, office, and other land use) and land use mix of two (residential, and other land use) are taken into consideration as indicators of land use diversity. We considered land use mix of two because the residential land ratio is usually higher (Mean = 0.340). Additionally, POI mix for three (residential, commercial, and office POI) is employed to measure built environment diversity. The entropy score derived from Shannon diversity (Manaugh and Kreider, 2013; Shannon, 1948) is applied to measure land use mix and POI mix as follows (Eq. (4)):

$$Mix = -\sum_{i=1}^{k}(p_i lnp_i)/lnk \tag{4}$$

in which p_i is the ratio of the i th category of land use or POI, and k depending on the number of different land uses or POI types present in station buffer.

- Design. Street length, and number of street intersection were taken as design indicators for street network. Street network is derived from Open Street Map and census data. Node that connected with more than three street segments is taken as an intersection. Moreover, number of metro station exit, number of bus station, number of bus line were also regarded as design factors. Both directions, if any, of each bus line are included.

3.4. Data analysis

As the dependent variables, metro commuting characteristics are represented by origin, destination trips, and mode share in each station. Metro ridership is classified into weekday and weekend, peak hour and non-peak hour, and boarding and alighting, because they varied by travel purpose and station service area's features. For instance, metro commuting origin trips might be higher in the residential-oriented metro station in the peak time. Built environment factors based on the 3Ds framework are regarded as predictors for outcomes of metro usage. All of the independent variables were collected near 2015.

Before developing regression models to investigate the relationship between built environment factors and outcomes of metro usage,

Table 3
Regression models to predict metro ridership and mode share for commuting trips. (N = 75 stations).

Explanatory variables	Origin trips		Destination trips		Mode share	
	β (p) OLS	β (p) SLM	β (p) OLS	β (p) SLM	β (p) OLS	β (p) SLM
Population density	−0.054, 0.714	−0.059, 0.654	−0.009, 0.954	−0.008, 0.949	−0.463, 0.003 ***	−0.374, 0.001 ***
Residential land ratio	1.588, 0.113	1.646, 0.065 *	0.692, 0.453	0.669, 0.423	−0.358, 0.694	0.217, 0.754
Commercial land ratio	−1.370, 0.341	−1.620, 0.226	0.974, 0.486	1.046, 0.420	−1.136, 0.406	−0.484, 0.640
Office land ratio	1.302, 0.301	1.391, 0.218	2.525, 0.040 **	2.504, 0.022 **	−0.541, 0.652	−0.611, 0.503
Commercial POI density	0.001, 0.963	0.001, 0.913	0.001, 0.967	0.001, 0.959	−0.001, 0.883	0.001, 0.919
Office POI density	0.004, 0.763	0.005, 0.731	0.031, 0.018 **	0.031, 0.008 ***	0.002, 0.858	−0.001, 0.952
Land use mix of four	1.214, 0.274	1.265, 0.203	1.110, 0.301	1.076, 0.265	−0.725, 0.492	−0.281, 0.725
POI mix for three	0.372, 0.572	0.374, 0.526	0.448, 0.481	0.445, 0.436	−0.266, 0.672	−0.076, 0.873
Number of metro exit	−0.061, 0.385	−0.063, 0.317	0.001, 0.985	0.001, 0.983	−0.071, 0.303	−0.048, 0.361
Number of bus station	0.038, 0.096 *	0.039, 0.058 *	0.028, 0.210	0.028, 0.160	0.026, 0.248	0.013, 0.443
Number of bus line	0.001, 0.896	0.001, 0.895	0.004, 0.359	0.004, 0.302	−0.006, 0.189	−0.001, 0.690
Street length	−0.059, 0.202	−0.057, 0.168	0.093, 0.283	0.090, 0.251	−0.078, 0.374	−0.054, 0.418
Number of street intersection	0.016, 0.079 *	0.016, 0.049 **	0.006, 0.387	0.006, 0.333	0.019, 0.024 **	0.016, 0.011 **
Pseudo R^2	0.440	0.443	0.735	0.738	0.497	0.639
Adjusted R^2	0.321	—	0.678	—	0.389	—

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. β = coefficient. p = p value. OLS = Ordinary Least Squares regression. SLM = Spatial Lag Model.

Table 4
Regression models to predict metro boarding ridership for both weekday and weekend. (N = 75 stations).

Explanatory variables	Weekday boarding ridership		Weekend boarding ridership	
	β (p) OLS	β (p) SLM	β (p) OLS	β (p) SLM
Population density	−0.012, 0.931	−0.031, 0.797	−0.046, 0.758	−0.063, 0.629
Residential land ratio	−0.020, 0.981	0.237, 0.749	0.263, 0.767	0.551, 0.485
Commercial land ratio	0.659, 0.604	−0.078, 0.944	1.851, 0.166	1.266, 0.285
Office land ratio	2.051, 0.067 *	2.414, 0.012 **	2.913, 0.015 **	3.345, 0.001 ***
Commercial POI density	0.002, 0.605	0.002, 0.525	0.003, 0.505	0.003, 0.424
Office POI density	0.013, 0.262	0.015, 0.130	0.009, 0.505	0.011, 0.340
Land use mix of four	0.451, 0.642	0.600, 0.478	0.081, 0.937	0.143, 0.872
POI mix for three	−0.359, 0.535	−0.361, 0.472	−0.901, 0.142	−0.920, 0.083 *
Number of metro exit	−0.028,0.667	−0.028, 0.613	0.006, 0.925	0.005, 0.920
Number of bus station	0.028, 0.18	0.028, 0.116	0.020, 0.356	0.020, 0.288
Number of bus line	0.003, 0.454	0.002, 0.464	0.004, 0.373	0.003, 0.381
Street length	0.014, 0.856	0.045, 0.514	0.002, 0.982	0.029, 0.692
Number of street intersection	0.011, 0.095 *	0.011, 0.057 *	0.013, 0.101	0.013, 0.061 *
Pseudo R^2	0.606	0.629	0.607	0.627
Adjusted R^2	0.522	—	0.524	—

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. β = coefficient. p = p value. OLS = Ordinary Least Squares regression. SLM = Spatial Lag Model.

variation inflation factor (VIF) was used to test the multicollinearity between the independent variables. Correlation matrix for all independent variables (see Appendix Table 1) shows that some variables are highly correlated such as population density, building floor area, and residential POI. We sequentially ruled out built environment variables that have the largest VIF above than 5. The VIF of independent variables before and after multicollinearity test are illustrated in Table 3. Finally, building floor area, residential POI density, and land use mix of two were excluded in final regression analysis.

This study firstly employed ordinary least square (OLS) regression (also as "linear regression") to investigate the association between built environment factors and outcomes of metro usage. OLS holds the basic assumption that the residual is random. However, the results (e.g., coefficient size, significance) of the OLS regression could be biased if spatial effects exist. For instance, stations that are closer to each other also will have a correlation of observations. Therefore, we also considered spatial regression including spatial lag model and spatial error model (Anselin, 1988; Anselin and Rey, 1991). We conducted pre-tests to determine which model is better. Results show that spatial lag model usually has a larger value of log likelihood and smaller value of Akaike info criterion (AIC). Hence, we also employed spatial lag model (SLM) using Geoda. SLM can be expressed as follows (Eq. (5)):

$$y = \rho W_y + X\beta + \varepsilon \tag{5}$$

where ρ is a spatial autocorrelation parameter, and W_y is a spatial weight matrix of the spatial lags for the dependent variables at nearby locations. Spatial weight was created in Geoda by K-Nearest neighbors of 4. Among the 75 stations, minimum neighbor is 1 and maximum neighbor is 4 in the spatial weight file.

Finally, we conduct the OLS regression and SLM to investigate the association between built environment characteristics and outcomes of metro usage. Notably, all of the dependent variables have been logarithmic transformed to achieve better normal distribution (see scatter plots between some variables in Appendix). The independent variables remain their original format. All of the 75 metro stations were included in our regression models. Regression coefficients (β) and p value were reported.

4. Results

Table 2 illustrates the summary statistics of dependent variables for outcomes of metro usage and independent variables for built environment factors. Descriptive statistics includes minimum (Min), maximum (Max), mean, and standard deviation (SD) of variables. For example, the mean value of metro commuting mode share for each station is 25.0% (SD = 22.1%) among the 75 stations. The mean value of population density (population size) in station service area is 16,770 persons, and mean number of street intersection is 44.773. Average ratio of residential land is 34.0%, while it is 11.6% and 12.7% for commercial and office land respectively. Diversity index is high in station buffer that the average value for land use mix of four and POI mix of three are 0.798 and 0.766 respectively.

Table 3 illustrates the associations between built environment factors and metro commuting ridership. The adjusted R^2 of OLS regression

Table 5
Regression models to predict weekday temporal metro ridership. (N = 75 stations).

Explanatory variables	Peak-hour boarding		Peak-hour alighting		Non-peak-hour alighting	
	β (p) OLS	β (p) SLM	β (p) OLS	β (p) SLM	β (p) OLS	β (p) SLM
Population density	0.075, 0.629	0.053, 0.700	0.022, 0.882	0.024, 0.857	−0.029, 0.848	−0.032, 0.804
Residential land ratio	0.476, 0.601	0.651, 0.430	−0.061, 0.946	0.085, 0.914	−0.273, 0.758	−0.029, 0.969
Commercial land ratio	−0.661, 0.631	−1.216, 0.337	1.145, 0.396	0.669, 0.585	1.427, 0.288	0.691, 0.560
Office land ratio	1.395, 0.246	1.618, 0.131	2.102, 0.076 *	2.305, 0.027 **	2.208, 0.061 *	2.638, 0.009 ***
Commercial POI density	0.002, 0.707	0.002, 0.601	−0.002, 0.625	−0.003, 0.550	0.002, 0.710	0.002, 0.679
Office POI density	−0.003, 0.823	−0.002, 0.835	0.026, 0.042 **	0.028, 0.011 **	0.017, 0.169	0.021, 0.051 *
Land use mix of four	0.717, 0.497	0.842, 0.369	0.772, 0.454	0.909, 0.320	0.664, 0.517	0.828, 0.350
POI mix for three	0.234, 0.709	0.236, 0.670	−0.383, 0.532	−0.374, 0.491	−0.615, 0.315	−0.578, 0.272
Number of metro exit	−0.067, 0.335	−0.069, 0.262	−0.014, 0.831	−0.016, 0.786	0.006, 0.925	0.002, 0.970
Number of bus station	0.043, 0.057 *	0.043, 0.027 **	0.029, 0.190	0.029, 0.134	0.019, 0.387	0.019, 0.312
Number of bus line	0.001, 0.868	0.001, 0.862	0.007, 0.091 *	0.007, 0.066 *	0.005, 0.222	0.004, 0.256
Street length	−0.101, 0.241	−0.087, 0.252	0.063, 0.456	0.083, 0.271	0.048, 0.566	0.082, 0.261
Number of street intersection	0.013, 0.082 *	0.012, 0.049 **	0.009, 0.214	0.009, 0.160	0.010, 0.156	0.010, 0.090 *
Pseudo R^2	0.452	0.463	0.692	0.702	0.647	0.671
Adjusted R^2	0.335	—	0.627	—	0.572	—

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. β = coefficient. p = p value. OLS = Ordinary Least Squares regression. SLM = Spatial Lag Model.

Table 6
Regression models to predict weekend temporal metro ridership. (N = 75 stations).

Explanatory variables	Peak-hour boarding		Peak-hour alighting		Non-peak-hour alighting	
	β (p) OLS	β (p) SLM	β (p) OLS	β (p) SLM	β (p) OLS	β (p) SLM
Population density	0.088, 0.563	0.063, 0.637	0.057, 0.718	0.063, 0.643	−0.074, 0.645	−0.081, 0.556
Residential land ratio	0.273, 0.762	0.506, 0.531	−0.008, 0.993	0.260, 0.753	0.169, 0.859	0.546, 0.512
Commercial land ratio	0.522, 0.701	−0.096, 0.937	3.030, 0.034 **	2.515, 0.044 **	2.364, 0.101	1.760, 0.159
Office land ratio	1.697, 0.154	2.017, 0.055 *	3.307, 0.009 ***	3.761, 0.001 ***	2.872, 0.025 **	3.427, 0.001 ***
Commercial POI density	0.004, 0.456	0.004, 0.322	−0.003, 0.566	−0.003, 0.483	0.003, 0.554	0.003, 0.479
Office POI density	−0.012, 0.349	−0.011, 0.291	0.016, 0.237	0.020, 0.092 *	0.010, 0.449	0.014, 0.240
Land use mix of four	0.320, 0.758	0.420, 0.646	0.208, 0.848	0.313, 0.740	0.650, 0.554	0.702, 0.459
POI mix for three	−0.159, 0.798	−0.172, 0.751	−1.117, 0.087 *	−1.123, 0.045 **	−1.030, 0.118	−1.001, 0.076 *
Number of metro exit	−0.037, 0.592	−0.039, 0.513	0.005, 0.948	−0.002, 0.967	0.057, 0.426	0.054, 0.380
Number of bus station	0.035, 0.120	0.035, 0.072 *	0.025, 0.289	0.025, 0.208	0.005, 0.826	0.004, 0.841
Number of bus line	0.003, 0.464	0.003, 0.427	0.008, 0.081 *	0.007, 0.074 *	0.007, 0.153	0.005, 0.157
Street length	−0.097, 0.256	−0.079, 0.294	−0.011, 0.903	0.011, 0.887	0.008, 0.933	0.035, 0.657
Number of street intersection	0.013, 0.067 *	0.013, 0.032 **	0.010, 0.254	0.010, 0.173	0.012, 0.155	0.013, 0.080 *
Pseudo R^2	0.432	0.452	0.635	0.656	0.633	0.661
Adjusted R^2	0.311	—	0.558	—	0.554	—

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. β = coefficient. p = p value. OLS = Ordinary Least Squares regression. SLM = Spatial Lag Model.

model is 0.321 for the commuting origin ridership, 0.678 for the commuting destination ridership and 0.389 for mode share. SLM achieves higher goodness of fit than OLS. For instance, R^2 of OLS for mode share is 0.497, while it reaches 0.639 in SLM. Metro commuting destination ridership is better explained than those of origin ridership by the built environment factors.

Built environment of residential land ratio, number of bus station, and number of street intersection are positively associated with metro commuting origin ridership in SLM. For the metro commuting destination ridership, office land ratio and office POI density have positive relationships. Station buffer area with higher number of street intersection tends to have higher metro mode share for commuting trips. However, population density is negatively associated with mode share for commuting trips.

Table 4 shows the regression results between built environment factors and daily metro boarding ridership on weekday and weekend. The adjusted R^2 of OLS regression model on weekday ridership is 0.522 and that for weekend is 0.524. SLM achieves better goodness of fit than OLS. Office land ratio is positively associated with both weekday and weekend ridership. POI mix is negatively associated with weekend daily ridership, but not with weekday ridership. Number of street intersections is positively associated with both weekday and weekend ridership.

Table 5 shows the results of regression models fitting metro ridership in different time periods (peak and non-peak hours) on weekday. The adjusted R^2 of OLS regression model are 0.335, 0.627, 0.572 of peak-hour boarding, alighting and non-peak-hour alighting ridership. SLM achieves higher goodness of fit than OLS. Built environment characteristics can explain peak-hour alighting ridership better than boarding. Number of bus station and number of street intersection are positively related to peak-hour boarding ridership. Office land ratio, office POI density, and number of bus line are positively related to peak-hour alighting ridership. During non-peak hours, office land ratio, office POI density, and number of street intersection are positively related to alighting ridership.

Table 6 illustrates the results of regression models fitting metro ridership in different time periods (peak and non-peak) on weekend. The adjusted R^2 of OLS regression model are 0.311, 0.558 and 0.554 of peak-hour boarding, alighting and non-peak-hour alighting ridership. SLM achieves better goodness of fit than OLS. Built environment characteristics can explain peak-hour alighting ridership better than boarding. Office land ratio, number of bus station and number of street interaction are positively related to peak-hour boarding ridership. Commercial land ratio, office land ratio, office POI density and number of bus line are positively related to peak-hour alighting ridership, while POI mix is negatively related to it. During non-peak hours, office land ratio and number of street intersection are positively related to alighting ridership and POI mix is negatively related to it.

5. Discussion

Previous studies have inconsistent findings about the associations between built environment characteristics around transit stations and transit usage (e.g., ridership). The inconsistence is largely due to different urban contexts and limited sample size of travel survey data. Research focusing on mode choice of TOD such as commuting mode share at station level is still scarce. To fill these gaps, we identified commuting trips and overall ridership using metro service with smart sard data (SCD) in Wuhan, China. The big data-based approach can comprehensively depict the majority of metro users' travel behaviors. The outcomes of metro usage are measured from the perspective of commuting origin and destination ridership, mode share, and boarding and alighting ridership during peak and non-peak hours both on weekday and weekend. Built environment characteristics are systematically evaluated based on density, diversity and design. The relationship between built environment factors around metro stations and metro usage were investigated with linear regression, and spatial lag model to mitigate spatial effects. The results demonstrate that some built environment characteristics were significantly related with metro usage.

5.1. Density

Expect for one caveat, our findings largely occur with previous findings that urban density in station service area positively impacts transit usage (Sung and Oh, 2011; Zhao et al., 2013).

In this study, density of residential land ratio is positively linked with commuting origin ridership in the station area. It is reasonable because an area with higher ratio of residential land is more likely to be a residential area for commuters. Higher office land ratio and office POI density can predict more commuting destination trips, which can be explained by potential employment destinations in these areas. Metro commuters tend to commute to office-oriented station area for work. Similarly, more office land and office POI is usually linked with higher ridership (e.g., daily boarding ridership, peak and non-peak-hour alighting ridership). The findings give us a message that metro usage is highly correlated with office-oriented station area in Wuhan. The same finding was revealed in Nanjing (Zhao et al., 2013).

However, the impacts of urban density on metro mode share for commuting trips are mixed. In areas with higher population density, people are less likely to use metro for commuting trips. This finding disagree with some studies that verify positive relation between residential density and rail mode share (Loo et al., 2010). One possible explanation is that the areas with higher population density are in the urban core of Wuhan. People in these areas may less rely on metro service due to metro crowdedness and they may have alternative transportation options, such as walk, bus or drive for commuting trips. On the other hand, the areas with lower population density are often in the urban perimeters, and alternative transportation options are either infeasible or inconvenience. Hence, commuters in the low-population-density areas more rely on metro service.

5.2. Diversity

Diversity is another indispensable built environment characteristic affects transit usage (Loo et al., 2010; Park et al., 2018). We used land use mix and POI mix to measure the degree of diversity within station buffer area. However, both land use mix and POI mix have no significant associations with metro commuting trips, mode share, and ridership. In some cases (e.g., daily boarding, peak and non-peak alighting ridership on weekend), POI mix is negatively associated with metro usage. The inconsistent findings from this study and others (Arrington and Cervero, 2008; Jun et al., 2015; Park et al., 2018) may due to the exclusive and separated land use zones in Wuhan. Different types of land use, such as, office or residential, are often not mixed together. For example, the metro ridership in area with sole residential land use can be higher than areas with mixed land use. At the same time, diversity may provide more destinations for local residents to reach by foot, resulting in less transit use (Kamruzzaman et al., 2014).

5.3. Design

Design is the third major built environment characteristics affecting transit usage (Ewing and Cervero, 2010; Nasri and Zhang, 2014). In this study, street intersection density is positively correlated with metro usage. Stations with more street intersections in the buffer have more commuting origin trips and higher mode share. Street intersections are also significantly related overall metro ridership. Other studies support the same conclusion (Durning and Townsend, 2015). Wuhan has undergone rapid urbanization recently; there are many old and gated communities with less street intersections. Such places might generate less metro commuters and ridership. It is highly likely that places with more street intersections have more middle-class residents who prefer metro. Furthermore, places with more street intersections tend to have better accessibility to metro stations and hence have higher metro usage.

In sum, the findings of this research have meaningful implications both for TOD planning and built environment design in transit service area. The 3Ds framework may exert diverse effects on metro ridership and metro mode share in different urban contexts. In this study, we found two of Ds, density and design has a positive effect on metro usage, while diversity has no effect. Hence, exploring the fine-grained relationship of metro usage and built environment factors around transit stations in different urban and social contexts warrants further research attention. With the background of rapid TOD planning and implementation in China, policymakers and urban planners should pay great attention to the roles of built environment around transit stations to maximize transit ridership and reduce auto usage.

This study also has some limitations. First, only commuting trips were identified using a data mining method based on time and frequency rules. More sophisticated method should be developed to extract metro commuting trips and trips for other activities or purposes, e.g., non-commuting trips, leisure activities. Second, metro commuter mode share should be operationally calculated by metro commuters divided by total commuters in station buffer, while this study replaced it with population size aged 19–59, which may cause bias. Future studies should implement more accurate method to estimate total commuters in station served area. Third, using consistent station buffer to measure station catchment area may be biased. For example, it could fail to include the people who live outside the buffer but take bus cycle to use metro. The station catchment area should be refined in the future. Fourth, the socio-demographic attribute of travelers and station served area is indispensable in shaping travel behavior. However, due to data unavailability, we cannot control for such attributes. Finally, future studies should also consider other transportation modes, e.g., walking, buses, private cars to depict the comprehensive relationship between built environment characteristics and travel mode choice at station level.

6. Conclusion

In this study, we measured metro mode share for commuting trips and metro ridership with smart card data. The associations between metro usage and built environment factors were investigated in Wuhan, China. Our study extends previous research by considering the impacts of built environment on mode share of commuting trips at station level. Our results suggest that population density is negatively associated with metro commuting mode share, while street intersection shows a positive relationship. Office-oriented urban function and street intersection are positively correlated with metro ridership The results of this study can benefit both present and future TOD planning to maximize metro ridership and commuting mode share with proper urban planning strategies. Researchers, policymakers and urban planners should explore the fine-grained relationship of metro usage and built environment factors in different urban and social contexts.

CRediT authorship contribution statement

Long Chen: Conceptualization, Methodology, Formal analysis, Visualization, Writing – original draft. **Yi Lu:** Conceptualization, Writing – review & editing, Supervision. **Yanfang Liu:** Writing – review & editing, Supervision. **Linchuan Yang:** Writing – review & editing. **Mingjun Peng:**. **Yaolin Liu:** Writing – review & editing.

Acknowledgement

The work described in this paper was fully supported by grants from the National Natural Science Foundation of China (No. 51778552).

Appendix A

Fig. A1 and Table A1.

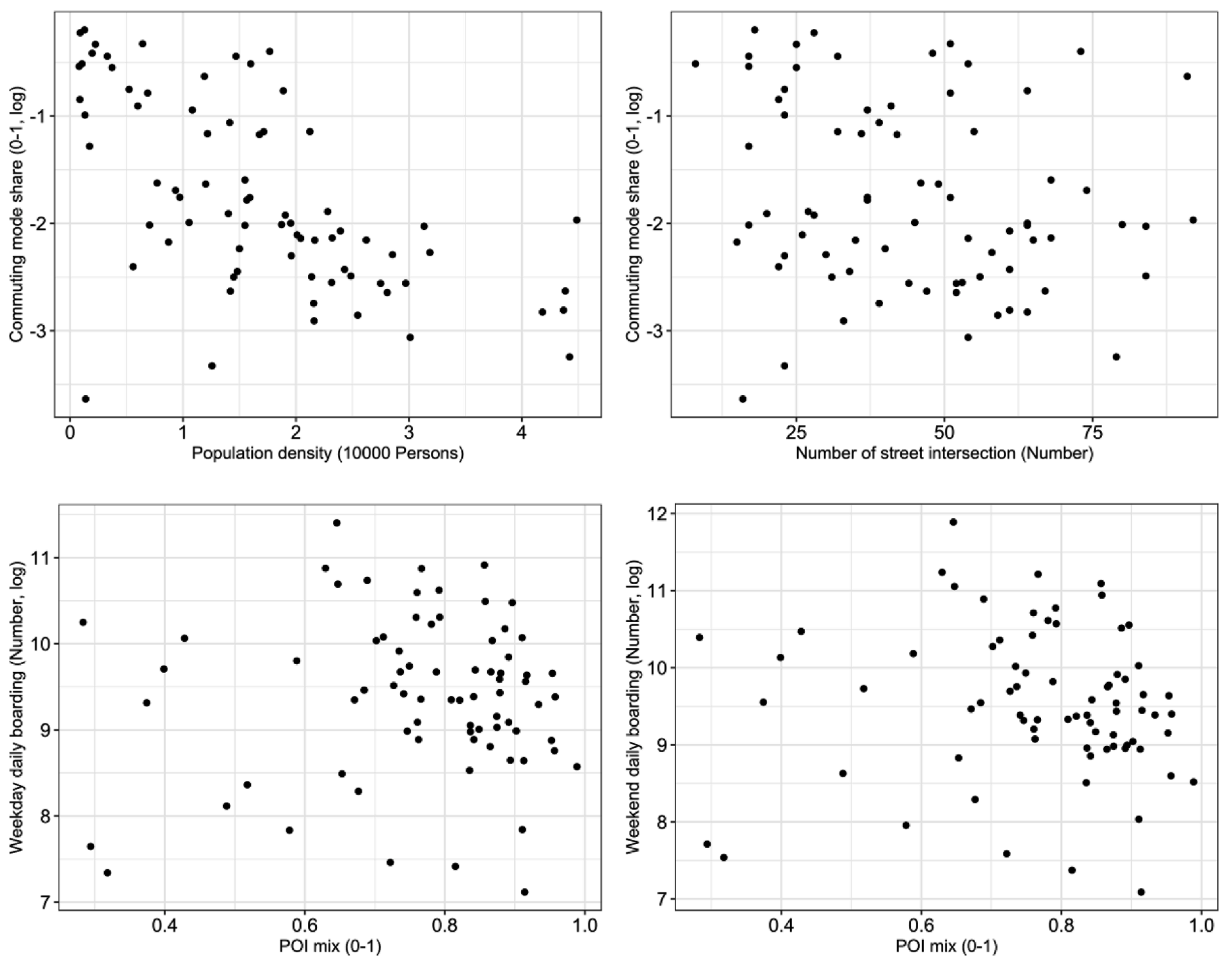

Fig. A1. Scatter plots.

Table A1
Correlation matrix for all independent variables of the built environment.

Variables	Density							
	Population density	Building floor area	Residential land ratio	Commercial land ratio	Office land ratio	Residential POI density	Commercial POI density	Office POI density
Building floor area	0.769	1						
Residential land ratio	0.535	0.504	1					
Commercial land ratio	−0.158	−0.036	−0.198	1				
Office land ratio	0.363	0.469	−0.016	−0.143	1			
Residential POI density	0.842	0.839	0.377	−0.027	0.394	1		
Commercial POI density	0.503	0.607	0.126	0.201	0.273	0.662	1	
Office POI density	0.549	0.715	0.012	0.138	0.421	0.750	0.726	1
Land use mix of four	0.429	0.544	0.288	0.417	0.513	0.449	0.402	0.458
Land use mix of two	0.545	0.508	0.870	−0.095	0.125	0.387	0.176	0.104
POI mix of three	0.239	0.293	0.260	−0.031	0.089	0.247	−0.244	0.091
Number of metro exit	0.229	0.378	−0.080	−0.084	0.378	0.297	0.253	0.466
Number of bus station	0.271	0.414	0.261	−0.103	0.226	0.281	0.217	0.223
Number of bus line	0.646	0.624	0.188	−0.024	0.259	0.569	0.448	0.570
Street length	0.573	0.630	0.257	−0.018	0.265	0.697	0.459	0.542
Number of street intersection	0.525	0.682	0.367	−0.026	0.535	0.626	0.509	0.582

	Diversity			Design				
	Land use mix of four	Land use mix of two	POI mix of three	Number of metro exit	Number of bus station	Number of bus line	Street length	Number of street intersection
Land use mix of two	0.536	1						
POI mix of three	0.182	0.271	1					
Number of metro exit	0.161	−0.057	0.028	1				
Number of bus station	0.289	0.287	0.182	0.301	1			
Number of bus line	0.360	0.220	0.184	0.268	0.430	1		
Street length	0.314	0.240	0.189	0.225	0.440	0.556	1	
Number of street intersection	0.471	0.348	0.176	0.449	0.508	0.488	0.533	1

References

An, D., Tong, X., Liu, K., Chan, E.H.W., 2019. Understanding the impact of built environment on metro ridership using open source in Shanghai. Cities 93, 177–187. https://doi.org/10.1016/j.cities.2019.05.013.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Springer, Dordrecht.

Anselin, L., Rey, S., 1991. Properties of Tests for Spatial Dependence in Linear Regression Models. Geographical Analysis 23 (2), 112–131. https://doi.org/10.1111/j.1538-4632.1991.tb00228.x.

Arrington, G., Cervero, R., 2008. Effects of TOD on Housing, Parking, and Travel. Transit Cooperative Research Program (TCRP) Report 128, published by. Transportation Research Board, Washington.

Bertolini, L., Curtis, C., Renne, J., 2012. Station Area projects in Europe and Beyond: Towards Transit Oriented Development? Built Environment 38 (1), 31–50. https://doi.org/10.2148/benv.38.1.31.

Bureau Statistics. (2016). 2015 Statistical Bulletin on Wuhan's National Economic and Social Development. http://tjj.wuhan.gov.cn/tjfw/tjgb/202001/t20200115_841062.shtml.

Calthorpe, P., 1993. The Next American Metropolis: Ecology, Community, and the American Dream. Princeton Architectural Press, Princeton, NJ.

Cardozo, O.D., García-Palomares, J.C., Gutiérrez, J., 2012. Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. Applied Geography 34, 548–558. https://doi.org/10.1016/j.apgeog.2012.01.005.

Cervero, R., Day, J., 2008. Suburbanization and transit-oriented development in China. Transport Policy 15 (5), 315–323. https://doi.org/10.1016/j.tranpol.2008.12.011.

Cervero, R., Murphy, S., Ferrell, C., Goguts, N., Tsai, Y.-H., Arrington, G.B., McKay, S., 2004. *Transit-Oriented Development in the United States: Experiences, Challenges, and Prospects*: Washington. Federal Transit Administration, DC.

Ding, C., Cao, X., Liu, C., 2019. How does the station-area built environment influence Metrorail ridership? Using gradient boosting decision trees to identify non-linear thresholds. Journal of Transport Geography 77, 70–78. https://doi.org/10.1016/j.jtrangeo.2019.04.011.

Duncan, M., 2010. The Impact of Transit-oriented Development on Housing Prices in San Diego. CA. *Urban Studies* 48 (1), 101–127. https://doi.org/10.1177/0042098009359958.

Durning, M., Townsend, C., 2015. Direct Ridership Model of Rail Rapid Transit Systems in Canada. Transportation Research Record 2537 (1), 96–102. https://doi.org/10.3141/2537-11.

Ewing, R., Cervero, R., 2010. Travel and the built environment: A meta-analysis. Journal of the American Planning Association 76 (3), 265–294. https://doi.org/10.1080/01944361003766766.

Frank, L.D., Mayaud, J., Hong, A., Fisher, P., Kershaw, S., 2019. Unmet Demand for Walkable Transit-Oriented Neighborhoods in a Midsized Canadian Community: Market and Planning Implications. Journal of Planning Education and Research 0739456X19831064. https://doi.org/10.1177/0739456X19831064.

Gan, Z., Yang, M., Feng, T., Timmermans, H.J.P., 2020. Examining the relationship between built environment and metro ridership at station-to-station level. Transportation Research Part D: Transport and Environment 82, 102332. https://doi.org/10.1016/j.trd.2020.102332.

Guerra, E., 2014. Mexico City's suburban land use and transit connection: The effects of the Line B Metro expansion. Transport Policy 32, 105–114. https://doi.org/10.1016/j.tranpol.2013.12.011.

Hasan, S., Schneider, C.M., Ukkusuri, S.V., González, M.C., 2013. Spatiotemporal Patterns of Urban Human Mobility. Journal of Statistical Physics 151 (1), 304–318. https://doi.org/10.1007/s10955-012-0645-0.

Huang, X., Cao, X., Yin, J., Cao, X., 2017. Effects of metro transit on the ownership of mobility instruments in Xi'an, China. Transportation Research Part D: Transport and Environment 52, 495–505. https://doi.org/10.1016/j.trd.2016.09.014.

Huang, X., Cao, X., Yin, J., Cao, X., 2019. Can metro transit reduce driving? Evidence from Xi'an, China. Transport Policy 81, 350–359. https://doi.org/10.1016/j.tranpol.2018.03.006.

Huang, J., Levinson, D., Wang, J., Zhou, J., Wang, Z.-J., 2018. Tracking job and housing dynamics with smartcard data. Proceedings of the National Academy of Sciences 115 (50), 12710–12715. https://doi.org/10.1073/pnas.1815928115.

Huang, J., Levinson, D., Wang, J., Jin, H., 2019. Job-worker spatial dynamics in Beijing: Insights from Smart Card Data. Cities 86, 83–93. https://doi.org/10.1016/j.cities.2018.11.021.

Jun, M.-J., Choi, K., Jeong, J.-E., Kwon, K.-H., Kim, H.-J., 2015. Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul. Journal of Transport Geography 48, 30–40. https://doi.org/10.1016/j.jtrangeo.2015.08.002.

Kamruzzaman, M., Baker, D., Washington, S., Turrell, G., 2014. Advance transit oriented development typology: case study in Brisbane, Australia. Journal of Transport Geography 34, 54–70. https://doi.org/10.1016/j.jtrangeo.2013.11.002.

Kamruzzaman, M., Baker, D., Washington, S., Turrell, G., 2016. Determinants of residential dissonance: Implications for transit-oriented development in Brisbane. International Journal of Sustainable Transportation 10 (10), 960–974. https://doi.org/10.1080/15568318.2016.1191094.

Li, Z., Han, Z., Xin, J., Luo, X., Su, S., Weng, M., 2019. Transit oriented development among metro station areas in Shanghai, China: Variations, typology, optimization and implications for land use planning. Land Use Policy 82, 269–282. https://doi.org/10.1016/j.landusepol.2018.12.003.

Lin, J.-J., Shin, T.-Y., 2008. Does Transit-Oriented Development Affect Metro Ridership?: Evidence from Taipei. Taiwan. *Transportation Research Record* 2063 (1), 149–158. https://doi.org/10.3141/2063-18.

Loo, B.P.Y., du Verle, F., 2017. Transit-oriented development in future cities: towards a two-level sustainable mobility strategy. International Journal of Urban Sciences 21 (sup1), 54–67. https://doi.org/10.1080/12265934.2016.1235488.

Loo, B.P.Y., Chen, C., Chan, E.T.H., 2010. Rail-based transit-oriented development: Lessons from New York City and Hong Kong. Landscape and Urban Planning 97 (3), 202–212. https://doi.org/10.1016/j.landurbplan.2010.06.002.

Lu, Y., Xiao, Y., Ye, Y., 2017. Urban density, diversity and design: Is more always better for walking? A study from Hong Kong. Preventive Medicine 103, S99–S103. https://doi.org/10.1016/j.ypmed.2016.08.042.

Lyu, G., Bertolini, L., Pfeffer, K., 2016. Developing a TOD typology for Beijing metro station areas. Journal of Transport Geography 55, 40–50. https://doi.org/10.1016/j.jtrangeo.2016.07.002.

Lyu, G., Bertolini, L., Pfeffer, K., 2020. How does transit-oriented development contribute to station area accessibility? A study in Beijing. International Journal of Sustainable Transportation 14 (7), 533–543. https://doi.org/10.1080/15568318.2019.1578841.

Manaugh, K., & Kreider, T. (2013). What is mixed use? Presenting an interaction method for measuring land use mix. *Journal of Transport and Land Use, 6*(1), 63-72. Retrieved from http://www.jstor.org/stable/26202648.

Nasri, A., Zhang, L., 2014. The analysis of transit-oriented development (TOD) in Washington, D.C. and Baltimore metropolitan areas. Transport Policy 32, 172–179. https://doi.org/10.1016/j.tranpol.2013.12.009.

Noland, R.B., DiPetrillo, S., 2015. Transit-oriented development and the frequency of modal use. Journal of Transport and Land Use 8 (2), 21–44. Retrieved from www.jstor.org/stable/26202711.

Pagliara, F., Papa, E., 2011. Urban rail systems investments: an analysis of the impacts on property values and residents' location. Journal of Transport Geography 19 (2), 200–211. https://doi.org/10.1016/j.jtrangeo.2010.02.006.

Pan, H., Zhang, M., 2008. Rail Transit Impacts on Land Use: Evidence from Shanghai. China. *Transportation Research Record* 2048 (1), 16–25. https://doi.org/10.3141/2048-03.

Pan, H., Li, J., Shen, Q., Shi, C., 2017. What determines rail transit passenger volume? Implications for transit oriented development planning. Transportation Research Part D: Transport and Environment 57, 52–63. https://doi.org/10.1016/j.trd.2017.09.016.

Papa, E., Bertolini, L., 2015. Accessibility and Transit-Oriented Development in European metropolitan areas. Journal of Transport Geography 47, 70–83. https://doi.org/10.1016/j.jtrangeo.2015.07.003.

Park, K., Ewing, R., Scheer, B.C., Tian, G., 2018. The impacts of built environment characteristics of rail station areas on household travel behavior. Cities 74, 277–283. https://doi.org/10.1016/j.cities.2017.12.015.

Pelletier, M.-P., Trépanier, M., Morency, C., 2011. Smart card data use in public transit: A literature review. Transportation Research Part C: Emerging Technologies 19 (4), 557–568. https://doi.org/10.1016/j.trc.2010.12.003.

Ratner, K.A., Goetz, A.R., 2013. The reshaping of land use and urban form in Denver through transit-oriented development. Cities 30, 31–46. https://doi.org/10.1016/j.cities.2012.08.007.

Shannon, C.E., 1948. A Mathematical Theory of Communication. Bell System Technical Journal 27 (3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Shen, Q., Chen, P., Pan, H., 2016. Factors affecting car ownership and mode choice in rail transit-supported suburbs of a large Chinese city. Transportation Research Part A: Policy and Practice 94, 31–44. https://doi.org/10.1016/j.tra.2016.08.027.

Singh, Y.J., Lukman, A., Flacke, J., Zuidgeest, M., Van Maarseveen, M.F.A.M., 2017. Measuring TOD around transit nodes – Towards TOD policy. Transport Policy 56, 96–111. https://doi.org/10.1016/j.tranpol.2017.03.013.

Chinese Standard. (2014). GB 50137-2011: Translated English of Chinese Standard. GB50137-2011: Code for classification of urban land use and planning standards of development land: https://www.chinesestandard.net.

Sung, H., Oh, J.-T., 2011. Transit-oriented development in a high-density city: Identifying its association with transit ridership in Seoul. Korea. *Cities* 28 (1), 70–82. https://doi.org/10.1016/j.cities.2010.09.004.

Vergel-Tovar, C.E., Rodriguez, D.A., 2018. The ridership performance of the built environment for BRT systems: Evidence from Latin America. Journal of Transport Geography 73, 172–184. https://doi.org/10.1016/j.jtrangeo.2018.06.018.

Walton, D., Sunseri, S., 2010. Factors Influencing the Decision to Drive or Walk Short Distances to Public Transport Facilities. International Journal of Sustainable Transportation 4 (4), 212–226. https://doi.org/10.1080/15568310902927040.

Wu, W., Hong, J., 2017. Does public transit improvement affect commuting behavior in Beijing, China? A spatial multilevel approach. Transportation Research Part D: Transport and Environment 52, 471–479. https://doi.org/10.1016/j.trd.2016.08.032.

Zhang, M., Wang, L., 2013. The impacts of mass transit on land development in China: The case of Beijing. Research in Transportation Economics 40 (1), 124–133. https://doi.org/10.1016/j.retrec.2012.06.039.

Zhao, J., Deng, W., Song, Y., Zhu, Y., 2013. What influences Metro station ridership in China? Insights from Nanjing. Cities 35, 114–124. https://doi.org/10.1016/j.cities.2013.07.002.

Zhao, J., Deng, W., Song, Y., Zhu, Y., 2014. Analysis of Metro ridership at station level and station-to-station level in Nanjing: an approach based on direct demand models. Transportation 41 (1), 133–155. https://doi.org/10.1007/s11116-013-9492-3.

Zhao, P., Li, S., 2018. Suburbanization, land use of TOD and lifestyle mobility in the suburbs: An examination of passengers' choice to live, shop and entertain in the metro station areas of Beijing. Journal of Transport and Land Use 11 (1), 195–215. https://doi.org/10.5198/jtlu.2018.1099.

Zhao, P., Yang, H., Kong, L., Liu, Y., Liu, D., 2018. Disintegration of metro and land development in transition China: A dynamic analysis in Beijing. Transportation Research Part A: Policy and Practice 116, 290–307. https://doi.org/10.1016/j.tra.2018.06.017.

Zhou, J., 2016. The transit metropolis of Chinese characteristics? Literature review, interviews, surveys and case studies. Transport Policy 51, 115–125. https://doi.org/10.1016/j.tranpol.2015.11.009.

Zhou, J., Long, Y., 2014. Jobs-Housing Balance of Bus Commuters in Beijing: Exploration with Large-Scale Synthesized Smart Card Data. Transportation Research Record 2418 (1), 1–10. https://doi.org/10.3141/2418-01.

Zhou, J., Yang, Y., Gu, P., Yin, L., Zhang, F., Zhang, F., Li, D., 2019. Can TODness improve (expected) performances of TODs? An exploration facilitated by non-traditional data. Transportation Research Part D: Transport and Environment 74, 28–47. https://doi.org/10.1016/j.trd.2019.07.008.