

Effects of the built environment on travel distance in bus-oriented, medium-sized cities in China

Xiaowei Li

School of Civil Engineering, Xi'an University of Architecture & Technology
lixiaowei@xauat.edu.cn

Lanxin Shi

School of Civil Engineering, Xi'an University of Architecture & Technology
lancyshek@gmail.com

Junqing Tang (corresponding author)

School of Urban Planning and Design, Shenzhen Graduate School, Peking University
junqingtang@pku.edu.cn

Jiaying Li

School of Urban Planning and Design, Shenzhen Graduate School, Peking University
joyinglee@stu.pku.edu.cn

Pengjun Zhao

School of Urban Planning and Design, Shenzhen Graduate School, Peking University
pengjun.zhao@pku.edu.cn

Qian Liu

School of Civil Engineering, Xi'an University of Architecture & Technology
2372924846@qq.com

Jun Chen

School of Civil Engineering, Xi'an University of Architecture & Technology
chenjuntom@126.com

Changxi Ma

School of Traffic and Transportation, Lanzhou Jiaotong University
machangxi@mail.lzjtu.cn

Abstract: The impact of the built environment and weather conditions on travel behavior has been widely studied. However, limited studies have focused on better understanding such effects in medium-sized cities with bus-oriented transit systems, particularly from a separate perspective of travelers' origins and destinations. We took Weinan, China, as a representative of second-tier cities in developing countries that concentrate on bus-oriented development strategies. New evidence of feature importance and nonlinear effects of crucial factors were revealed by an interpretable machine learning-based approach combining XGBoost and Shapley Additive Explanation (SHAP) with multi-source data. Most key factors were critical at both origins and destinations, such as the density of residential and commercial facilities. However, several important factors, such as road density and boarding time, had strong imbalanced effects on travel behavior. These findings provide novel insights and empirical implications to support urban planning strategies in medium-sized cities.

Keywords: Travel distance, bus transit systems, built environment, medium-sized cities, machine learning, big data

Article history:

Received: September 5, 2023
Received in revised form:
March 18, 2024
Accepted: March 29, 2024
Available online: May 21, 2024

1 Introduction

Urban transportation systems play a pivotal role in shaping the sustainability and livability of cities. In recent years, medium-sized cities have emerged as crucial focal points for urban development and transportation planning. Unlike large metropolitan areas, which often possess extensive metro transit networks, medium-sized cities typically lack such infrastructure and heavily rely on bus-oriented development strategies. Concurrently, road transport has become the leading contributor to CO₂ emissions and anthropogenic air pollutants, with transport-related CO₂ emissions expected to increase by 57% globally between 2005 and 2030 far outstripping emissions from other sectors (Bongardt et al., 2010). The promotion of public transit has served as an effective strategy in contemporary society for achieving global climate goals (Agreement, 2015) and facilitating low-carbon green growth. However, the feasibility and sustainability of this strategy in medium-sized cities that adopt bus-based systems have not been thoroughly examined, including knowledge of the extent to which the external environment impacts travel behavior of bus riders, as well as the potential imbalanced effects at the locations of origin and destination. Therefore, understanding the built environment factors that influence travel behavior in these cities is of paramount importance for developing effective and sustainable planning and transportation policies.

Promoting public transportation is crucial for mitigating traffic congestion, reducing greenhouse gas emissions, and improving overall urban mobility. Particularly, in medium-sized cities, where private car ownership may be lower, bus-oriented transit systems provide a cost-effective and adaptable solution, addressing the diverse transportation needs of growing populations. Bus-oriented transit systems serve as the most extensively employed form of public transportation and offer numerous advantages to social welfare. On a regional scale, buses are the least energy- and GHG-intensive modes compared with railways and aircraft (Schäfer & Yeh, 2020). On a community scale, developing bus transit can optimize citizens' mobility patterns and alleviate congestion (Allen et al., 2019; Nguyen-Phuoc et al., 2018; Rong et al., 2022). Additional benefits include boosting profits and reducing operating costs (Matzler et al., 2003), which contribute significantly to social equity (Tuan et al., 2022). In light of the essential role of the bus transit system, it is necessary to enhance bus route rationalization and service performance to encourage more residents to opt for buses as their preferred mode of transportation. Thus, identifying key factors and analyzing the potential influences of the built environment on bus riders' travel behavior can support local departments in proposing guidelines for transport mode shifting and developing optimal strategies for dynamic bus allocation in the long term.

One critical determinant commonly used for comprehending travel patterns and mobility flow is "travel distance," which pertains to the spatial separation between a pair of origin-destination (OD) locations. This factor not only serves as a revenue source for public transport operators but also offers insights into travel demand and travel behavior (Park et al., 2019). In the realm of mobility research, travel distance yields valuable information about what key factors conditioning riders' decision-making processes when choosing whether to use a bus service and determining where to start or end their trips. Moreover, an in-depth comprehension of public-transport-travel behavioral patterns is required to uncover how travel distances vary across diverse locations, which may indicate imbalanced effects resulting from the variation in the built environment.

The built environment, comprising various man-made surroundings created for human activities, plays a crucial role in shaping travel behavior. For medium-sized cities experiencing rapid urban sprawl, uneven regional planning and inadequate resources can lead to heterogeneity in the distribution of population hotspots and disparities in the

accessibility at origins and destinations. Consequently, the spatial distribution of land uses, transportation infrastructure, and urban form could disproportionately influence travel distances and travel mode choices (Ewing & Cervero, 2010; Handy et al., 2002). These imbalanced effects on travel behavior can create variation in accessibility, service quality, and ridership, potentially influencing people's willingness to utilize public transportation. Recognizing the importance of this knowledge can facilitate the optimization of bus operating networks (Baumgarte et al., 2021) and the design of a more efficient and equitable transportation system that can attract more potential citizens adopt bus services. Yet, the unbalanced impacts of the built environment on travel behavior at origins and destinations have not been thoroughly investigated using quantitative methods and multi-source data. This study takes an approach by analyzing travel distances from the lens of splitting OD pairs, allowing an exploration of the potential imbalanced effects the built environment may exert on individuals' choices regarding boarding and alighting.

This study aims to fill in gaps and explores the impacts of the built environment at OD locations, together with weather conditions, on the travel distances of bus passengers. Weinan City (China)—a medium-sized city without a metro transit system that has implemented distinctive bus-oriented development strategies—was taken as a typical representative example. We collected multi-source data for the research, encompassing smart card transactions, bus GPS information, bus station details, road infrastructure, points-of-interest (POI) and weather data. The time series under investigation spans two consecutive weeks from 12 to 30 November 2018. The novel interpretable machine learning-based approach that combines XGBoost, and Shapley Additive Explanation (SHAP) in conjunction with spatial-temporal big data, was employed to investigate the importance of various impact factors, disentangle intricate nonlinear influences, and analyze heterogeneous interactions between the built environment and weather conditions on travel distances. The main contributions of this study can be summarized as follows.

(1) We consider the combined effects of the built environment and dynamic weather conditions on bus passengers' travel distances. More importantly, we differentiate the built environment, represented by POI data, at both origin and destination locations to explore the relationships between the spatial density of various POI types and travel distance. This study contributes novel insights into the potentially imbalanced influences of the built environment at origin and destination locations in different urban areas.

(2) This paper applies an interpretable machine learning (IML) method that combines XGBoost and SHAP to disentangle the nonlinear influences and heterogeneous interactions of the built environment and weather conditions on travel distances.

(3) Taking a bus-oriented medium-sized city in China as an illustrative case study, we present new empirical evidence to enhance our understanding of bus transit system development and urban planning for underdeveloped areas. Furthermore, our study provides inspiration for similar medium-sized cities in other regions or countries to improve the competitiveness of sustainable transport modes.

2 Literature review

Existing studies on passengers' travel distance have widely investigated the intrinsic influencing characteristics of travel behavior, such as individual attributes (Dédélé et al., 2020; Ko et al., 2019; Mishra et al., 2017), and travel habits (Baumgarte et al., 2021). For example, with respect to individual attributes, research has shown that factors such as gender, age, and socio-economic status are important determinants of the frequency and distance of daily travel (Reichert et al., 2016). Abbasi et al. (2022) found that people in

their 50s travel longer distances, and women travel 1.4 kilometers more than men on average. Similarly, passengers with higher incomes and education levels travel longer distances more frequently than those with lower education levels (Holz-Rau et al., 2013). Travel habits constitute other intriguing aspects in understanding travel behavior, such as travel purposes, travel frequency, and preferences for specific modes of transportation. For instance, Baumgarte et al. (2021) found that the majority of passengers utilized shared cars at night as long-distance travelers. While these studies have provided valuable insights into individual-level factors affecting travel distances, they leave a notable gap in comprehending the role of urban built environment factors in shaping travel behavior. This is particularly relevant in the context of medium-sized cities undergoing rapid growth, characterized by the swift development and diversification of infrastructure.

The exploration of the potential to influence travel demands through modifying the built environment has received considerable attention in urban planning research. Researchers have recognized the built environment as a pivotal element that shapes peoples' travel behavior (Böcker et al., 2013; Ding et al., 2019; Yang et al., 2022; Yu et al., 2019). The most prevalent characterizations of the built environment are often associated with the "6Ds" indicators: Density, Diversity, Design, Destination accessibility, Distance to transit, and Demand management (Ao et al., 2018; Cervero & Kockelman, 1997; Ogra & Ndebele, 2014). In particular, Density, Diversity, and Design have been underscored in previous studies as some of the most significant influencers of travel behavior, especially in the context of medium-sized cities (Cao et al., 2009; Hu et al., 2018). These three dimensions, among others, are integral to urban vitality, encompassing aspects such as land use patterns, building density, traffic infrastructure, and road design (Jiang et al., 2022; Montgomery, 1998), which, in turn, impact the quality, quantity, and price of travel for human activities (Chatman, 2005).

A comprehensive review found that travel distance is primarily a function of the built environment (Ewing & Cervero, 2010). When residential, employment, entertainment and living service facilities within a region are closely situated, residents' travel distances can be greatly reduced (Yue et al., 2017). The integration of diverse land uses, including residential, employment, and recreational areas, contributes to the reduction in travel distances and fosters the use of public transport (Cervero & Duncan, 2006; Frank et al., 2008). Leck (2011) found that residential density, employment density, and the degree of land-use mix are inversely related to vehicle miles traveled (VMT) through an evaluation of the statistical significance of relationships between travel distances and the built environment. Jain and Tiwari (2019) highlighted the importance of diversity in origins, positing that a dense and diverse area is likely to result in reduced travel distances. Recently, the causal relationships between travel behavior outcomes and the built environment factors have been subjected to considerable interest. However, qualitatively determining how changes in specific built environment factors lead to specific interventions, treatments, or actions influences remains challenging. In this study, we utilize POI data to characterize the built environment. POI data is a type of emerging large-scale big data (Wang et al., 2022), with each data point typically referring to a type of element of physical location entities on electronic maps associated with the urban economy and residents' lives (Yan et al., 2021). The density and diversity of the built environment can be determined by measuring the number and degree of mixing of various POI types.

With recognition that interactions of the built environment and weather conditions could potentially influence various environmental outcomes, researchers have yet to sufficiently examine their joint impacts on travel distances. Weather has been extensively discussed as a key determinant of individual travel choices (Bi et al., 2022; Böcker et al.,

2013; Li et al., 2021; Liu et al., 2017). Adverse weather conditions, such as rainfall and humidity, can negatively impact traffic volume, transit bus performance, and even severely disrupt transit lines (Kamga & Yazıcı, 2014; Ma et al., 2019; Yin et al., 2016). Notably, bus riders are more vulnerable to outdoor environmental exposure and disruptions during transit compared to car drivers and metro users (Tang et al., 2020). In general, citizens' travel behavior is adaptive to dynamic weather conditions in real-time. Concurrently, extreme weather events could also pose considerable risks to the built environment (Hallegatte et al., 2013). It is intriguing to determine the joint effects and ascertain if one factor exerts greater dominance than the other. Such granular information can support bus assignments, route optimization and long-term planning strategies.

The nonlinear relationships between the density of the built environment and the intensity of weather conditions on human travel activities are widely acknowledged (Böcker et al., 2013; Yang et al., 2022; Yu et al., 2019). Traditional linear regression methods, which rely on linear or pre-determined nonlinear hypotheses, often fail to capture, or inaccurately estimate these intricate associations. Even when using state-of-the-art machine learning (ML) techniques, many models struggle to identify threshold effects, establish causality, and investigate interactions (Koushik et al., 2020). Utilizing the data-driven mechanisms of machine learning algorithms and corresponding interpretation techniques enables the analysis of feature importance, interaction effects, and partial dependence graphs of variables (Gao et al., 2021; Tu et al., 2021; Wei, 2022; Yang et al., 2021). Interpretable machine learning (IML) methods offer increased confidence in model outcomes, delivering higher prediction accuracy and superior performance in elucidating complex relationships. In this study, IML methods were employed to gain valuable insights into the associations between various impact factors and travel distance, allowing for more informed decision-making as well as increased the reliability of results (Lundberg & Lee, 2017; Ribeiro et al., 2016). XGBoost model developed by Chen and Guestrin (2016) is based on the Gradient Lifting Decision Tree (GBDT) algorithm proposed by Friedman (2001). Key contributions of using XGBoost and SHAP in this research include the examination of feature importance, heterogeneous interaction effects, and nonlinear effects. Previous studies have demonstrated the distinct capabilities of XGBoost in addressing nonlinear relationship questions for spatial-temporal travel behavior analyses. For instance, Ji et al. (2022) examined the nonlinear relationship and interaction effects between the built environment and cycling distance using XGBoost and SHAP, while Wu et al. (2022) employed XGBoost to analyze and predict the relationship between flight delays and meteorological conditions.

Existing literature has primarily focused on large cities and failed to quantitatively reveal the intricate relationships between built environment factors and travel behavior, particularly, in the context of second-tier medium-sized cities (Ding et al., 2018, 2019; Durning & Townsend, 2015; Tao et al., 2020; Yu et al., 2019). In contrast to larger cities, medium-sized cities typically lack metro networks and heavily rely on bus-oriented development strategies. The role of these expanding, medium-sized urban areas in this field of research has been largely overlooked in the literature. Cities vary in size, population, road network, transit infrastructure and socioeconomic development level; thus, the built environment and travel demands in medium-sized and large cities are distinctively different (Li et al., 2020; Proboste et al., 2020). Owing to the relatively smaller population, less-developed traffic infrastructure, lower car ownership and limited financial revenue, residents' mobility is constrained, and large-capacity public transport—such as the intra-city metro—is rarely available in medium-sized cities, where traditional bus transit systems play a backbone role in out-of-home trips. It is apparent that results and experiences from megacities can hardly provide appropriate guidance for

medium-sized cities. Another potential consequence is that travel behaviors in different urban areas have imbalanced effects as medium-sized cities are more unevenly developed. For example, a common feature in underdeveloped cities is the job-housing imbalance, which has been indicated as the major causative factor determining commuting time and vehicle miles traveled (Cervero, 1989; Sultana, 2002). This job-housing relationship may subsequently result in an imbalanced impact on travel demands across different areas. However, no clear findings are available concerning the following questions: What is the nature of these imbalanced impacts? To what extent are they unbalanced?

Based on the above literature review, three clear gaps can be identified as follows, and they are the key issues we address in this paper.

(1) Previous studies have mainly emphasized the effects of individual attributes, travel habits and weather conditions on travel distance, besides, only concentrating on a limited set of factors. A more comprehensive factor analysis is required. studies exploring the combined effects of the built environment and weather conditions on travel distance—especially from an imbalanced perspective that distinguishes the origins and destinations (OD) of travel—remain relatively underdeveloped.

(2) The influence of these factors on bus travel distance in medium-sized cities may be nonlinear, with possible heterogeneous interactions between variables; nevertheless, the precise nature of nonlinear relationships among factors in such cases remain unclear.

(3) While existing studies have provided abundant evidence concerning the analysis of travel distance and its influencing mechanisms within megacities worldwide, there persists an urgent need to further understand the ways in which urban density and other built environment factors affect the travel behavior of public transport users in the context of medium-sized self-contained cities in developing countries.

3 Data and material

3.1 Study area

The case selected for this study is Weinan, a typical medium-sized prefecture-level city in Shaanxi Province, China, covering a total area of 13,134 km² (Liu et al., 2020). The city lies in the eastern part of the Guanzhong Plain, adjacent to the southern edge of the Loess Plateau, and falls within the lower reaches of the Wei River. As one of the primary cradles of Chinese civilization, the city is a favorable agroecological zone in northwest China, characterized as the “root of China, source of culture, the holy land of rivers and mountains, and a city of humanity.” Weinan encompasses 2 municipal districts and 7 counties, among which the Linwei district serving as the economic and political center, being the most urbanized area of the city.

Weinan has prioritized a bus-oriented public transportation system as its primary urban mobility strategy, actively developing a coordinated urban public transportation network (The People’s Government of Weinan City, 2023). By 2018, the city has operated 24 bus lines and 377 stops, with a service radius covering 35.10 km² of the main districts (Li et al., 2020). Compared to other medium-sized Chinese cities lacking dense bus infrastructure, Weinan achieved 100% bus stop coverage within 500 meters of its main urban area by 2020 (The People’s Government of Linwei District, Weinan City, 2020). In urban areas, bus operations commence around 6:10 am, with buses running at an average of 10-minute intervals. Service frequency also varies by time of day, dynamically adjusting to address rush hour fluctuations and seasonal variations based on real-time passenger flow data (Weinan City Transportation Bureau, 2020). From 2016 to

2018, the total annual bus passenger volume in Weinan City increased by 10.71%, while the national average of China decreased by 5.58% (Tianjin Kuangwei Company, 2019).

In this study, the selected time frame from November 12th to 30th, 2018, avoids major holidays to minimize potential impacts on travel behavior. Our focus on the Linwei district, an area with a well-developed public transportation network, aims to provide targeted insights into the core urban travel patterns. Considering that the average distance between bus stations is approximately 500 meters, we have selected a 500-meter radius grid for research analysis. This grid size allows us to capture an appropriate number of relevant factors within the immediate vicinity of each bus station. The study regions and bus lines are depicted in Fig 1. Weinan’s primary districts are served by an 187.6 kilometer bus network, yielding a density of 2.37 km/km² and annual ridership of 50.18 million. This extensive bus infrastructure and high utilization again underscores Weinan’s bus-oriented urban mobility focus.

3.2 Data description

The dataset for this research included bus smart card data, bus GPS data, bus station information, road information, points of interest (POI), and weather data. The Advanced Public Transportation Systems (ATS) data used in this research was provided by the Weinan City Bus Company (Chen & Yang, 2013). The dataset covered a period of two consecutive weeks, specifically from 12 to 30 November 2018.

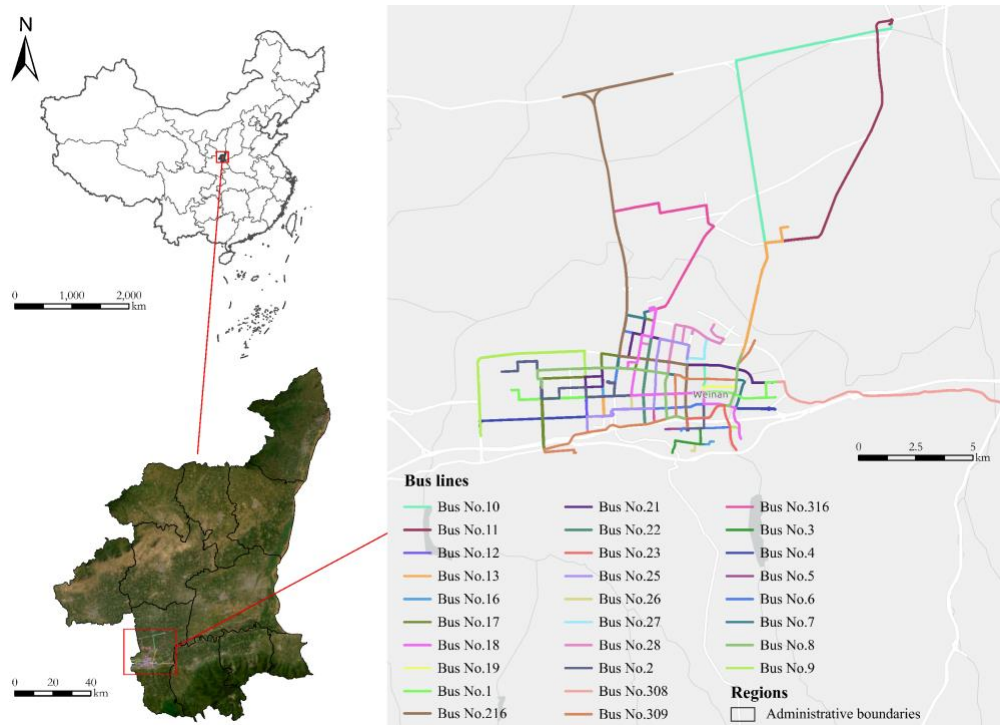


Figure 1. Weinan’s bus route network

(1) Bus smart card data

A sample of the bus smart card data is listed in Table A1. The bus smart card data contains information of bus users’ travel date, boarding swiping time, self-numbering, bus route, bus registration number, terminal number, and smart card number. These data

elements are commonly used for data filtering and association. The OD locations of trips are identified based on the boarding and alighting bus stations, respectively. It is important to note that due to the "one-ticket" operational nature of the bus system, where passengers are only required to swipe their cards upon boarding, no specific information regarding the alighting stations is available in the dataset. In the present study, the estimation of alighting stops was carried out using a methodology established in our previous research (Chen et al., 2018). A concise explanation and further details of this methodology can be found in Section 4.3 of the Methods.

(2) Bus data

A sample of the bus GPS data can be found in **Table A2**. All buses operating on the 24 lines managed by the Weinan Bus Company are equipped with onboard GPS systems. The bus GPS data comprises information, such as the date, time, route, self-numbering, bus registration number, running mileage, instantaneous bus speed, as well as the longitude and latitude coordinates of the bus lines. Samples of bus station data is provided in **Table A3**.

(3) Road information data

Samples of road information data area is provided in **Table A4**. We employ road density as a representation of the "Design" component within the "6Ds" metrics for the built environment. Road density, to a considerable extent, reflects the diversity of internal connectivity choices, indicating the convenience of residents in traversing or reaching different areas. The calculation formula for road density is expressed as follows in Equation (1).

$$d = \sum L_n / A \quad (1)$$

where d represents the road density (km/km^2); L_n represents the length of the n th road in the region; A denotes the total area of the region.

(4) Bus station POI data

The original bus station POI data can be found in **Table A5**. In this study, we utilize the Amap open platform to collect POI data for bus stations (Amap open platform, 2022). The POI data are reclassified into five categories based on distinct varieties: commercial land, science, and education, residential, office, and life services. Taking the bus station as the center point and setting a 500-meter search radius, we gather data according to the POI coding rules. Ten feature values are selected, including the number of commercial, science and education, residential, office, and life service POIs. To characterize the urban built environment, we use density as the primary metric. Finally, these features are sequentially denoted as "origin business density, origin science and education density, origin residential density, origin office density, origin life service density, destination business density, destination science and education density, destination residential density, destination office density, and destination life service density."

In this study, the degree of land use diversity is characterized by the density of commercial, science and education, residential, office, and life service sectors at both origins and destinations. The diversity at the origin/destination is calculated using Equation (2):

$$H_u = - \frac{\sum_{t=1}^T p_{tu} \ln p_{tu}}{\ln N_u} \quad (2)$$

where H_u represents the land use mix in research unit u ; N_u is the number of land use types within research unit u ; P_{tu} is the proportion of land use type t within research unit u ; and T is the total number of land use types.

(5) Weather data

The original weather data is presented in **Table A6**. A Python-based web crawler was utilized to extract historical weather information on an hourly basis from (<http://lishi.tianqi.com/weinan/201811.html>), including temperature, humidity, precipitation, and visibility. These crawled weather fields function as indicators to characterize the dynamic weather conditions corresponding to the boarding swipe time of each individual travel record.

3.3 Data pre-processing

For the pre-processing stage of the analysis, we implemented a data fusion treatment. Smart card, bus GPS, and bus station data were linked based on relevant fields. The data-association relationship is depicted in **Fig 2**. Subsequently, these parameters were integrated to estimate the travel distance of bus passengers.

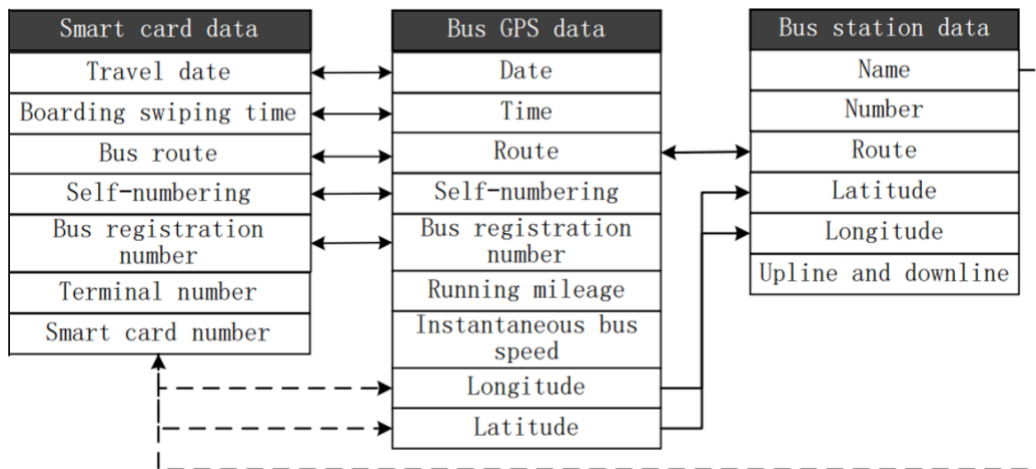


Figure 2. Data-association relationship

The bus operating period in Weinan is from 6 am to 10 pm. Passengers' smart card swiping records were sorted and matched with bus GPS data to ensure consistency between swiping time and bus GPS time. Based on the matching results, passenger boarding time and corresponding GPS coordinates were obtained. The bus GPS coordinates with the lowest instantaneous bus speed were considered when swiping, ensuring the most accurate coordinate selection. Besides, the misalignment distance was calculated using bus coordinates when swiping the card and bus station coordinates. The closest bus station is where the passenger boards the bus. It is important to note that round trips were filtered out and not considered in this study. Finally, the bus's self-numbering and route were determined by matching smart card data with bus station data. Based on the boarding station and station number, the bus's running head direction (upline or downline) was identified.

To enhance the computational efficiency while maintaining an acceptable level of accuracy, in this paper, we define the travel distance as the cumulative total of the Euclidean distances between each consecutive station within the trip's origin (boarding) and destination (alighting) stations. This methodology is different from the conventional practice of directly computing the Euclidean distance between the boarding and alighting stations. Instead, we have chosen a stepwise approach of calculating distances between

successive stations, based on the understanding that urban transit routes are not straight lines, but rather, a series of connected segments between adjacent stations. Our approach is consistent with the ground-truth of the road network in Weinan, enabling a closer approximation of the actual bus route. Furthermore, a rigorous validation is conducted, and the results present that the discrepancies between our estimated travel distance with the actual bus travel distance less than 200 meters. Thus, here we calculate the travel distance as Equations (3)–(6):

$$DLng = Lng2 - Lng1 \quad (3)$$

$$DLat = Lat2 - Lat1 \quad (4)$$

$$Dis1 = \arcsin(\sqrt{\sin(DLat/2)^2 + \cos(lat1) \times \cos(lat2) \times \sin(DLng/2)^2}) \quad (5)$$

$$Dis2 = Dis1 \times 2 \times 6731 \quad (6)$$

where $DLng$ represents the difference between the longitudes of two adjacent stations; $DLat$ represents the latitude difference between the two adjacent stations; $Dis1$ denotes an intermediate variable for distance calculation, and $Dis2$ is the final distance between the two adjacent stations, calculated using the Earth's average radius (6731 km).

The collected data were normalized and processed through the following steps: (1) outlier processing for checking missing values and duplicate values; (2) text tagging for digitally translating the text information and labeling all character factors. **Table 1** presents the descriptive analysis of the basic data and brief definitions for each element. In this study, boarding time was selected during peak hours between 6am and 8am. The description of the variables and their correlation between variables was visualized using pair diagrams (**Fig 3**), with the diagonal line representing the distribution of the corresponding variable.

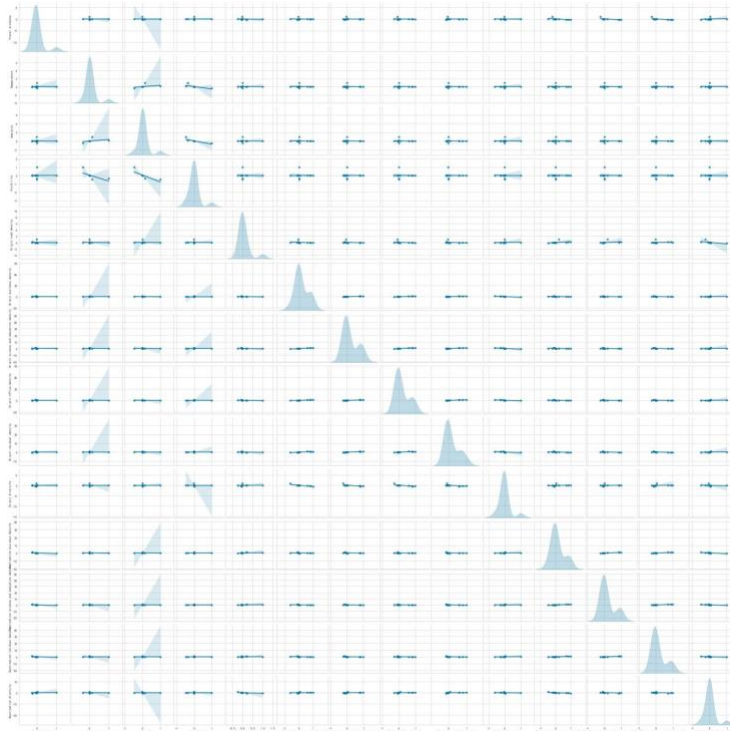


Figure 3. Pair diagrams showing correlated variations between variables

Table 1. Descriptive statistics of variables

	Variables	Definitions	O bs	Me an	Std .	M in	M ax
Weather	Travel distance (m)	The Euclidean distance between boarding and alighting stations.	7, 82 2	419 4.6 1	213 6.3 9	10 00	11 50 0
	Temperature (°C)	The average temperature of one day within the study time frame.	7, 82 2	3.8 0	2.4 2	- 0. 8	7. 9
	Humidity (%rh)	The average humidity of one day within the study time frame.	7, 82 2	75. 45	15. 16	44	98
	Precipitation (mm)	The average precipitation of one day within the study time frame.	7, 82 2	0.0 1	0.0 5	0	0. 3
	Visibility (m)	The average visibility of one day within the study time frame.	7, 82 2	5.7 7	6.0 9	0	26
Built environment (Density)	Origin business density (pcs/500*500m ²)	The business land density in the area where a trip begins.	7, 82 2	550 .95	471 .06	1	19 20
	Origin science and education density (pcs/500*500m ⁴)	The science and education land density in the area where a trip begins.	7, 82 2	55. 16	45. 48	1	23 9
	Origin office density (pcs/500*500m ²)	The office land density in the area where a trip begins.	7, 82 2	70. 08	48. 86	1	26 6
	Origin living density (pcs/500*500m ²)	The living service land density in the area where a trip begins.	7, 82 2	256 .03	198 .26	1	92 1
	Origin residential density (pcs/500*500m ²)	The residential land density in the area where a trip begins.	7, 82 2	53. 68	48. 65	1	49 7
	Destination business density (pcs/500*500m ²)	The business land density in the area where a trip ends.	7, 82 2	791 .92	609 .04	1	19 20
	Destination science and education density (pcs/500*500m ²)	The science and education land density in the area where a trip ends.	7, 82 2	75. 89	55. 78	1	19 4
	Destination office density (pcs/500*500m ²)	The office land density in the area where a trip ends.	7, 82 2	91. 43	67. 18	3	26 6
	Destination living density (pcs/500*500m ²)	The living service land density in the area where a trip ends.	7, 82 2	359 .02	259 .73	1	92 1
	Destination residential density (pcs/500*500m ²)	The residential land density in the area where a trip ends.	7, 82 2	73. 27	58. 65	1	32 6
Built environment (Diversity)	Origin diversity	The degree of land mixing in the area where a trip begins.	7, 82 2	0.7 5	0.0 9	0. 38	0. 98
	Destination diversity	The degree of land mixing in the area where a trip ends.	7, 82 2	0.7 3	0.0 9	0. 38	0. 96
Built environment (Design)	Origin road density (km/500*500m ²)	The road density in the area where a trip begins.	7, 82 2	4.1 2	0.9 5	0. 93	10 .5 7
	Destination road density (km/500*500m ²)	The road density in the area where a trip ends.	7, 82 2	4.1 2	0.9 5	0. 93	10 .5 7

4 Methods

4.1 XGBoost

XGBoost is a widely used machine learning algorithm, particularly in the field of transport, where it has been applied to predict future traffic flow on road sections (Dong et al., 2018), traffic speed considering weather effects (Du et al., 2020), and study macroscopic influences of traffic fatalities (Jiang et al., 2021). XGBoost offers advantages in accuracy, efficiency, scalability, and robustness, making it suitable for various machine learning tasks. The algorithm employs an adaptive gradient boosting framework for training and optimizing models, resulting in high accuracy. Its optimization techniques, multi-threading support, and distributed computing capabilities make it efficient and scalable. XGBoost effectively handles missing values and outliers, showcasing robustness. Additionally, it carries out feature selection and prevents overfitting, which enhances both model performance and interpretability. Consequently, XGBoost has demonstrated exceptional performance in practical applications.

As introduced by Chen and Guestrin (2016), the XGBoost algorithm is a novel implementation for gradient-boosting machines, specifically designed for categorical regression trees. This technique is founded on the principle of “boosting,” which amalgamates several weak learners through supplementary training methodologies to construct strong learners (Fan et al., 2018). XGBoost effectively mitigates overfitting and optimizes the model’s computational capacity by minimizing the objective function, thereby facilitating the integration of prediction and regularization components while preserving the highest achievable processing velocity. During the training process, XGBoost automatically conducts similarity estimation. To address the limitations of weak learners, the initial learner is fit to the entirety of the input dataset, whereas the subsequent model is fit to the residuals. **Fig 4** illustrates the XGBoost schematic. This iterative method is executed multiple times until the predetermined stopping condition is met. The summation of the predictions for each learner culminates in the final model prediction. The ensuing equation delineates the generic function for the phase prediction: (Alabdullah et al., 2022):

$$g_i^n = \sum_{k=1}^K g_k(x_i) = g_i^{(p-1)} + g_i(x_i) \quad (7)$$

where i represents the i th sample; k is the k th tree; $g_p(x_i)$ is the learner at phase p , g_i^p and g_i^{p-1} are the prediction at phases p and $p-1$, and x_i is the input variable.

To avoid overfitting while losing model computation, XGBoost created an analytical formulation to evaluate the “merits” of the original function-based model:

$$\text{Obj}^{(p)} = \sum_{k=1}^n \ell(\bar{y}_i, y_i) + \sum_{k=1}^p \sigma(g_i) \quad (8)$$

where ℓ denotes the loss function; n indicates the number of observations; and σ represents the regularization term, which is described as follows.

$$\sigma(g) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (9)$$

where ω denotes the vector scores in the leaves; γ is the minimum loss necessary to further divide the leaf node, and λ represents the regularization parameters. Chen and Guestrin provide a detailed illustration of XGBoost (Chen & Guestrin, 2016).

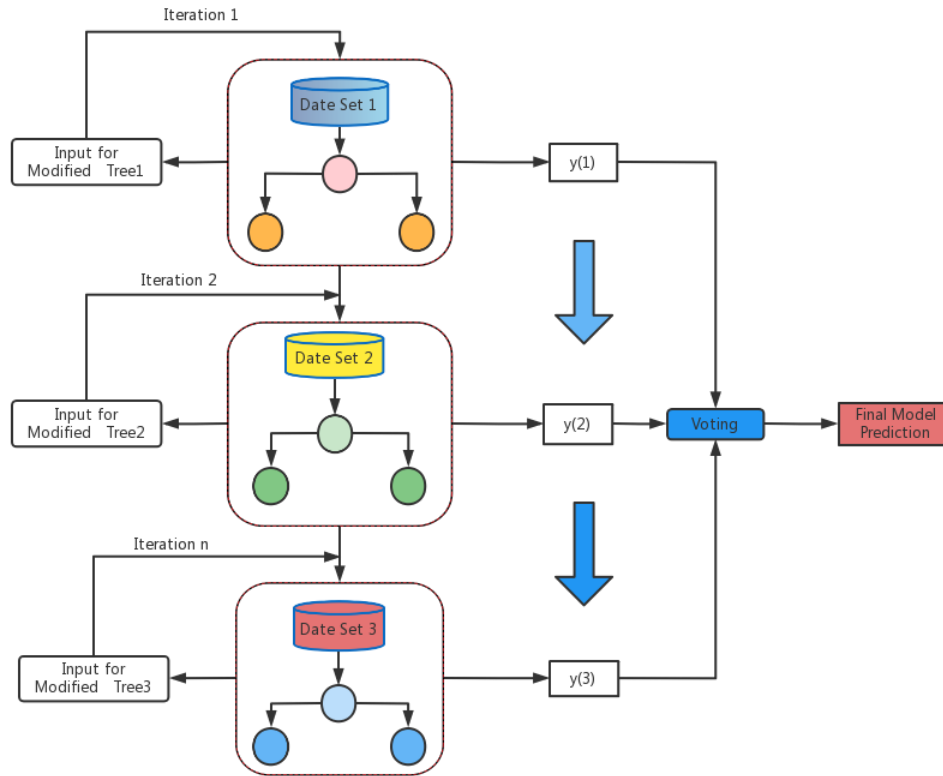


Figure 4. Illustration of XGBoost regression

The validity of the model hinges on an amalgamation of crucial features. To enhance performance and accuracy, the grid search method was employed for optimizing XGBoost hyperparameters. This approach evaluates the performance of all specified hyperparameters and each combination of their values, subsequently selecting the optimal hyperparameter value. During the grid search for hyperparameter optimization, a subset of data points is entirely concealed within the model as a “test set” to augment accuracy and mitigate potential overfitting. The models were appraised using several prevalent statistical indicators, including Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), Coefficient of determination (R^2), Root Mean Squared Logarithmic Error (RMSLE), and Mean Absolute Percentage Error (MAPE). It is essential to note that the ideal values for R^2 are 1, while for MAE, MSE, RMSE, and RMSLE, the ideal values are 0. In this study, we constructed the XGBoost regression model utilizing the pycaret package 2.0 in Python 3.8.

4.2 SHAP

SHAP is a method for elucidating the performance of machine learning models and explaining individual predictions (Mangalathu et al., 2020), drawing upon the synthesis of game theory and the additive local interpretation (Lundberg & Lee, 2017). This approach can interpret feature importance scores derived from intricate training models and deliver an interpretable prediction for a test sample. For instance, inputs are regarded as participants, and predictions assume the role of expenses. SHAP calculates each player’s contribution to the game and creates the input value z after being simplified by

$r = h_r z$ mapping r to z . Based on z , the original model $f(r)$ can be approximated by a linear function of binary variables (Wang et al., 2022), which is formulated as

$$f(r) = u(z) = \varphi_0 + \sum_{j=1}^M \varphi_j z_j \quad (10)$$

where $z = \{0,1\}^M$; M is the input feature number, $\varphi_0 = f(h_r(0))$; j is the j th feature, and φ_j is the feature attribute value (Wang et al., 2022):

$$\varphi_j = \sum_{S \in F \setminus \{j\}} \frac{|S|!(M-|S|-1)!}{M!} [f_r(S \cup \{j\}) - f_r(S)] \quad (11)$$

$$f_r(s) = f(h_r^{-1}(z)) = E[f(r)|r_s] \quad (12)$$

where F is the non-zero set of inputs in z , and S is the subset of F obtained after excluding the M th feature from F . φ_j denotes a uniform measure of the additive feature attributes, referred to as the SHAP value. To date, SHAP is the only method possessing the properties of local accuracy, missingness, and consistency. As the exact computation of $E[f(r)|r_s]$ is quite challenging, several approximation methods have been developed, including the tree SHAP method used in this study. Considering the maximum depth of any tree D , the complexity of computing $E[f(r)|r_s]$ with tree SHAP is $O(TLM^2)$, which reduces the computational complexity from a high-order exponential to a quadratic level (Lundberg & Lee, 2017).

Lundberg and Lee propose the SHAP value as a fundamental tool for developing an IML model (Lundberg & Lee, 2017). Briefly, SHAP is a method employed to demonstrate the relative contribution of each input variable in generating the final output variable. This notion bears resemblance to parametric analysis, where all other variables are held constant while one variable is altered to observe the impact of the changed variable on the target attribute (Ji et al., 2022). The SHAP method utilizes the SHAP value and variable independence, which is a numerical technique capable of calculating the contribution of each variable to the production of outcomes. For every sample model, a predicted value was generated, and a SHAP value was assigned to each feature in the sample. Consequently, SHAP not only supplies the magnitude of feature influence but also conveys the positive and negative aspects of feature influence in each sample. Moreover, SHAP can be applied to investigate the heterogeneous interaction effects between features and the nonlinear effects of features. The nonlinear effects of features on the model, also known as the partial dependence plot, offer a visualization of the relationship between the predicted outcome of a target model and multiple input features while maintaining other features constant. This plot reveals the impact of a feature on the model's prediction as a function of another feature, thereby enabling the visualization of nonlinear relationships between features and an enhanced comprehension of how they interact to influence the model's predictions. The nonlinear effects of features serve as a potent tool for feature engineering and can assist in gaining insights into the behavior of complex machine learning models.

In this paper, we adopt two additional analytical tools from the SHAP analysis, namely the SHAP force diagram and SHAP heatmap, respectively. The global interpretation only provides the relative importance of the contributing variables and their effects on the target variables, such as how the impacts on the target variables change as each input factor increases or decreases. Individual interpretations in the form of SHAP are necessary to optimize the input variable values.

SHAP force diagrams offer interpretability of single-sample predictions by converting SHAP values into forces, where each feature value acts as a force that increases or decreases the predicted value. In the SHAP force diagram, the predictions start from a baseline, a constant that represents the model, and the base value is the average of the model output and training data. Each attribute value is denoted by an arrow that increases or decreases the predicted value. The number below the arrow represents the feature

variable's SHAP value in that instance. Features that push the predicted value higher to the right are displayed in red, while features that push the predicted value lower to the left are shown in blue. Longer arrows signify a greater impact on the feature output. The middle number represents the final predicted value for a single sample. SHAP force plots can be used for interpreting specific sample predictions and analyzing error samples.

The SHAP heatmap is a method that employs supervised clustering to visualize the overall structure of a dataset using 2D heatmaps. Supervised clustering involves clustering data points not by their original attribute values but by their SHAP values. It groups samples together for the same reasons and with the same model output. When plotting, the SHAP value matrix is passed to the heatmap plotting functions. In the resulting graph, there are instances on the X-axis, model inputs on the Y-axis, and coded SHAP values on the color scale. Above the heatmap matrix are the predictions of the instance models, and the grey dashed line is the baseline, representing the average prediction for all instances. The black bars on the right also indicate the global importance of the model inputs.

5 Results

5.1 Model construction

We established an XGBoost regression model using Pycaret 2.0, automates the process of removing multicollinearity between variables and selecting the optimal set of independent variables for modeling during training. The values of important parameters before and after hyperparameter tuning are presented in **Table 2**. Additionally, the performance of the two models was validated through a 5-fold cross-validation test, as shown in **Table 3**. The results indicate that the performance of the XGBoost model after hyperparameter tuning was significantly better than that before tuning.

To further demonstrate the effectiveness of the XGBoost model, its performance was compared with that of other machine learning models, such as Random Forest, LightGBM, KNN, and Linear Regression, using the same dataset. The XGBoost model outperformed the other models in terms of accuracy and fitting performance, as shown in **Table 4**. **Appendix B** provides a detailed illustration of the generation process and prediction performance of the developed XGBoost model. The superior performance of the XGBoost model compared to other ML models highlights its effectiveness in this particular application, making it a suitable choice for further analysis and interpretation using techniques such as SHAP analysis.

5.2 Variables interpretation

Fig 5 displays the impact of the top 22 features on travel distances. The average SHAP value in all samples represents the overall impact of each variable on travel distance, indicating the average influence of each variable on the travel distance. The results show that the POI information at the locations of OD is of particular importance in determining travel distance. **Fig 6** illustrates the interactions between the variables, demonstrating the heterogeneous influence of one variable on another in different directions. The nonlinear effects of the POI and weather variables on passengers' travel distances were presented in **Fig 7**.

Table 2. Hyperparameter for XGBoost models

Hyperparameter	Value	
	Before tuning	After tuning
base_score	0.5	0.5
colsample_bytree	1	1
enable_categorical	False	False
gamma	0	0
gpu_id	-1	-1
importance_type	None	None
learning_rate*	0.3	0.4
max_depth	6	6
min_child_weight*	1	3
missing	Nan	Nan
monotone_constraints	()	()
n_estimators*	100	270
random_state	4774	4774
reg_alpha*	0	2
reg_lambda*	1	2
scale_pos_weight*	1	22.0
subsample*	1	0.9

Table 3. Five-fold cross-validation result

XGBoost models	5-Fold	MAE	MSE	RMSE	R ²	RMSLE	MAPE
Before tuning	0	481.2239	828470.9375	910.2038	0.8202	0.2248	0.1434
	1	530.1092	957384.8125	978.4604	0.7956	0.2383	0.1475
	2	462.5461	738043.0625	859.0943	0.8293	0.2255	0.1424
	3	518.9788	1042089.6250	1020.8279	0.7688	0.2813	0.1802
	4	537.2076	1007047.5000	1003.5176	0.7890	0.2670	0.1751
	Mean	506.0131	914607.1875	954.4208	0.8006	0.2473	0.1578
	SD	29.0714	114211.4287	60.7299	0.0218	0.0228	0.0164
After tuning	0	450.3609	782173.3750	884.4056	0.8302	0.2329	0.1360
	1	488.8621	895203.8125	946.1521	0.8089	0.2320	0.1360
	2	437.5330	735868.7500	857.8279	0.8298	0.2296	0.1313
	3	480.8282	930493.0625	964.6207	0.7935	0.2757	0.1689
	4	486.2679	926501.1250	962.5493	0.8059	0.2574	0.1604
	Mean	468.7704	854048.0250	923.1111	0.8137	0.2455	0.1465
	SD	20.8323	79899.1873	43.7477	0.0143	0.0182	0.0151

Table 4. Comparison of five machine learning algorithms

Model	MAE	MSE	RMSE	R ²	RMSLE	MAPE
XGBoost	468.7704	854048.0250	923.1111	0.8137	0.2455	0.1465
RF	486.6732	921572.8431	958.9392	0.7990	0.2451	0.1505
lightGBM	692.5497	1222969.0485	1105.1184	0.7329	0.2938	0.2203
KNN	809.7348	2019067.4500	1418.3175	0.5593	0.3482	0.2394
LR	1635.5589	4213722.2500	2052.4762	0.0796	0.5288	0.5364

Among all variables investigated, destination residential density had the most significant effect on travel distance. (1) As residential density at destinations increases, travel distance correspondingly decreases. This relationship can be primarily attributed to the increased availability and accessibility of services and amenities in areas with higher residential density. In those areas, there is usually a greater concentration of infrastructure such as shops, schools, healthcare facilities, and public transit systems. Demands from a more populous environment often leads to a more compact urban form, which in turn enables individuals to fulfill their travel needs within shorter distances. The negative correlation between residential density and travel distance is further substantiated by the nonlinear variation diagram. This finding is also consistent with the research conducted by Levinson and Kumar (1997) on automobile commuters in U.S. cities. They indicated that the residential density in the vicinity of the tripmakers' home as a critical determinant of travel distance, with higher densities correlating with shorter travel distances. This correlation might be even more pronounced in larger cities or megacities. (2) A heterogeneous interaction is observed between the residential density at destinations and the office density at origins when the SHAP value of destination residential density is below 200 (Fig 6a). In areas with high office density at the origin, when the destination residential density value is between 50 and 200, travel distance is greatly reduced. This may be attributed to office workers preferring to live close to their workplaces and being less likely to make long-distance journeys for recreational purposes. (3) Furthermore, when the destination residential density value is high, an increase in precipitation has a positive effect on travel distance (Fig 6j). This is possible because road congestion may become more severe as rainfall increases, and residents of medium-sized cities may give up bus and choose other modes of transportation at short distances, such as walking, which leads to a longer average distance traveled by buses on rainy days.

Origin office density and destination business density are identified as the second- and third-most significant factors, respectively. (1) Both these factors exhibit negative impacts on travel distance, indicating a decrease in travel distance as their values increase. This trend could perhaps be attributed to the preference of individuals in medium-sized cities to work near their homes, thereby minimizing commuting time, which resonates with findings of Ding et al. (2017), which discovered that increasing employment opportunities in residential areas could effectively reduce travel distances. However, paradoxically, the sprawl of urban areas and the accumulation of capital could lead to lengthier and more time-consuming commuting behaviors (Engelfriet & Koomen, 2018). This is a common issue in many of today's large cities that lack clear clusters of businesses and other facilities. (2) Travel distance sharply declines when the destination business density exceeds 1,500, indicating large commercial centers attract people living nearby. This observation is indeed sensible, as business development typically targets meeting the needs of the local community, which in turn attracts more individuals from the surrounding area. (3) the interaction between origin office density and destination business density (Fig 6c) shows that high value of origin office density positively impacts

travel distance. However, this positive effect diminishes as destination business density increases, becoming negative when the value of destination business density exceeds 1,250. This could be because people finishing work are less likely to visit large commercial buildings far away, opting instead for nearby commercial clusters.

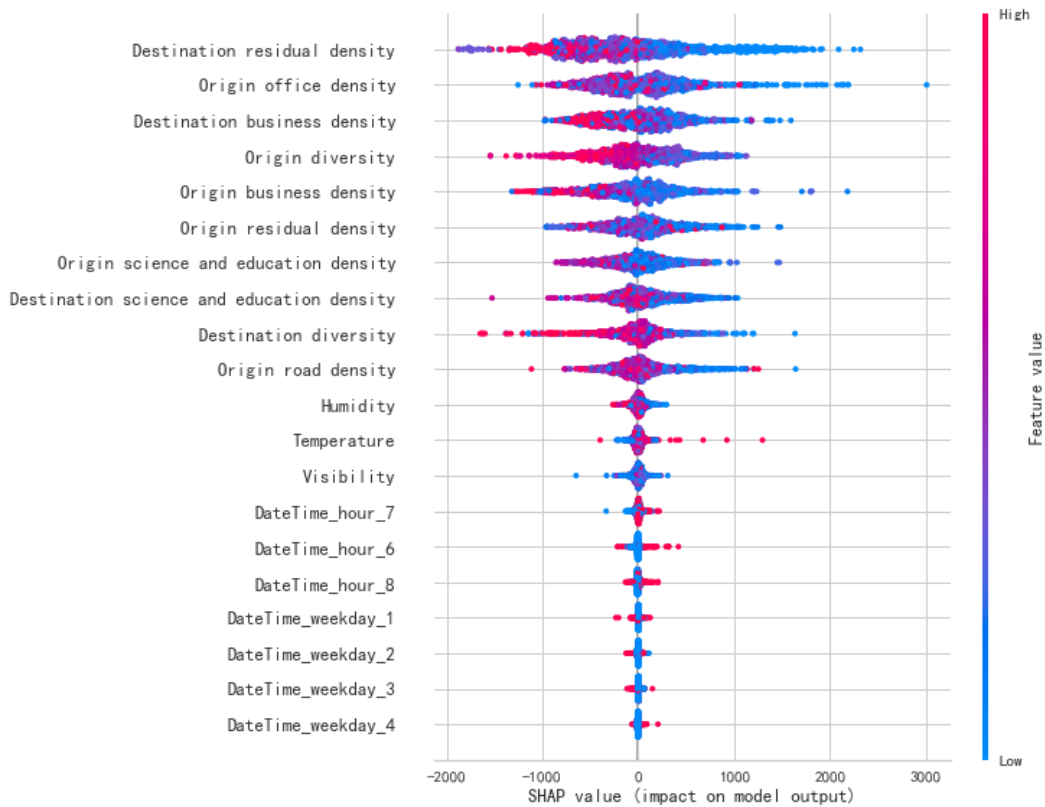


Figure 5. The SHAP values of variables (Note that “Datetime_weekday_0” represents Monday, and “Datetime_weekday_6” represents Sunday)

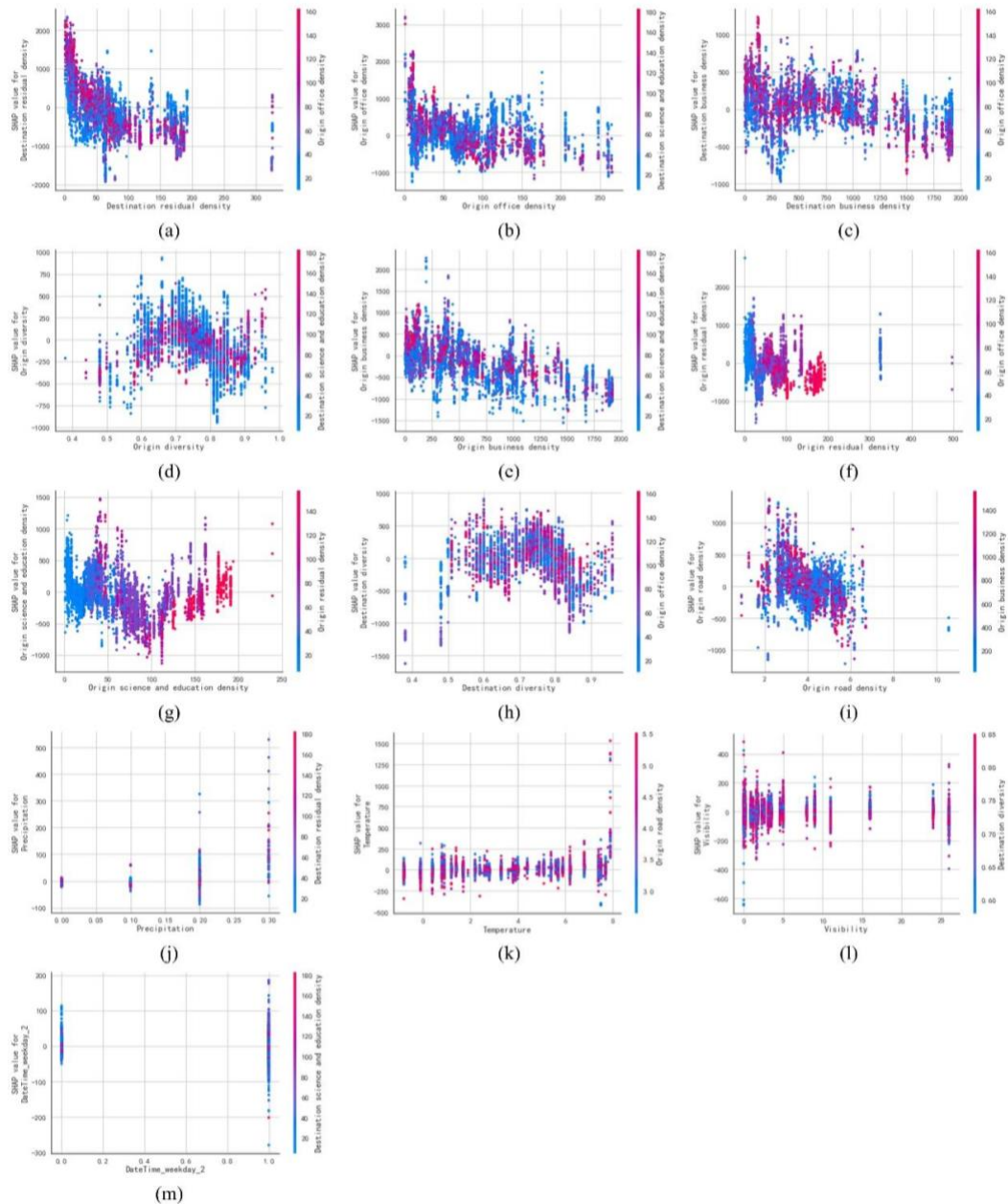


Figure 6. Heterogeneous interactions between factor pairs

Origin diversity, origin business density, origin residential density, and density of science and education at origin were found to have negative impacts on travel distance. Appropriately enhancing the values of these factors can result in shorter travel distances for travelers. (1) When the value of origin diversity is below 0.55, it has a strong inhibitory impact on travel distance. This effect peaks at a diversity value of 0.86. It is in line with Jain and Tiwari (2019) which suggests that origin diversity is a determining factor affecting travel distance. High origin diversity is often associated with mixed land-use areas, allowing people to meet their needs locally and thus reducing travel distance. This finding is also true in the context of large cities, according to a study by Zhang and Zhao (2017), on Beijing, China found that high land-use diversity and a balanced jobs-

housing ratio significantly reduced commuting travel. Similarly, Ding et al. (2017), in their study of the Baltimore metropolitan area, suggested that high diversity, though slightly insignificant, directly reduces car travel distance in large cities. However, they also noted that high diversity might increase the likelihood of forming travel chains for various activities, potentially leading to an overall increase in travel distance. (2) In contrast, origin business density does not exhibit a significant nonlinear effect on travel distance. (3) Origin residential density has a negative effect on travel distance. However, when the value of residential density is less than 20, which suggests people might live in relatively scattered areas, the demand for long-distance trips increases dramatically. (4) Analyses of heterogeneous interactions reveal that the density of science and education at destinations has unstable interaction effects with origin diversity (Fig 6d) and origin business density (Fig 6c). Additionally, the density of science and education at origins interacts with the origin's residential density (Fig 6g). The negative effect is most prominent when the value of the density of science and education at origins reaches 100. This implies that individuals originating from areas with advanced science, education, and cultural facilities may exhibit a lower propensity to engage in long-distance travel primarily motivated by educational purposes.

The effects of science and education density at destinations, destination diversity, and origin road density on travel distance are less pronounced than the previously analyzed variables. (1) For science and education density at destinations, the nonlinear plots (Fig 7i) show that the negative effect of science and education density at destinations on travel distance initially strengthens and then weakens, and the inhibitory effect on travel distance is strongest when its value is between 75 and 125. Interestingly, its impact on travel distance is the greatest on Wednesdays (Fig 6m), and `weekday_2` represents Wednesdays. (2) Regarding destination diversity (Fig 7k), when its value is below 0.5, the travel distance is relatively short. However, areas with rich features greatly attract people to travel when the corresponding diversity value is above 0.5. (3) Origin road density also plays an essential role in model predictions (Fig 7b). Well-connected road networks tend to facilitate shorter travel distances. As road network density increases, the probability of individuals opting for long-distance trips decreases. This is plausible given that denser networks usually provide more flexible and available travel routes. Conversely, in regions with underdeveloped road networks, individuals often face limited route choices and are forced to transfer at different stations. These findings are in line with observations from developed cities. For instance, a study conducted in Shanghai, China, revealed that destination road density has a negative correlation with commute distance by transit, but a positive correlation with cycling commute distance (Sun et al., 2017). (4) Moreover, heterogeneous interactions exist between origin road density and temperature (Fig 6k). When the origin road density is high, travel distance increases as the temperature increases. This conclusion aligns with a survey conducted during winter.

Weather-related variables had a relatively minor effect on travel distance compared to the built environment variables. (1) The most influential weather variable is humidity, followed by temperature. In contrast, the impact of visibility on travel distance is not particularly pronounced. (2) A heterogeneous interaction can be observed between precipitation and destination residential density (Fig 6j). When the destination residential density is high, an increase in precipitation leads to a longer travel distance. This relationship is also evident from the nonlinear plot (Fig 7a), which shows that travel distance increases with increasing precipitation. This trend may be attributed to the inclination of individuals to opt for public transport during rainy days for their long-distance commutes.

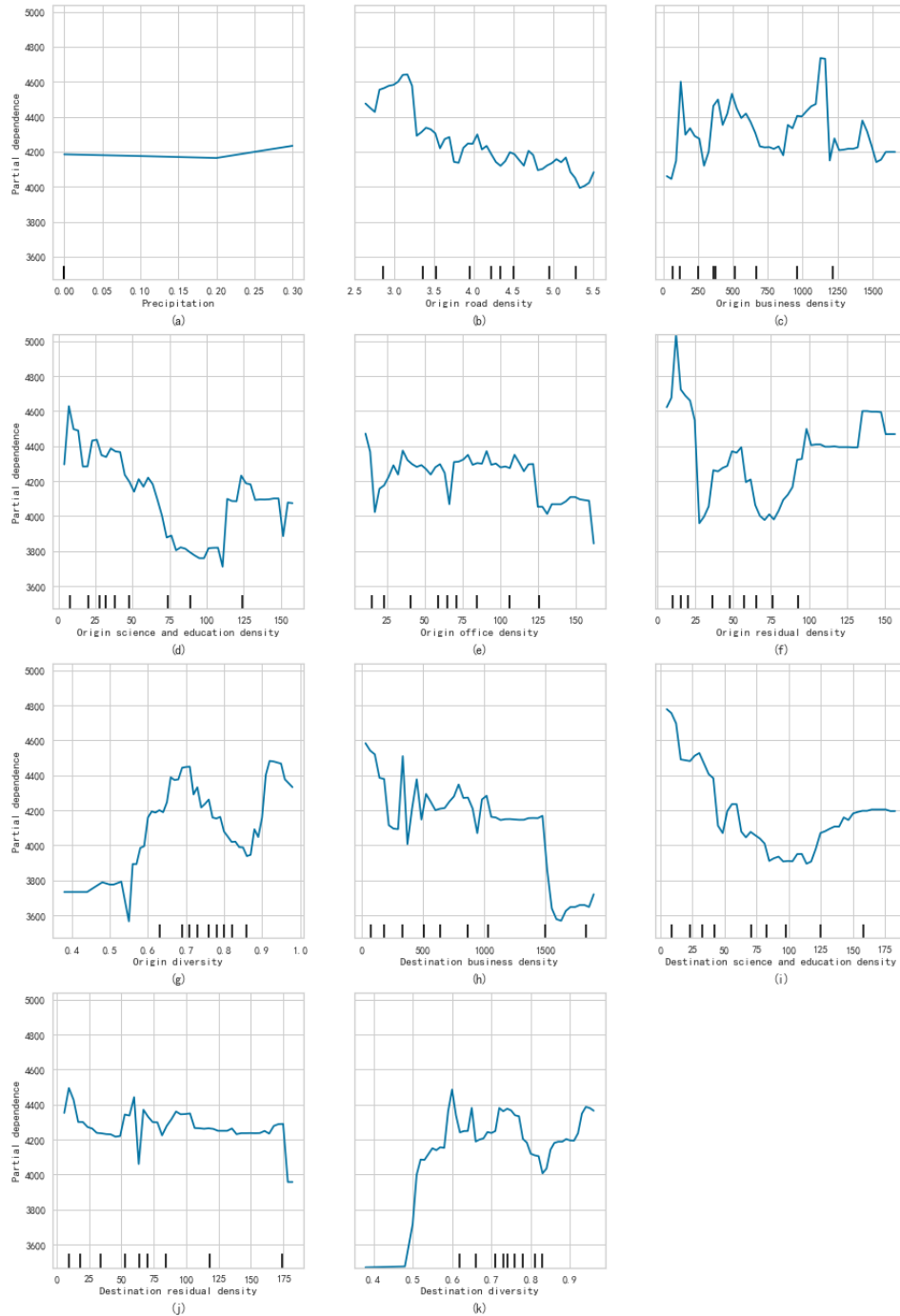


Figure 7. Nonlinear effects of factors on travel distance

Among all the time-series variables, the morning peaks at 6am, 7am and 8am (DateTime_hour_6, DateTime_hour_7, and DateTime_hour_8) had the most significant effect on travel distance. (1) The impact at 7 am is the greatest. This may be because commuting accounts for a large portion of people's trips, and between 7 am and 8 am is the peak period for commuting to work and school, during which time travel distance remarkably increases. One possible reason for this is that long mandatory travel is

clustered in the morning peak hours, as opposed to other Irregular leisure trips, which have a low probability of being clustered at a specific time of the day. (2) During weekdays, the influence at peak times is more intense. Weekdays—especially from Tuesdays to Fridays—have a much more noticeable impact on travel distance than weekends. These results are intuitive, as there are mandatory travel purposes on weekdays, such as commuting to work, which leads to people taking buses more frequently.

5.3 Individual SHAP interpretation and heatmap

Fig 8 displays the SHAP force plot for the first three samples from the dataset of 7,822 samples. Taking the first sample as an example, the base value of the model was 4,196, and the output value of the travel distance was 3,586.38. The values of different variables in this sample are also shown in plot (Fig 8b), such as destination diversity (= 0.74), origin office density (= 17.0), origin science and education density (=4.0), origin residual density (=5.0), destination residual density (=187.0), and origin diversity (= 0.87). The red variables of the model, such as destination diversity, origin office density, origin science and education density, and origin residential density, push the predicted values to the right. In contrast, the blue variables, such as destination residential density and origin diversity, push the predicted values down to the left. Origin office density had the most significant impact on the prediction of this sample. **Fig 9** illustrates the overall impact of the nine variables and the remaining 16 variables on the model output. Samples with large effects on destination residential density were grouped. The high predictive value on the right side of $f(x)$ was mainly correlated with destination residential density and origin office density.

6 Discussion and conclusion

This study explores the impacts of the built environment at OD locations together with weather conditions on the travel distances of bus passengers, utilizing multi-source big data of public transport in Weinan, a bus-oriented medium-sized city in western China. A spatiotemporal database of bus travel was constructed by correlating and transferring the data hierarchy. The joint influential effects of the built environment and weather conditions on bus passengers' travel behavior were analyzed using the XGBoost model and SHAP values. This research presents several novel implications for medium-sized cities, with the main conclusions as follows.

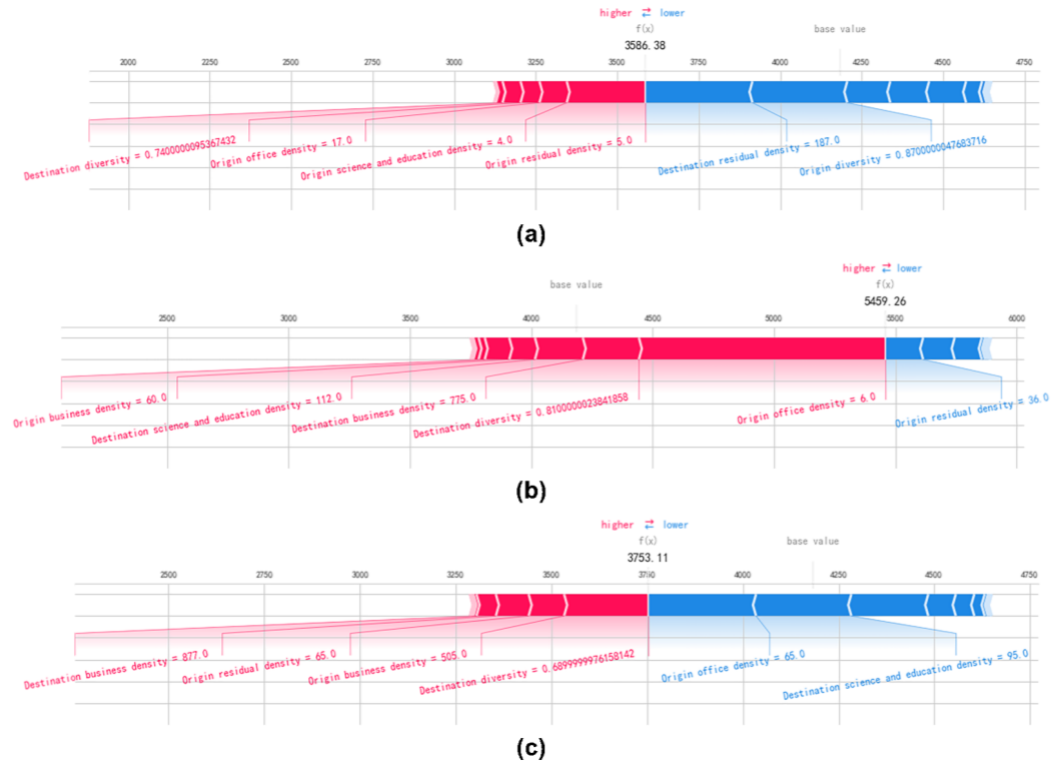


Figure 8. Individual SHAP force plots



Figure 9. SHAP heatmap

The factors associated with the origin location, ranked in order of their influencing strength, are found to be origin office density, origin diversity, origin business density, origin residential density, origin science and education density, and origin road density. On the other hand, the factors related to the destination location, ranked in terms of their importance, include destination residential density, destination business density, destination science and education density, and destination diversity. Notably, the analysis indicates that most factors associated with the built environment exert a more pronounced influence on travel behavior in comparison to weather features and temporal features. This finding underscores the substantial impact of the surrounding built environment on individuals' travel choices and patterns. Additionally, the study reveals that the salient features of the built environment differ between origins and destinations, with distinct orders of importance. This indicates a clear imbalance in the effects of these factors. Understanding the varying impacts of the built environment on travel behavior at different locations can inform targeted interventions and policies aimed at optimizing transportation systems in medium-sized cities. By recognizing the specific factors that influence travel behavior at origins and destinations, urban planners and policymakers can develop tailored strategies to address the unique needs of different areas. This may include improving accessibility to public transport, enhancing connectivity between residential and commercial areas, or promoting mixed-use development to reduce travel distances and encourage sustainable transportation modes. Ultimately, such targeted interventions can contribute to more efficient, equitable, and environmentally friendly transportation systems in medium-sized cities.

The heterogeneous and nonlinear results also provide several insights (1) Working people in medium-sized cities may prefer to work close to home, valuing comfort over lengthy commutes. For home-based travel, people prefer taking buses; therefore, bus companies can design bus routes that are more convenient for commuting between residential and office areas. (2) The development of regional businesses should prioritize the needs of neighborhood residents. For example, large-scale commercial buildings are more attractive to people. Appropriately improving origin diversity can effectively shorten travel distances. Constructing a mix of commercial, office, science, and education facilities would greatly enhance an area's carrying capacity while reducing the likelihood of long-distance travel. (3) Individuals who originate from areas with advanced science, education, and cultural facilities will likely reduce their long-distance travel for education. Urban planners should consider these facilities when designing residential areas to minimize travel distances for educational purposes. (4) When the regional road network is underdeveloped, people have fewer travel routes to choose from and may not be able to select the shortest route for their trips. As the nearby road network becomes denser and road conditions improve, the probability of people choosing long-distance travel for outings will also increase. Bus companies can set up more bus lines in areas with a higher population density to meet the travel needs of residents.

Given that the survey period took place during winter in China, it is observed that individuals are more inclined to undertake longer-distance trips as the temperature increases. Consequently, (5) in response to this pattern, the bus company can consider increasing the frequency of bus rides during periods of temperature rise in winter to satisfy the passengers' travel demand. (6) Further, during weekday morning peak hours, specifically at 6 am, 7 am, and 8 am, there is a notable increase in public transport usage, primarily driven by mandatory commuting activities such as going to work or school. The peak impact on travel distance is observed between 7 am and 8 am. To alleviate congestion during this critical morning peak period, enhancing the frequency of bus services during these time slots would further enhance the attractiveness of the bus as a mode of transportation for commuters.

The application of POI data can indeed help promote integrated urban land-use planning in medium-sized cities in developing countries for the optimization of dynamic public transport management. For example, (7) Strive for a balance between jobs and housing in an area, which significantly reduces commuting distances. This can be achieved when residential density is between 50 and 200, and office density is between 120 and 160. People living in medium-sized cities are more likely to prefer working close to dwellings rather than spending a lot of time commuting. (8) Encourage the construction of commercial complexes with a business density of more than 1,500 near high-density residential areas. This strategy can significantly shorten travel distances. (9) Establish areas with developed science and education facilities near high-density residential areas to reduce the waste of traffic resources caused by long-distance travel for educational purposes. This impact is most prominent when the origin education density reaches 100. Simultaneously, the establishment of areas with developed science and education near high-density residential areas will greatly reduce the waste of traffic resources caused by long-distance travel for educational purposes. (10) Increase origin diversity in urban land use to encourage more trips using public transport. When the value of origin diversity is between 0.7 and 0.85, travel distance decreases with the increase in origin diversity. Urban land use departments should consider setting up rich land use functions instead of relying on single-function land use clusters. These findings provide a comprehensive understanding of the multifaceted impacts of the built environment on passengers' travel behavior at OD locations in the context of medium-sized, bus-oriented cities. Historically, much of the existing research has neglected the asymmetrical effects of the built environment, thereby overlooking valuable information with significant implications for urban planning and land-use development. This research reveals that the imbalanced outcomes of pivotal factors influencing travel distance at OD locations likely reflect a deeper job-housing imbalance, a pervasive issue in numerous urban centers. It has been demonstrated that creating balanced job-housing opportunities can effectively reduce car usage and promote non-motorized travel (Wang & Zhou, 2017). Similarly, increased mixed land-use, compact community development, and improved accessibility to residential, commercial, and job locations are associated with a decrease in personal travel (Auld & Zhang, 2013; Naess, 2010; Sun et al., 2017). The conclusions drawn from this study align with prior study conducted in larger cities, indicating that urban expansion into denser areas may precipitate an increase in long-distance private trips (Tennøy, Gundersen, et al., 2022; Tennøy, Knapskog, et al., 2022). Consequently, it is of utmost importance for medium-sized cities to consider these factors, such as origin office density and destination business density, during their developmental phase. The deployment strategic spatial planning to create or retain distinct clusters of businesses and other facilities can contribute to a more efficient urban form and alleviate the adverse effects of job-housing imbalance.

In light of these observations, transit-oriented development (TOD) emerges as a viable recommendation for the case city under investigation, considering its notable compatibility with high-density urban growth, contribution to affordable housing, and potential to mitigate automobile-reliant land-use patterns in medium-sized cities. Consequently, decision-makers are encouraged to adopt more proactive approaches, such as fostering behavioral shifts in urban residents through the meticulous design of land-use patterns and the promotion of TOD initiatives.

The comprehension of the interplay between the built environment, weather conditions, and bus passengers' travel behavior is instrumental for improving public transit services, as well as devising more effective bus allocation and scheduling plans. However, it is crucial to recognize that citizens' preferences for adopting public transit services are subject to continuous evolution. For example, shared bikes have emerged as a popular transportation mode among public transit users in China, expanding the options for shared mobility. In recent years, the shared mobility market has undergone significant transformations, with some companies, like EVCAR and Mobike, consolidating or facing bankruptcy. Concurrently, unpredictable increases in user fees could gradually diminish user satisfaction, making the long-term market share uncertain. Still in the post-pandemic era, we are yet to fully understand the lasting effects the pandemic will have on our travel behavior. Public transit users in medium-sized cities might be more sensitive to these dynamics due to the lack of diverse transportation options and the monotonous urban forms compared to large metropolitan areas. Medium-sized cities tend to have fewer alternative transit modes, and less variation in land use patterns across neighborhoods. This combination of limited choices and relative homogeneity in the built environment could amplify the influence of weather conditions and other factors on public transit use in these settings. Nonetheless, as Ding et al. (2022) suggest, the fundamental relationships between built environment features and transit use are not likely to change drastically. Therefore, the implications derived from this study should remain relevant for medium-sized cities over an extended period.

We also acknowledge several limitations of this study that provide opportunities for future research. (1) Data period and update: The analysis was based on multi-source bus data for only 15 days in 2018. Expanding the dataset to include updated bus information across a longer timeframe would enable more robust analysis of factors influencing passengers' travel distance in the future. (2) Residential proximity: The bus smart card data only records where passengers boarded, but this does not definitively indicate that they are residents living nearby. This limitation precludes accurately analyzing job-housing relationships and connections between living conditions and bus usage patterns. (3) POI data: The bus station POI data used in this study represents the quantity and diversity of POIs within a 500-square meter radius of bus stops. However, it is the fact that the impact of POIs on travel behavior is not solely determined by their number, but also by their size. Our reliance on the count of POIs to characterize the built environment may cause representativeness issues. Future research should consider both dimensions—quantity and size—of POIs when investigating the mechanisms of travel behavior influenced by the built environment. (4) Representativeness of the built environment: due to the characteristics of POIs data, our study specifically focuses on density, design, and diversity, three of the "6D" indicators of the built environment. However, there is an absence in the exploration of the remaining "6Ds" indicators, namely destination accessibility, distance to transit, and demand management. This presents an opportunity for future research to investigate and address these aspects, enabling a more comprehensive understanding of how the built environment influences travel behavior. (5) Socioeconomic and demographic profiles: In our examination of important factors influencing travel behavior and distance, we primarily considered variables benchmarked in weather and the built environment. Future studies should broaden this scope to explore a wider array of main effects and interactions, for instance, the intersection of socioeconomic and demographic profiles of travelers—information that can be gleaned from resources like Census or Origin-Destination zone data—and built environment factors could be a significant determinant of travel behavior and distance, thereby contributing to a more rounded understanding of travel behavior and distance.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NO:51878166; NO:71701046, and 52002282), Shaanxi Natural Science Foundation Youth Project (NO: 2017JQ5086), Shaanxi Education Department Special Science and Technology Project Science (NO:19JK0477), and Technology Foundation Project of Xi'an University of Architectural Science and Technology (NO: ZR18021 and QN1711).

Author contribution

Xiaowei Li: Conceptualization, method, Software, writing—original draft preparation. Lanxin Shi: Software, visualization, writing—original draft preparation. Junqing Tang: Conceptualization, investigation, supervision, writing—original draft preparation, writing—reviewing and Editing. Jiaying Li: Investigation, visualization, writing—reviewing and editing. Qian Liu: Software, visualization. Jun Chen: Validation. Changxi Ma: Validation.

Appendices

Appendices available as supplemental files at <https://doi.org/10.5198/jtlu.2024.2427>.

References

- Abbasi, S., Ko, J., & Kim, J. (2022). Carsharing travel distance and its associated factors: A case study of Seoul, South Korea. *Journal of Cleaner Production*, 362, 132380.
- Agreement, P. (2015). *Paris Agreement*. Retrieved from https://unfccc.int/sites/default/files/english_paris_agreement.pdf
- Alabdullah, A. A., Iqbal, M., Zahid, M., Khan, K., Amin, M. N., & Jalal, F. E. (2022). Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis. *Construction and Building Materials*, 345, 128296.
- Allen, J., Muñoz, J. C., & Rosell, J. (2019). Effect of a major network reform on bus transit satisfaction. *Transportation Research Part A: Policy and Practice*, 124, 310–333.
- Amap open platform. (2022). Retrieved from <https://lbs.amap.com/demo/javascript-api/example/callapp/poisearch>
- Ao, Y., Chen, C., Yang, D., & Wang, Y. (2018). Relationship between rural built environment and household vehicle ownership: An empirical analysis in rural Sichuan, China. *Sustainability*, 10(5), 1566.
- Auld, J., & Zhang, L. (2013). Inter-personal interactions and constraints in travel behavior within households and social networks. *Transportation*, 40(4), 751–754.
- Baumgarte, F., Brandt, T., Keller, R., Röhrich, F., & Schmidt, L. (2021). You'll never share alone: Analyzing carsharing user group behavior. *Transportation Research Part D: Transport and Environment*, 93, 102754.
- Bi, H., Ye, Z., & Zhu, H. (2022). Data-driven analysis of weather impacts on urban traffic conditions at the city level. *Urban Climate*, 41, 101065.
- Böcker, L., Prillwitz, J., & Dijst, M. (2013). Climate change impacts on mode choices and travelled distances: A comparison of present with 2050 weather conditions for the Randstad Holland. *Journal of Transport Geography*, 28, 176–185.
- Bongardt, D., Breithaupt, M., & Creutzig, F. (2010). *Beyond the fossil city: Towards low carbon transport and green growth*. New York: United Nations Center for Regional Development.
- Cao, X., Mokhtarian, P. L., & Handy, S. L. (2009). Examining the impacts of residential self-selection on travel behavior: A focus on empirical findings. *Transport Reviews* 29(3), 359–395.
- Cervero, R. (1989). Jobs-housing balancing and regional mobility. *Journal of the American Planning Association*, 55(2), 136–150.
- Cervero, R., & Duncan, M. (2006). Which reduces vehicle travel more: Jobs-housing balance or retail-housing mixing? *Journal of the American Planning Association*, 72(4), 475–490.
- Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 2(3), 199–219.
- Chatman, D. G. (2005). *How the built environment influences non-work travel: Theoretical and empirical essays*. Los Angeles: University of California, Los Angeles.
- Chen, J., & Yang, D. (2013). Estimating smart card commuters origin-destination distribution based on APTS data. *Journal of Transportation Systems Engineering and Information Technology*, 13(4), 47–53.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

- Chen, J., Lvu, Y. & Cui, M. (2018). Estimating alighting stops of smart card public transportation passengers based on travel patterns. *Journal of Xi'an University of Architecture & Technology*, 50(1), 23–29.
- Dėdelė, A., Miškinytė, A., Andrušaitytė, S., & Nemaniiūtė-Gužienė, J. (2020). Dependence between travel distance, individual socioeconomic and health-related characteristics, and the choice of the travel mode: A cross-sectional study for Kaunas, Lithuania. *Journal of Transport Geography*, 86, 102762.
- Ding, C., Wang, D., Liu, C., Zhang, Y., & Yang, J. (2017). Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. *Transportation Research Part A: Policy and Practice*, 100, 65–80.
- Ding, C., Cao, X., & Næss, P. (2018). Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. *Transportation Research Part A: Policy and Practice*, 110, 107–117.
- 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.
- Ding, C., Liu, T., Cao, X., & Tian, L. (2022). Illustrating nonlinear effects of built environment attributes on housing renters' transit commuting. *Transportation Research Part D: Transport and Environment*, 112, 103503.
- Dong, X., Lei, T., Jin, S., & Hou, Z. (2018). Short-term traffic flow prediction based on XGBoost. *2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS)*, 854–859.
- Du, Q., Yin, F., & Li, Z. (2020). Base station traffic prediction using XGBoost-LSTM with feature enhancement. *IET Networks*, 9(1), 29–37.
- Durning, M., & Townsend, C. (2015). Direct ridership model of rail rapid transit systems in Canada. *Transportation Research Record*, 2537(1), 96–102.
- Engelfriet, L., & Koomen, E. (2018). The impact of urban form on commuting in large Chinese cities. *Transportation*, 45(5), 1269–1295.
- Ewing, R., & Cervero, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*, 76(3), 265–294.
- Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., ... & Xiang, Y. (2018). Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Conversion and Management*, 164, 102–111.
- Frank, L., Bradley, M., Kavage, S., Chapman, J., & Lawton, T. K. (2008). Urban form, travel time, and cost relationships with tour complexity and mode choice. *Transportation*, 35(1), 37–54.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232.
- Gao, K., Yang, Y., & Qu, X. (2021). Examining nonlinear and interaction effects of multiple determinants on airline travel satisfaction. *Transportation Research Part D: Transport and Environment*, 97, 102957.
- Hallegatte, S., Green, C., Nicholls, R. J., & Corfee-Morlot, J. (2013). Future flood losses in major coastal cities. *Nature Climate Change*, 3(9), 802–806.
- Handy, S. L., Boarnet, M. G., Ewing, R., & Killingsworth, R. E. (2002). How the built environment affects physical activity. *American Journal of Preventive Medicine* 23(2), 64–73.

- Holz-Rau, C., Scheiner, J., & Sicks, K. (2013). Travel distances in daily travel and long-distance travel: What role is played by urban form? *Environment and Planning A*, 46(2), 488–507.
- Hu, H., Xu, J., Shen, Q., Shi, F., & Chen, Y. (2018). Travel mode choices in small cities of China: A case study of Changting. *Transportation Research Part D: Transport and Environment*, 59, 361–374.
- Jain, D., & Tiwari, G. (2019). Measuring density and diversity to model travel behavior in Indian context. *Land Use Policy*, 88, 104199.
- Ji, S., Wang, X., Lyu, T., Liu, X., Wang, Y., Heinen, E., & Sun, Z. (2022). Understanding cycling distance according to the prediction of the XGBoost and the interpretation of SHAP: A non-linear and interaction effect analysis. *Journal of Transport Geography*, 103, 103414.
- Jiang, F., Ma, L., Broyd, T., & Chen, K. (2021). Digital twin and its implementations in the civil engineering sector. *Automation in Construction*, 130, 103838.
- Jiang, Y., Han, Y., Liu, M., & Ye, Y. (2022). Street vitality and built environment features: A data-informed approach from fourteen Chinese cities. *Sustainable Cities and Society*, 79, 103724.
- Kamga, C., & Yazıcı, M. A. (2014). Temporal and weather related variation patterns of urban travel time: Considerations and caveats for value of travel time, value of variability, and mode choice studies. *Transportation Research Part C: Emerging Technologies*, 45, 4–16.
- Ko, J., Lee, S., & Byun, M. (2019). Exploring factors associated with commute mode choice: An application of city-level general social survey data. *Transport Policy*, 75, 36–46.
- Koushik, A. N., Manoj, M., & Nezamuddin, N. (2020). Machine learning applications in activity-travel behavior research: A review. *Transport Reviews*, 40(3), 288–311.
- Leck, E. (2011). The impact of urban form on travel behavior: A meta-analysis. *Berkeley Planning Journal*, 19(1), 37–58.
- Levinson, D. M., & Kumar, A. (1997). Density and the journey to work. *Growth and Change*, 28(2), 147–172.
- Li, X., Fan, J., Wu, Y., Chen, J., & Deng, X. (2020). Exploring influencing factors of passenger satisfaction toward bus transit in small-medium city in China. *Discrete Dynamics in Nature and Society*, 2020, 1–11.
- Li, X., Ma, Q., Wang, W., & Wang, B. (2021). Influence of weather conditions on the intercity travel mode choice: A case of Xi'an. *Computational Intelligence and Neuroscience*, 2021, 1–15.
- Liu, C., Susilo, Y. O., & Karlström, A. (2017). Weather variability and travel behavior—What we know and what we do not know. *Transport Reviews*, 37(6), 715–741.
- Liu, Y., So, E., Li, Z., Su, G., Gross, L., Li, X., ..., & Wu, L. (2020). Scenario-based seismic vulnerability and hazard analyses to help direct disaster risk reduction in rural Weinan, China. *International Journal of Disaster Risk Reduction*, 48(4), 101577.
- Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777.
- Ma, L., Xiong, H., Wang, Z., & Xie, K. (2019). Impact of weather conditions on middle school students' commute mode choices: Empirical findings from Beijing, China. *Transportation Research Part D: Transport and Environment*, 68, 39–51.
- Mangalathu, S., Hwang, S.-H., & Jeon, J.-S. (2020). Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Engineering Structures*, 219, 110927.

- Matzler, K., Sauerwein, E., & Heischmidt, K. (2003). Importance-performance analysis revisited: The role of the factor structure of customer satisfaction. *The Service Industries Journal*, 23(2), 112–129.
- Mishra, G. S., Mokhtarian, P. L., Clewlow, R. R., & Widaman, K. F. (2017). Addressing the joint occurrence of self-selection and simultaneity biases in the estimation of program effects based on cross-sectional observational surveys: Case study of travel behavior effects in carsharing. *Transportation*, 46(1), 1–29.
- Montgomery, J. (1998). Making a city: Urbanity, vitality and urban design. *Journal of Urban Design*, 3(1), 93–116.
- Naess, P. (2010). Residential location, travel, and energy use in the Hangzhou Metropolitan Area. *Journal of Transport and Land Use*, 3(3), 27–59.
- Nguyen-Phuoc, D. Q., Currie, G., Gruyter, C. D., Kim, I., & Young, W. (2018). Modelling the net traffic congestion impact of bus operations in Melbourne. *Transportation Research Part A: Policy and Practice*, 117, 1–12.
- Ogra, A., & Ndebele, R. (2014). The role of 6Ds: Density, diversity, design, destination, distance, and demand management in transit oriented development (TOD). Paper presented at the Neo-International Conference on Habitable Environments, Oct. 31–Nov. 2, Jalandhar, Punjab, India.
- Park, S., Yang, Y., & Wang, M. (2019). Travel distance and hotel service satisfaction: An inverted U-shaped relationship. *International Journal of Hospitality Management*, 76, 261–270.
- Proboste, F., Muñoz, J. C., & Gschwender, A. (2020). Comparing social costs of public transport networks structured around an open and Closed BRT corridor in medium sized cities. *Transportation Research Part A: Policy and Practice*, 138, 187–212.
- Reichert, A., Holz-Rau, C., & Scheiner, J. (2016). GHG emissions in daily travel and long-distance travel in Germany—Social and spatial correlates. *Transportation Research Part D: Transport and Environment*, 49, 25–43.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Model-agnostic interpretability of machine learning. Retrieved from <https://arxiv.org/pdf/1606.05386>
- Rong, R., Liu, L., Jia, N., & Ma, S. (2022). Impact analysis of actual traveling performance on bus passenger's perception and satisfaction. *Transportation Research Part A: Policy and Practice*, 160, 80–100.
- Schäfer, A. W., & Yeh, S. (2020). A holistic analysis of passenger travel energy and greenhouse gas intensities. *Nature Sustainability*, 3(6), 459–462.
- Sultana, S. (2002). Job/housing imbalance and commuting time in the Atlanta Metropolitan Area: Exploration of causes of longer commuting time. *Urban Geography*, 23(8), 728–749.
- Sun, B., Ermagun, A., & Dan, B. (2017). Built environmental impacts on commuting mode choice and distance: Evidence from Shanghai. *Transportation Research Part D: Transport and Environment*, 52, 441–453.
- Tang, T., Liu, R., & Choudhury, C. (2020). Incorporating weather conditions and travel history in estimating the alighting bus stops from smart card data. *Sustainable Cities and Society*, 53, 101927.
- Tao, T., Wu, X., Cao, J., Fan, Y., Das, K., & Ramaswami, A. (2020). Exploring the nonlinear relationship between the built environment and active travel in the Twin Cities. *Journal of Planning Education and Research*, 43(3), 0739456X2091576.
- Tennøy, A., Gundersen, F., & Øksenholt, K. V. (2022). Urban structure and sustainable modes' competitiveness in small and medium-sized Norwegian cities. *Transportation Research Part D: Transport and Environment*, 105, 103225.

- Tennøy, A., Knapskog, M., & Wolday, F. (2022). Walking distances to public transport in smaller and larger Norwegian cities. *Transportation Research Part D: Transport and Environment*, *103*, 103169.
- The People's Government of Linwei District, Weinan City. (2020). *The comprehensive transportation development plan for the 14th Five-Year Plan period and the long-term objectives for 2035 in Weinan City*. Retrieved from <http://www.linwei.gov.cn/zfxxgk/fdzdgnr/ghjh/1628689084346040321.html>
- The People's Government of Weinan City. (2023). Weinan City Public Transport Group Co., Ltd. Retrieved from <https://www.weinan.gov.cn/zfxxgk/fdzdgnr/ggqsydw/jt/1643812198829346818.html>
- Tianjin Kuangwei Company. (2019). *Weinan City (municipal district) public bus operation quantity and passenger volume: Three-year data report*. Weinan, China: Weinan City.
- Tu, M., Li, W., Orfila, O., Li, Y., & Gruyer, D. (2021). Exploring nonlinear effects of the built environment on ridesplitting: Evidence from Chengdu. *Transportation Research Part D: Transport and Environment*, *93*, 102776.
- Tuan, V. A., Truong, N. V., Tetsuo, S., & An, N. N. (2022). Public transport service quality: Policy prioritization strategy in the importance-performance analysis and the three-factor theory frameworks. *Transportation Research Part A: Policy and Practice*, *166*, 118–134.
- Wang, D., & Zhou, M. (2017). The built environment and travel behavior in urban China: A literature review. *Transportation Research Part D: Transport and Environment*, *52*, 574–585.
- Wang, R., Wu, J., & Qi, G. (2022). Exploring regional sustainable commuting patterns based on dockless bike-sharing data and POI data. *Journal of Transport Geography*, *102*, 103395.
- Wei, M. (2022). Investigating the influence of weather on public transit passenger's travel behavior: Empirical findings from Brisbane, Australia. *Transportation Research Part A: Policy and Practice*, *156*, 36–51.
- Weinan City Transportation Bureau. (2020). *Reply to proposal No. 138 of the first session of the Sixth Municipal Committee of the Chinese People's Political Consultative Conference*. Retrieved from <https://jtj.weinan.gov.cn/zwxgk/fdzdgnr/tajybl/1628290069586542594.html>
- Weinan City Transportation Bureau. (2018). *The record of transportation work in Weinan in 2018*. Retrieved from <https://jtj.weinan.gov.cn/hdjl/zxft/1628289823745802241.html>
- Wu, Y., Mei, G., & Shao, K. (2022). Revealing influence of meteorological conditions and flight factors on delays using XGBoost. *Journal of Computational Mathematics and Data Science*, *3*, 100030.
- Yan, L., Wang, D., Zhang, S., & Ratti, C. (2021). Understanding urban centers in Shanghai with big data: Local and non-local function perspectives. *Cities*, *113*, 103156.
- Yang, C., Chen, M., & Yuan, Q. (2021). The application of XGBoost and SHAP to examining the factors in freight truck-related crashes: An exploratory analysis. *Accident Analysis & Prevention*, *158*, 106153.
- Yang, H., Zheng, R., Li, X., Huo, J., Yang, L., & Zhu, T. (2022). Nonlinear and threshold effects of the built environment on e-scooter sharing ridership. *Journal of Transport Geography*, *104*, 103453.
- Yin, J., Yu, D., Yin, Z., Liu, M., & He, Q. (2016). Evaluating the impact and risk of pluvial flash flood on intra-urban road network: A case study in the city center of Shanghai, China. *Journal of Hydrology*, *537*, 138–145.

- Yu, L., Xie, B., & Chan, E. H. W. (2019). Exploring impacts of the built environment on transit travel: Distance, time and mode choice, for urban villages in Shenzhen, China. *Transportation Research Part E: Logistics and Transportation Review*, 132, 57–71.
- Yue, Y., Zhuang, Y., Yeh, A. G. O., Xie, J.-Y., Ma, C.-L., & Li, Q.-Q. (2017). Measurements of POI-based mixed use and their relationships with neighborhood vibrancy. *International Journal of Geographical Information Science*, 31(4), 658–675.
- Zhang, M., & Zhao, P. (2017). The impact of land-use mix on residents' travel energy consumption: New evidence from Beijing. *Transportation Research Part D: Transport and Environment*, 57, 224–236.