The Journal of Transport and Land Use jtlu.org
Vol. 17 No. 1 [2024] pp. 241–267

JTLU

The nonlinear impact of cycling environment on bicycle distance: A perspective combining objective and perceptual dimensions

Yantang Zhang
School of Transportation Science and Engineering, Harbin Institute of Technology
21B932009@stu.hit.edu.cn

Xiaowei Hu
School of Transportation Science and Engineering, Harbin Institute of Technology
xiaowei_hu@hit.edu.cn

Abstract: Extending cycling distances is crucial for sustainable urban transport development and plays a role in encouraging the shift from motorized vehicles to public transport. However, there is a lack of research examining the combined impacts of both objective and perceived aspects of the cycling environment on cycling distance, and the existence of threshold effects remains unclear. This study uses 2019 cycling data from Shenzhen, China, employing the XGBoost algorithm to uncover the relative importance and thresholds of objective and perceived factors in the cycling environment. The results indicate that population density (24.8%), road network density (15.2%), the proportion of recreational facilities (9.1%), perceived accessibility (8.0%), and comfort (8.6%) hold high relative importance in predicting cycling distance. Also, maintaining road network density between 3 to 6 km/km2 and increasing the population density to exceed 22,000 people/km2 proves effective in extending cycling distances. Land use demonstrates a threshold effect, with cycling distances increasing when the recreational facilities share exceeds 8%, transport facilities share remains below 25%, and commercial facilities share stays below 30%. Perceived metrics exhibit a clear threshold effect. The study identifies that perceived safety indicates a psychological bottleneck in increasing cycling distance. Perceived accessibility is positively correlated with cycling distance when accessibility is at a low level, while comfort shows a positive correlation with cycling distance when comfort is at a high level. These findings can contribute to refining land planning and prioritizing resource allocation for organizations aiming to promote non-motorized travel and design bicycle-friendly environments.

Keywords: Cycling distance, cycling environment, land use, perception, non-linearity

Article history:
Received: October 9, 2023
Received in revised form: December 28, 2023
Accepted: January 11, 2024
Available online: March 22, 2024

https://doi.org/10.5198/jtlu.2024.2434

The *Journal of Transport and Land Use* is the official journal of the World Society for Transport and Land Use (WSTLUR) and is published and sponsored by the University of Minnesota Center for Transportation Studies.

1 Introduction

Bicycles offer an eco-friendly and cost-efficient means of transportation, not only easing traffic congestion but also fostering public health. They are especially preferred for short trips and feeder transport services, as emphasized by Ji et al. (2022), Kraus and Koch (2021), and Qu et al. (2022) The distance covered while cycling, identified as a key factor in predicting cycling behavior (Buchel et al., 2022; Ji et al., 2022), has garnered significant attention. Numerous studies suggest that longer cycling distances are more likely to substitute motor vehicle trips and, due to effective integration with public transport, can successfully redirect short car trips towards public transportation (Ji et al., 2022). This transition directly reduces car dependence, contributing to the alleviation of urban congestion and air pollution (Ji et al., 2022; Kraus & Koch, 2021; Ospina et al., 2020; Qu et al., 2022; Zhao et al., 2020). Hence, it is crucial to comprehend the influencing factors and their impact on the traveled distance.

Mainstream literature indicates that objective attributes like land-use elements and street design in the cycling environment significantly impact riding distance (Chen et al., 2022; Feng & Li, 2016; Ferenchak & Marshall, 2021; Kim et al., 2012; Ospina et al., 2020; Wang et al., 2018; Wei & Zhu, 2023). However, most studies predicting riding distance typically concentrate solely on objective cycling environments. Rarely do they investigate the influence of cycling environments on riding distance from both objective and perceptual dimensions (Guo & He, 2021; Zhang et al., 2023). Perceived variables, such as comfort, perceived safety and perceived accessibility, are explanatory factors distinct from physical attributes and have a more direct impact in influencing riding distance (Echiburu et al., 2021; Guo & He, 2021).

Moreover, the impact mechanisms of cycling environments on riding distance have not been sufficiently addressed in modeling techniques. Current modeling approaches, such as multiple linear regression (MLR) and geographically weighted regression (GWR), presume a linear relationship between explanatory variables and the response variable. However, the threshold effects produced by elements in the cycling environment suggest that the linear assumption may inherently compromise some explanatory power (Chen, 2022). In summary, for a more profound understanding of the mechanisms through which cycling environments exert influence, there is an urgent need to investigate the nonlinear and threshold effects of both objective and perceptual cycling environments on riding distance.

To unravel the underlying mechanisms between cycling environments and riding distance across both objective and perceptual dimensions, more intriguing questions should be posed. For instance: 1)What is the relative importance of objective and perceptual dimension cycling environment variables in predicting riding distance? 2)Do cycling environment variables within objective and perceptual dimensions exhibit nonlinear relationships with riding distance? 3)What policy recommendations would the potential mechanisms between cycling environment and cycling distance provide for improving cycling distance and shaping a friendly cycling environment?

To address these questions, our study employed extensive bike-sharing data gathered in the primary urban area of Shenzhen, China. We introduced a modeling framework based on eXtreme Gradient Boosting (XGBoost), a high-performance gradient boosting decision tree (GBDT) algorithm, to elucidate the nonlinear relationships between cycling environments in both objective and perceptual dimensions and bike travel distances. These findings provide a cohesive plan to actively promote the development of urban cycling.

The contributions of this paper are as follows:

- It serves as another demonstration of how urban transportation can effectively integrate with nonlinear modeling techniques to address urban-related issues. Non-linear modelling techniques from machine learning are applied to reveal the non-linear relationship between cycling environment and cycling distance, with particular emphasis on the threshold effect of the perceived cycling environment.
- From the existing literature, this is the first paper that simultaneously uses land use (including transportation land, residential land, commercial land, recreational land, mixed land use, etc.) and subjective perceptual measures (including comfort, perceived safety, and perceived accessibility) to predict urban bike travel distance.

The rest of the paper is organized as follows. Section 2 provides a literature review of the variables affecting cycling distance, Section 3 describes the study data and the XGBoost algorithm, Section 4 presents the calculation results and discusses them, Section 5 answers the questions in the introduction and summarizes the findings.

2 Literature review

In this section, we conducted a literature review to explore the influence of both objective and perceived environmental factors on bicycle distance. Additionally, we addressed the necessity for research into the nonlinear relationship between the subjective and objective dimensions of the cycling environment and its impact on cycling distance. Concerning objective factors, there is a widespread belief that socio-economic indicators are closely linked to cycling distance (Feng & Li, 2016). However, in special situations such as the pandemic period, this correlation may not be significant (Schaefer et al., 2021). Social demographic indicators and population density also correlate with cycling distance. For instance, personal income is positively correlated with daily cycling distance (Nielsen et al., 2013). Li and Xu (2022) observed a significant decrease in cycling distance in densely populated areas, attributed by Cervero and Duncan (2003) to safety issues and collision risks in such areas. Land use emerges as a key determinant in predicting cycling distance (Zhao et al., 2020). Zhao et al. (2020) found that commercial and residential land use, along with road network density, were negatively correlated with cycling distance. Conversely, leisure facilities, especially those with aquatic features, showed a positive correlation with cycling distance. Kim and Lee (2023) and Jiao et al. (2022) similarly concluded that leisure facilities, like parks, can increase cycling distance, while areas with higher land-use intensity and building density result in decreased cycling distance. Similar findings were reported by Kabak et al. (2018) and Wang et al. (2018), suggesting that parks, convenience service areas, transport hubs, and residential areas all contribute to extended cycling distances. However, Ji et al. (2022) demonstrated that land-use patterns may not be as influential as expected and incorporated topological measurements of the road network in his calculations and found that road network patterns played a more significant role in predicting cycling distances. Furthermore, while most studies indicate that land diversity can significantly reduce cycling distance (Cervero & Duncan, 2003), Ospina et al. (2020) arrived at an opposing conclusion. The cycling environment variables discussed above, encompassing social attributes and land use, showcase inconsistent effects on cycling distance. Ji et al. (2022) investigated the link between land use and cycling distance using XGBoost and SHAP and highlighted the existence of a non-linear relationship between the cycling environment and cycling distance. The non-linear results suggest that cycling distance is only affected when the explanatory variables reach a certain level. Therefore, this may be an important factor contributing to the inconsistent results of the above studies. Prior studies may have focused on specific ranges of explanatory variables, either before or after the thresholds (Cheng et al., 2022; Galster, 2018; Zhuang et al., 2022). The general understanding of the non-linear relationship between objective variables of the cycling environment, such as land use, and cycling distance remains unresolved. Therefore, in terms of objective dimensions, this paper draws on the research by Ji et al. (2022) and uses the XGBoost algorithm to further explore the non-linear relationship between the cycling environment and cycling distance.

The subjective perception dimension of the cycling environment focuses on individuals' cognitive and emotional responses to the built environment, including buildings, public spaces, and urban landscapes (Guo & He, 2021; Kerr et al., 2016; Kim & Lee, 2023). As an intrinsic determinant, perception has a more profound effect on cycling distance than objective variables (Banerjee et al., 2022; Echiburu et al., 2021; Guo & He, 2021). Chen et al. (2022) discovered that cyclists' perceptions of the quality of the cycling environment were more likely to lead to positive cycling intentions, encouraging them to explore a wider range of destinations and thereby increase the distance traveled. Specifically, Guo and He (2021) emphasized the more direct impact of perceived accessibility on cycling use, and Chen et al. (2018) explored the impact of bicycle collisions on cycling use in urban cycling environments. Kerr et al. (2016) demonstrated a positive correlation between perceived land diversity and cycling. Both Han et al. (2018) and Wang et al. (2018) mentioned that comfort plays an important role in influencing cycling behavior. The impression given by green spaces and leisure facilities has also been suggested to promote cycling trips (Porter et al., 2020). Kim & Lee (2023), Williams et al. (2022), and Yasir et al. (2022) also highlight the influence of subjective perceptual factors (e.g., opinions, attitudes, etc.) on cycling. However, few existing studies have explored the influence of perceptual factors on distance ridden. This gap has also been noted in studies by Guo and He (2021) and Winters et al. (2010). Limited studies include Chataway et al. (2014), who identified perceived safety as a factor influencing riding distance, and Banerjee et al. (2022), who emphasized the role of the perceived riding experience in prolonging riding distance. Furthermore, Golledge (1997) highlighted residents' interest in perceived thresholds during movement from a spatial behavioral perspective. This suggests that subjective perceptions of the riding environment may also have a threshold effect on riding distance. For example, Fitch et al. (2022) found that comfort only came into play when the riding environment was at a high level of comfort. This suggests that metrics related to the comfort of the cycling environment, such as quality of service, greenness, and landscape aesthetics, do not have a significant effect on cycle use when they are not at a high level. However, it remains unclear whether these perceptual indicators have a threshold effect on riding distance. Therefore, the non-linear relationship between the subjective perceptual dimensions of the cycling environment and its effect on cycling distance needs to be further verified.

3 Methodology

3.1 Study context and data

3.1.1 Study context

The city of Shenzhen, in the south of Guangdong Province, is a special economic zone in China. The city has a total area of 1997.47 square kilometers and a resident population of 17,681,600. The city has formed seven hotspot riding areas, including Futian, Luohu, Nanshan, Bao'an Centre, Longhua Centre, Fuhai Street and along Longgang Avenue. By

2021, the city's registered users of shared riding reached 27.7 million, with an average daily riding volume of about 1.38 million, which better meets the public's "last kilometer" connectivity and short-distance travel needs and plays a positive role in restoring travel during the pandemic, easing urban traffic congestion, and building a green travel system.

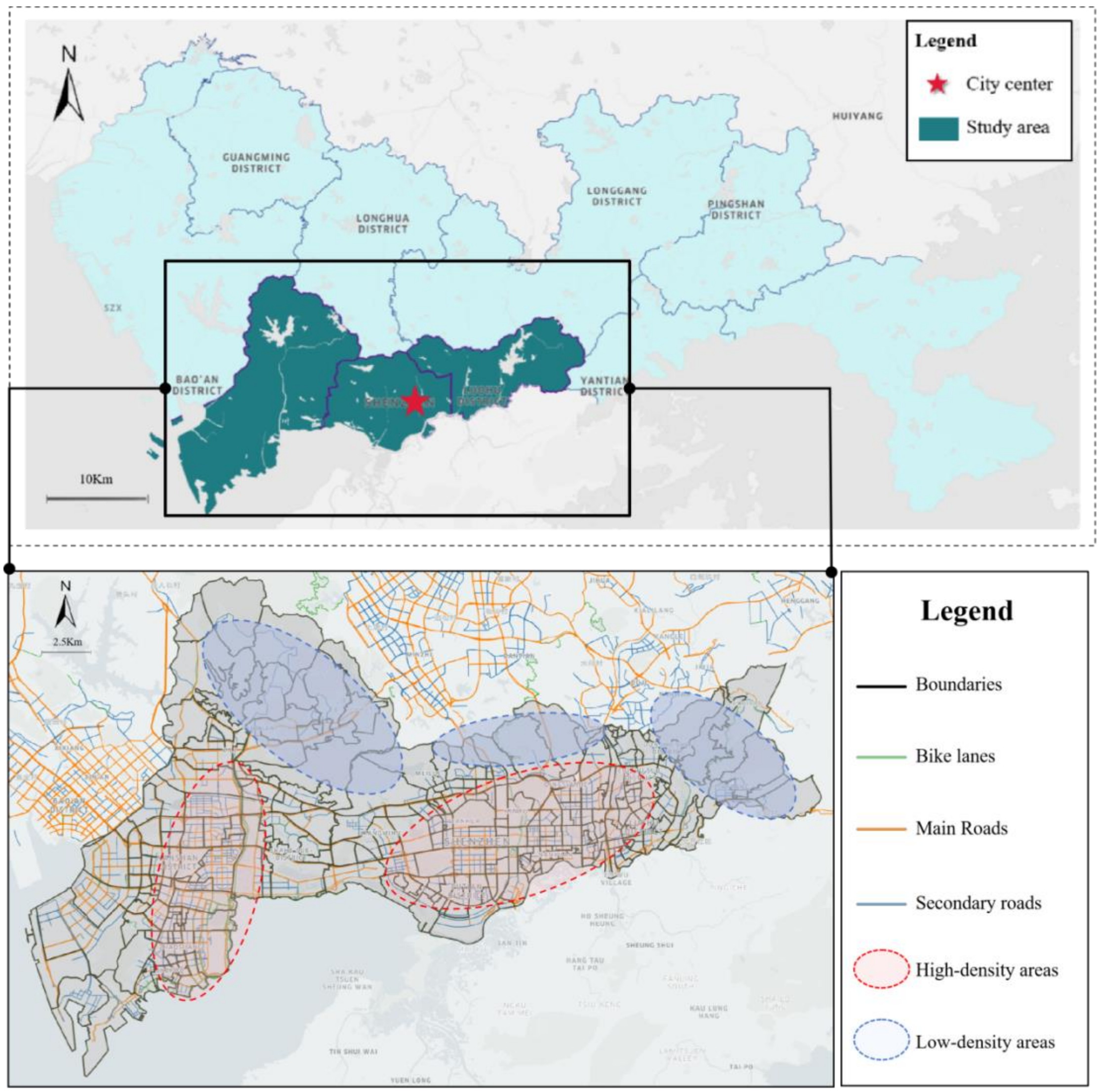


Figure 1. Study scope and distribution of cycling facilities

Also, Shenzhen possesses distinct administrative boundaries that demarcate its urban areas, including Nanshan, Luohu, and Futian districts, from its suburban regions (i.e., other areas). Among these districts, Nanshan District stands out as a high-tech industrial hub in Shenzhen, hosting numerous science and technology enterprises and research and development institutions. The predominant land use in this area is technological, research and development, and residential, with a proportion of cultural, leisure, and public facilities. The cycling environment is generally favorable, featuring extensive cycle paths and minimal traffic congestion. Nanshan District also offers numerous parks and leisure facilities, enhancing the appeal of cycling and leisure activities. Futian District serves as the administrative center of Shenzhen, housing a substantial number of government agencies and business offices. The district boasts various residential areas, including high-end residential projects. The land use in Futian District is primarily business, office, and residential, with cultural, leisure, and public facilities also present. The area is well-suited for cycling, characterized by extensive cycle lanes and relatively low traffic volumes. Luohu District, one of the earliest urban areas in Shenzhen, holds a rich historical and cultural heritage. With early development, the district thrives in commercial and tourism industries, notably in places like Dongmen Commercial Pedestrian Street. Additionally, Luohu District features residential neighborhoods and industrial parks. The land use in this area includes commercial, residential, and industrial elements, along with cultural, educational, and public facilities. Overall, these main urban areas exhibit higher population densities and economic activity levels, setting them apart significantly from the suburbs in terms of transportation demand and employment opportunities. Studying urban areas helps to elucidate the intrinsic relationship between cycling distances and population and economic activity. Urban areas typically have well-developed transport networks, including subways, buses, and roads. The diverse land-use patterns in urban regions may impact cycling distance, and studying urban areas aids in uncovering this relationship. Moreover, Shenzhen's urban and suburban areas differ in topography, the built environment, and cycling distance. Considering both urban and suburban areas together may result in the observation of numerous outliers, complicating the identification of the intrinsic pattern between the cycling environment and cycling distance. Given these considerations, this paper focuses on the main urban areas of Shenzhen, namely Nanshan District, Luohu District, and Futian District, as the study area. The scope of the study and the results of the data visualization are shown in Fig. 1.

In the primary urban areas of Shenzhen, specifically, the Luohu, Futian, and Nanshan districts, bicycle usage extends beyond dedicated bicycle lanes to encompass main roads, secondary roads, and other thoroughfares. As illustrated in Figure 1, bicycle lanes are represented by green lines, main roads are depicted by earthy yellow lines, and secondary roads are denoted by blue lines. The coverage of these cycle lanes, which includes roads where cycling activities occur (encompassing not only dedicated cycle lanes but also main and secondary roads), is relatively extensive. In the primary urban areas, the cycle lane network essentially spans most residential areas and commercial centers. Although cycle lanes are present in all districts, their precise length and density may vary. For instance, there are two high-density cycle lane networks in the central south of Nanshan District, the eastern side of Futian District. These areas are notably rich in transport, commercial, and leisure facilities, serving as commercial centers and densely populated regions in the primary urban areas of Shenzhen. Conversely, in some relatively new residential and industrial areas and large parks, the construction of cycle lanes may be relatively scarce, such as in the northern part of Nanshan and Futian districts and the eastern side of Luohu District. It's clear that while bike lanes are relatively well established in Shenzhen's major urban areas, they don't fully cover the entire city.

3.1.2 Data sources and variables selection

The unit of analysis for this study was the strip buffer obtained from each cycling route. Calculations based on Ji et al. (2022) for cycling data, a buffer with a radius of 400 meters was established based on the cycling routes. The choice of a 400-meter radius aligns with common practice in defining a threshold for users' perception of their surroundings and has been used in several empirical studies (Ma et al., 2014). The Geopandas library in Python was employed to construct the buffer, capturing the environmental elements surrounding each route. The unit sample contains fundamental information, including the dependent variable, i.e., the cycling distance derived from the riding trajectory. Additionally, it includes a set of independent variables encompassing both objective and perceived elements of the cycling environment. These variables consist of population density, road network density, land mix entropy, the share of transport facilities, the share of commercial facilities, the share of residential facilities, the share of leisure facilities, perceived accessibility, perceived safety, and comfort. Below are the data sources and detailed descriptions of the variables used in the study.

1) Riding data. This paper uses high-quality cycling data collected by the bike-sharing company Meituan from 1 January to 7 January 2019. The cycling data consisted of 7 million trajectory data points per day. The average click frequency of the trajectory data is 5 seconds. These trajectory data can accurately reflect the riding trajectory and are suitable for this study. In this paper, a total of 190,000 data points were collated by restricting the start position, end position, and tracking point position of the cycling track data to the scope of the study, resulting in 514 cycling tracks, and the field names of the cycling trajectory data include user ID, bicycle ID, unlock time, lock time, unlock latitude and longitude, lock latitude and longitude, and tracking point latitude, longitude, and tracking time.

The dependent variable in our study is the cycling distance, obtained from cycling trajectory data using Geographic Information System (GIS) tools. Given the significant characteristic of detouring behavior in bicycle use, we use trajectory data for calculating the actual riding distance, as opposed to Euclidean or Manhattan distances, providing a more realistic reflection of bicycle use. In this article, a ride trip represents the complete riding process from the starting point to the destination. The data points included in a trip include the latitude and longitude and time stamp of the starting and ending points and the tracking points. Out of 190,000 data points, we identified and screened 514 real trajectories within the study area. Several pre-processing methods for the trajectory data were employed, including the exclusion of trajectories with start or arrival points outside the study area. Additionally, we excluded trips with ride durations of less than 30 seconds or more than 1 hour, as well as trips shorter than 100 meters or longer than 4 kilometers, following the suggestions of Shen et al. (2018). It is essential to note that in this study, we did not consider multi-destination or multi-purpose riding trajectories. Due to the lack of information about mobile phone riding users and mobile phone mobility data, we faced challenges in targeting or identifying the purpose of riding. Although some methods, such as detecting the duration of the bicycle’s stay near Points of Interest (POI), can be used to identify the purpose of riding in a rough manner, this coarse-grained destination identification may not be accurate.

2) Socio-demographic data. Population density is found to have a strong correlation with cycling distance and is a key variable in predicting cycling distance. The population data involved in this paper are mainly derived from the data of the 6th National Census of China. The data type is grid data.

We derived the independent variable, population density, from socio-demographic data. As our unit sample is a strip buffer, we use spatial connectivity in Geopandas to link the grid data information to the unit sample. In cases where the unit sample covers more than one grid, we calculated the average population density by incorporating all the relevant grids as the population density of the unit sample.

3) Land-use data. The data is mainly POI data crawled from Amaps, a dataset containing 17 categories, as shown in Table 1.

We obtained two independent variables from the land-use data, the percentage of facilities on each type of land, and the land mix entropy. For the ratio of facilities on each type of land use, recognizing that overuse of less informative feature variables in the computation process can hinder model fitting, especially with limited samples, and that higher feature dimensions can exacerbate model overfitting, we took steps to screen and filter the 17 types of POIs. A mapping approach is shown in Table 1, and similarly, we used the geopandas library to perform spatial set counting on the POI data, based on which we obtained four key explanatory variables, i.e., the share of transport, the share of residential, the share of commercial, and the share of leisure, where the share of each type of land-use facility is the ratio of the corresponding number of POIs to the total number of POIs in each sample.

Land-use mix entropy is an indicator of land diversity, with higher values indicating a richer urban function, and is calculated as follows: Equation 1, where p_i represents the areal percentage of the i th pattern of land use; and n is the total number of land-use patterns.

$$Q_{entropy} = -(\sum_{i=1}^{n} p_i * \ln(p_i))/\ln(n) \tag{1}$$

Table 1. The POI categories

Code	POI Category	Category assigned	Code	POI Category	Category assigned
1	Car Service	Transport	2	Car Sales	Transport
3	Car Repair	Transport	4	Motorcycle Service	Transport
5	Restaurant	Leisure	6	Shopping	Commercial
7	Daily Life	-	8	Sports	Leisure
9	Hospital Related	-	10	Hotel Related	Leisure
11	Tourist Attraction	Leisure	12	Residence	Residence
13	Governmental Organization	-	14	Education Related	-
15	Transport Related	Transport	16	Finance Service	Commercial
17	Enterprises	Commercial			

4) Street data. Street design mainly includes the density of the road network, which comes from the Open Street Map（OSM), and mainly includes the main roads, secondary roads, bike lanes and side roads of all administrative districts in Shenzhen.

We obtained the independent variable, road network density, from street data. The density of the road network is expressed by the ratio of the length of the road route to the area of the study unit, the strip buffer.

5) Questionnaire data. We conducted an on-site questionnaire survey from July 1st to August 30th, 2019. To mitigate the influence of weather on survey results, we excluded days with extreme weather conditions and selected days conducive to cycling for the survey. We recruited 15 students from our research group to collectively distribute the questionnaires, aiming to gather opinions from the users of the Meituan Bike Sharing System about the cycling environment along 514 bicycle lanes.

We use simple random sampling to conduct surveys within the survey area of each route to ensure that each Meituan BikeShare user has an equal opportunity to participate in the questionnaire survey. This helps reduce sample selection bias and increases the representativeness of the respondents. It is worth noting that due to the existence of different brands of bike users, such as Qingju Bicycle and Mobike, the respondents selected after simple random sampling may not necessarily be Meituan BikeShare users. To avoid this bias, we confirm whether the respondents are Meituan BikeShare users by asking “Have you used Meituan BikeShare to ride on this route?” If not, we consider replacing the sample until the respondents of the randomly sampled samples are all

Meituan BikeShare users. Also, each participant was offered a prize valued at 3.5 RMB to incentivize participation. Surveyors interviewed cycling users along the riding routes at approximately every 300 meters. In total, we collected 3,013 questionnaires, out of which 2,471 were deemed valid.

Table 2. Survey statistics on socio-demographic and perceived measures of cycling environments

Variable	Category and code	Percentage/ Average
Social-demographic characteristics (N=2471)		
Gender	Female = 0	46.63%
	Male = 1	53.37%
Age	Under 25 = 1	31.43%
	26–35 = 2	53.24%
	36–45 = 3	10.39%
	Over 46 = 4	4.94%
Income (monthly)	<4,999 = 1	11.09%
	5,000–9,999 = 2	45.27%
	10,000–14,999 = 3	22.67%
	>15,000 = 4	20.97%
Perceptual measures		
Perceived accessibility	Q1: Do you find it easier to reach your destination when you pass this route?	59.1
	Q2: Do you find it easier to reach nearby bus and metro stations when you pass this route?	54.8
Perceived safety	Q3: Do you feel at risk of collision when travelling through this route?	73.3
	Q4: Do you think you feel safe riding on this route to score perceived safety?	70.6
Comfort	Q5: When you pass this route, you feel relaxed and happy.	63.2
	Q6: When you pass this route, you feel comfortable.	71.4

The questionnaire was divided into two parts, as indicated in Table 2. The first part required respondents to provide basic information such as gender and age. We observed a higher proportion of male respondents (53.37%) compared to female respondents (46.63%). Additionally, the majority of participants were young, with 84.67% of them being under the age of 35. Furthermore, a significant percentage of respondents fell into the middle-income bracket, with 67.94% reporting earnings between 5,000 RMB and 14,999 RMB. The data we surveyed is roughly consistent with the information of Meituan bike-sharing users surveyed in the Report on Sharing Bicycles and Motorcycles in Major Cities of China released by the China Internet Data Consulting website (https://www.199it.com/). For example, the report shows that male users account for 51.6%, slightly higher than female users, and more than 80% of users are younger than 25 years old and 26 to 35 years old. More information and details are not described in this article due to space limitations in the above report.

We can obtain social statistical data of the surveyed users from the results of the questionnaire survey, as well as independent variables related to the perception dimensions of the cycling environment, including perceived accessibility, perceived safety, and comfort. We used the concept of perceived accessibility to measure the perceived ease with which a cyclist can reach a destination within the area covered by the cycling track. We selected perceived accessibility, perceived safety, and comfort as perceived variables of the riding environment. Specifically. We did this by asking Question 1: Do you find it easier to reach your destination when you pass this route? and Question 2: Do you find it easier to reach nearby bus and metro stations when you pass this route? to assess users' perceived accessibility to the cycling environment of the corresponding cycling routes, these questions were adapted from Guo and He (2021) and Scheepers et al. (2016). Similarly, we assessed the perceived accessibility of the route by asking Question 3: Do you feel at risk of collision when travelling through this route? and question 4: Do you think you feel safe riding on this route to score perceived safety, questions about safety we drew on research by Chan et al. (2019), Guo et al. (2023) and Keppner et al. (2023). Finally, we scored perceived comfort by asking Question 5: When you pass this route, you feel relaxed and happy; and Question 6: When you pass this route, you feel comfortable. The design of the question on comfort was provided by the study of Fitch et al. (2022). Respondents were asked to rate each question out of 100 and the scores obtained were finally normalized. Preliminary statistical analyses of the survey data indicate that cycling users generally do not perceive the perceived accessibility of the cycling environment in which they live to be good (mean rating of 0.591 for question 1 and 0.548 for question 2). On the other hand, surveyed cyclists expressed relatively positive perceptions of the comfort and perceived safety of the cycling environment (mean score of 0.733 for question 3, 0.706 for question 4, 0.632 for question 5, and 0.714 for question 6). We did not choose to use the 5-scale and 7-scale Likert scales because this paper favors continuous results over discrete ratings, which helps to explain the non-linear results.

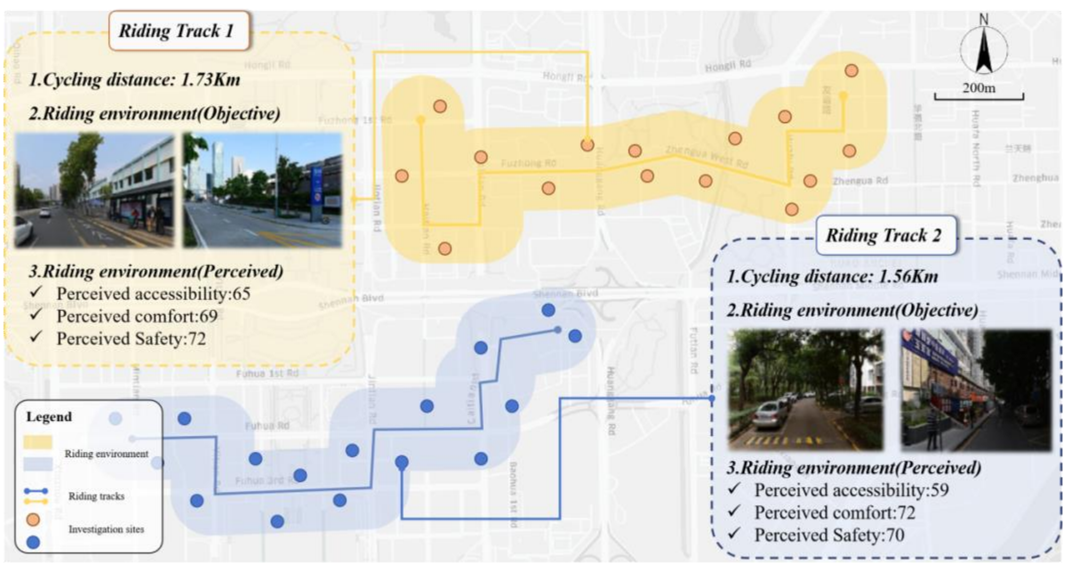


Figure 2. Schematic of sample data with buffer based on riding trajectory

Overall, we obtained a study unit containing both dependent and independent variables as shown in Figure 2 (this is a schematic diagram; the extent of the buffer in the figure does not represent the actual extent). Both objective and perceptual dimensions of cycling environment indicators were covered on the buffer of cycling trajectories, and descriptive statistics for each variable are presented in Table 3.

Table 3. Statistics on input data for model calculations

	Definition	Mean	Std
Dependent variable			
Cycling distance	Distance travelled on the ride (m)	1255.231	291.729
Independent variables			
Population density	Ratio of population size to area within the study unit (thousand people /km^2)	18.050	9.130
Road Network Density	Ratio of road length to area within the study unit (km/km^2)	4.050	3.620
Land-use mix entropy	An entropy index of four land uses: Land-use mixing entropy is an indicator of land diversity, and a higher entropy value means a richer urban function.	0.540	0.110
Transport facility	Ratio of the number of POIs related to transport facilities to the total number of POIs.	0.230	0.078
Leisure facility	Ratio of the number of POIs related to leisure facilities to the total number of POIs.	0.074	0.041
Residential facility	Ratio of the number of POIs related to residential facilities to the total number of POIs.	0.186	0.061
Commercial facility	Ratio of the number of POIs related to commercial facilities to the total number of POIs.	0.308	0.041
Perceived accessibility	Perceived accessibility of the destination score (from 0 to 1).	0.570	0.041
Perceived safety	Riding route perceived safety score (from 0 to 1).	0.720	0.069
Comfort	Riding route comfort score (from 0 to 1).	0.673	0.088

3.2 Nonlinear model

Most of the previous literature used global models such as multiple linear regression (MLR), geographically weighted regression (GWR) and their derived regression models such as multi-scale GWR (MGWR) to determine the effects of various explanatory variables on cycling behavior (Alcorn & Jiao, 2019; Shen et al., 2018; Zhang et al., 2017). Alcorn and Jiao (2019) and Zhang et al. (2017) used MLR to investigate the effect of the built environment on station-level cycling distance, and Shen et al. (2018) used spatial autoregressive modelling to investigate shared cycling distance considering spatial dependence. However, recent studies have noted a non-linear relationship between environmental attributes and cycling distance (Ji et al., 2022). The agglomeration effect and diminishing returns effect in urban economics help to demonstrate potential non-linear effects in the environment (Galster, 2018). Empirical studies have also increasingly recognized the importance of identifying and understanding the non-linear effects of the built environment on urban mobility. Cheng et al. (2022) combined bike-sharing with urban rail travel modes using quantile regression to reveal a strong link between the built environment and last-mile travel. It is noteworthy that Ji et al. (2022) adopted an efficient nonlinear regression model and machine learning model to predict the distance of cycling, namely the XGBoost algorithm (Chen & Guestrin, 2016).

Specifically, XGBoost is an improvement and extension of GBDT. According to Elith et al. (2008), GBDT is a family of ensemble models based on decision trees that incorporates multiple decision trees and gradient boosting methods. Each decision tree in GBDT learns from previous trees and has an impact on them, thus reducing the loss of the model and ultimately providing an optimal function. Based on the structure of GBDT, XGBoost was proposed (Chen & Gerstrin, 2016) on the basis of optimizing the objective function. This algorithm is superior to traditional generalized linear regression in several aspects. First, it provides better fitting than traditional models. Second, it can handle various data types, such as continuous variables and categorical variables. Third, it can flexibly handle missing data and is not affected by outliers. Fourth, it helps to solve multi-collinearity problems. More importantly, XGBoost can explain the nonlinear correlation and other irregular associations between variables. XGBoost method also has some disadvantages. For example, like other machine learning techniques, the XGBoost method cannot perform significance testing or provide coefficients and confidence intervals for independent variables. However, XGBoost can give the relative importance of independent variables, which reflects their impact on the dependent variable. The commonly used partial correlation plot (PDP) can visualize the relationship between independent variables and dependent variables. All these shortcomings can be corrected. Therefore, this study uses the XGBoost model to determine the relative importance and non-linear association of land-use elements and perceived elements to cycling distance and employs Partial Dependency Plots (PDP) for interpretation.

The processed dataset including both independent x_i and dependent variables y_i, with 10 characteristics per sample, including population density, road network density, land-use mix entropy, et al. The tree integration model uses k additive functions to estimate the target value ($\hat{y}$), as shown in Equation 2.

$$\hat{y} = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i),\ f_k \in F \tag{2}$$

Where y is the dependent variable (regression value, indicating the distance ridden), x_i is the independent variable, k is the number of number functions, f_k is the independent tree structure and F is the tree space.

To minimize the objective function, the function $\mathrm{L}(\phi)$ is constructed and its mathematical expression is shown in Equation 3.

$$\mathrm{L}(\phi) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \tag{3}$$

Where $\sum_i l(y_i, \hat{y}_i)$ represents the sum of the loss values for i samples and $\sum_k \Omega(f_k)$ represents the sum of the complexity for k trees.

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda w_i^2 \tag{4}$$

For the complexity Ω, we give a more detailed formula as shown in Equation 4, where T is the number of leaves and w_i is the weight of leaf i. The optimal weight w_i^* is calculated as shown in Equation 5, and the corresponding optimal values are estimated from Equation 5 to Equation 7. The biggest difference with GBDT is that XGBoost optimizes the objective function using Taylor's second-order expansion, which approximates the value of the objective function more accurately and accelerates convergence to better handle non-convex problems.

In calculating the optimal value of w_i by computing w_i^*, g_i and h_i are used to denote the first-order gradient and second-order gradient statistics of the i-th sample at the t-1 th iteration in terms of the brought true value y_i and the predicted value at the t-1st iteration $\hat{y}^{t-1}$, respectively.

$$w_i^* = -\frac{\sum_{i \in I_j} \partial_{\hat{y}}^{t-1} l\left(y_i, \hat{y}^{t-1}\right)}{\sum_{i \in I_j} \partial_{\hat{y}^{t-1}}^2 l\left(y_i, \hat{y}^{t-1}\right) + \lambda} \tag{5}$$

$$g_i = \partial_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}^{t-1}\right) \tag{6}$$

$$h_i = \partial_{\hat{y}^{(t-1)}}^2 l\left(y_i, \hat{y}^{t-1}\right) \tag{7}$$

This section provides a brief description of the optimization of the objective function of the XGBoost algorithm, which is the most important key point for understanding XGBoost, and for space limitation, one can refer to Chen and Guestrin (2016) for more technical details of XGBoost.

4 Results and discussion

XGBoost and the Sklearn package in Python are used to develop the model. We use learning curve and grid search methods to find the combination of hyperparameters in the XGBoost algorithm that guarantees the best parameters fit. Five-fold cross-validation was used to limit overfitting and reduce generalization error, with 20% of the data sample used as the test set and other parameters chosen as default. The final iterative results show that the cardinality fit of the model can reach 0.694 when n_estimators is 76, subsample is 0.95, learning rate is 0.3, maximum depth is 6, alpha is 0.4, Lambda is 0.7 and Gamma is 0.9. As can be seen from Table 4, the inclusion of perceptual elements increases the explanatory power of the model and the XGBoost model fits better compared to the linear model. Next, we interpret the relative importance of the features and the non-linear relationship between the features and the explanatory variables as a way of revealing the effect of the cycling environment on cycling distance.

Table 4. Comparison of the XGboost model and basic linear regression model

Model	Metrics	XGBoost	Linear
Models incorporating perceptual elements	R^2	0.694	0.327
	MAE	0.113	0.221
	MSE	0.024	0.063
Models non-incorporating perceptual elements	R^2	0.432	0.213
	MAE	0.194	0.295
	MSE	0.073	0.092

4.1 Relative importance

Relative importance is one of the tools used by machine learning models to explain the predictive contribution of feature variables, it measures the value and contribution of features in the construction of boosted decision trees in the model. The more an attribute is used to construct a decision tree in a model, the relatively higher its relative importance is. Specifically, the larger an attribute's performance measure for split-point improvement (the closer it is to the root node), the larger the weight, meaning that it is selected by more boosting trees, and therefore, the higher the relative importance of the feature attribute. In this paper, if the relative importance of a feature attribute is higher, it means that this feature contributes substantially to the prediction of riding distance and is strongly correlated with riding distance.

As demonstrated in Table 5, we obtained the relative importance of all feature attributes. The relative importance of population density is the highest, at 24.8%, indicating that population density has the greatest contribution to the prediction of cycling distance and has a strong explanatory effect on cycling distance. This means that compared to other variables, cycling distance can achieve the most significant change when population density variables can undergo effective changes. This indicates that the priority indicator for extending cycling distance is population density. Meanwhile, road network density exerts a relative importance of about 15.2%. Leisure and leisure facilities emerge as crucial determinants, wielding relative importance of approximately 9.1%. This may mean that cycling through sports and leisure facilities is more enjoyable than passing through buildings or factories, which in turn has an obvious impact on cycling distance (Alcorn & Jiao., 2023; Fitch et al., 2022). Conversely, the influence of residential and commercial facilities on cycling journeys is less pronounced, suggesting that commercial centers in Shenzhen may not be conducive to cycling. In addition, we found no strong correlation between transport facilities and cycling distance. While previous studies have often emphasized the strong link between transport facilities and cycling distance (Ji et al., 2022), this discrepancy may stem from differences in the density of facilities within the study area, as compared to the main urban area of Shenzhen, which is close to saturation in terms of the share of transport facilities. The cumulative contribution of perceptual variables in our analysis amounts to 23.8%. This underscores the substantial explanatory power of perceived factors, particularly perceived accessibility and comfort, in predicting cycling distance. Overall, our findings suggest that both objective and perceptual variables exhibit strong relative importance (51.3% and 23.8%) in predicting cycling distance.

Table 5. Relative importance of explanatory variables

Variables	Rank	Relative importance
Population density	1	0.248
Road Network Density	2	0.152
Leisure facility	3	0.091
Comfort	4	0.086
Perceived Accessibility	5	0.080
Transport facility	6	0.077
Perceived Safety	7	0.072
Land-use mix entropy	8	0.070
Commercial facility	9	0.064
Residential facility	10	0.059

4.2 Non-linear relationships

The PDP plots indicate a non-linear relationship between the overwhelming majority of independent variables and the cycling distance. This observation enables us to examine intricate trends and pinpoint the threshold as well as the effective range of influence.

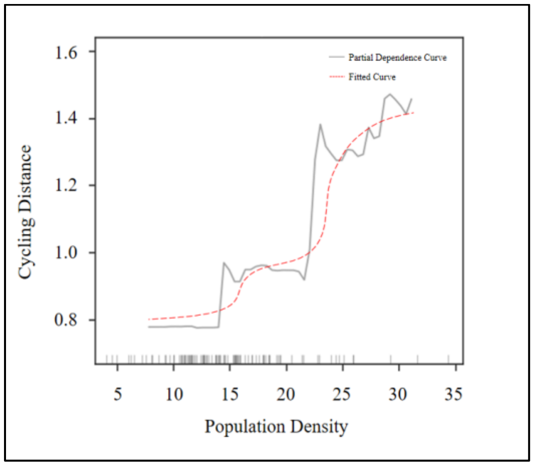


Figure 3. Non-linear relationship between population density and cycling distance

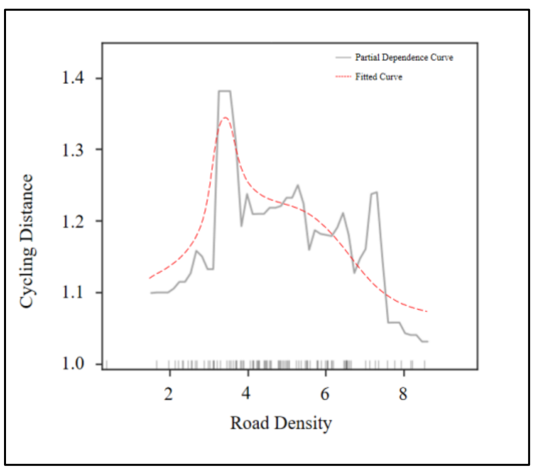


Figure 4. Non-linear relationship between road density and cycling distance

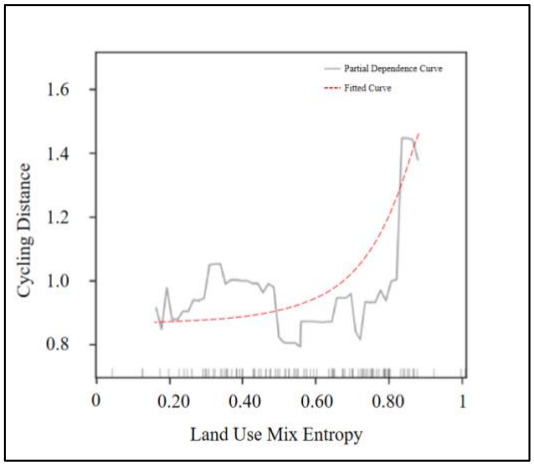


Figure 5. Non-linear relationship between land-use mix entropy and cycling distance

As shown in Figure 3, as expected, population density is positively correlated with cycling distance and has a strong threshold effect, which is consistent with the findings of Chen et al. (2021). Specifically, the relationship between population density and cycling distance in Shenzhen has a clear three-stage gradient, when the population density reaches 14,000/km^2, the cycling distance rises from 0.8km to about 1km, and when it exceeds 22,000/km^2, the cycling distance is able to reach 1.4km. This indicates that population density is not simply linearly and positively correlated with cycling distance. This suggests that when population densities are roughly 14,000 people/km^2 and 22,000 people/km^2, the overall average cycling distance in the city can be promoted. On the non-threshold plane, attempting to increase the average riding distance by increasing population density is not advisable, as non-linear relationships show that this increase in population density does not have a positive correlation with riding distance within these non-threshold ranges.

It is worth noting that although we found a threshold effect of population density on riding distance, the magnitude of change in riding distance was not as large. On the one hand, the reason for this comes from the data processing, as the sample sizes for longer riding distances are further away from the sample center, they may be considered discrete values and excluded, and the presentation of the PDP plot as a global visualization may also lose some of the longer riding distances; On the other hand, in sifting through the available data, we did not select a large study area and excluded some trips across the study area, which is an important reason for the small variation in riding distances in the results. However, this does not mean that the smaller change in cycling distance is meaningless; this result suggests that if the population density in the study area reaches 14k/km^2, the overall cycling distance in the city will theoretically increase by 200 meters. According to statistics, the daily cycling volume of shared bikes in Shenzhen is about 1.38 million. We also know that cycling can replace driving a car for 100km, thus reducing carbon dioxide emissions by about 20.1kg (Chen et al., 2020). Assuming that you only drive once a day, the increased distance is used to replace the distance travelled by car, which means a reduction of 554.8 tons of carbon emissions per day. This is a considerable figure and has a positive impact on environmental protection and sustainable urban transport development.

The non-linear relationship driven by road network density is in the shape of an inverted U as depicted in Figure 4. We discover a remarkable threshold effect at 3km/km^2

and 7km/km^2, a finding agreeing with Shao et al. (2022). However, these findings are inconsistent with Chen and Ye (2021), whose prediction of road network density depicts only the first half of an 'inverted U.' The results of this paper suggest that higher road network density does not have a complete positive correlation with cycling distance, which may be due to the fact that high road network densities increase the risk of riding collisions, and the accessibility generated by high road network densities is overshadowed by the risk that they create, which in turn reduces the cycling distance. As depicted in Figure 5, the non-linearity of land diversity is apparent. This is similar to the findings of Ding et al. (2019) and Zhao et al. (2018), where land diversity was found to have a substantial positive correlation and threshold effect. The positive threshold effect is pronounced at high levels of land-use mixing, around 0.8.

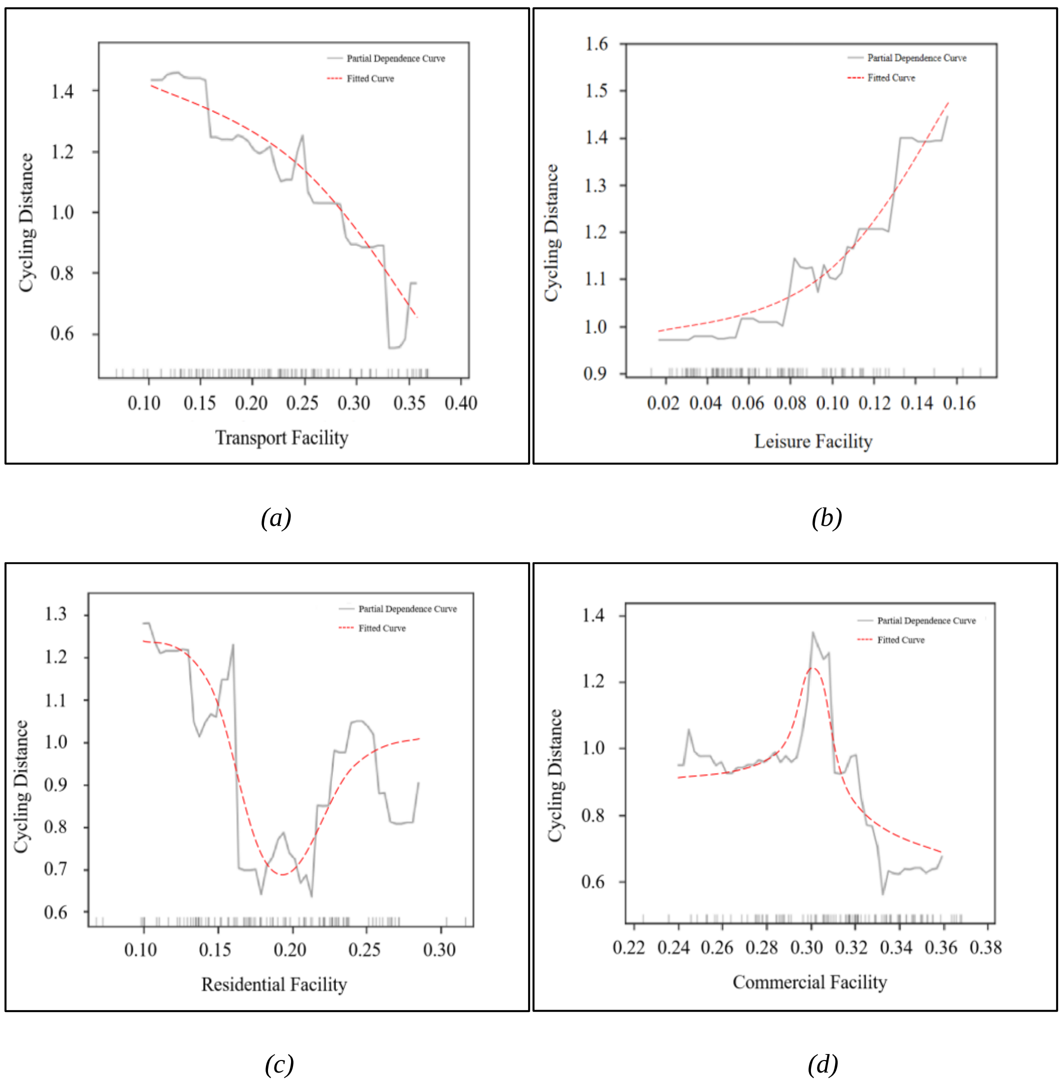


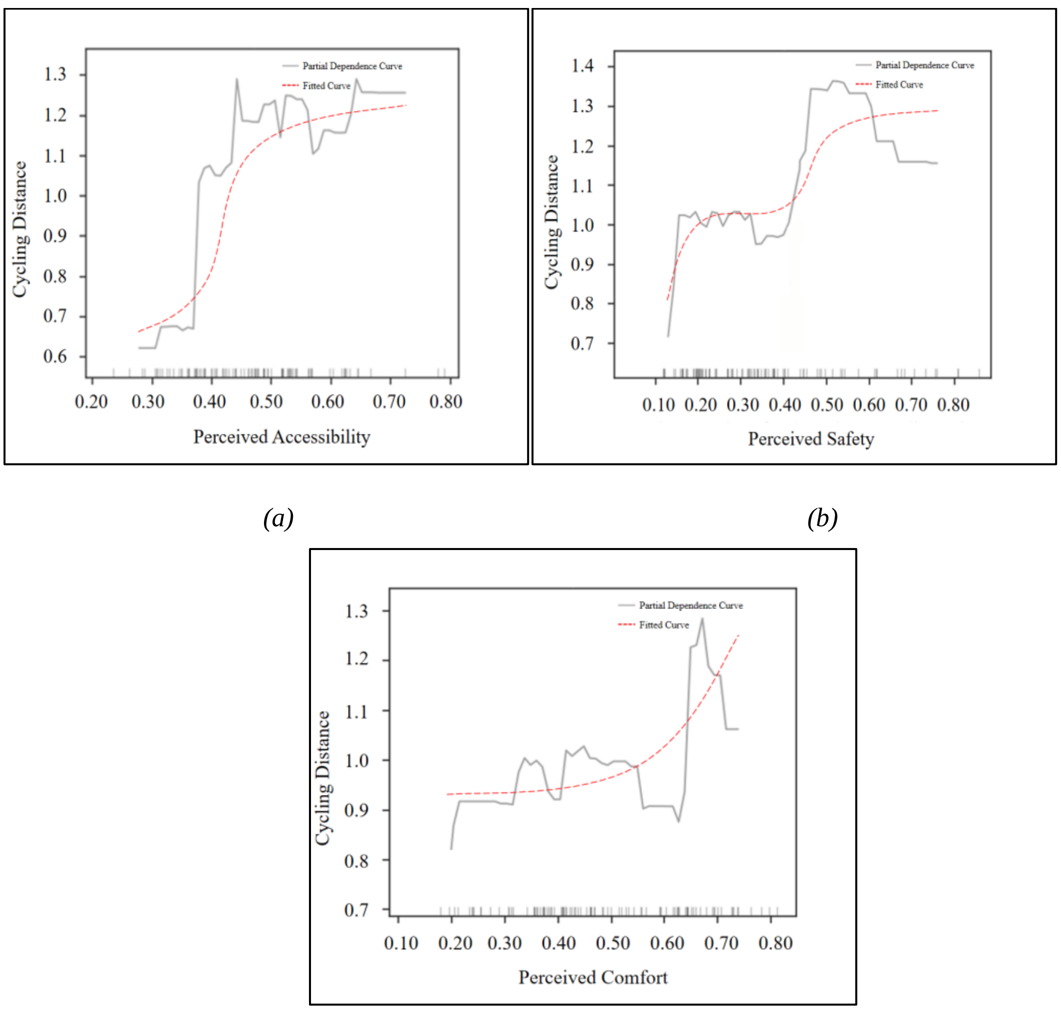
Figure 6. Non-linear effects of land uses on cycling distance

Figure 6(a) illustrates the non-linear impact of the transport facilities. As evidenced by previous studies (Chen & Ye, 2021; Wu, et al., 2021), a high percentage of land use for transport facilities contributes to cycling behavior. However, an interesting point was found in our results, where an increase in the share of transport facilities had a negative correlation with cycling distance. There is a remarkable threshold effect as the distance cycled gradually decreases from 1.4km to 0.7km when the transport land share is 0.15 and 0.33 respectively. This may be due to the high density of transport facilities in Shenzhen city district. Ding et al. (2019) suggest that the saturation of land-use density suppresses cycling behavior. This assertion was confirmed through empirical evidence in this article, revealing a negative correlation between a disproportionate share of transport facilities and cycling distance. Wu et al. (2021) also pointed out that the saturation of bicycle facilities does not consistently lead to increased bicycle usage. Although increasing the share of transportation facilities is often crucial to promoting the use of bicycles (Chen et al., 2021), the negative nonlinear relationship between the share of transportation facilities and the distance travelled by bicycles reminds land planners that too many transportation facilities may indeed inhibit the distance travelled by bicycles. The unreasonable growth of transportation facilities has compressed the share of other land uses (Ding et al., 2019). Leisure facilities, residential facilities, and commercial facilities are positively correlated with the distance travelled by bicycles to varying degrees. Therefore, due to the compression of these land-use shares, it is likely that the growth of the distance travelled by bicycles will be restricted again. This emphasizes the importance of a reasonable land-use structure. In addition, with the increase in the proportion of transportation facilities, the substitutability between bicycles and other modes of transportation may also increase. The increase in the proportion of transportation facilities represents opportunities to use corresponding transportation tools. Previous studies have emphasized the substitutability between bicycles and public transportation (Saltykova et al., 2022; Sun & Zacharias, 2017), especially for medium-distance and long-distance travel, which is also one of the reasons for the decrease in the distance travelled by bicycles.

As shown in Figure 6(b), there is a positive correlation between leisure facilities, with the proportion increasing from 0.02 to 0.16, and the cycling distance increasing by about 400 meters. The robust positive correlation between leisure facilities and cycling distance indicates that cycling users are more likely to extend their cycling distance near green spaces, parks, entertainment and other leisure facilities. Moreover, it exhibits a certain threshold effect, with a remarkable increase in cycling distance when the proportion reaches 0.08 and 0.13. It can be seen that although the average proportion of leisure facilities is not high, about 0.1, compared to transportation facilities, residential facilities, and commercial facilities, the proportion of leisure facilities is the only land use attribute that has a completely monotonic positive correlation with cycling distance. This emphasizes the key role of leisure facilities in increasing cycling distance, which has been mentioned by Kim and Lee, Wei and Xhu (2023), and (2023) Zhao et al. (2020).

The non-linear relationship between the residential share and cycling distance is quite obvious, with a negative relationship in a slightly “U” shape as illustrated in Figure 6(c). This negative correlation is mainly due to the higher share of dwellings occupying the traffic service area, thus reducing the opportunity to extend cycling distances, which is similar to that of Ding et al. (2019). Interestingly, the negative correlation reverses when the proportion of residential land use is greater than 0.21, a difference that may arise from the form of data samples for empirical analyses. Established studies commonly use traffic analysis zones, e.g. Chen and Ye, (2021), and Ding et al. (2019), but this form of the data sample is not conducive to extracting the influence of the cycling process by the environmental elements and lacks the detection of such finer relationship in the whole process of cycling. According to our observation, the coverage area of cycling trajectories has a higher proportion of residences near the starting point, in the case of a smaller proportion of residences, the residential facilities are more clustered at the starting location, which would lead to similar findings as Ding et al. (2019). However, as the proportion of residential facilities within the riding trajectory coverage area increases, residential facilities are not only clustered at the starting point but also pass by residential facilities in the area along the riding route. This suggests that cyclists are more likely to move within higher proportions of residential neighborhoods because areas with higher proportions of residential neighborhoods are safer, as also mentioned by Chen et al. (2018), Guo et al. (2023), and Marshall and Ferenchak (2019).

The percentage of commercial land use exhibited a strong non-linear relationship with cycling distance, and the first half of the obtained non-linear relationship is similar to previous studies, with Chen and Ye (2021) similarly emphasizing that cycling distance increases as the percentage of commercial land use increases. As shown in Figure 6(d), when the percentage of commercial land is at 0.23 to 0.30, there is a substantial positive correlation with cycling distance and a threshold effect when the percentage reaches 0.29. However, when the percentage of commercial land use reaches about 0.32, the riding distance decreases. This suggests that a higher proportion of commercial land use can limit the increase in cycling distances, as often areas with a higher proportion of commercial land use represent the end point of the ride, which can lead to an earlier termination of cycling activity.



(a) *(b)* *(c)*

Figure 7. Non-linear effects of perceived measures on cycling distance

Perceived accessibility, defined as the perceived potential for engaging in spatial dispersal opportunities by Pot et al. (2021), represents a psychological evaluation of one's ability to reach a specific destination. Existing research has demonstrated that perceived accessibility is highly correlated with bicycle use (Scheepers et al., 2016). However, previous studies have not extensively explored the specific levels of perceived accessibility that most effectively promote cycling distance (Samadzad et al., 2023). Our results found a positive correlation between perceived accessibility and riding distance and showed a threshold effect, as illustrated in Figure 7(a). An obvious threshold is when the perceived accessibility score reaches 0.4, the riding distance shows a strong positive correlation with perceived accessibility, with the riding distance increasing from 0.7km to 1.2km, and this positive correlation is no longer evident when the perceived accessibility is greater than 0.45, and a bottleneck stage is reached. This suggests that perceived accessibility yields substantial benefits in terms of extended cycling distances. It comes as no surprise that heightened perceived accessibility leads cyclists to believe it's easier to reach various destinations, including those farther away. Notably, however, this positive correlation seems to be more pronounced at lower levels of perceived accessibility.

Figure 7(b) illustrates that the perceived safety of the cycling environment also has a positive correlation with cycling distance, corroborating the findings of Guo et al. (2023), Milakis et al. (2017), and Shaer et al. (2021). Also, our results reveal a threshold effect. The non-linear results are very similar to binary classification, indicating that perceived safety is decisive in influencing an individual's choice of riding distance, which may be a psychological bottleneck. That is to say, gradually improving perceived safety to increase riding distance is not always effective because it is linear thinking. This conclusion confirms the existence of the perception threshold mentioned by Golledge et al. (1997) and also supports the work of Mandic et al. (2016). The threshold effect of the non-linear relationship between perceived safety and cycling distance confirms the differential impact of perceived safety and insecurity on cycling distance. When the perceived level of safety is below a certain level, individuals may be unwilling to choose a longer ride distance due to concerns about personal safety. On the contrary, when the safety threshold is exceeded, individuals may be more willing to try longer cycling distances. In this article, this threshold is close to 0.45. Specifically, the perceived safety of medium-low and medium-high (<0.4; >0.5) is not negatively or positively correlated with cycle distance. A positive correlation only occurs when the threshold is reached. Although many previous studies have mainly emphasized that a safer bicycle environment will attract more bicycle users (Sultana et al., 2018), our findings go deeper and reveal the threshold effect between perceived safety and bicycle distance.

Established studies have consistently highlighted the influence of the comfort of the cycling environment on bicycle usage. For instance, Alcorn and Jiao (2023) argue that more comfortable cycling facilities exhibit a positive correlation with the use of shared bikes, a finding that our results corroborate, as depicted in Figure 7(c). Moreover, Young et al. (2022) contend that the comfort of the cycling environment not only encourages greater cycling activity but also extends the distance. Our study aligns with this perspective. Figure 7(c) further illustrates the non-linear effect of comfort on the distance ridden. We observe that only higher levels of comfort are associated with longer cycling distances, while lower levels of comfort have a limited effect on the distance. This threshold is roughly around 0.65. These results are consistent with those of Fitch et al. (2022), who similarly argued that high levels of comfort are what have a more substantial effect on bicycle use but did not find a threshold effect between both.

5 Conclusions

In this study, we offer a fresh perspective by integrating both objective and perceived dimensions to investigate the effects of the riding environment on riding distance. Initially, we assessed the relative importance of objective and perceived riding environment factors on riding distance, emphasizing the predictive role of each explanatory variable. Subsequently, through the computation and visualization of partial dependency plots, we determined the effective range and non-linear relationships of cycling environment factors, with a particular focus on identifying whether there is a threshold effect for the perceived cycling environment dimension. These insights are crucial for planning to enhance average urban cycling distances, improve sustainable transport development, and design bicycle-friendly environments.

Specifically, our findings suggest that population density, road network density, recreational facilities, perceived accessibility and comfort are closely related to cycling distance and all show threshold effects in their non-linear relationships with cycling distance.

(1) Population density had the highest relative importance in predicting cycling distance at 24.8%. Cycling distance was positively correlated with population density when the population densities were approximately 14,000 persons/km^2 and 22,000 persons/km^2. However, this positive correlation diminishes as the population density approaches 24,000 people/km^2.

Taking the main urban area of Shenzhen as an example, policymakers should consider optimizing the distribution of population density in urban planning. For example, in areas where the population density is lower than 14,000 people/km^2, the population density can be increased moderately through measures such as increasing housing and commercial facilities. For areas where the population density has reached or exceeded 22,000 people/km^2, the focus should be on investing in the improvement and expansion of cycling infrastructures, such as increasing cycling lanes, cycling parking facilities, and maintenance points in order to satisfy the higher cycling demand.

(2) The contribution of road network density follows closely with about 15.2%. There is an inverted U-shaped relationship between road network density and cycling distance. We speculate that when the road network density reaches 3 km/km^2, the safety risks of a dense road network system may outweigh the accessibility benefits of increased cycling distance.

(3) Among the various land-use attributes, recreational facilities are an important factor with a relative importance of about 9.1%. When its proportion reaches 8 to 13 per cent, there is a significant positive correlation between the two. Therefore, there is a need to increase recreational facilities such as parks, green spaces, and cultural and entertainment venues in urban planning, and the integration of urban greenways or bicycle paths is a promising initiative.

(4) The contribution of perceived variables should not be overlooked, with perceived accessibility and comfort having a strong explanatory power for cycling distance, at 8% and 9.1%, respectively. There is also a threshold effect for perceptual variables. When perceived accessibility is around 0.4, it has a significant positive correlation with riding distance. Perceived safety was positively correlated with riding distance only when it was near the threshold value of 0.45. Comfort is positively correlated with riding distance only when it reaches a high level (> 0.65).

In terms of policy, there are a number of measures that can be taken to improve perceived safety in the cycling environment. Measures such as improving road conditions and visibility, increasing safety patrols in cycling hotspots, and providing real-time traffic information and safety campaigns can help to break through the bottleneck of perceived safety for cyclists, thereby increasing the distance travelled. Perceived accessibility has the highest relative importance, suggesting that urban planning should not only focus on the design of cycle lanes but also on accessibility in the cycling experience of users. It is worth noting that perceived accessibility is a positive correlation with cycling distance occurs in the range of lower levels of perceived accessibility, and on the contrary, it is positively correlated with cycling distance only when comfort is at a high level. Therefore, this discrepancy reminds us to implement progressive resource allocation strategies. Where resources are limited, priority should be given to projects that improve perceived accessibility, thereby increasing the overall perceived accessibility of the urban cycling environment. Where resources are sufficient, long-term planning should focus on improving the overall amenity of cycle paths in order to maximize the use of resources.

Several limitations and issues warrant further discussion and validation. Firstly, the results obtained in this study reflect correlation rather than causation, thereby weakening the interpretability of the findings to some extent. Secondly, due to data limitations, we did not identify the multiple purposes of riding, potentially reducing readers' understanding of the riding trajectory data. Finally, our study design introduced some time bias between the questionnaire data and the trajectory data we utilized. Although Qiao and Yeh (2023) argues that this bias is acceptable, it is undeniable that bias may impact the results. In future research, we plan to employ a more thoughtful methodology and research design, along with richer data to enhance and deepen our understanding of the relationship between the riding environment and riding behavior. This may involve utilizing mobile phone signaling data and GPS tracking of riding users for research purposes. Additionally, we express a keen interest in investigating shared e-bikes, as our findings indicate that the change in the riding distance of shared bikes, despite the existence of a threshold, has overall variations that are not large. In contrast, shared e-bikes, being non-human powered vehicles, exhibit greater variation in riding distance. Due to space limitations, we did not present our results regarding spatial location information. Investigating the spatial heterogeneity of feature thresholds affecting riding distance is a future direction in this line of research.

There are several limitations and issues that deserve further discussion and validation. Firstly, the results of this study reflect correlation rather than causation, thus somewhat diminishing the interpretability of the findings. Second, due to data limitations, we did not identify the multiple purposes of cycling, which may reduce the reader's understanding of the cycling trajectory data. In addition, the investigator may have had a subjectivity bias in selecting respondents for the survey, which may also reduce the interpretability of the results. Finally, our study design introduced some temporal bias between the questionnaire data and the trajectory data. Although Qiao et al. (2023) argues that this bias is acceptable, it is undeniable that the bias may affect the results of the study. In future studies, we plan to use a more thoughtful methodology and research design, as well as richer data, to improve and deepen our understanding of the relationship between the cycling environment and cycling behavior. This may involve conducting research using mobile phone signal data and GPS tracking of cycling users. In addition, we expressed a keen interest in studying shared e-bikes as our findings suggest that changes in cycling distance on shared bikes generally vary little despite the presence of thresholds. In contrast, shared e-bikes, as a non-human-powered mode of transport, showed greater variability in cycling distance. Due to space constraints, we do not present results regarding spatial location information. Investigating the spatial heterogeneity of feature thresholds affecting riding distance is a future direction for this line of research.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 52272332), Heilongjiang Provincial Natural Science Foundation (Grant No. YQ2021E031), and the Fundamental Research Funds for the Central Universities (Grant No. HIT.OCEF.2022026). In addition, this work could not have been improved during the revision process without the contribution of two anonymous reviewers, who are especially thanked for their careful review and dedication.

References

Alcorn, L. G., & Jiao, J. (2023). Bike-sharing station usage and the surrounding built environments in major Texas cities. *Journal of Planning Education and Research, 43*(1), 122–135.

Banerjee, A., Łukawska, M., Jensen, A. F., & Haustein, S. (2022). Facilitating bicycle commuting beyond short distances: Insights from existing literature. *Transport Reviews*, *42*(4), 526–550.

Büchel, B., Marra, A. D., & Corman, F. (2022). COVID-19 as a window of opportunity for cycling: Evidence from the first wave. *Transport Policy, 116*, 144–156.

Cervero, R., & Duncan, M. (2003). Walking, bicycling, and urban landscapes: Evidence from the San Francisco Bay Area. *American Journal of Public Health, 93*(9), 1478–1483.

Chan, E. T., Schwanen, T., & Banister, D. (2021). The role of perceived environment, neighbourhood characteristics, and attitudes in walking behaviour: Evidence from a rapidly developing city in China. Transportation, 48, 431–454.

Chataway, E. S., Kaplan, S., Nielsen, T. A. S., & Prato, C. G. (2014). Safety perceptions and reported behavior related to cycling in mixed traffic: A comparison between Brisbane and Copenhagen. *Transportation Research Part F: Traffic Psychology and Behavior*, *23*, 32–43.

Chen, J., Zhou, D., Zhao, Y., Wu, B., & Wu, T. (2020). Life cycle carbon dioxide emissions of bike sharing in China: Production, operation, and recycling. *Resources, Conservation and Recycling, 162,* 105011.

Chen, X. (2022). Predicting college students' bike-sharing intentions based on the theory of planned behavior. *Frontiers in Psychology, 13,* 836983.

Chen, E., & Ye, Z. (2021). Identifying the nonlinear relationship between free-floating bike sharing usage and built environment. *Journal of Cleaner Production, 280,* 124281.

Chen, P., Sun, F., Wang, Z., Gao, X., Jiao, J., & Tao, Z. (2018). Built environment effects on bike crash frequency and risk in Beijing. *Journal of Safety Research, 64*, 135–143.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

Chen, Z., Van Lierop, D., & Ettema, D. (2022). Dockless bike-sharing's impact on mode substitution and influential factors. *Journal of Transport and Land Use*, *15*(1), 71–93.

Cheng, L., Wang, K., De Vos, J., Huang, J., & Witlox, F. (2022). Exploring non-linear built environment effects on the integration of free-floating bike-share and urban rail transport: A quantile regression approach. *Transportation Research Part A: Policy and Practice*, *162*, 175–187.

Ding, C., Cao, X., Dong, M., Zhang, Y., & Yang, J. (2019). Non-linear relationships between built environment characteristics and electric-bike ownership in Zhongshan, China. *Transportation Research Part D: Transport and Environment, 75*, 286–296.

Echiburú, T., Muñoz, J. C., & Hurtubia, R. (2021). The role of perceived satisfaction and the built environment on the frequency of cycle-commuting. *Journal of Transport and Land Use, 14*(1), 171–196.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813.

Feng, P., & Li, W. (2016). Willingness to use a public bicycle system: An example in NanJing City. *Journal of Public Transportation, 19*(1), 84–96.

Ferenchak, N. N., & Marshall, W. E. (2021). Bicycling facility inequalities and the causality dilemma with socioeconomic/sociodemographic change. *Transportation Research Part D: Transport and Environment, 97,* 102920.

Fitch, D. T., Carlen, J., & Handy, S. L. (2022). What makes bicyclists comfortable? Insights from a visual preference survey of casual and prospective bicyclists. *Transportation Research Part A: Policy and Practice, 155*, 434–449.

Galster, G. C. (2018). Nonlinear and threshold effects related to neighborhood: Implications for planning and policy. *Journal of Planning Literature, 33*(4), 492–508.

Garrard, J., Rose, G., & Lo, S. K. (2008). Promoting transportation cycling for women: The role of bicycle infrastructure. *Preventive Medicine, 46*(1), 55–59.

Golledge, R. G. (1997). *Spatial behavior: A geographic perspective*. New York: Guilford Press.

Guo, X., Tavakoli, A., Angulo, A., Robartes, E., Chen, T. D., & Heydarian, A. (2023). Psycho-physiological measures on a bicycle simulator in immersive virtual environments: How protected/curbside bike lanes may improve perceived safety. *Transportation Research Part F: Traffic Psychology and Behavior, 92*, 317–336.

Guo, Y., & He, S. Y. (2021). Perceived built environment and dockless bikeshare as a feeder mode of metro. *Transportation Research Part D: Transport and Environment, 92,* 102693.

Han, Y., Li, W., Wei, S., & Zhang, T. (2018). Research on passenger’s travel mode choice behavior waiting at bus station based on SEM-logit integration *Model. Sustainability, 10*(6), 1996.

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.

Jiao, J., Lee, H. K., & Choi, S. J. (2022). Impacts of COVID-19 on bike-sharing usages in Seoul, South Korea. *Cities, 130,* 103849.

Kabak, M., Erbaş, M., Cetinkaya, C., & Özceylan, E. (2018). A GIS-based MCDM approach for the evaluation of bike-share stations. *Journal of Cleaner Production, 201*, 49–60.

Keppner, V., Krumpoch, S., Kob, R., Rappl, A., Sieber, C. C., Freiberger, E., & Siebentritt, H. M. (2023). Safer cycling in older age (SiFAr): Effects of a multi-component cycle training. A randomized controlled trial. *BMC Geriatrics, 23*(1), 1–9.

Kerr, J., Emond, J. A., Badland, H., Reis, R., Sarmiento, O., Carlson, J., ... & Natarajan, L. (2016). Perceived neighborhood environmental attributes associated with walking and cycling for transport among adult residents of 17 cities in 12 countries: The IPEN study. *Environmental Health Perspectives, 124*(3), 290–298.

Kim, D. J., Shin, H., Im, H., and Park, J. (2012). Factors influencing travel behaviors in bikesharing. Paper presented at the Transportation Research Board’s 91st Annual Meeting, January 22–26, Washington, DC.

Kim, J., & Lee, S. (2023). Determining factors affecting public bike ridership and its spatial change before and after COVID-19. *Travel Behavior and Society, 31*, 24–36.

Kraus, S., & Koch, N. (2021). Provisional COVID-19 infrastructure induces large, rapid increases in cycling. *Proceedings of the National Academy of Sciences, 118*(15), e2024399118.

Li, Q., & Xu, W. (2022). The impact of COVID-19 on bike-sharing travel pattern and flow structure: Evidence from Wuhan. *Cambridge Journal of Regions, Economy and Society, 1*5(3), 477–494.

Ma, L., Dill, J., & Mohr, C. (2014). The objective versus the perceived environment: what matters for bicycling? *Transportation, 41*, 1135–1152.

Mandic, S., Williams, J., Moore, A., Hopkins, D., Flaherty, C., Wilson, G., ... & Spence, J. C. (2016). Built environment and active transport to school (BEATS) study: Protocol for a cross-sectional study. *BMJ Open*, *6*(5), e011196.
Marshall, W. E., & Ferenchak, N. N. (2019). Why cities with high bicycling rates are safer for all road users. *Journal of Transport & Health, 13,* 100539.
Milakis, D., Efthymiou, D., & Antoniou, C. (2017). Built environment, travel attitudes and travel behavior: Quasi-longitudinal analysis of links in the case of Greeks relocating from US to Greece. *Sustainability, 9*(10), 1774.
Nielsen, T. A. S., Olafsson, A. S., Carstensen, T. A., & Skov-Petersen, H. (2013). Environmental correlates of cycling: Evaluating urban form and location effects based on Danish micro-data. *Transportation Research Part D: Transport and Environment, 22*, 40–44.
Ospina, J. P., Botero-Fernández, V., Duque, J. C., Brussel, M., & Grigolon, A. (2020). Understanding cycling travel distance: The case of Medellin City (Colombia). *Transportation Research Part D: Transport and Environment, 86*, 102423.
Porter, A. K., Kohl III, H. W., Pérez, A., Reininger, B., Gabriel, K. P., & Salvo, D. (2020). Bikeability: Assessing the objectively measured environment in relation to recreation and transportation bicycling. *Environment and Behavior*, *52*(8), 861–894.
Pot, F. J., van Wee, B., & Tillema, T. (2021). Perceived accessibility: What it is and why it differs from calculated accessibility measures based on spatial data. *Journal of Transport Geography, 94,* 103090.
Qiao, S., & Yeh, A. G. O. (2023). Understanding the effects of environmental perceptions on walking behavior by integrating big data with small data. *Landscape and Urban Planning, 240,* 104879.
Qu, T., Gates, T. J., Xu, C., Seguin, D., & Kay, J. (2022). The disparate impact of COVID-19 pandemic on walking and biking behaviors. *Transportation Research Part D: Transport and Environment, 112,* 103494.
Saltykova, K., Ma, X., Yao, L., & Kong, H. (2022). Environmental impact assessment of bike-sharing considering the modal shift from public transit. *Transportation Research Part D: Transport and Environment*, *105,* 103238.
Samadzad, M., Nosratzadeh, H., Karami, H., & Karami, A. (2023). What are the factors affecting the adoption and use of electric scooter sharing systems from the end user's perspective? *Transport Policy*, *136*, 70–82.
Schaefer, K. J., Tuitjer, L., & Levin-Keitel, M. (2021). Transport disrupted–Substituting public transport by bike or car under Covid 19. *Transportation Research Part A: Policy and Practice*, *153*, 202–217.
Scheepers, C. E., Wendel-Vos, G. C. W., Van Kempen, E., De Hollander, E., van Wijnen, H. J., Maas, J., ... & Schuit, A. J. (2016). Perceived accessibility is an important factor in transport choice—Results from the AVENUE project. *Journal of Transport & Health*, *3(*1), 96–106.
Shaer, A., Rezaei, M., Rahimi, B. M., & Shaer, F. (2021). Examining the associations between perceived built environment and active travel, before and after the COVID-19 outbreak in Shiraz city, Iran. *Cities*, *115*, 103255.
Shao, Q., Zhang, W., Cao, X. J., & Yang, J. (2022). Nonlinear and interaction effects of land use and motorcycles/e-bikes on car ownership. *Transportation Research Part D: Transport and Environment, 102*, 103115.
Shen, Y., Zhang, X., & Zhao, J. (2018). Understanding the usage of dockless bike sharing in Singapore. *International Journal of Sustainable Transportation, 12*(9), 686–700.
Sultana, S., Kim, H., Pourebrahim, N., & Karimi, F. (2018). Geographical assessment of low-carbon transportation modes: A case study from a commuter university. *Sustainability*, *10*(8), 2696.

Sun, G., & Zacharias, J. (2017). Can bicycle relieve overcrowded metro? Managing short-distance travel in Beijing. *Sustain, Cities Soc.*, *35*, 323–330.

Wang, K., Akar, G., & Chen, Y. J. (2018). Bike sharing differences among millennials, Gen Xers, and baby boomers: Lessons learnt from New York City's bike share. *Transportation Research Part A: Policy and Practice, 116,* 1–14.

Wei, B., & Zhu, L. (2023). Exploring the impact of built environment factors on the relationships between bike sharing and public transportation: A case study of New York. *ISPRS International Journal of Geo-Information*, *12*(7), 293.

Williams, N. L., Wassler, P., & Ferdinand, N. (2022). Tourism and the COVID-(Mis) infodemic. *Journal of Travel Research*, *61*(1), 214–218.

Winters, M., Teschke, K., Grant, M., Setton, E. M., & Brauer, M. (2010). How far out of the way will we travel? Built environment influences on route selection for bicycle and car travel. *Transportation Research Record*, *2190*(1), 1–10.

Wu, C., Kim, I., & Chung, H. (2021). The effects of built environment spatial variation on bike-sharing usage: A case study of Suzhou, China. *Cities, 110*, 103063.

Yasir, A., Hu, X., Ahmad, M., Alvarado, R., Anser, M. K., Işık, C., ... & Khan, I. A. (2022). Factors affecting electric bike adoption: Seeking an energy-efficient solution for the post-COVID era. *Frontiers in Energy Research, 9*, 817107.

Young, E., Kastner, P., Dogan, T., Chokhachian, A., Mokhtar, S., & Reinhart, C. (2022). Modeling outdoor thermal comfort along cycling routes at varying levels of physical accuracy to predict bike ridership in Cambridge, MA. *Building and Environment, 208*, 108577.

Zhang, Y., Thomas, T., Brussel, M., & Van Maarseveen, M. (2017). Exploring the impact of built environment factors on the use of public bikes at bike stations: Case study in Zhongshan, China. *Journal of Transport Geography, 58,* 59–70.

Zhang, Y., Kasraian, D., & van Wesemael, P. (2023). Built environment and micro-mobility. *Journal of Transport and Land Use, 16*(1), 293–317.

Zhao, Y., Ke, S., Lin, Q., & Yu, Y. (2020). Impact of land use on bicycle usage. *Journal of Transport and Land Use*, *1*3(1), 299–316.

Zhao, C., Nielsen, T. A. S., Olafsson, A. S., Carstensen, T. A., & Meng, X. (2018). Urban form, demographic and socio-economic correlates of walking, cycling, and e-biking: Evidence from eight neighborhoods in Beijing. *Transport Policy*, *64*, 102–112.

Zhuang, C. G., Li, S. Y., & Tan, Z. Z., Gao, F., & Wu, Z. (2022). Nonlinear and threshold effects of traffic condition and built environment on dockless bike sharing at street level. *Journal of Transport Geography, 102,* 103375.