Article:

Chen, Y., Wadud, Z., Jia, S. et al. (2026) Exploring the heterogeneous nonlinear effects of the built environment on metro ridership considering bike-sharing catchment areas. Journal of Transport Geography, 136. 104763. ISSN: 0966-6923

https://doi.org/10.1016/j.jtrangeo.2026.104763

Exploring the heterogeneous nonlinear effects of the built environment on metro ridership considering bike-sharing catchment areas

Abstract: Existing research on the relationship between the built environment and metro ridership mostly uses pedestrian catchment areas (PCA) such as the 800m buffer, ignoring bike-sharing catchment areas (BSCA). The heterogeneous nonlinear effects of the built environment in different types of stations also tend to be overlooked. Here, we used bike-sharing trip data to identify the BSCA of each metro station in Beijing. A Fuzzy C-means (FCM) clustering-based explainable machine learning framework was proposed to explore the nonlinear effects of the built environment, including land use, transport, accessibility and socio-economic factors on four kinds of metro ridership under PCA and BSCA, respectively. The heterogeneity of the nonlinear relationship among different types of stations was also investigated. We found that BSCA-based models outperform PCA-based models. The nonlinear effects of the built environment on metro ridership change with various catchment areas. These effects tend to have various trends and thresholds in different types of stations. Furthermore, residential and employment-oriented stations are more sensitive to catchment areas. This paper emphasizes the role of BSCA and heterogeneous planning targets in different types of stations in TOD (Transit-oriented development) planning.

Keywords: Urban rail transit; Metro ridership; Built environment; Land use; Machine learning; Bike-sharing; Accessibility; Transit-oriented development

1 Introduction

Transit-oriented development (TOD) has been implemented in many cities to address growing environmental challenges and promote sustainable urban growth (Cervero and Kockelman, 1997; Jamme et al., 2019). Recently, many Chinese cities have undergone rapid development of urban rail transit systems. By the end of 2024, 258 metro lines were in operation across 41 cities, covering 9,306 km (China Association of Metros, 2025). However, the rapid expansion of metro networks, coupled with an inadequate integration between urban rail transit and surrounding land use, has led to a lack of ridership and high operating costs (Li et al., 2020; Wang et al., 2023a; Ding et al., 2025). For example, 56% of the metro cities in China have not met the requirement specified in Document No. 52 for a ridership intensity of 7,000 passenger trips per km per day (State Council of the People's Republic of China, 2018). In Beijing, the operating cost of metro is 10.27 yuan per trip, while fare income is less than 5 yuan per trip in 2023 (People's Public Transportation, 2024). Therefore, most metro systems need a significant subsidy from the local government, hindering the sustainable growth of cities.

To address these problems and advance sustainable TOD, it is important to explore the relationship between the built environment and metro ridership. A clearer understanding of this relationship can inform integrated transport and land use planning, improve ridership, and support the high-quality development of urban metro systems. Furthermore, under the context of emerging shared mobility services, such as bike-sharing (Fishman, 2016), the interplay between metro and other transportation modes within a multimodal system has significant effects on metro ridership (Peng et al., 2023). Therefore, it is also essential to consider shared mobility systems in TOD planning.

Since the introduction of Direct ridership models (DRMs) in TOD planning (Kuby et al., 2004; Cervero, 2006), a growing number of studies have explored the effects of the built environment on metro ridership (Loo et al., 2010; Durning and Townsend, 2015; Zhao et al., 2014; Ding et al., 2019; Gao et al., 2024). Initially, regression models were applied due to the advantages of operability and interpretability. Such models are easy to implement, provide clear parameter estimates, and allow straightforward hypothesis testing, which makes them highly suitable for identifying the linear effects of built environment factors on metro ridership. However, these methods ignore complex nonlinear relationships (Van Wee and Handy, 2016). To address this, Ding et al. (2019) introduced gradient boosting decision trees (GBDT) in DRMs, machine learning models are now gaining increased popularity due to their strength in investigating the nonlinear and threshold effects of the built environment on metro ridership (Gan et al., 2020; Shao et al., 2020; Du et al., 2022; Liu et al., 2023c). Thus, we can find the most effective ranges of built environment variables to inform transport and land use planning in station catchment areas.

Despite the increasing use of DRMs and advanced modeling approaches, several key research gaps remain. Notably, most studies rely on pedestrian catchment areas (PCA), typically the 800m (0.5 mile) buffer, as the study unit, based on early foundational work (Kuby et al., 2004; Cervero, 2006). However, bike-sharing has been proven to be one of the most important modes in accessing and egressing metro stations (Guo et al., 2021; Zhan et al., 2023a). According to the relevant report, there are 31% bike-sharing trips are generated around metro stations in peak hours in major cities in China, and the ratio in Beijing is 45% (China Academy of Urban Planning and Design, 2023). Yet, few studies consider the bike-sharing catchment areas (BSCA) when exploring the relationship between the built environment and metro ridership. Secondly, although some studies have incorporated density of bike stations (Chen et al., 2019; Wang et al., 2022a; Zhu et al., 2019) as built environment factors, existing studies tend to ignore the effects of the bike-sharing supply and bike-sharing accessibility on metro ridership. Thirdly, few studies have investigated the heterogeneous nonlinear effects of the built environment on metro ridership among different types of metro stations with various spatiotemporal ridership patterns. Therefore, it is also unknown whether the effects of the built environment differ in various types of stations and if heterogeneous planning targets should be set based on metro station characteristics.

Hence, this manuscript answers the following research questions:

(1) Compared with PCA, how does BSCA affect the nonlinear relationship between the built environment and metro ridership?

(2) How do bike-sharing supply and bike-sharing accessibility affect metro ridership?

(3) Do the effects of the built environment on metro ridership differ between different types of metro stations?

To answer these questions, taking Beijing as an example, this paper used bike-sharing data to identify the BSCA of metro stations. Then, a clustering-based explainable machine learning method was applied to explore the heterogeneous nonlinear relationship between the built environment and metro ridership.

2 Literature review

2.1 The catchment areas of metro stations

Station catchment areas are used as the study unit in DRMs (Ding et al., 2019). Thus, determining the catchment areas is the first key issue in exploring the relationship between metro ridership and the built environment. Traditionally, the most commonly used measure is a radial buffer, typically 800m, (Du et al., 2022; Gan et al., 2020; Huang et al., 2016; Li et al., 2020; Shi et al., 2018; Su et al., 2022b; Yang et al., 2023a), representing a suitable walking distance (Kuby et al., 2004; Cervero, 2006). To avoid overlapping areas in adjacent buffer areas, the Thiessen polygon was introduced to delineate the individual catchment areas of each station (Gao et al., 2022; Wang et al., 2023a). Moreover, as buffer areas do not consider actual road networks, isochronous based on walking networks are used to represent the catchment areas (Huang et al., 2022; Li et al., 2023; Liu et al., 2023a).

However, most studies investigating the relationship between the built environment and metro ridership only focus on the PCA of metro stations, few studies have considered BSCA, which is different from the PCA in both size and shape due to the longer cycling travel distances and the configuration of the road network (Zuo et al., 2020). For example, Guo et al. (2021) defined BSCA as the areas with a spatial range of 800-1500m buffer. Wu et al. (2021) delineated the BSCA by incorporating the origins and destinations of integrated metro and bike-sharing trips without consideration of overlapping areas. Similarly, Zhan et al. (2023) used the 85^{th} percentile of riding distance of the integrated bike-sharing trips in each metro station as the buffer size. Differentiated buffers are then used to represent BSCA.

Different catchment areas relate to the measurement of built environment factors and thus affect the relationship between the built environment and metro ridership. Studies comparing the results of the relationship between metro ridership and the built environment under different buffers (Sung et al., 2014; Liu et al., 2023c; Wang et al., 2022b) have found that the results are sensitive to the buffer size. Li et al. (2023) compared 17 types of different catchment areas and found that relatively small and non-overlapping study units can better interpret metro ridership. These findings call for more discussion about the choice of catchment areas, especially considering the lack of research on BSCA. Although one study has used BSCA based on bike-sharing trip data (Liu et al., 2023b), we still do not know how BSCA improves our understanding of the effects of the built environment on metro ridership. Therefore, comparing the relationship between the built environment and metro ridership by using PCA and BSCA enriches the existing body of knowledge.

2.2 Built environment factors

We review four types of the most used built environment factors in existing research, including land use, transport, accessibility, and socio-economic.

2.2.1 Land use

Regarding land use, many studies have found that high-density development in metro stations' catchment areas promotes ridership (Kuby et al., 2004; Cervero, 2006; Loo et al., 2010; Jun et al., 2015; Ding et al., 2019; Yang et al., 2023b). Density is also the first factor in "5Ds" framework including density, diversity, design, distance to transit, and destination accessibility (Cervero and Kockelman, 1997; Ewing and Cervero, 2010). Population density and employment density within PCA are the most commonly selected factors of density and have been shown to have significant positive effects on metro ridership. Except for traditional census data (Ding et al., 2019; Liu et al., 2016; Zhu et al., 2019), more open data such as Worldpop (Du et al., 2023; Gao et al., 2022; Su et al., 2022a), mobile phone signal data (Liu et al., 2023; Wang et al., 2023b; Yang et al., 2023b), and POI (Point of Interest) data (Ding et al., 2023; Li et al., 2022; Liu et al., 2023a) are used to calculate the population density and employment density. These emerging data types show the advantages of frequent updates, high resolution, and good flexibility for aggregation in different scales. In addition, the density factors of various urban services such as healthcare, education, shopping, restaurant, and park tend to be adopted to show the land use characteristics (Chen et al., 2022a; Ding et al., 2023; Li et al., 2022; Wang et al., 2023a). The floor area or floor area ratio (FAR) is also usually applied to indicate the development intensity (Li et al., 2020; Liu et al., 2023a; Wang et al., 2023b; Yan et al., 2023).

As an important component of the TOD characteristic, diversity is chosen in most studies. A popular method to estimate diversity is to calculate entropy based on different types of land use (Deng et al., 2022; Gao et al., 2022; Yang et al., 2023a). The findings of the effects of land use diversity on metro ridership are relatively complex. Although there have been many studies that have shown that higher diversity can facilitate metro ridership (Gao et al., 2022b; Wang et al., 2023a; Yang et al., 2023b), however, some research identified an inverse relationship between diversity and ridership (Chen et al., 2022a; Yang et al., 2023a).

2.2.2 Transport

Internal transport attributes refer to the characteristics of metro stations. First, the factors about the station's location in the metro network tend to be used, e.g. distance to center, closeness centrality, and betweenness centrality (Chen et al., 2019; Ding et al., 2023; Yang et al., 2023a). The second type of factor refers to the facilities at the station, such as the number of entrances (Li et al., 2020; Liu et al., 2023a; Liu et al., 2023b).

External transport attributes mainly include the intermodal connection with the metro and other transport modes such as walking, cycling, bus, and taxi. The common factors include road density (Chen et al., 2022a; Liu et al., 2023b; Yang et al., 2023a; Yang et al., 2023b), pedestrian road density (Lin and Shin, 2008; Su et al., 2022a; Zhu et al., 2019), intersection density (Chen et al., 2022a; Ding et al., 2019; Gan et al., 2020; Zhu et al., 2019), number of bus stops and bus lines (Chen et al., 2019; Li et al., 2020; Liu et al., 2023a; Su et al., 2022a; Wang et al., 2023a), density of parking lots (An et al., 2019; Chen et al., 2019; Shi et al., 2018), and density of bike stations (Chen et al., 2019; Wang et al., 2022a; Zhu et al., 2019).

It is noted that many of these factors are usually divided into the design and distance to transit aspects of the "5Ds" framework in existing research (Pang et al., 2023; Yan et al., 2023; Yang et al., 2023a; Yang et al., 2023b). But there is no clear standard. Therefore, we group these factors in transport. Moreover, few studies consider the number of available shared bicycles in past studies, ignoring the critical built environment factor for representing the transport supply for the integrated use of bike-sharing and metro.

2.2.3 Accessibility

Accessibility has been found to greatly affect travel behavior (Ewing and Cervero, 2010). Nevertheless, it is less used in existing studies exploring the effects of the built environment on metro ridership. Some studies defined accessibility as the distance or travel time to the city center or sub-center (Liu et al., 2023a; Peng et al., 2023; Yang et al., 2023a; Yang et al., 2023b), ignoring the land use nature of accessibility (Geurs and van Wee, 2004). Some other studies used the number of various types of urban services within the PCA as the accessibility indicator (Du et al., 2022), which is more concerned with density. In general, these accessibility metrics cannot capture the integration between transport and land use, and ignore people's travel behavior using the metro system. Therefore, a better accessibility factor is needed to explore how accessibility affects metro ridership.

2.2.4 Socio-economic

Socio-economic factors are another type of important factors often considered in previous research, including age (Gao et al., 2022; Liu et al., 2016; Zhu et al., 2019), occupation (Yan et al., 2023), income (Su et al., 2022a), and race (Ding et al., 2019; Liu et al., 2016). For instance, Gao et al. (2022b) used three age groups, including youth, middle-aged, and older adults. They found that senior citizens (aged over 60) are less willing to use the metro system because of the defects of their physical functions. Meanwhile, housing price is another essential socio-economic factor, which has been widely applied for its good availability and operability (Li et al., 2023b; Pang et al., 2023; Peng et al., 2023; Yang et al., 2023a; Yang et al., 2023b).

In general, all these built environment factors are measured within the PCA in existing research, leading to a certain degree of miscalculation of actual characteristics of the built environment around metro stations. On the other hand, some key built environment factors, such as the number of available shared bicycles and accessibility, still need to be further explored.

2.3 Modeling approaches

Regression models and machine learning models are often used to explore the effects of the built environment on metro ridership. Regression models such as linear regression model (Cervero, 2006; Kuby et al., 2004; Lee et al., 2013; Lin and Shin, 2008; Loo et al., 2010; Zhao et al., 2014, 2013), Poisson regression model (Choi et al., 2012), and negative binomial regression model (Ramos-Santiago and Brown, 2016) are widely applied in existing research. For instance, Cervero (2006) used the linear regression model and found that a 10% growth in population density within the PCA of stations could lead to a 1.9% increase of daily transit ridership. However, a limitation of these models is that they ignore the spatial heterogeneity between stations. To address this limitation, geographically weighted regression (GWR) has been applied (Cardozo et al., 2012; Jun et al., 2015; Shi et al., 2018;

Zhu et al., 2019; Chen et al., 2019; Li et al., 2020; Su et al., 2022b; Wang et al., 2023a; Li et al., 2023a; Gao et al., 2024). Cardozo et al. (2012) first utilized GWR as DRM in the Madrid Metro. The findings demonstrated improved performance of the GWR compared to the linear regression model, which was attributed to the ability to identify spatial variability of the local coefficients of built environment factors.

Nevertheless, these regression models often assume a predefined (linear or log-linear) relationship between the built environment and metro ridership. Ding et al. (2019) first used GBDT to investigate the nonlinear effects of station-area built environment elements on daily Metrorail ridership. They found that all built environment factors show threshold effects on ridership, indicating effective ranges of planning targets. Since then, machine learning models have become mainstream to explore the relationship between the built environment and ridership (Gan et al., 2020; Shao et al., 2020; Du et al., 2022; Su et al., 2022a; Li et al., 2023b; Liu et al., 2023c; Yang et al., 2024b; Xi et al., 2024). Among these studies, the eXtreme Gradient Boosting (XGBoost) model is gaining increased importance as it is based on GBDT and introduces a regularization term in the objective function to control the complexity of the model to prevent over-fitting (Liu et al., 2023c; Wang et al., 2023c; Gu and Dou, 2024; Liu et al., 2024; Xi et al., 2024).

Despite the good performance of machine learning models, these models may suffer from potential overfitting, especially when applied to limited or context-specific datasets. Overfitting can reduce the generalizability of findings and result in misleading policy recommendations if not properly validated. Meanwhile, the black box approach of machine learning models leads to insufficient interpretability. It poses a higher requirement for adopting regularization techniques, cross-validation, and the integration of interpretable machine learning frameworks, which can help mitigate overfitting risks and enhance the robustness of policy-relevant insights. Thus, many studies began to use SHAP (Shapley Additive exPlanations) to interpret the nonlinear results of machine learning models (Du et al., 2022; Liu et al., 2023c; Peng et al., 2024; Su et al., 2022a; Yang et al., 2024b). SHAP can give interpretation from a global or local perspective by calculating the contribution of each feature of a single sample to the model output. It facilitates the analysis of heterogeneous effects of built environment factors across stations, which has not been explored in existing studies. Hence, it is of great significance to use interpretable machine learning models to investigate how the nonlinear effects of the built environment are different in various types of stations.

2.4 Nonlinear relationship between built environment and bike-sharing usage

Apart from metro ridership, recent studies increasingly show that the built environment has nonlinear and threshold effects on bike-sharing ridership. Chen and Ye (2021) applied gradient boosting regression trees (GBRT) to explore the nonlinear relationship between dockless bike-sharing usage and land use mix, transport facility, and accessibility variables. The results showed that population density and employment density are the two most significant variables. From street-level, Caigang et al. (2022) identified clear nonlinear effects in how traffic conditions and built-environment indicators shape dockless bike-sharing ridership by using GBDT. While Chen et al. (2024) further investigated pronounced threshold effects in population, bus, and subway densities for origin-destination bike-sharing ridership by using XGBoost.

Meanwhile, many studies focus on the integrated use of metro and bike-sharing, which can also be seen as a special form of metro ridership. Cheng et al. (2022) used quantile regression to show that the built-environment effects vary across quantiles in both the significance levels and magnitudes of coefficients. Song et al. (2023) showed nonlinear and interactive influences of built environment factors on metro-oriented dockless bike-sharing imbalances. Recent work shows that the effects of the built environment vary by space, time, user, and mode. Considering spatial dependence, Zhang et al. (2025) found that nonlinear effects on the integrated use of metro and bike-sharing differ by time of day and transfer flow. Dai et al. (2025) developed a user-ID–based clustering to distinguish commuting scenarios and show that treating all integrated trips as one category can misstate nonlinear effects. Li et al. (2025) classified Shenzhen stations into commercial, residential, and office types within a "3Cs" (convenience, comfort, caution) framework and found that the determinants of integrated use of metro and bike-sharing during the morning peak are station-type specific, evidencing place-level heterogeneity of nonlinear effects. Heterogeneity also arises from modal difference. Liu et al. (2025) compared the factors shaping the choice between shared bikes and shared e-bikes for metro access, indicating that the built environment conditions users' mode selection for first/last-mile links. Complementing this, Zhou et al. (2025) applied random forests (RF) to compare nonlinear built-environment effects on metro-integrated bike-sharing and ride-sourcing, showing that key predictors and threshold patterns differ across these two shared modes integrated with the metro.

In general, these studies have provided valuable insights into the nonlinear relationship between the built environment and bike-sharing ridership. However, it is still unknown how bike-sharing usage will shape BSCA and how BSCA will affect the heterogeneous nonlinear effects of the built environment on metro ridership. Therefore, in this paper, we aim to focus on the metro ridership and explore the nonlinear relationship change between PCA and BSCA.

3 Material and methods

The research framework of this study is shown in Fig. 1. First, we collect station level metro ridership data, bike-sharing trip data, and the built environment data, including land use, transport, accessibility, and socio-economic data. Then, we propose an identification method based on the bike-sharing trip data to identify the BSCA of each metro station. Next, we train 8 XGBoost models of four types of metro ridership in PCA and BSCA and interpret them with clustering-based SHAP to explore the heterogeneous nonlinear relationship in different types of stations.

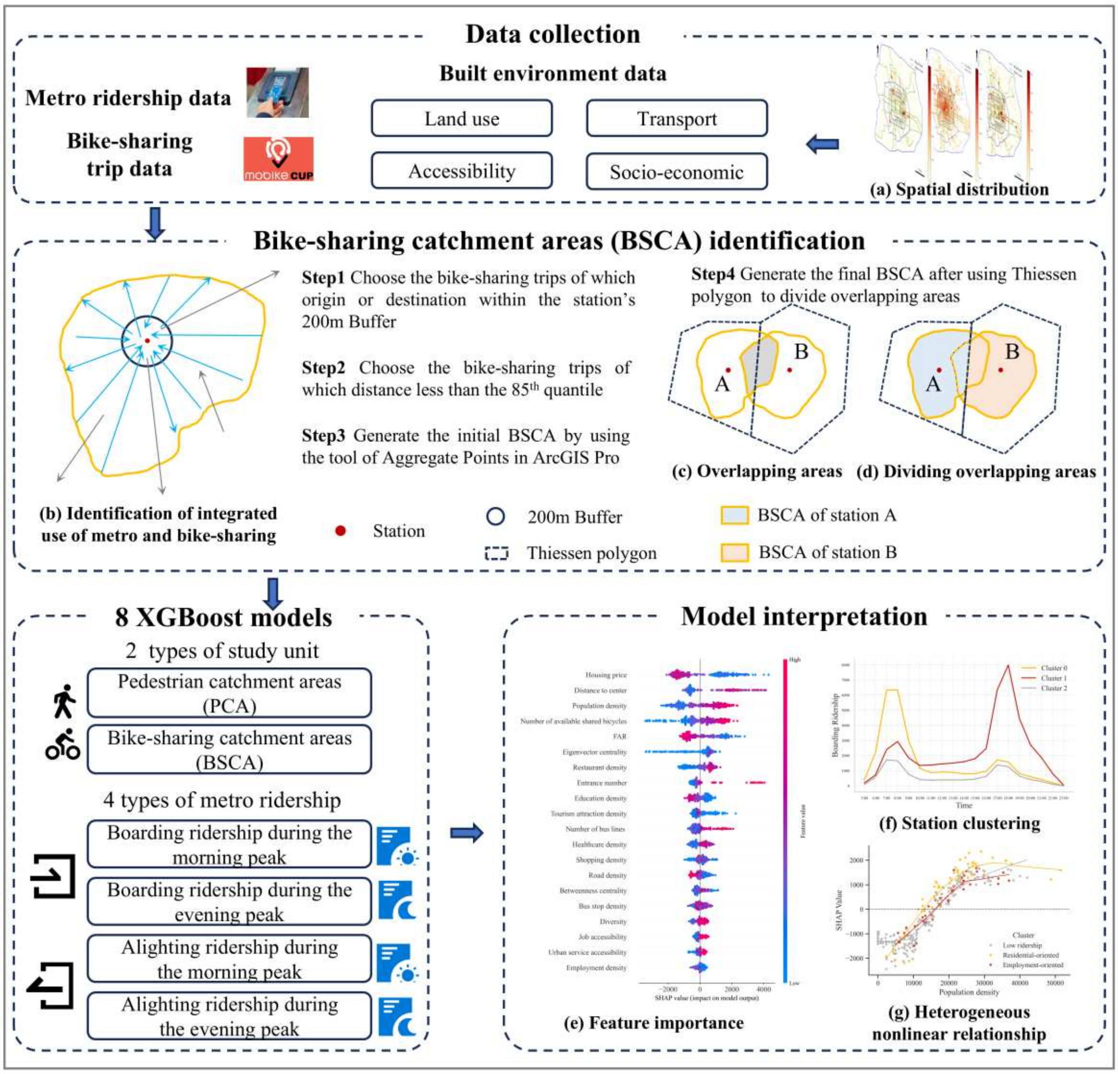


Fig. 1. Research framework

3.1 Study area

We select Beijing as the study area. As shown in Fig. 2, Beijing has 16 districts, with a total area of 16,800 km^2, covering a population of over 21 million. Among these districts, Dongcheng, Xicheng, Chaoyang, Haidian, Fengtai, and Shijingshan are six central districts. According to the Beijing Transport Annual Report (Beijing Transport Institute, 2018), as of the end of 2017, there were 22 urban rail transit lines and 370 urban rail transit stations. After removing other types of urban rail transit stations and reduplicated transfer stations, we select 302 metro stations in the study area. The metro network has a total operational mileage of 608 km. The metro network mainly serves the areas within the 6th ring road of Beijing, covering about 2,267 km^2. The residential population within the 6th ring road accounts for 78% and 80% of Beijing's total population and employment, respectively. The intensely developed urban fabric makes Beijing an ideal case study for researching the relationship between the built environment and metro ridership.

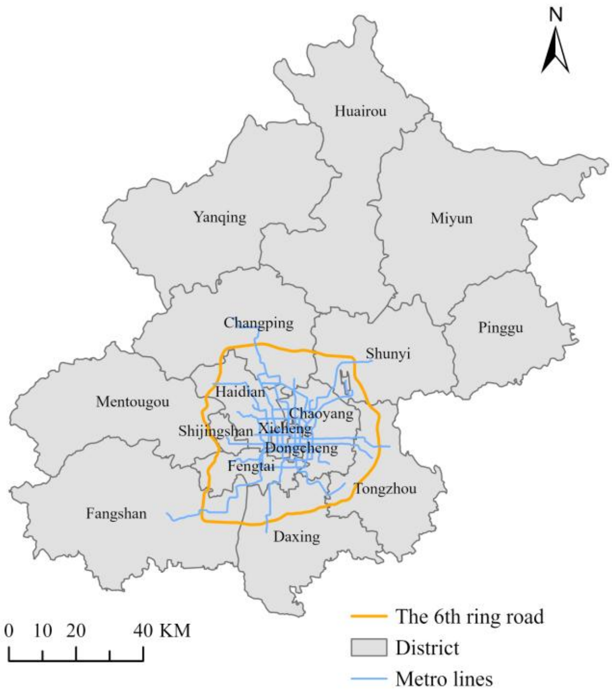
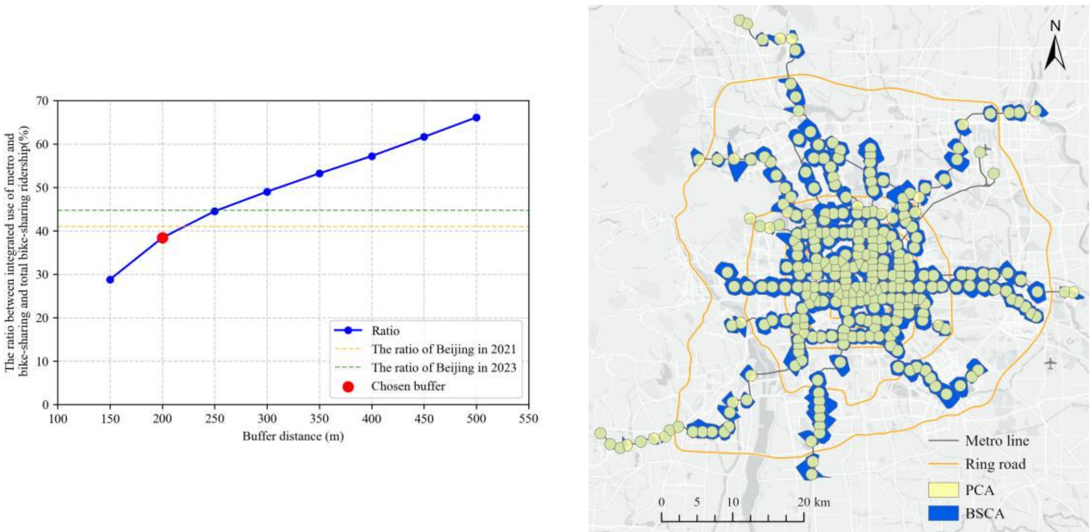


Fig. 2. Study area

3.2 Identification of bike-sharing catchment areas (BSCA)

We used one week of Mobike bike-sharing trip data in 2017 to identify the bike-sharing catchment areas. Mobike is one of the biggest bike-sharing companies in China, which also has the largest market share in Beijing in 2017. First, a total of 648,946 trips during morning and evening peak hours on weekdays were chosen. Second, we filtered trip records of which the origin or destination is within the 200m buffer of a station as the potential integrated use of metro and bike-sharing. In existing research, the buffer size for identification of the integrated use of metro and bike-sharing usually ranges from 50m to 500m, considering the travel behaviour of parking shared-bikes (Chen et al., 2022b; Dai et al., 2025; Fu et al., 2023; Guo et al., 2021; Lv et al., 2024; Sun et al., 2025; Zhan et al., 2023; Zhou et al., 2025; Zhu et al., 2024; Cheng et al., 2022). The 200m buffer was chosen in this study for several reasons. The original bike-sharing trip data was aggregated into the 150m grid. Thus, a larger buffer was needed to ensure each station include the girds of bike-sharing trips. In addition, as shown in Fig. 3(a), we measured the ratio between identified integrated use of metro and bike-sharing in different buffers and total bike-sharing ridership. According to bike-sharing reports in China (China Academy of Urban Planning and Design, 2021; China Academy of Urban Planning and Design, 2023), this ratio of Beijing was 41.0% in 2020 and 44.7% in 2023. Considering the increased popularity of metro and bike-sharing usage in recent years, we selected the 200m buffer, for which the ratio is 38.4%. The average distance of these identified trips is 1.2 km, which is also close to that in the above reports. Third, we excluded long-distance trips based on the 85th percentile of travel distance following existing research (Wu et al., 2021; Zhan et al., 2023; Wang et al., 2025). Fourth, for the remaining trips, we aggregated origin-destination (OD) points using ArcGIS's "Aggregate Points" tool to generate the initial catchment areas for each station. For stations with less than 10 bike-sharing trips (e.g., airport stations), we retained the traditional PCA instead, following the existing research (Zhan et al., 2023). Finally, the overlapping areas of all stations were assigned using Thiessen polygon, and we obtained the differentiated bike-sharing catchment areas of each station.

In this paper, we choose the 800m buffer to represent PCA for several reasons. First, based on the most early research of Kuby et al. (2004) and Cervero (2006), most research on the effects of the built environment on metro ridership have used 800m buffer as the PCA (Cardozo et al., 2012; Deng et al., 2022; Du et al., 2022; Gan et al., 2020; Huang et al., 2024; Li et al., 2020; Li et al., 2024; Su et al., 2022a; Wang et al., 2023a; Xi et al., 2024; Yang et al., 2023a; Yang et al., 2024a; Yang et al., 2023b; Zhao et al., 2014; Zhu et al., 2024b). Second, the 800m buffer is widely used because it is close to the suitable walking distance. And the walking time is about 10 minutes, which is also close to the average travel time of bike-sharing trips (China Academy of Urban Planning and Design, 2021). Therefore, using the 800m buffer to represent PCA makes the comparison between the nonlinear effects of the built environment on metro ridership under PCA and BSCA more comparable. As shown in Fig. 3 (b), we compare the obtained BSCA with traditional PCA. The former is much larger than the latter, especially in suburban stations. On average, compared with PCA, the areas and radius of BSCA increase 48% and 20%, respectively. The bigger areas can effectively consider the built environment for the integrated use of metro and bike-sharing. BSCA also shows more spatial heterogeneity as the actual bike-sharing trip data can reflect travel behavior in different areas. Moreover, BSCA has the advantage of excluding inaccessible areas such as rivers and mountains.



(a) The choice of buffer (b) BSCA vs PCA

Fig. 3. Identification procedure and spatial distribution of BSCA

3.3 Data and variable

3.3.1 Dependent variable

We used the metro ridership data of three weekdays during one week in January 2018 from the Beijing Subway, including daily 6,085,654 trips within the metro system. It was a typical week with good weather, which is suitable for travel by cycling and the metro. According to the Beijing Transport Annual Report (Beijing Transport Institute, 2019), the average boarding ridership on a weekday in Beijing in 2018 was 6,340,000. We chose each station’s boarding ridership and alighting ridership during the morning peak hours (7:00-9:00) and evening peak hours (17:00-19:00) as the dependent variable. As shown in Table 1, the ridership during the morning peak is much larger than that of the evening peak.

Table 1. Descriptive statistics of variables

Variables		Description	PCA mean	PCA std	BSCA mean	BSCA std
Dependent						
Boarding ridership during the morning peak		Passengers entering the station from 7:00 to 9:00	5373.45	4626.12	5373.45	4626.12
Boarding ridership during the evening peak		Passengers entering the station from 17:00 to 19:00	4458.03	4890.58	4458.03	4890.58
Alighting ridership during the morning peak		Passengers leaving the station from 7:00 to 9:00	4767.58	5418.84	4767.58	5418.84
Alighting ridership during the evening peak		Passengers leaving the station from 17:00 to 19:00	3863.29	2968.21	3863.29	2968.21
Independent						
Land use	Population density	Population per km^2	15639.79	10289.76	15694.25	9945.88
	Employment density	Employment POIs per km^2	56.18	64.25	54.46	58.32
	Healthcare density	Healthcare POIs per km^2	26.63	21.67	26.10	20.24
	Education density	Healthcare POIs per km^2	61.64	56.24	59.71	48.70
	Restaurant density	Restaurant POIs per km^2	96.66	90.12	92.02	82.97
	Shopping density	Shopping POIs per km^2	185.71	218.93	176.64	193.92
	Tourism attraction density	Tourism attraction POIs per km^2	8.43	16.30	7.93	14.72
	FAR	The ratio of the total floor area of all buildings within the catchment areas and the total catchment areas	1.47	0.61	1.47	0.58
	Diversity	Land use mix	1.40	0.21	1.42	0.17
Transport	Distance to center	The network distance to Qianmen station in metro system (m)	16561.89	11004.51	16561.89	11004.51
	Betweenness centrality	It is calculated by identifying all shortest paths between every pair of stations in the network and determining how frequently the target station appears on these paths.	0.05	0.04	0.05	0.04
	Eigenvector centrality	It computes the centrality of a node by adding the centrality of its predecessors. It is higher when the station links more important stations.	0.03	0.05	0.03	0.05
	Entrance number	Number of entrances of the station	4.17	1.91	4.17	1.91
	Road density	Length of road per km^2 (km/km^2)	11.43	5.46	11.25	5.32
	Number of bus lines	Number of bus lines	24.46	14.99	30.29	17.37
	Bus stop density	Number of bus stops per km^2	7.07	3.96	7.09	3.72
	Number of available shared bicycles	Number of available shared bicycles during peak hours	1735.73	1229.08	2116.46	1503.25
Accessibility	Job accessibility	Number of employment-type POIs that can be reached within 45 minutes by using the metro system	5950.62	4960.75	13569.68	8413.63
	Urban service accessibility	Number of other five types of POIs that can be reached within 45 minutes by using the metro system	37647.54	29701.53	87530.76	53241.37
Socio-economic	Housing price	The average housing price of residential communities (CNY/m^2)	72828.99	29555.65	72441.04	29810.40

3.3.2 Independent variable

Regarding land use, we first used population density, employment density, and the density of the other five key urban services; healthcare, education, restaurant, shopping, and tourism attraction. Population data was obtained by using data from 2020 Beijing's 7th National Population Census data and residential community data from Lianjia (https://bj.lianjia.com/), one of China's largest real estate companies. The latter comprises 8,923 records that include community name, geographic coordinates, housing price, number of households, and building count. We proportionally distributed the census tract-level population to communities based on the share of households. The resulting community-level population was then aggregated in each metro station's catchment areas. The average housing price of residential communities was also selected as the socio-economic factor.

Employment, healthcare, education, restaurant, shopping, and tourism attraction data were collected from the POI data of Amap (https://ditu.amap.com/) in 2018. FAR (Floor area ratio) was calculated based on the 2020 building data obtained from the open data from Che et al. (2024). Diversity was determined by the Shannon diversity index, which can be calculated as follows:

$$Land\,use\,mix = -\sum_{i=1}^{n} p_i \ln p_i \tag{1}$$

where p_i represents the proportion of i^{th} POI and n is the number of POI types.

In terms of transport, we first chose three indicators related to the station's location in a metro system including distance to the city center, betweenness centrality, and eigenvector centrality. Then, we incorporated transport facilities within and outside the metro system. The data of entrance, bus stop and bus line were collected from Amap. Road density was calculated based on 2018 OSM data . We used the number of available shared bicycles to represent the supply of bike-sharing around each station. All bike-sharing trips with an origin or destination located within the station's catchment area were selected. The total number of available shared bicycles during peak hours was then calculated according to the unique ID of bicycles.

For accessibility, we defined job accessibility and urban service accessibility as the ease of reaching employment and the other five kinds of urban services by using urban rail transit, respectively. Considering that previous studies tend to ignore the travel time during the first and last mile, an improved cumulative opportunity method was proposed to calculate the accessibility based on travel time during the full trip chain following the research of Chen et al. (2025). It can be illustrated as:

$$A_i^m = \sum_j O_j f\left(t_{ij}^m\right) \tag{2}$$

$$f\left(t_{ij}^m\right) = \begin{cases} 1, & if\ t_{ij}^m \le t_0 \\ 0, & if\ t_{ij}^m > t_0 \end{cases} \tag{3}$$

where A_i^m is accessibility of station i under particular catchment areas (m=1 for PCA, m=2 for BSCA), representing the total number of opportunities that can be reached within the time threshold; O_j is the number of POI at station j and $f\left(t_{ij}^m\right)$ is the binary impedance function based on travel time considering full trip chain; t_{ij}^m includes the average travel time from the residential communities within catchment areas to origin station i, the travel time from station i to station j spent within the metro system, and the average travel time from destination station j to POI within catchment areas. The time spent in the metro system was calculated based on the shortest path, using the Dijkstra algorithm. The time for access and egress travel under PCA and BSCA was measured by the walking and cycling route planning API of Amap, respectively; 45 min was chosen for t_0 as the average commuting time in Beijing is 47 min (China Academy of Urban Planning and Design, 2020).

3.4 Clustering-based interpretable machine learning method

3.4.1 Station clustering method

In this study, we applied the Fuzzy C-means (FCM) clustering algorithm to identify typical temporal ridership patterns across metro stations. Unlike conventional clustering, FCM allows each station to simultaneously belong to multiple clusters with varying membership degrees, which better reflects the multifunctional and overlapping roles of urban transit nodes (Liu et al., 2019; Mirzaei et al., 2025). Average hourly boarding and alighting series (5:00–23:00) across three weekdays were used to construct a representative "typical day" profile for each station. These time-series vectors were used as input features for clustering. The fuzziness parameter was set to m = 2.0, with a maximum of 1000 iterations and a convergence threshold of 1e-6. Cluster assignments were determined by the maximum membership degree, while the full membership distribution was retained to assess classification uncertainty. The optimal number of clusters was selected based on the fuzzy partition coefficient (FPC). FCM enabled us to capture differences in daily demand dynamics while preserving the intrinsic fuzziness of station functions, providing a more nuanced characterization than rigid partitioning.

3.4.2 XGBoost model

The eXtreme Gradient Boosting (XGBoost) model (Chen and Guestrin, 2016) was selected in this study for its good performance in analyzing the relationship between the built environment and metro ridership among machine learning models (Liu et al., 2023c; Wang et al., 2023c; Gu and Dou, 2024; Liu et al., 2024; Xi et al., 2024). It has several advantages. First, being non-parametric, tree ensembles learn nonlinear response functions without having to pre-specify functional forms. Second, by growing trees sequentially, the model naturally captures high-order interactions among built-environment variables that are difficult to enumerate a priori. Third, the split-based mechanism induces data-driven thresholds. Fourth, XGBoost is robust to outliers and multicollinearity in the predictors: tree splits depend on order statistics rather than magnitudes, extreme values are isolated in small leaves, and the algorithm can simply choose one of several correlated features at a split; shrinkage, L1/L2 regularization, minimum child weight, and column/row subsampling further stabilize the model. Finally, XGBoost handles missing values natively and scales well to medium-sized tabular data, which fits our setting.

In this study, 80% of the dataset was used for training, and 20% was used for testing. To reduce the risk of overfitting, we employed five-fold cross-validation on the training set. The grid search strategy was adopted to tune the key hyperparameters, including the number of estimators, maximum tree depth, learning rate, minimum child weight, and the L2 regularization term on weights. The optimal parameter set was selected according to the highest average R^2 across the folds.

3.4.3 SHapley Additive exPlanations

We applied the Shapley Additive exPlanations (SHAP) proposed by Lundberg and Lee (2017) to interpret the XGBoost models to address the black box problem of machine learning models. It has been widely used in the research of exploring the nonlinear relationship between the built environment and metro ridership (Du et al., 2022; Liu et al., 2023c; Peng et al., 2024; Su et al., 2022a; Yang et al., 2024b). Based on cooperative game theory, SHAP calculates the marginal contribution of each feature by assessing the change in model output when the feature is added to different combinations of inputs. It can quantify the contribution of each built environment variable to the output ridership in the XGBoost model from the global perspective. As for the local perspective, SHAP produces a contribution vector for every station, enabling cluster analyses without changing the model. We therefore pre-define station clusters and stratify SHAP dependence plots by cluster to study how marginal effects differ across station types. This lets us compare how the same variable operates across station types, and to detect cluster-based thresholds where marginal effects strengthen or saturate.

To explore the change of nonlinear relationship in models with different catchment areas, we defined a new SHAP value change index (SVCI), which can be calculated as follows:

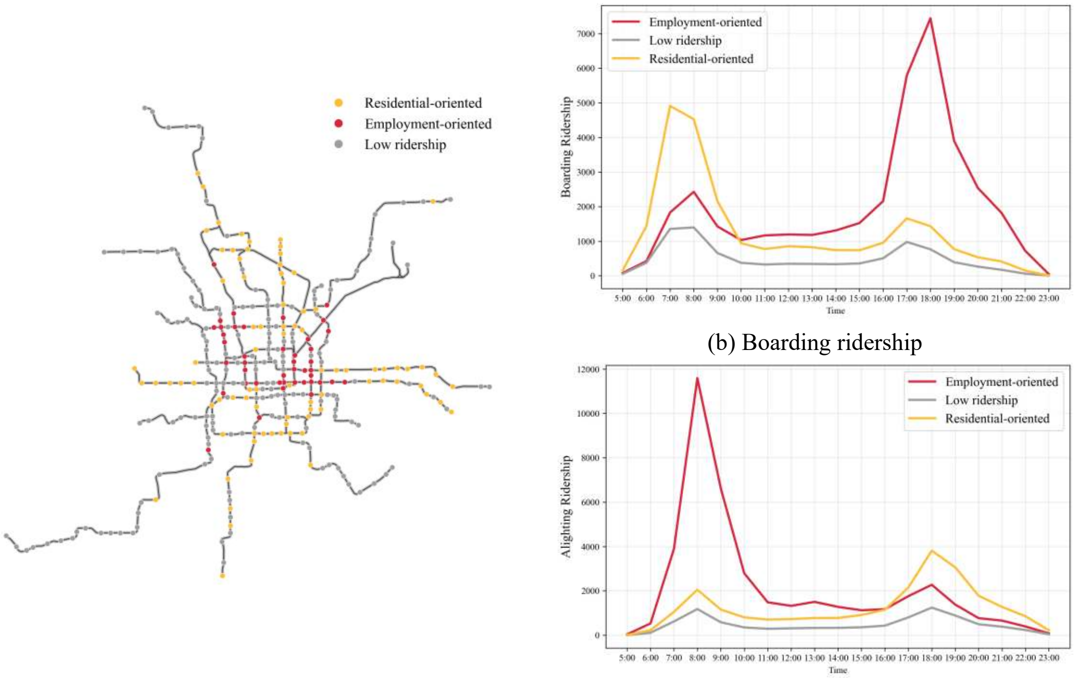
$$SVCI_j = \frac{\sum_{i=1}^{n} | \phi_j^{(i,c)} - \phi_j^{(i,w)} |}{n} \tag{4}$$

where $SVCI_j$ is the SHAP value change of built environment feature j; $\phi_j^{(i,c)}$ and $\phi_j^{(i,w)}$ represents the Shapley value of feature j corresponding to the i^{th} sample derived from the XGBoost model with BSCA and PCA, respectively.

4 Results

4.1 Clustering results

The results of FCM clustering indicate that 3 clusters have the highest FPC. Therefore, we selected 3 clusters of metro stations. The spatial distribution of different clusters is shown in Fig. 4 (a), and hourly boarding ridership and alighting ridership distributions of different clusters are presented in Fig. 4(b) and Fig. 4(c). The first cluster is residential-oriented, showing very high boarding ridership during the morning peak and alighting ridership during the evening peak. They tend to concentrate on some large residential areas, such as Huilongguan and Tiantongyuan in the northern part of the Haidian district and distribute at the end of metro lines. The second cluster can be viewed as employment-oriented stations, presenting extremely high alighting ridership during the morning peak and boarding ridership during the evening peak. These stations concentrate on places that provide large amounts of employment, such as the CBD and Zhongguancun. At last, the ridership of the third type of station is relatively low throughout the day. These stations tend to be widely distributed in the suburban areas and some central areas. The number of stations in residential-oriented, employment-oriented, and low ridership clusters is 84, 43, and 175. It indicates that most ridership concentrates on residential-oriented and employment-oriented stations, while the ridership of a large number of stations is low.



(b) Boarding ridership

(a) Spatial distribution

(c) Alighting ridership

Fig. 4. The identification of optimal clusters and spatiotemporal pattern of different types of station

4.2 Model performance

The best hyperparameters of 8 XGBoost models with different types of metro ridership and catchment areas are shown in Table 2. The Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and coefficient of determination (R^2) are used for the model evaluation. To evaluate the robustness of the models, we repeated each experiment ten times with different random seeds under the same set of hyperparameters, and recorded the resulting performance metrics. We got multiple paired samples of model performance for PCA models and BSCA models across four types of metro ridership. Then we conducted paired significance tests on the differences between PCA in each metric. For each case, we first assessed the normality of the paired differences using the Shapiro-Wilk test. If the normality assumption was satisfied, a one-tailed paired t-test was applied; otherwise, we used the Wilcoxon signed-rank test.

The results show that the models with BSCA significantly outperform the models with PCA. Compared with the PCA-based models, the Test R^2 in the models with BSCA increased by an average of 6% among the four kinds of ridership, and the Test MAE and RMSE decreased by an average of 15% and 17%, respectively. It means that the models with BSCA can better capture the relationship between metro ridership and the built environment. Because BSCA provides bigger catchment areas for bike-sharing, it leads to more accurate calculations of different built environment factors. We also compare the model performance of XGBoost and the other two popular machine learning models, including RF and GBDT (Table B.1). The results indicate that XGBoost has better model performance, confirming the advantage of XGBoost in exploring the nonlinear relationship between the built environment and metro ridership.

Table 2. Performance of the XGBoost models (Statistical significance was tested for all Δ(%) values. Improvements in Test R^2, MAE, and RMSE were statistically significant ($p < 0.01$).)

Model		N_ estimators	Max_ depth	Learning_rate	Test R^2	Δ (%)	Test MAE	Δ (%)	Test RMSE	Δ (%)
Boarding ridership during the morning peak	PCA	300	3	0.05	0.7284 ± 0.0046	-	1823.76 ± 19.57	-	2451.44 ± 21.00	-
	BSCA	500	3	0.02	0.7466 ± 0.0038	+2.5%	1634.58 ± 23.32	-10.4%	2205.18 ± 16.66	-10.0%
Boarding ridership during the evening peak	PCA	400	4	0.02	0.8850 ± 0.0045	-	922.64 ± 15.24	-	1262.77 ± 24.93	-
	BSCA	300	3	0.05	0.9298 ± 0.0021	+5.1%	788.31 ± 25.95	-14.6%	986.89 ± 14.90	-21.8%
Alighting ridership during the morning peak	PCA	200	4	0.04	0.8178 ± 0.0051	-	1472.02 ± 21.10	-	2292.51 ± 32.18	-
	BSCA	300	3	0.03	0.9071 ± 0.0013	+10.9%	1140.14 ± 18.41	-22.5%	1636.73 ± 11.50	-28.6%
Alighting ridership during the evening peak	PCA	200	3	0.04	0.6849 ± 0.0033	-	1080.98 ± 14.68	-	1289.56 ± 6.66	-
	BSCA	200	3	0.04	0.7317 ± 0.0065	+6.8%	965.77 ± 7.88	-10.7%	1190.02 ± 14.42	-7.7%

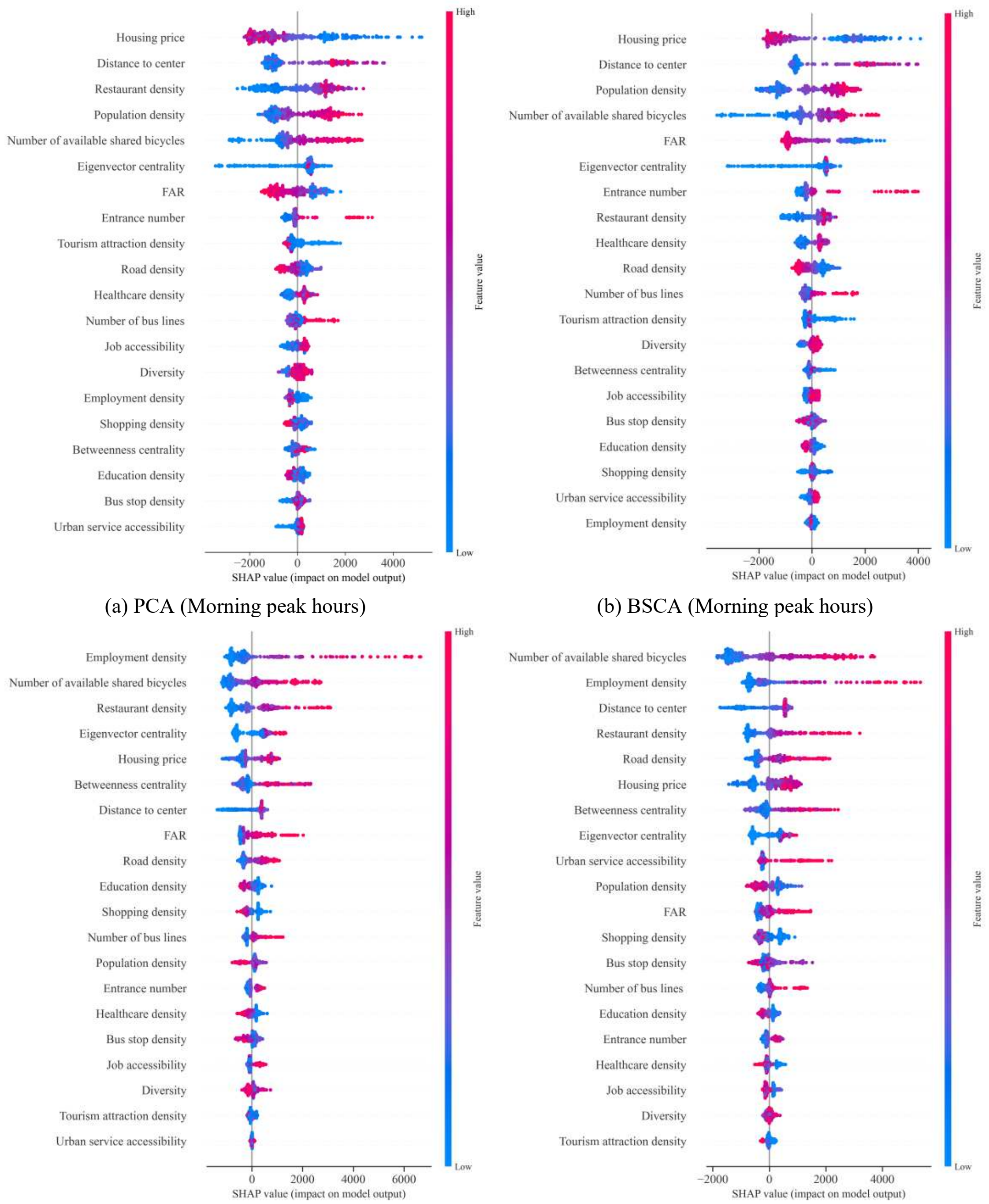
4.3 Importance of features

4.3.1 Boarding ridership

Fig. 5 and Fig. 6 show the feature importance of the models of boarding ridership and alighting ridership under different catchment areas. SHAP values above 0 represent the feature's positive effect on model output, while values below zero indicate a negative contribution. The color represents the magnitude of the feature value, with red indicating higher values and blue indicating lower values.

For the boarding ridership during the morning peak, housing price is the most important built environment variable. Stations with lower housing prices have higher ridership. It may be attributed to the fact that people in suburban areas with low housing prices rely on the metro system to commute, indicating the importance of socio-economic. Population density, distance to center, and FAR are also important for interpreting metro ridership. The large population in outer areas provides sufficient potential commuting ridership for the metro system during the morning peak. FAR was found to negatively affect the ridership, as these large residential communities outside the PCA tend to have relatively low development density due to the dispersed layout. The FAR of residential-oriented stations under PCA and BSCA is 1.55 and 1.52, respectively, indicating that the FAR outside the PCA is relatively low. Furthermore, it is noted that the number of available shared bicycles is more important than other transport facilities such as the number of bus lines, bus stop density and road density. It demonstrates that bike-sharing contributes more to addressing the first-mile problem than other modes.

Regarding boarding ridership during the evening peak, the number of available shared bikes is the most essential feature for explaining ridership, followed by employment density, distance to center, and restaurant density. It indicates that the stations with high density of employment and restaurants in central areas are most connected to trip generation in the evening peak. Bike-sharing provides good opportunities for people to access stations from these places. Road density and FAR also positively affect boarding ridership, indicating that high development density and developed road networks promote metro ridership.



(a) PCA (Morning peak hours) (b) BSCA (Morning peak hours)

(c) PCA (Evening peak hours) (d) BSCA (Evening peak hours)

Fig. 5. Feature importance of the models for boarding ridership

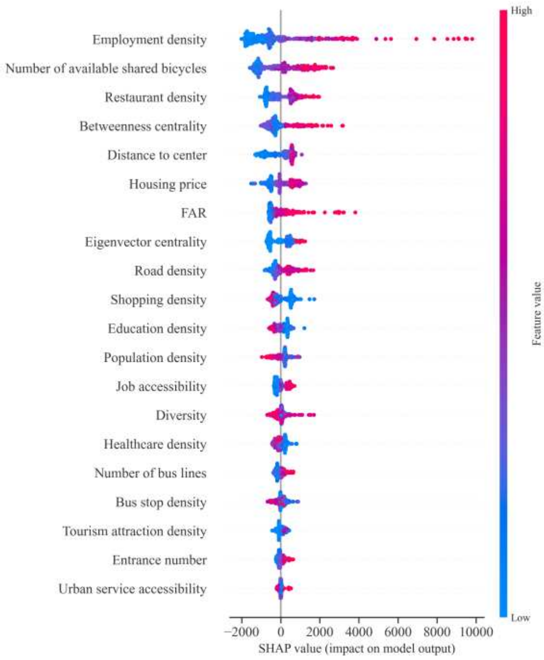
There are some differences between the feature importance of models under PCA and BSCA. For boarding ridership during the morning peak, population density, number of available shared bicycles, and FAR are more highly ranked after considering BSCA. It shows that these features are more important than those we obtain from the PCA-based model. Many large residential communities with relatively low development density in suburban areas are far from the metro station and outside the PCA. Travelers in these areas have to use bike-sharing to access metro stations. The PCA-based model will underestimate the population density and ignore the potential demand by cycling for the metro system (Fig. A.1 (a, b, e, f)). As for the evening peak, the importance of the number of available shared bicycles, distance to center, and road density is higher in the BSCA-based model. It indicates that we should place greater emphasis on these built environment factors contributing to the metro trip generation during the evening peak within BSCA.

4.3.2 Alighting ridership

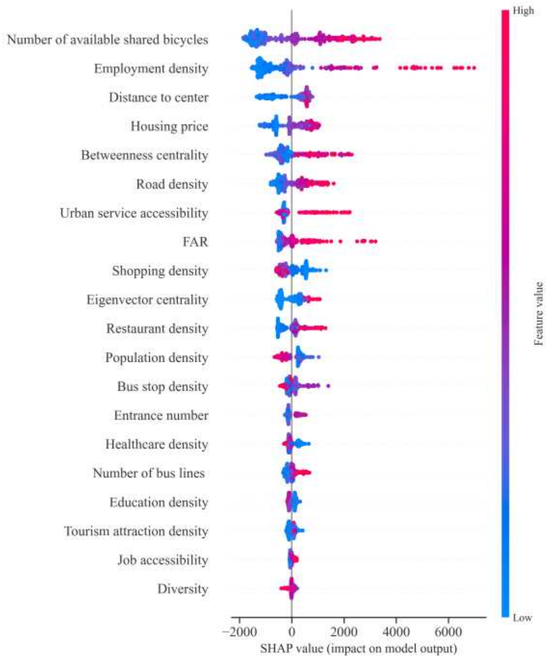
As for the alighting ridership during the morning peak (Fig. 6 (a, b)), the number of available shared bicycles and employment density are the two most essential features. It indicates that commuting for employment is still the most important purpose among all trip purposes during the morning peak, and bike-sharing also serves as the key facility in the last mile of commuting. Furthermore, the stations with more essential locations in the metro network, higher housing prices, and higher FAR tend to have more alighting ridership. Moreover, as with boarding ridership, diversity has little effect on alighting ridership because the origin and destination in peak hours tend to be concentrated in large residential and employment clusters with low diversity.

In terms of alighting ridership during the evening peak, restaurant density contributes the most to the model output, followed by the number of available shared bicycles, FAR, housing price, and distance to the center. Compared with the feature importance in morning boarding ridership, the importance of restaurant density is much higher, indicating the diverse trip purposes in the evening hours. Bike-sharing also serves as an essential solution to the last mile of this type of travel.

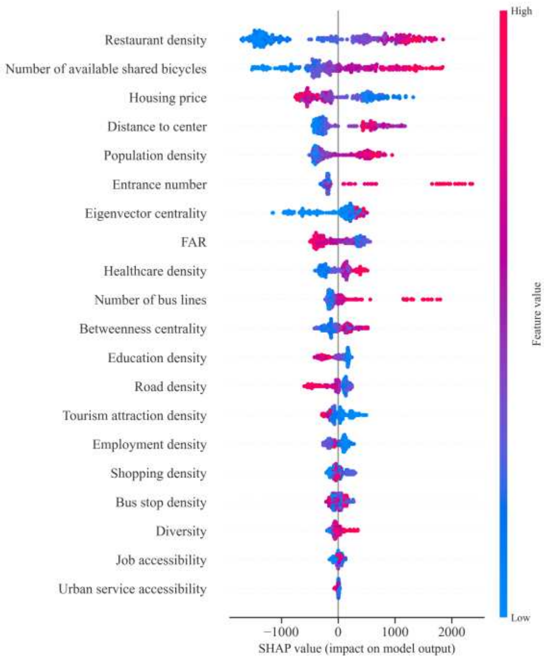
There are also some changes for alighting ridership when using different catchment areas. During the morning peak, first, in the model with BSCA, the number of available shared bicycles surpasses employment density, becoming the most important feature. This can be attributed to the fact that bike-sharing serves various trip purposes during the evening peak hours, and its supply is underestimated in PCA. Second, urban service accessibility ranks much higher in models with BSCA than in PCA. Urban service accessibility under BSCA increases significantly due to the greater first-mile mobility offered by cycling compared to walking, especially for stations located in central areas (Fig. A.1 (g-h)). Good accessibility to these amenities by urban rail transit can promote more ridership generation in stations. During the evening peak, FAR is more highly ranked after considering BSCA, indicating BSCA can better capture the movement of people from high development density places to low development density places.



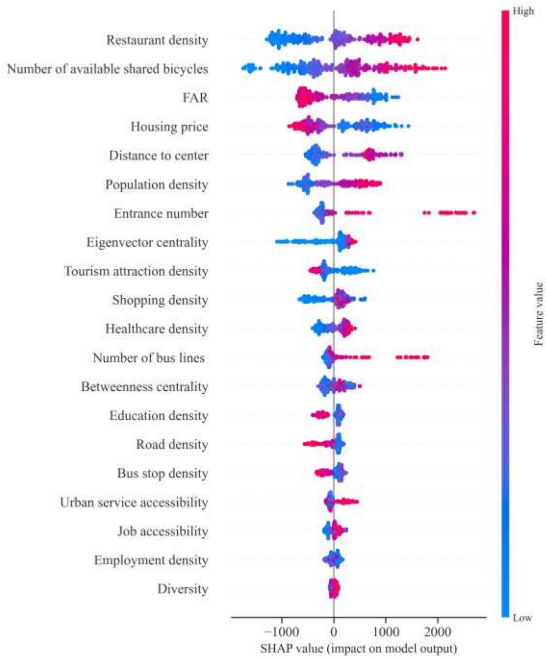
(a) PCA (Morning peak hours)



(b) BSCA (Morning peak hours)



(c) PCA (Evening peak hours)



(d) BSCA (Evening peak hours)

Fig. 6. Feature importance of the models for alighting ridership

4.4 Nonlinear effects analysis

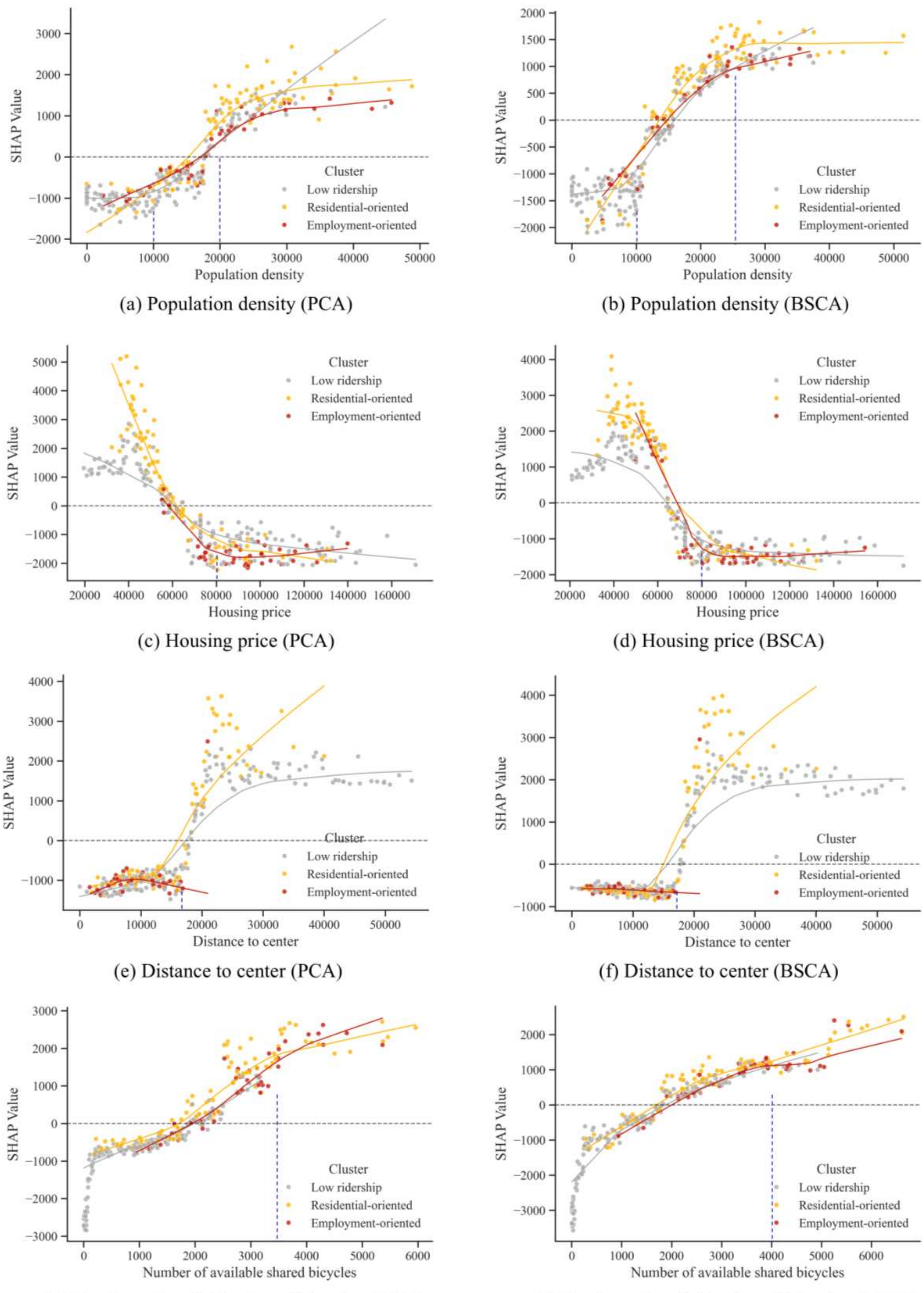
Fig. 7, Fig. 8, Fig. C.1, and Fig. C.2 show the nonlinear relationship between the built environment and different types of metro ridership under PCA and BSCA. We selected four built environment features with either high overall importance or substantial change across catchment areas. Locally weighted scatterplot smoothing (LOWESS) was used in the scatter plots to enhance the visualization of trend changes and visually demonstrate the threshold effects of built environment variables on metro ridership (Liu et al., 2025).

4.4.1 Boarding ridership

For boarding ridership during the morning peak, in general, population density has a strong positive effect on metro ridership as more population provides sufficient travel demand for the metro system (Fig. 7(a-b)). But the 3 types of stations show very different trends. For low ridership stations, population density can contribute to higher metro ridership only when it exceeds about 10,000 people per km^2. It underlines the importance of the demand foundation for these stations. Other types of stations do not show low thresholds, which is different from existing research (Ding et al., 2019; Yang et al., 2023b; Liu et al., 2023b). Meanwhile, there are also some differences in nonlinear relationships from various catchment areas. When population density increases to about 20,000 people per km^2, the model with PCA shows that higher population density in residential and employment-oriented stations cannot significantly increase metro ridership anymore. However, the key threshold is much higher in the model with BSCA, especially for residential-oriented stations, reaching about 25,000. This difference indicates that the model under BSCA underlines higher TOD planning targets for population density, which also means that the model under PCA may underestimate the role of high population density in residential-oriented stations.

Housing price also has a significant threshold effect (Fig. 7(c-d)). When it increases over about 50,000 CNY/m^2, the SHAP value dramatically decreases and remains steady until it reaches 80,000 CNY/m^2, suggesting that metro use diminishes significantly in high-income areas under both PCA and BSCA. Regarding distance to center, it has positive effects on metro ridership in residential-oriented and low ridership stations when ranging from 15 to 30 km (Fig. 7(e-f)). Although existing research has similar findings (Yang et al., 2023b), we also find that the distance to center does not have a significant relationship with boarding ridership for employment-oriented stations. In terms of the number of available shared bicycles, it has a great positive effect on metro ridership as it addresses the problem of the first and last mile (Fig. 7(g-h)). However, in the model with BSCA, the number of available shared bicycles has trivial effects when it exceeds 4,000. It suggests that the oversupply of bike-sharing cannot contribute to increasing metro ridership.

As illustrated in Fig. 8, the results of boarding ridership during the evening peak show employment density has a threshold effect in employment-oriented stations when it exceeds 200 counts per km^2. However, there is no significant threshold for other types of stations (Fig. 8(a-b)). For those stations, both models show the relationship between employment and ridership with a slow initial increase followed by accelerated growth when employment exceeds 70 counts per km^2. As for the number of available shared bicycles, there are evident threshold effects in residential and employment-oriented stations (Fig. 8(c-d)). And the thresholds in models with BSCA are larger. Because PCA-based model underestimates the supply of bike-sharing. In terms of restaurant density, the threshold in employment-oriented stations is 300 counts per km^2 (Fig. 8(e-f)).



(a) Population density (PCA)
(b) Population density (BSCA)
(c) Housing price (PCA)
(d) Housing price (BSCA)
(e) Distance to center (PCA)
(f) Distance to center (BSCA)
(g) Number of available shared bicycles (PCA)
(h) Number of available shared bicycles (BSCA)

Fig. 7. Clustering-based SHAP dependence plots for boarding ridership during the morning peak under PCA and BSCA

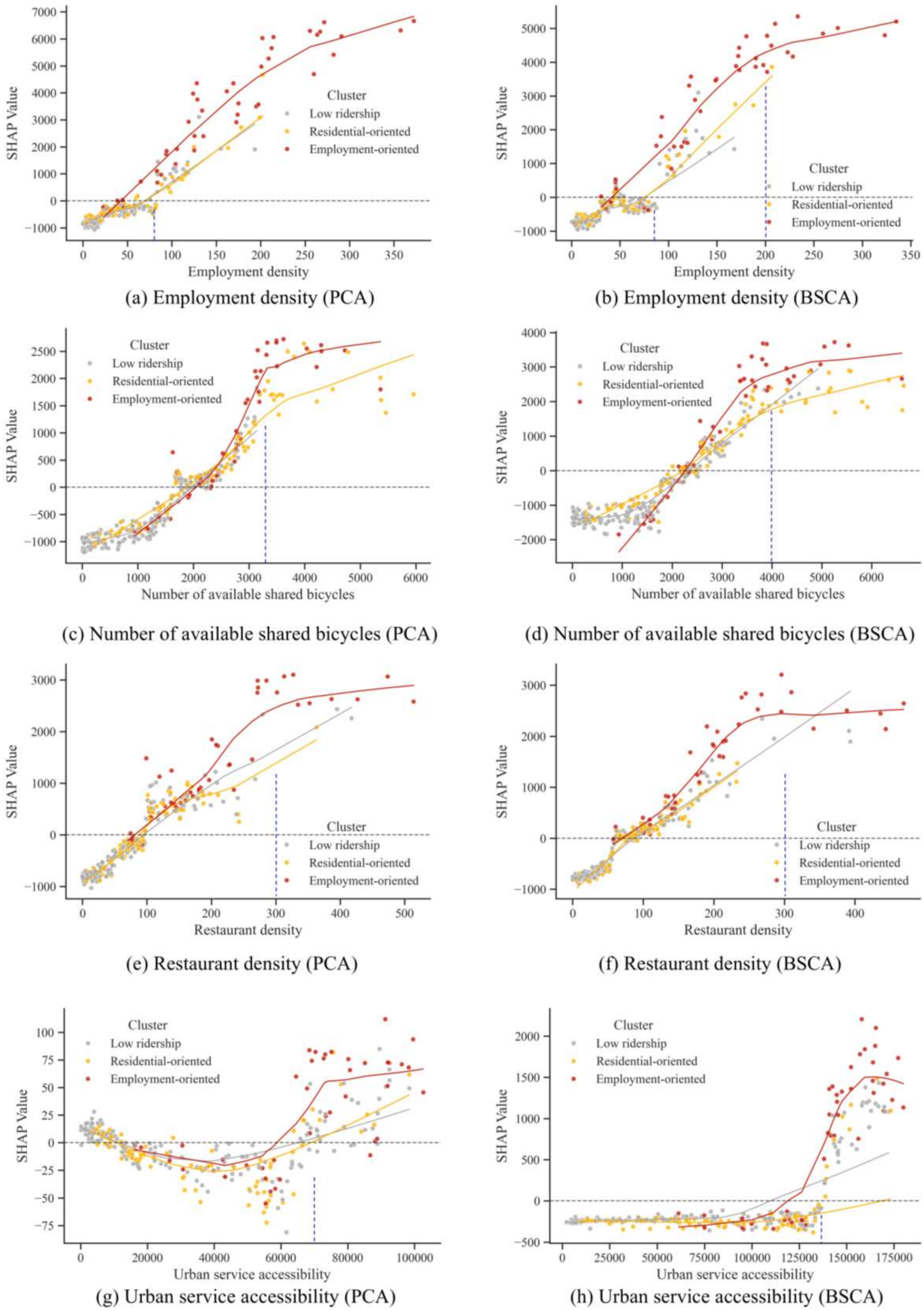


Fig. 8. Clustering-based SHAP dependence plots for boarding ridership during the evening peak under PCA and BSCA

Regarding urban service accessibility, the nonlinear relationship between it and boarding ridership during the evening peak presents huge differences in the two models (Fig. 8(g-h)). In the model with PCA, urban service accessibility is within 100,000 and does not significantly affect ridership. In the model with BSCA, urban service accessibility increases dramatically due to bigger catchment areas and higher mobility during the full trip chain (Fig. A.1(i-j)). Although the effects of urban service accessibility in residential-oriented and low ridership stations are limited, it significantly increases the metro ridership of employment-oriented stations when it exceeds about 130,000. This is likely because good accessibility to urban services such as restaurants and shopping can attract many passengers after work.

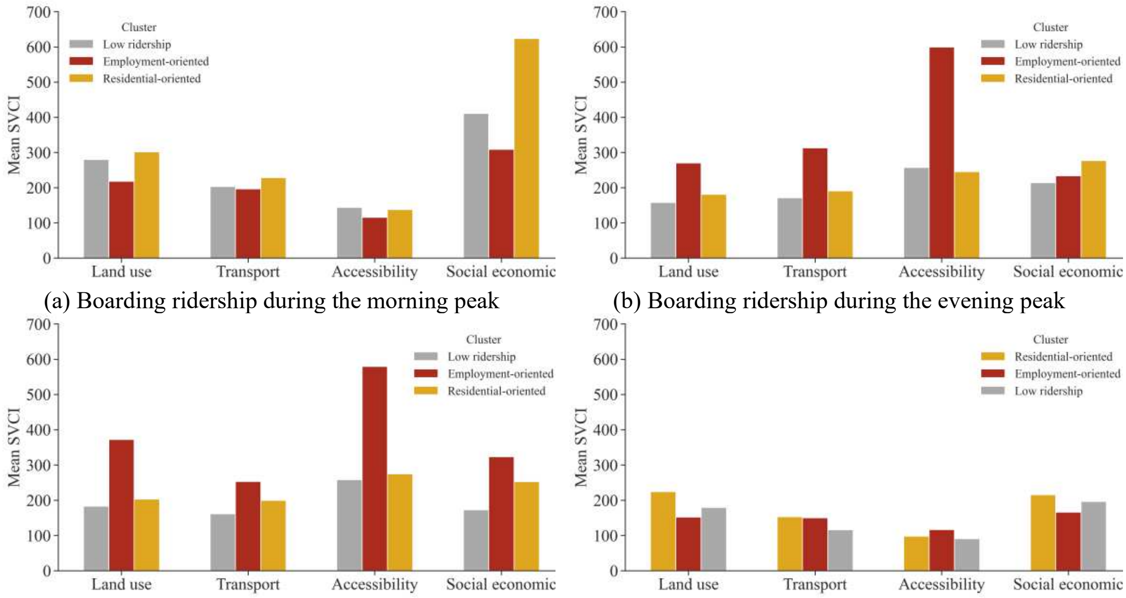
4.4.2 Alighting ridership

As shown in Fig. C.1, in the models for alighting ridership during the morning peak, the nonlinear effects of employment density and number of available shared bicycles are similar to those of boarding ridership during the evening peak, suggesting the commuting pattern of these stations in the morning and evening (Fig. C.1(a-d)). For FAR, it has various effects on 3 types of stations (Fig. C.1(e-f)). FAR can largely promote ridership in residential-oriented and low ridership stations when it exceeds 1.5. While a FAR higher than 2.3 can greatly improve ridership at employment-oriented stations. The results underline the importance of high development density in the heterogeneous TOD planning for various stations. Regarding bus stop density, it contributes more in the model with BSCA and presents opposite trends (Fig. C.1(g-h)). In the BSCA model, when bus stop density is higher than around 8 stops per km^2, it has negative effects on all types of stations. It is noted that the relationship is different from the average linear positive effects suggested in previous research (Chen et al., 2019). A plausible explanation is that the high level of bus service can lead to competition between the bus and the metro. In particular, dense bus networks may provide comparable accessibility at a lower cost or with fewer transfers, thereby reducing passengers' incentive to choose the metro.

Regarding the models for alighting ridership during the evening peak (Fig. C.2), the nonlinear relationship between population density and ridership shows similar patterns to that in the models for boarding ridership during the morning peak (Fig. C.2(a-b)). The model with BSCA has higher thresholds. For restaurant density (Fig. C.2(c-d)), the effects in residential stations are stronger than those in other stations. And restaurant density has trivial effects on metro ridership when it surpasses about 200 counts per km^2. In the BSCA model, road density has negative effects on metro ridership for residential-oriented and low ridership stations when it exceeds about 10-12 km per km^2 (Fig. C.2(e-f)). It is different from the insignificant relationship between road density and boarding ridership in some studies (Chen et al., 2019; Yang et al., 2023b). It indicates that a certain level of road density can contribute to alighting ridership, but higher road density may facilitate alternative travel modes such as private cars and buses, thereby decreasing potential alighting ridership. The effects of the number of bus lines are limited when it is less than 40, suggesting that ample bus service also serves as an important way of addressing the last mile problem in the evening peak (Fig. C.2(g-h)).

4.4.3 Change of the nonlinear relationship

To explore the change of the nonlinear relationship in different catchment areas, Fig. 9 shows the mean SVCI of four types of built environment features in different station clusters. First, the SVCI of the model for alighting ridership during the evening peak is the lowest, but models for the other 3 types of metro ridership show high sensitivity to the choice of catchment areas. Second, the results also reveal that the employment and residential-oriented stations are more sensitive to catchment areas. In the models for boarding ridership during the morning peak and alighting ridership during the evening peak, the SVCI of residential-oriented stations is higher than others. In the models for boarding ridership during the evening peak and alighting ridership during the morning peak, the SVCI of employment-oriented stations is much higher than others. These two types of stations are most related to trip generation and attraction and are also very likely to change in BSCA. Third, in terms of the four types of built environment factors, land use and socio-economic factors tend to change more dramatically in models for boarding ridership during the morning peak and alighting ridership during the evening peak because BSCA incorporates more residential communities with lower housing price outside PCA. Therefore, these findings indicate that the planning targets from the traditional model may be less accurate, especially for residential-oriented and employment-oriented stations, and BSCA is of great significance for effective TOD planning.



(a) Boarding ridership during the morning peak (b) Boarding ridership during the evening peak

(c) Alighting ridership during the morning peak (d) Alighting ridership during the evening peak

Fig. 9. The mean SHAP value change index (SCVI) of four types of built environment factors

5 Discussion

5.1 PCA VS BSCA

To answer the first research question, we identify BSCA and compare the nonlinear effects of the built environment on metro ridership from models with PCA and BSCA. First, we found that the models with BSCA outperform the models with PCA. The average R^2 of the testing dataset in all kinds of ridership is 0.83, much higher than that of models with PCA. It underscores the advantage of using BSCA in exploring the relationship between the built environment and metro ridership, which has not been discovered in existing research (Liu et al., 2023b). Although some studies focusing on the relationship between the built environment and integrated use of metro and bike-sharing underline the use of BSCA (Zhan et al., 2023; Wang et al., 2025; Zhu et al., 2024), this paper investigates how BSCA improves the model performance and better explains metro ridership. Theoretically, our findings highlight the critical limitation in traditional TOD research. Different from previous research focusing on the various buffer sizes of PCA (Wang et al., 2022b; Liu et al., 2023c; Li et al., 2023a), we find that the reliance on PCA fails to capture the actual effective catchment area expanded by micromobility. By delineating the BSCA, this study bridges the gap between static buffer and dynamic multimodal accessibility, proving that the integrated use of metro and bike-sharing fundamentally reshapes the spatial scale of metro demand.

Second, regarding the key factors in the built environment that affect metro ridership, we found that population density, employment density, restaurant density, FAR, and housing price are the most important factors for explaining metro ridership. This is in line with the results of previous studies (Cervero, 2006; Andersson et al., 2021; Wang et al., 2023; Yang et al., 2023b). However, the magnitude and ranking of these factors differ notably under BSCA. For example, population density and FAR rank higher under the BSCA model for boarding ridership during morning peak, which also rank higher than the findings from existing research (Liu et al., 2023c; Yang et al., 2023b). Meanwhile, compared with the model with PCA, the model with BSCA has a higher threshold for population density. These differences can be attributed to the built environment change in the larger spatial scale of BSCA that tends to be ignored in previous studies. Our findings highlight that higher planning targets of population density within BSCA should be considered in practical TOD planning.

Third, regarding our second research question, we also found that the number of available shared bicycles is one of the most important variables among all built environment factors across all models. The results are different from those of previous research that only considers the effects of bike stations (Chen et al., 2019; Wang et al., 2022a; Zhu et al., 2019), underlining the importance of bike-sharing supply in explaining metro ridership. Meanwhile, the number of available shared bicycles ranks higher in the BSCA models than in the PCA models. These new findings highlight the important role of bike-sharing in promoting metro ridership. The traditional PCA model will underestimate the importance of the bike-sharing supply. In terms of bike-sharing accessibility, urban service accessibility ranks highly in the BSCA models for boarding ridership during the morning peak and alighting ridership during the morning peak. It also shows very different nonlinear trends in the BSCA models. These findings enrich previous studies that ignore bike-sharing accessibility (Liu et al., 2023a; Peng et al., 2023; Du et al., 2022), highlighting the important role of bike-sharing in improving accessibility and boosting metro ridership.

5.2 Heterogeneity in different types of stations

To answer the third research question, we propose a clustering-based interpretable machine learning method to explore the heterogeneity of nonlinear effects in different types of stations. We found that the effects of the built environment on metro ridership vary across station types, showing distinct trends and threshold patterns that highlight a heterogeneous nonlinear relationship. The inclusion of clustering extends the knowledge of existing research which has only explored the nonlinear relationship in all stations (Liu et al., 2023c; Peng et al., 2024; Xi et al., 2024). Theoretically, this challenges the "one-size-fits-all" assumption in TOD planning, which often presumes the same relationship between built environment factors and ridership across all stations. Our results demonstrate that such simplifications fail to capture the complex spatial heterogeneity where station function changes travel behavior mechanisms. Practically, it implies that applying uniform density targets across the city is inefficient. Since the nonlinear thresholds of the built environment vary significantly by station type, the uniform policy may lead to resource mismatch.

Furthermore, we also found that the SHAP value of built environment factors changes more dramatically in residential and employment-oriented stations. It means that the nonlinear relationship between the built environment and metro ridership in these stations is more sensitive to catchment areas. Thus, TOD planning for these specific station types requires a more expansive spatial scope (BSCA) to accurately estimate potential demand, whereas traditional PCA might suffice for city-center stations where walking dominates.

5.3 Policy implications

Since the beginning of research on TOD planning based on the built environment two decades ago (Kuby et al., 2004; Cervero, 2006), most subsequent studies followed their use of PCA. However, under the new context of bike-sharing in recent years, we show that BSCA extends the PCA and can better interpret the relationship between the built environment and metro ridership. Therefore, BSCA should be more carefully considered in future TOD practice. First, the planning targets for key transport and land use factors should be determined in BSCA rather than only in PCA. It is noted that the obtained planning targets reflect Beijing's urban structure and mobility patterns and should not be applied directly to other cities without adaptation. For other urban contexts, policymakers should calibrate these thresholds by considering their own demographic, spatial, and transport conditions. Second, our findings suggest that planning authorities need to take more concrete and coordinated actions. Urban planning departments should refine zoning regulations to optimize population and employment density and other urban services around metro stations. Transport management agencies should strengthen the supply and allocation of shared bicycles, ensuring seamless connections between bike-sharing and metro services, particularly in high-demand catchment areas. Meanwhile, inter-agency coordination is needed to integrate infrastructure investment (e.g., bike lanes, parking facilities) with land use and metro development.

Considering the nonlinear effects of the built environment on metro ridership differ in different types of stations, we should also consider the heterogeneity of TOD planning in different types of stations. The planning targets of key transport and land use factors need to be carefully set in various kinds of stations according to their nonlinear relationship. For example, for residential-oriented stations, a higher population density threshold requires more bike-sharing supply in BSCA. For employment-oriented stations, policy should prioritize higher development density than other types of stations and allocate enough shared bicycles. The most important aim of the low-ridership stations should first meet the minimum threshold of population density to ensure enough potential demand.

5.4 Limitations and future directions

This study has several limitations. First, regarding catchment areas, we only choose the most used

800m buffer to represent the PCA. More typical buffers, such as 1000m and 400m, can be used in future research to explore the effects of PCA's buffer. The BSCA identified by bike-sharing data ignores the integrated use of other kinds of bicycles and metros, and we only use the data of one of the biggest bike-sharing companies in Beijing due to data limitations. The bike-sharing trip data from more bike-sharing companies should be used to better identify the BSCA and provide a more accurate built environment of bike-sharing supply. Moreover, we ignore the heterogeneity of bike-sharing travel behaviour of different demographic groups or trip types. Thus, it is still necessary to explore how different identification of catchment areas will affect the differences of the nonlinear relationship between PCA and BSCA. We also noted that other transportation modes, such as bus, taxi, ridesourcing, private vehicles, and e-scooters, are also essential in addressing the first and last-mile problem. Thus, these catchment areas should also be explored in future studies.

Second, regarding metro ridership, this paper mainly uses ridership during peak hours from three weekdays during one week in 2018. Ridership from other time periods and other types of metro ridership, such as weekday off-peak ridership and weekend ridership, can be used to explore the temporal heterogeneity of the nonlinear relationship if more data is available in the future. Meanwhile, we only focus on station-level metro ridership in this paper. More research can be done to focus on the station-to-station level metro ridership.

Third, about the built environment data, there are some time mismatches between some data, such as FAR data, bike-sharing data, and metro ridership data. More consistent data should be used in the future to avoid the effects of mismatched data on results. Furthermore, we only use the mean walking or cycling travel time when measuring accessibility factors. More granular calculations can be done to better represent accessibility. More socio-economic factors, such as age and income, can be used in the modeling if the data is available.

Fourth, when it comes to the modeling approach, we only use the machine learning models in this study. More other models can be used to test the effect of the model choice on results.

Fifth, the case study from Beijing may be different from other cities, and future studies can be done in more cities to confirm the generality of our research framework and findings.

6 Conclusion

Exploring the relationship between the built environment and metro ridership is essential in understanding how transport and urban planning can be improved to realize the sustainable growth of cities. Considering the increasingly popular integrated use of bike-sharing and metro, this study identified the BSCA and employed a clustering-based interpretable machine learning approach to compare the heterogeneous nonlinear effects of the built environment on metro ridership against PCA.

Our research has three primary conclusions. First, the BSCA model has better model performance than the PCA model. This confirms that the reliance on PCA fails to capture the effective catchment areas expanded by micromobility, whereas BSCA successfully bridges the gap between static planning boundaries and dynamic multimodal accessibility.

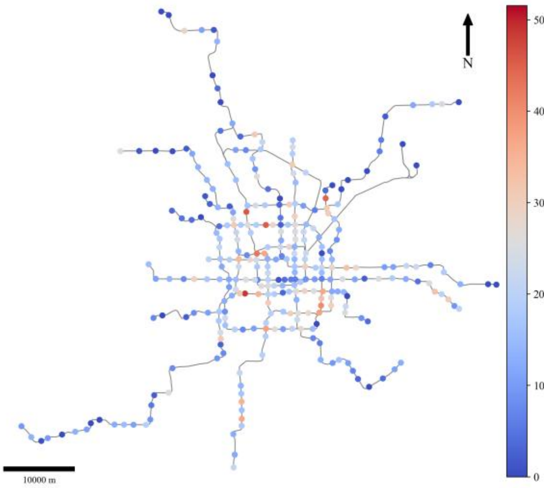
Second, the effects of the built environment on metro ridership exhibit distinct threshold effects and ranking shifts under the BSCA framework. We found that the thresholds for key density indicators, specifically population density, are notably higher within the BSCA compared to the PCA. Furthermore, the number of available shared bicycles emerged as a critical predictor across all models, which is often underestimated in traditional studies. This suggests that ignoring the bike-sharing supply leads to a biased understanding of metro ridership drivers.

Third, the nonlinear relationship between the built environment and ridership has significant spatial heterogeneity. Our results demonstrate that the employment and residential-oriented stations are more sensitive to catchment areas. Consequently, the mechanisms influencing travel behavior vary by station function, necessitating differentiated planning strategies to challenge the “one-size-fits-all” planning assumption.

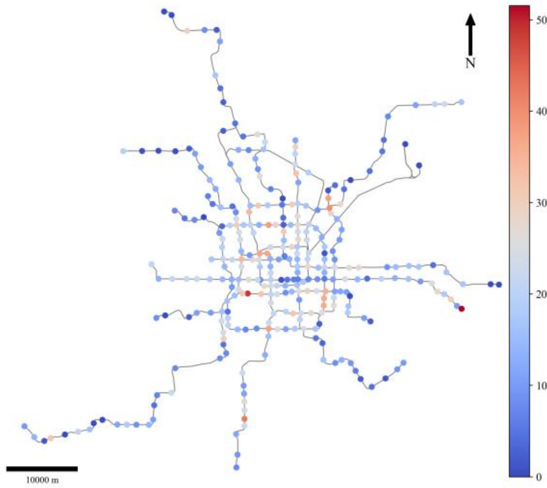
The findings advocate for a paradigm shift in TOD planning from PCA-based to BSCA-based approaches. Policymakers should calibrate planning targets and infrastructure allocation based on the expanded scope of BSCA, particularly for residential-oriented and employment-oriented stations. Achieving this requires strengthened inter-agency coordination to align land use zoning with transport resource allocation.

Future research should validate these findings across diverse urban contexts beyond Beijing and explore temporal heterogeneity using longitudinal data. Additionally, incorporating other feeder modes such as e-scooters and ride-sourcing will further refine the understanding of multimodal catchment areas, ultimately contributing to more resilient and sustainable urban mobility systems.

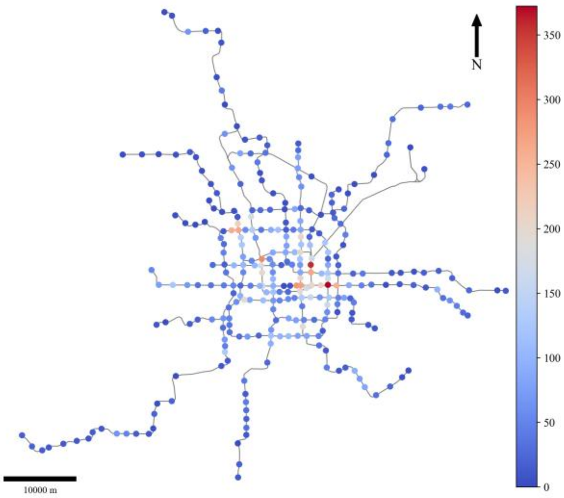
Appendix A



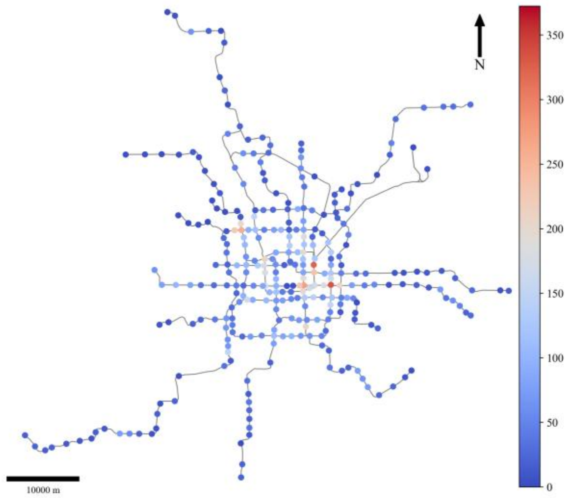
(a) Population density (PCA)



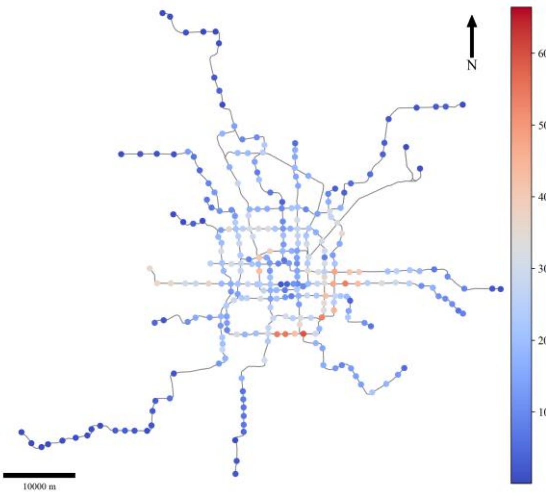
(b) Population density (BSCA)



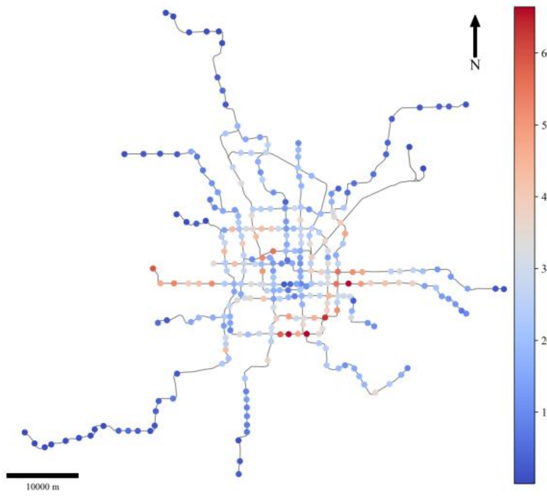
(c) Employment density (PCA)



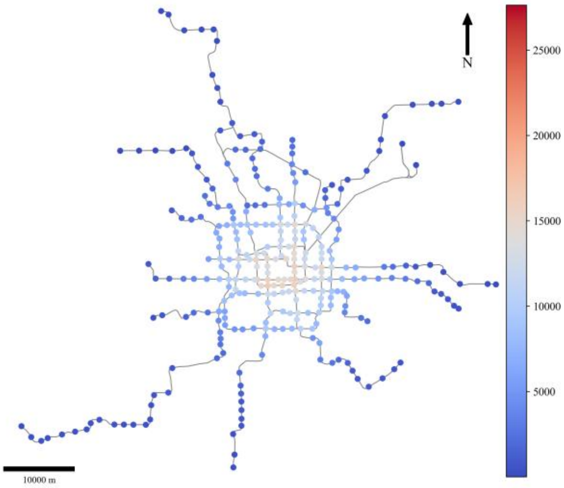
(d) Employment density (BSCA)



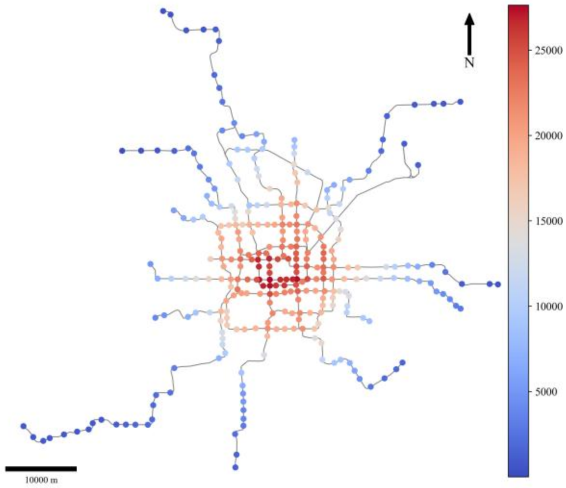
(e) Number of available shared bicycles (PCA)



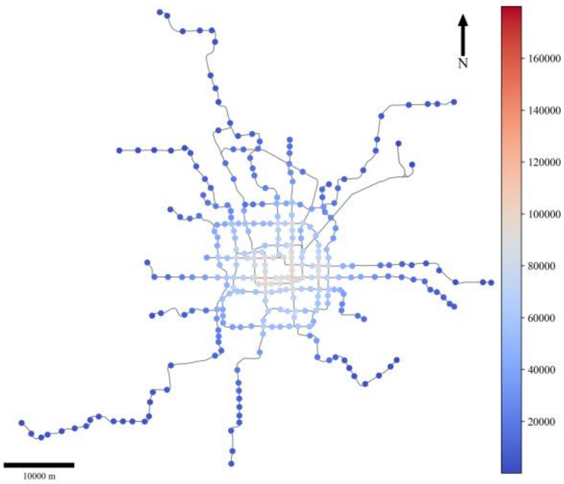
(f) Number of available shared bicycles (BSCA)



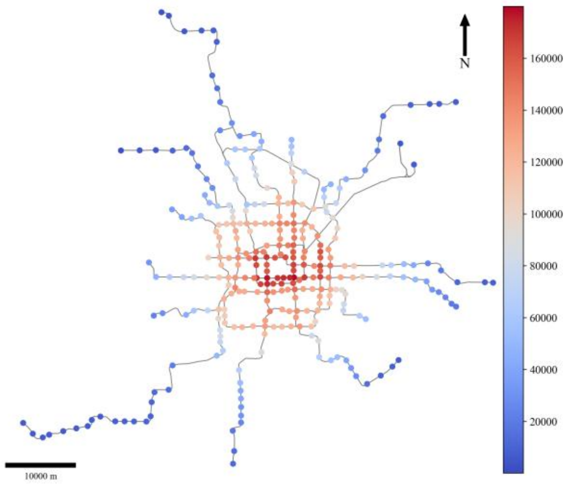
(g) Job accessibility (PCA)



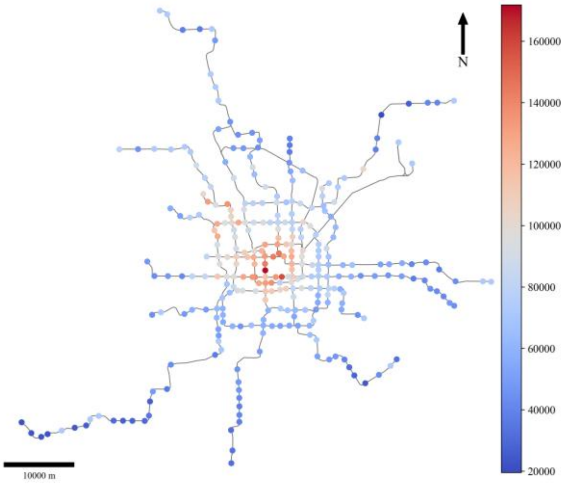
(h) Job accessibility (BSCA)



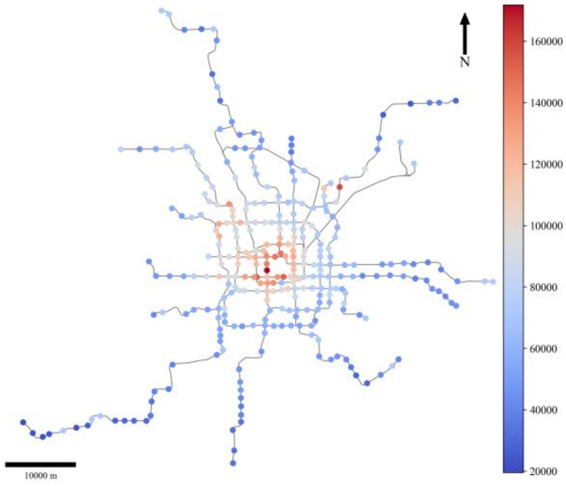
(i) Urban service accessibility (PCA)



(j) Urban service accessibility (BSCA)



(k) Housing price (PCA)



(l) Housing price (BSCA)

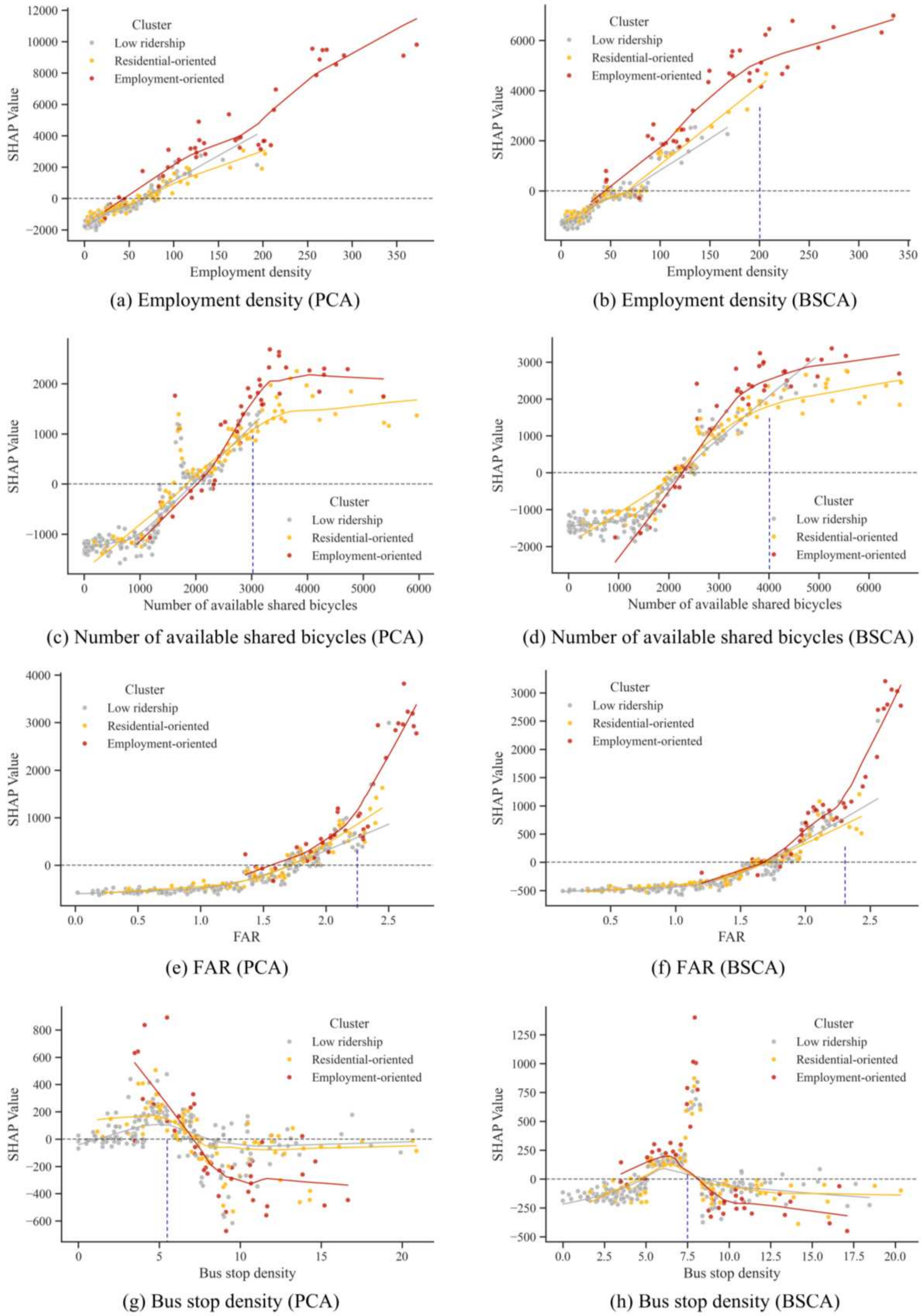
Fig. A.1. The spatial distribution of several key built environment factors

Appendix B

Table B.1. Performance of the RF and GBDT models

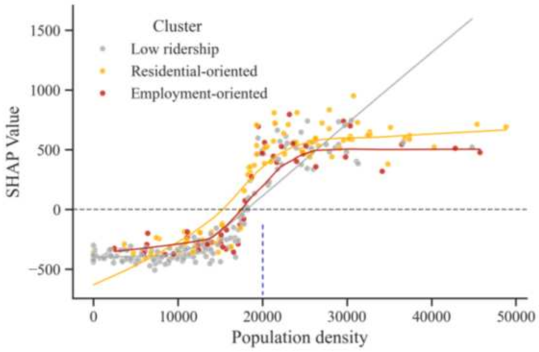
Model		RF			GBDT		
		Test R^2	Test MAE	Test RMSE	Test R^2	Test MAE	Test RMSE
Boarding ridership during the morning peak	PCA	0.5719 ± 0.0034	2232.11 ± 17.39	3179.51 ± 12.63	0.6303 ± 0.0066	1624.32 ± 28.56	2230.91 ± 19.89
	BSCA	0.5728 ± 0.0043	2275.07 ± 16.99	3176.16 ± 15.87	0.6558 ± 0.0066	1665.91 ± 31.09	2152.49 ± 20.58
Boarding ridership during the evening peak	PCA	0.7567 ± 0.0038	1578.72 ± 16.77	2358.07 ± 18.19	0.7771 ± 0.0073	1521.95 ± 31.30	2256.72 ± 36.97
	BSCA	0.7582 ± 0.0021	1551.03 ± 16.05	2351.03 ± 9.97	0.8420 ± 0.0078	1318.38 ± 44.61	1899.56 ± 46.57
Alighting ridership during the morning peak	PCA	0.7837 ± 0.0023	1776.99 ± 23.59	2555.35 ± 13.41	0.8424 ± 0.0040	1621.11 ± 24.93	2181.02 ± 27.51
	BSCA	0.7677 ± 0.0025	1841.98 ± 11.97	2648.53 ± 14.33	0.8449 ± 0.0055	1586.25 ± 46.22	2163.31 ± 38.59
Alighting ridership during the evening peak	PCA	0.5508 ± 0.0039	1280.94 ± 8.94	1562.82 ± 6.76	0.6138 ± 0.0113	1140.86 ± 25.96	1448.98 ± 21.14
	BSCA	0.5833 ± 0.0080	1224.14 ± 14.39	1505.15 ± 14.63	0.6721 ± 0.0056	1086.35 ± 15.63	1335.24 ± 11.47

Appendix C

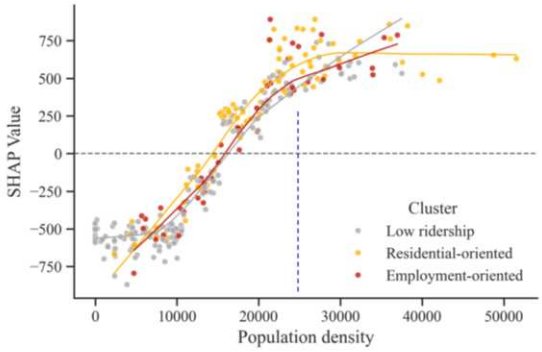


(a) Employment density (PCA)
(b) Employment density (BSCA)
(c) Number of available shared bicycles (PCA)
(d) Number of available shared bicycles (BSCA)
(e) FAR (PCA)
(f) FAR (BSCA)
(g) Bus stop density (PCA)
(h) Bus stop density (BSCA)

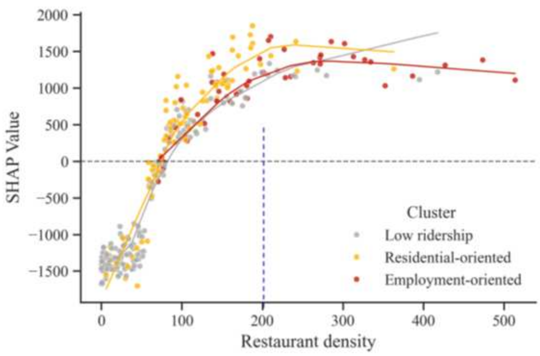
Fig. C.1. Clustering-based SHAP dependence plots for alighting ridership during the morning peak under PCA and BSCA



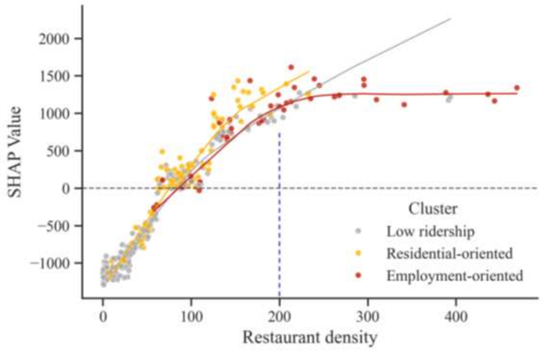
(a) Population density (PCA)



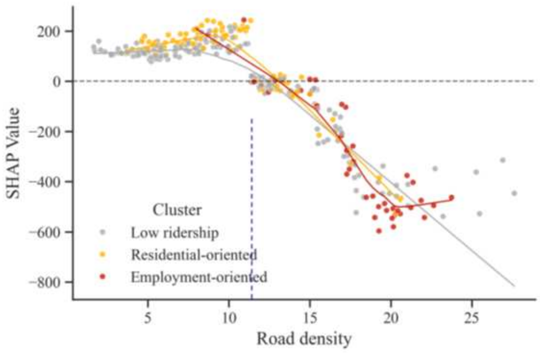
(b) Population density (BSCA)



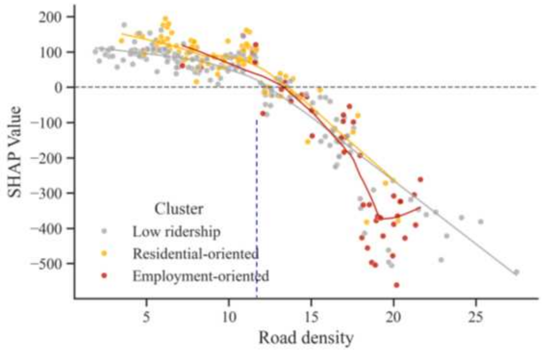
(c) Restaurant density (PCA)



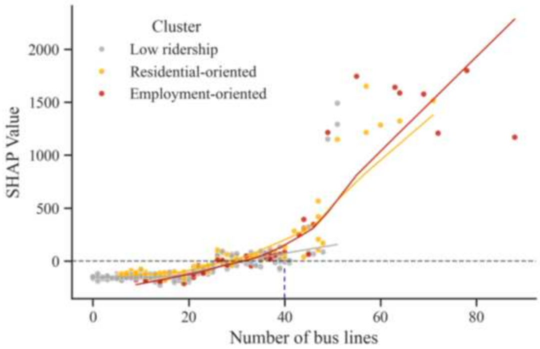
(d) Restaurant density (BSCA)



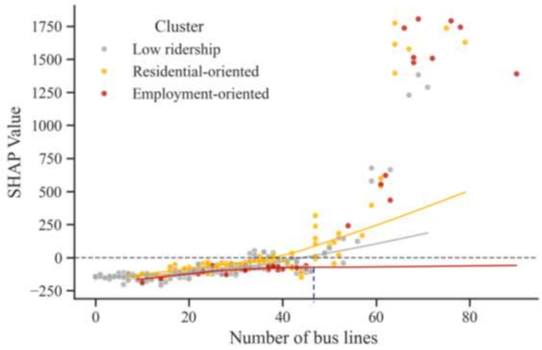
(e) Road density (PCA)



(f) Road density (BSCA)



(g) Number of bus lines (PCA)



(h) Number of bus lines (BSCA)

Fig. C.2. Clustering-based SHAP dependence plots for alighting ridership during the evening peak under PCA and BSCA

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

Andersson, D.E., Shyr, O.F., Yang, J., 2021. Neighbourhood effects on station-level transit use: Evidence from the Taipei metro. Journal of Transport Geography 94, 103127. https://doi.org/10.1016/j.jtrangeo.2021.103127

Beijing Transport Institute, 2019. 2019 Beijing transport annual report. [online] Available at: https://www.bjtrc.org.cn/List/index/cid/7/p/1.html [Accessed 30 Apr. 2025].

Beijing Transport Institute, 2018. 2018 Beijing transport annual report. [online] Available at: https://www.bjtrc.org.cn/List/index/cid/7/p/1.html [Accessed 30 Apr. 2025].

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

Caigang, Z., Shaoying, L., Zhangzhi, T., Feng, G., Zhifeng, W., 2022. Nonlinear and threshold effects of traffic condition and built environment on dockless bike sharing at street level. Journal of Transport Geography 102, 103375. https://doi.org/10.1016/j.jtrangeo.2022.103375

Cervero, R., 2006. Alternative approaches to modeling the travel-demand impacts of smart growth. Journal of the American Planning Association 72, 285–295. https://doi.org/10.1080/01944360608976751

Cervero, R., Kockelman, K., 1997. Travel demand and the 3Ds: Density, diversity, and design. Transportation Research Part D: Transport and Environment 2, 199–219. https://doi.org/10.1016/S1361-9209(97)00009-6

Che, Y., Li, X., Liu, X., Wang, Y., Liao, W., Zheng, X., Zhang, X., Xu, X., Shi, Q., Zhu, J., Zhang, H., Yuan, H., Dai, Y., 2024. 3D-GloBFP: the first global three-dimensional building footprint dataset. Earth System Science Data 16, 5357–5374. https://doi.org/10.5194/essd-16-5357-2024

Chen, E., Ye, Z., 2021. Identifying the nonlinear relationship between free-floating bike sharing usage and built environment. Journal of Cleaner Production 280, 124281. https://doi.org/10.1016/j.jclepro.2020.124281

Chen, E., Ye, Z., Wang, C., Zhang, W., 2019. Discovering the spatio-temporal impacts of built environment on metro ridership using smart card data. Cities 95, 102359. https://doi.org/10.1016/j.cities.2019.05.028

Chen, M., Wang, T., Liu, Z., Li, Y., Tu, M., 2024. Nonlinear and threshold effects of the built environment on dockless bike-sharing. Sustainability 16, 7690. https://doi.org/10.3390/su16177690

Chen, L., Lu, Y., Liu, Yanfang, Yang, L., Peng, M., Liu, Yaolin, 2022a. Association between built environment characteristics and metro usage at station level with a big data approach. Travel Behaviour and Society 28, 38–49. https://doi.org/10.1016/j.tbs.2022.02.007

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Presented at the KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Chen, W., Liu, X., Chen, X., Cheng, L., Wang, K., Chen, J., 2022b. Exploring year-to-year changes in station-based bike sharing commuter behaviors with smart card data. Travel Behaviour and Society 28, 75–89. https://doi.org/10.1016/j.tbs.2022.02.005

Chen, Y., Xu, Q., Jia, S., Wei, R., Sun, S. and Li, W., 2025. Job accessibility of urban rail transit considering access and egress travel. Journal of Zhejiang University (Engineering Science) 59(2), 310–318. https:/

/doi.org/10.3785/j.issn.1008-973X.2025.02.009
Cheng, L., Wang, K., De Vos, J., Huang, J., Witlox, F., 2022. Exploring non-linear built environment effects on the integration of free-floating bike-share and urban rail transport: A quantile regression approach. Transportation Research Part A: Policy and Practice 162, 175–187. https://doi.org/10.1016/j.tra.2022.05.022
China Association of Metros, 2025. Urban rail transit statistics 2024. [online] Available at: https://www.camet.org.cn/xytj/tjxx/660844283682885.shtml [Accessed 30 Apr. 2025].
China Academy of Urban Planning and Design, 2020. China principal cities commuting monitoring report. [online] Available at: https://huiyan.baidu.com/cms/report/2020tongqin/ [Accessed 30 Apr. 2025].
China Academy of Urban Planning and Design, 2021. 2021 China principal cities sharing bikes and sharing electric bikes riding report. [online] Available at: https://www.100ec.cn/detail--6603409.html [Accessed 30 Apr. 2025].
China Academy of Urban Planning and Design, 2023. 2023 annual report on sharing bikes and sharing electric bikes riding in major Chinese cities. [online] Available at: https://hulianhutongshequ.cn/upload/tank/report/2023/202310/4/47ba190b550b4f5dbe63f2d41be346a7.pdf [Accessed 30 Apr. 2025].
Choi, J., Lee, Y.J., Kim, T., Sohn, K., 2012. An analysis of metro ridership at the station-to-station level in Seoul. Transportation 39, 705–722. https://doi.org/10.1007/s11116-011-9368-3
Dai, W., Chen, X., Ning, J., Chen, W., 2025. Comparisons of built environment effects on the integration of dockless bike-sharing and metro transit across commuting scenarios: A user ID-based clustering identification method. Travel Behaviour and Society 40, 101047. https://doi.org/10.1016/j.tbs.2025.101047
Deng, X., Zhang, J., Liao, S., Zhong, C., Gao, F., Teng, L., 2022. Interactive impacts of built environment factors on metro ridership using Geodetector: From the perspective of TOD. ISPRS International Journal of Geo-Information 11, 623. https://doi.org/10.3390/ijgi11120623
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
Ding, C., Liu, T., Yao, B., Zhang, Y., Qiao, X., 2023. Understanding the time-dependent effect of built environment attributes on station-level metro ridership uncertainty in Beijing: A big data analytic approach. Tunnelling and Underground Space Technology 137, 105148. https://doi.org/10.1016/j.tust.2023.105148
Ding, F., Tang, Y., Wang, Y., Zhao, Z., 2025. Unraveling the network effects in station ridership growth patterns under metro network expansion. Journal of Transport Geography 125, 104205. https://doi.org/10.1016/j.jtrangeo.2025.104205
Du, Q., Huang, Y., Zhou, Y., Guo, X., Bai, L., 2023. Impacts of a new urban rail transit line and its interactions with land use on the ridership of existing stations. Cities 141, 104506. https://doi.org/10.1016/j.cities.2023.104506
Du, Q., Zhou, Y., Huang, Y., Wang, Y., Bai, L., 2022. Spatiotemporal exploration of the non-linear impacts of accessibility on metro ridership. Journal of Transport Geography 102, 103380. https://doi.org/10.1016/j.jtrangeo.2022.103380
Durning, M., Townsend, C., 2015. Direct ridership model of rail rapid transit systems in Canada. Transportation Research Record 2537, 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, 265–294. https://doi.org/10.1080/01944361003766766
Fishman, E., 2016. Bikeshare: A Review of Recent Literature. Transport Reviews 36, 92–113. https://doi.or

g/10.1080/01441647.2015.1033036
Fu, C., Huang, Z., Scheuer, B., Lin, J., Zhang, Y., 2023. Integration of dockless bike-sharing and metro: Prediction and explanation at origin-destination level. Sustainable Cities and Society 99, 104906. https://doi.org/10.1016/j.scs.2023.104906
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
Gao, F., Han, C., Yang, L., Liang, J., He, X., Li, F., 2024. Analyzing spatiotemporal distribution patterns of metro ridership: Comparison between common-class and business-class carriage service. Journal of Transport Geography 115, 103798. https://doi.org/10.1016/j.jtrangeo.2024.103798
Gao, F., Yang, L., Han, C., Tang, J., Li, Z., 2022. A network-distance-based geographically weighted regression model to examine spatiotemporal effects of station-level built environments on metro ridership. Journal of Transport Geography 105, 103472. https://doi.org/10.1016/j.jtrangeo.2022.103472
Geurs, K.T., van Wee, B., 2004. Accessibility evaluation of land-use and transport strategies: review and research directions. Journal of Transport Geography 12, 127–140. https://doi.org/10.1016/j.jtrangeo.2003.10.005
Gu, Y., Dou, M., 2024. Nonlinear and threshold effects on station-level ridership: Insights from disproportionate weekday-to-weekend impacts. ISPRS International Journal of Geo-Information 13, 365. https://doi.org/10.3390/ijgi13100365
Guo, Y., Yang, L., Lu, Y., Zhao, R., 2021. Dockless bike-sharing as a feeder mode of metro commute? The role of the feeder-related built environment: Analytical framework and empirical evidence. Sustainable Cities and Society 65, 102594. https://doi.org/10.1016/j.scs.2020.102594
Huang, J., Chen, S., Xu, Q., Chen, Y., Hu, J., 2022. Relationship between built environment characteristics of TOD and subway ridership: A causal inference and regression analysis of the Beijing subway. Journal of Rail Transport Planning & Management 24, 100341. https://doi.org/10.1016/j.jrtpm.2022.100341
Huang, X., Cao, X. (Jason), Cao, X., Yin, J., 2016. How does the propensity of living near rail transit moderate the influence of rail transit on transit trip frequency in Xi'an? Journal of Transport Geography 54, 194–204. https://doi.org/10.1016/j.jtrangeo.2016.05.016
Huang, Y., Zhang, Z., Xu, Q., Dai, S., Chen, Y., 2024. Causality between multi-scale built environment and rail transit ridership in Beijing and Tokyo. Transportation Research Part D: Transport and Environment 130, 104150. https://doi.org/10.1016/j.trd.2024.104150
Jamme, H.-T., Rodriguez, J., Bahl, D., Banerjee, T., 2019. A twenty-five-year biography of the TOD concept: From design to policy, planning, and implementation. Journal of Planning Education and Research 39, 409–428. https://doi.org/10.1177/0739456X19882073
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
Kuby, M., Barranda, A., Upchurch, C., 2004. Factors influencing light-rail station boardings in the United States. Transportation Research Part A: Policy and Practice 38, 223–247. https://doi.org/10.1016/j.tra.2003.10.006
Lee, S., Yi, C., Hong, S.-P., 2013. Urban structural hierarchy and the relationship between the ridership of the Seoul Metropolitan Subway and the land-use pattern of the station areas. Cities 35, 69–77. https://doi.org/10.1016/j.cities.2013.06.010
Li, L., Zhong, L., Ran, B., Du, B., 2024. Analysis of the relationship between metro ridership and built envi

ronment: A machine learning method considering combinational features. Tunnelling and Underground Space Technology 144, 105564. https://doi.org/10.1016/j.tust.2023.105564
Li, M., Kwan, M.-P., Hu, W., Li, R., Wang, J., 2023a. Examining the effects of station-level factors on metr o ridership using multiscale geographically weighted regression. Journal of Transport Geography 113, 103720. https://doi.org/10.1016/j.jtrangeo.2023.103720
Li, S., Lyu, D., Huang, G., Zhang, X., Gao, F., Chen, Y., Liu, X., 2020. Spatially varying impacts of built en vironment factors on rail transit ridership at station level: A case study in Guangzhou, China. Journal o f Transport Geography 82, 102631. https://doi.org/10.1016/j.jtrangeo.2019.102631
Li, X., Xing, G., Qian, X., Guo, Y., Wang, W., Cheng, C., 2022. Subway Station Accessibility and Its Impac ts on the Spatial and Temporal Variations of Its Outbound Ridership. Journal of Transportation Enginee ring, Part A: Systems 148, 04022106. https://doi.org/10.1061/JTEPBS.0000766
Li, Y., Li, J., Yu, Z., Li, S., Li, A., 2025. Exploring the relationship between the built environment and bike-sharing usage as a feeder mode across different metro station types in Shenzhen. Land 14, 1291. https://doi.org/10.3390/land14061291
Li, Z., Gao, F., Xiao, C., Tang, J., 2023b. The effects of the urban built environment on public transport ride rship: similarities and differences. Travel Behaviour and Society 33, 100630. https://doi.org/10.1016/j.tbs.2023.100630
Lin, J.-J., Shin, T.-Y., 2008. Does Transit-oriented development affect metro ridership?: Evidence from Taip ei, Taiwan. Transportation Research Record 2063, 149–158. https://doi.org/10.3141/2063-18
Liu, B., Xu, Y., Guo, S., Yu, M., Lin, Z., Yang, H., 2023a. Examining the nonlinear impacts of origin-destin ation built environment on metro ridership at station-to-station level. ISPRS International Journal of G eo-Information 12, 59. https://doi.org/10.3390/ijgi12020059
Liu, C., Erdogan, S., Ma, T., Ducca, F.W., 2016. How to increase rail ridership in Maryland: Direct ridershi p models for policy guidance. Journal of Urban Planning and Development 142, 04016017. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000340
Liu, M., Liu, Y., Ye, Y., 2023b. Nonlinear effects of built environment features on metro ridership: An integr ated exploration with machine learning considering spatial heterogeneity. Sustainable Cities and Societ y 95, 104613. https://doi.org/10.1016/j.scs.2023.104613
Liu, S., Yao, E., Li, B., 2019. Exploring urban rail transit station-level ridership growth with network expan sion. Transportation Research Part D: Transport and Environment 73, 391–402. https://doi.org/10.1016/j.trd.2018.04.006
Liu, X., Chen, X., Gao, Y., Pan, H., Tian, M., 2024. Should transit-oriented development consider station ag e effects? Transportation Research Part D: Transport and Environment 133, 104273. https://doi.org/10.1016/j.trd.2024.104273
Liu, X., Chen, X., Tian, M., De Vos, J., 2023c. Effects of buffer size on associations between the built envir onment and metro ridership: A machine learning-based sensitive analysis. Journal of Transport Geogra phy 113, 103730. https://doi.org/10.1016/j.jtrangeo.2023.103730
Liu, X., Yu, Q., Bian, W., Yu, H., Zhang, C., Liu, Xiang, Jiang, H., Luo, X., 2025. Unraveling spatiotempora l dynamics of ridesharing potential: Nonlinear effects of the built environment. Transportation Researc h Part D: Transport and Environment 139, 104594. https://doi.org/10.1016/j.trd.2025.104594
Liu, Y., Tang, R., Shi, Z., He, M., Cheng, L., 2025. Shared mobility choices in metro connectivity: shared bi kes versus shared e-bikes. Transportation. https://doi.org/10.1007/s11116-025-10612-x
Loo, B.P.Y., Chen, C., Chan, E.T.H., 2010. Rail-based transit-oriented development: Lessons from New Yor k City and Hong Kong. Landscape and Urban Planning 97, 202–212. https://doi.org/10.1016/j.landurb

plan.2010.06.002
Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems (NeurIPS) 30. https://doi.org/10.1109/1.2020048
Lv, H., Li, H., Sze, N.N., Cheng, L., Bai, X., 2024. How do walking environments along the connecting path affect the disparity in public bicycle usage among multiple docking stations around tube stations? Travel Behaviour and Society 34, 100661. https://doi.org/10.1016/j.tbs.2023.100661
Pang, L., Jiang, Y., Wang, J., Qiu, N., Xu, X., Ren, L., Han, X., 2023. Research of metro stations with varying patterns of ridership and their relationship with built environment, on the example of Tianjin, China. Sustainability 15, 9533. https://doi.org/10.3390/su15129533
People's public transportation, 2024. Urban Rail Transit: Mileage, Passenger Flow, and Development Trends. [online] Available at: https://mp.weixin.qq.com/s/LbzxDG0Nhv7YZCss0ZmRUw [Accessed 20 August 2025].
Peng, B., Wang, T., Zhang, Y., Li, C., Lu, C., 2024. Spatially varying effect mechanism of intermodal connection on metro ridership: Evidence from a polycentric megacity with multilevel ring roads. ISPRS International Journal of Geo-Information 13, 353. https://doi.org/10.3390/ijgi13100353
Peng, B., Zhang, Y., Li, C., Wang, T., Yuan, S., 2023. Nonlinear, threshold and synergistic effects of first/last-mile facilities on metro ridership. Transportation Research Part D: Transport and Environment 121, 103856. https://doi.org/10.1016/j.trd.2023.103856
Ramos-Santiago, L.E., Brown, J., 2016. A comparative assessment of the factors associated with station-level streetcar versus light rail transit ridership in the United States. Urban Studies 53, 915–935. https://doi.org/10.1177/0042098015571057
Mirzaei, E., Heidarzadeh, E., Hashemi, A., Dowlatshahi, M.B., Kheyroddin, R., 2025. Impact analysis of positive utility and built environment on car use for non-work trips: an ensemble approach. Transportation Research Part A: Policy and Practice 199, 104589. https://doi.org/10.1016/j.tra.2025.104589
State Council of the People's Republic of China, 2018. Notice on further strengthening the planning and management of urban rail transit construction (Document No. 52). Beijing: State Council. [online] Available at: https://www.gov.cn/gongbao/content/2018/content_5309432.htm [Accessed 20 August. 2025].
Shao, Q., Zhang, W., Cao, X., Yang, J., Yin, J., 2020. Threshold and moderating effects of land use on metro ridership in Shenzhen: Implications for TOD planning. Journal of Transport Geography 89, 102878. https://doi.org/10.1016/j.jtrangeo.2020.102878
Shi, Z., Zhang, N., Liu, Y., Xu, W., 2018. Exploring Spatiotemporal Variation in Hourly Metro Ridership at Station Level: The Influence of Built Environment and Topological Structure. Sustainability 10, 4564. https://doi.org/10.3390/su10124564
Song, Y., Luo, K., Shi, Z., Zhang, L., Shen, Y., 2023. Nonlinear influence and interaction effect on the imbalance of metro-oriented dockless bike-sharing system. Sustainability 16, 349. https://doi.org/10.3390/su16010349
Su, S., Wang, Z., Li, B., Kang, M., 2022a. Deciphering the influence of TOD on metro ridership: An integrated approach of extended node-place model and interpretable machine learning with planning implications. Journal of Transport Geography 104, 103455. https://doi.org/10.1016/j.jtrangeo.2022.103455
Su, S., Zhao, C., Zhou, H., Li, B., Kang, M., 2022b. Unraveling the relative contribution of TOD structural factors to metro ridership: A novel localized modeling approach with implications on spatial planning. Journal of Transport Geography 100, 103308. https://doi.org/10.1016/j.jtrangeo.2022.103308
Sun, X., Cui, X., Gao, Y., 2025. How urban form shapes bike-metro integration: A multidimensional typology of usage patterns and accessibility inequities. Travel Behaviour and Society 41, 101080. https://doi.

org/10.1016/j.tbs.2025.101080
Sung, H., Choi, K., Lee, S., Cheon, S., 2014. Exploring the impacts of land use by service coverage and station-level accessibility on rail transit ridership. Journal of Transport Geography 36, 134–140. https://doi.org/10.1016/j.jtrangeo.2014.03.013
Van Wee, B., Handy, S., 2016. Key research themes on urban space, scale, and sustainable urban mobility. International Journal of Sustainable Transportation 10, 18–24. https://doi.org/10.1080/15568318.2013.820998
Wang, J., Dong, C., Shao, C., Bao, J., Wan, F., 2025. Built environment effects on dockless bikesharing–metro integration: A spatial nonlinear analysis. Travel Behaviour and Society 40, 101003. https://doi.org/10.1016/j.tbs.2025.101003
Wang, J., Wan, F., Dong, C., Yin, C., Chen, X., 2023a. Spatiotemporal effects of built environment factors on varying rail transit station ridership patterns. Journal of Transport Geography 109, 103597. https://doi.org/10.1016/j.jtrangeo.2023.103597
Wang, J., Zhang, N., Peng, H., Huang, Y., Zhang, Y., 2022a. Spatiotemporal heterogeneity analysis of influence factor on urban rail transit station ridership. Journal of Transportation Engineering, Part A: Systems 148, 04021115. https://doi.org/10.1061/JTEPBS.0000639
Wang, J., Zhao, M., Ai, T., Wang, Q., Liu, Y., 2023b. Revealing the influence of the fine-scale built environment on urban rail ridership with a semiparametric GWPR model. ISPRS International Journal of Geo-Information 12, 218. https://doi.org/10.3390/ijgi12060218
Wang, Z., Li, S., Li, Y., Liu, D., Liu, S., Chen, N., 2023c. Investigating the nonlinear effect of built environment factors on metro station-level ridership under optimal pedestrian catchment areas via the machine learning method. Applied Sciences 13, 12210. https://doi.org/10.3390/app132212210
Wang, Z., Song, J., Zhang, Y., Li, S., Jia, J., Song, C., 2022b. Spatial heterogeneity analysis for influencing factors of outbound ridership of subway stations considering the optimal scale range of “7D” built environments. Sustainability 14, 16314. https://doi.org/10.3390/su142316314
Wu, X., Lu, Y., Gong, Y., Kang, Y., Yang, L., Gou, Z., 2021. The impacts of the built environment on bicycle-metro transfer trips: A new method to delineate metro catchment area based on people's actual cycling space. Journal of Transport Geography 97, 103215. https://doi.org/10.1016/j.jtrangeo.2021.103215
Xi, Y., Hou, Q., Duan, Y., Lei, K., Wu, Y., Cheng, Q., 2024. Exploring the spatiotemporal effects of the built environment on the nonlinear impacts of metro ridership: Evidence from Xi'an, China. ISPRS International Journal of Geo-Information 13, 105. https://doi.org/10.3390/ijgi13030105
Yan, J., Wan, Q., Feng, J., Wang, J., Hu, Y., Yan, X., 2023. The non-linear influence of built environment on the school commuting metro ridership: The case in Wuhan, China. ISPRS International Journal of Geo-Information 12, 193. https://doi.org/10.3390/ijgi12050193
Yang, H., Lu, Y., Wang, J., Zheng, Y., Ruan, Z., Peng, J., 2023a. Understanding post-pandemic metro commuting ridership by considering the built environment: A quasi-natural experiment in Wuhan, China. Sustainable Cities and Society 96, 104626. https://doi.org/10.1016/j.scs.2023.104626
Yang, H., Peng, J., Lu, Y., Wang, J., Yan, X., 2024a. Nonlinear impact of built environment on people with disabilities’ metro use behavior. Applied Geography 169, 103323. https://doi.org/10.1016/j.apgeog.2024.103323
Yang, H., Zhang, Q., Wen, J., Sun, X., Yang, L., 2024b. Multi-group exploration of the built environment and metro ridership: Comparison of commuters, seniors and students. Transport Policy 155, 189–207. https://doi.org/10.1016/j.tranpol.2024.06.020
Yang, L., Yu, B., Liang, Y., Lu, Y., Li, W., 2023b. Time-varying and non-linear associations between metro r

idership and the built environment. Tunnelling and Underground Space Technology 132, 104931. https://doi.org/10.1016/j.tust.2022.104931

Zhan, Z., Guo, Y., Noland, R.B., He, S.Y., Wang, Y., 2023. Analysis of links between dockless bikeshare and metro trips in Beijing. Transportation Research Part A: Policy and Practice 175, 103784. https://doi.org/10.1016/j.tra.2023.103784

Zhang, Y., Meng, Y., Chen, X.-J., Liu, H., Gong, Y., 2025. The nonlinear effect of the built environment on bike–metro transfer in different times and transfer flows considering spatial dependence. Sustainability 17, 251. https://doi.org/10.3390/su17010251

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, 133–155. https://doi.org/10.1007/s11116-013-9492-3

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

Zhou, X., Yang, C., Yang, L., Zhang, Y., 2025. Nonlinear effects of the built environment on metro-integrated bikesharing and ridesourcing usage. Transportation Research Part D: Transport and Environment 146, 104898. https://doi.org/10.2139/ssrn.5100393

Zhu, B., Hu, S., Kaparias, I., Zhou, W., Ochieng, W., Lee, D.-H., 2024a. Revealing the driving factors and mobility patterns of bike-sharing commuting demands for integrated public transport systems. Sustainable Cities and Society 104, 105323. https://doi.org/10.1016/j.scs.2024.105323

Zhu, H., Peng, J., Dai, Q., Yang, H., 2024b. Exploring the long-term threshold effects of density and diversity on metro ridership. Transportation Research Part D: Transport and Environment 128, 104101. https://doi.org/10.1016/j.trd.2024.104101

Zhu, Y., Chen, F., Wang, Z., Deng, J., 2019. Spatio-temporal analysis of rail station ridership determinants in the built environment. Transportation 46, 2269–2289. https://doi.org/10.1007/s11116-018-9928-x

Zuo, T., Wei, H., Chen, N., Zhang, C., 2020. First-and-last mile solution via bicycling to improving transit accessibility and advancing transportation equity. Cities 99, 102614. https://doi.org/10.1016/j.cities.2020.102614