Journal of Cleaner Production 471 (2024) 143444

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Understanding human mobility and trip demand through sparse trajectories of private e-bikes

Peixiao Wang [a,b], Hengcai Zhang [a,b,*], Shifen Cheng [a,b], Feng Lu [a,b,c], Tong Zhang [d], Zeqiang Chen [e]

[a] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing, 100101, China

[b] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, 100049, China

[c] Fujian Collaborative Innovation Center for Big Data Applications in Governments, Fuzhou, 350003, China

[d] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, 430079, China

[e] National Engineering Research Center for Geographic Information System, China University of Geosciences, Wuhan, 430074, China

ARTICLE INFO

Handling Editor: Giovanni Baiocchi

Keywords:
Human mobility
Private e-bike trajectories
e-bike trip
Model interpretability
Driving mechanisms

ABSTRACT

Understanding human mobility and trip demand through e-bike trajectories is crucial for urban planning, environmental enhancement, and sustainable development. However, existing studies predominantly focus on shared (e−)bike trips, neglecting private e-bike trips. With the recent availability of sparse trajectory for private e-bikes, we established a novel analysis framework to reveal human mobility and trip demand in Wuhan, China. First, we propose a two-step method for extracting trip behavior from sparse trajectories of private e-bikes, involving the identification of staying areas and the generation of e-bike trips. Second, we establish a spatial random forest method to capture the nonlinear relationship between private e-bike trips and driving factors. Finally, we use the interpretable SHAP method to reveal the driving mechanisms of e-bike trips and explore the impact of various factors on these trips. The results indicate that (1) trip distances of private e-bikes follow a lognormal distribution, with an Adj. R-Square of 0.99, while trip times exhibit a Hill distribution, with an Adj. R-Square of 0.95; (2) Private e-bike trips are not commonly employed to address the first/last mile problem in public transportation and are more frequently used for daily commuting needs, with over 65% of these trips covering distances greater than 1 km or lasting longer than 5 min; (3) private e-bike trips positively correlate with the density of POIs like *Hospital*, *School*, and *Transportation Station*. However, compared to shared (e−)bike trips, *Transportation Station Density*, especially *Metro Station Density*, is less important for private e-bike trips; and (4) private e-bike trips are also positively correlated with *Congestion Level* and *House Price*, meaning that areas with severe traffic congestion or high housing prices tend to have more private e-bike trips. This study provides a new framework for understanding private e-bike trip patterns, also helping authorities better grasp the factors influencing e-bike trip demand.

1. Introduction

The escalation in motor vehicle ownership has exacerbated issues of traffic congestion and environmental pollution, thus imperiling the sustainable progression of urban transportation (S. Cheng et al., 2023; McCaffery et al., 2021). In response, many countries have regarded electric bicycles (e-bikes) as a viable remedy to attain zero carbon emissions objectives, actively endorsing them as the preferred mode of transportation for human mobility (Guidon et al., 2020; McKenzie, 2020; McQueen et al., 2020). It is estimated that e-bike ownership around the world will continue to grow in the future, for example, China's annual sales of e-bikes soared to 54 million in 2023 (IResearch Consulting and Master Lu, 2023).

The proliferation of e-bikes has led to an explosive surge in trajectory data, offering novel avenues for scrutinizing human mobility and trip demand through the lens of e-bikes (Fukushige et al., 2021; Fyhri and

* Corresponding author. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing, 100101.China.

E-mail addresses: wpx@lreis.ac.cn (P. Wang), zhanghc@lreis.ac.cn (H. Zhang), chengsf@lreis.ac.cn (S. Cheng), luf@lreis.ac.cn (F. Lu), zhangt@whu.edu.cn (T. Zhang), chenzeqiang@cug.edu.cn (Z. Chen).

https://doi.org/10.1016/j.jclepro.2024.143444
Received 3 July 2024; Received in revised form 3 August 2024; Accepted 19 August 2024
Available online 21 August 2024

Fearnley, 2015; Z. Guo et al., 2023; McKenzie, 2020; Rich et al., 2021; Z. Zhang et al., 2023). Presently, extensive studied have been conducted on understanding human trips through e-bike trajectories, encompassing analyses of hotspots and cold spots in e-bike usage (Xu et al., 2023), exploration of connecting trips between e-bikes and subways/buses (Liu et al., 2023; Zhou et al., 2023), the elucidation of driving factors behind e-bike trips (Bi et al., 2022; Ding et al., 2019; Y. Yu et al., 2022).

Although considerable studies have been conducted on human trips using e-bikes, there are still gaps. From the research object, existing studies predominantly center on trips using shared e-bikes (Choi et al., 2023; Z. Zhang et al., 2023), rather than trips using private e-bikes (Y. Yu et al., 2022). Compared to shared e-bikes, private e-bikes have significant advantages in flexibility, convenience, and freedom. However, current research on trips using private e-bikes is very limited, leaving their spatiotemporal patterns—such as mobility patterns and trip demand—largely unexplored (Guidon et al., 2020; McKenzie, 2020). Regarding methodology, the current approach primarily relies on simplistic applications of classical machine learning models, particularly for modeling trip demand (Bi et al., 2022; S. Yu et al., 2021). The classical machine learning model assumes that e-bike trips are independent and identically distributed samples, which contradicts the spatial correlation and heterogeneity of e-bike trips.

In recent years, to enhance the management of private e-bikes, some cities have installed monitoring stations to collect their trajectories. Unlike GPS trajectories of shared e-bikes, the trajectories collected by monitoring stations exhibit significant sparsity, thereby safeguarding personal privacy. The collection of private e-bike trajectories has enabled numerous studies. Given this context, we established a novel analysis framework based on sparse trajectories of private e-bikes. The proposed framework extracts the trip behavior of private e-bikes from sparse trajectories and employs interpretable data-driven techniques to uncover their spatiotemporal patterns. Specifically, this study aims to answer the following three questions: (1) How to extract private e-bike trips from sparse trajectories? (2) How can we model the trip demand of private e-bikes while accounting for spatial correlation and heterogeneity? (3) What are the mobility patterns and driving factors behind private e-bike trips? This study fills the gap in current research on private e-bike trips and enriches the literature on human mobility. In addition, we open-sourced the extracted trip data to facilitate research on private e-bike trips.

The remainder of this study is organized as follows. Section 2 reviews the relevant literature; Section 3 describes the study area and data sources of this study; Section 4 presents a detailed description of the proposed framework for e-bikes trip analysis; Section 5 analyzes the spatiotemporal impacts of the built environment, traffic conditions, and socioeconomic factors on private e-bike trips; In Section 6, the discussion and summary are presented.

2. Related works

This study primarily focuses on human mobility for green trips, particularly private e-bike trips. Unfortunately, there has been limited research on private e-bike trips. Consequently, we primarily reviewed research related to shared (e−)bike trips.

In recent years, the establishment of shared (e−)bike systems has infused new vitality into green trips (Venkadavarahan et al., 2023). Shared (e−)bike trips often yield a wealth of dense trajectories thanks to the advantages of GPS positioning technology (Choi et al., 2023; Plazier et al., 2017). By integrating order data with dense trajectory data, researchers can easily extract the trip features of shared (e−)bikes, such as trip distance, duration, origin, and destination (Ji et al., 2022; Tang et al., 2024). At present, related scholars have conducted extensive studies on shared (e−)bike trips (Bieliński et al., 2021).

For shared bike trips, Xu et al. (2023) found that the trip distances of shared bikes follow a log-normal distribution, with approximately 90% of trips concentrated within 1.3 km. Ji et al. (2022) and Zhou et al. (2023) found a significant positive correlation between sharing bike trips and metro station density, particularly in downtown areas. Above finding suggests that shared bikes have become an effective solution for addressing the first/last mile problem in public transportation (van Kuijk et al., 2022; Yen et al., 2023; Zuo et al., 2020). Many commuters utilize shared bikes for connecting trips during workdays, especially from shared bikes to the metro(L. Cheng et al., 2019; Fu et al., 2023; Zhou et al., 2023). Regarding driving mechanism modeling, Yu et al. (2022) and Zhou et al. (2023) established the relationship between shared bike trips and their driving factors using ordinary least squares and geographically weighted regression, respectively. To capture the nonlinear relationship between trips and driving factors, many scholars have also employed classical machine learning models, such as gradient boosting decision trees(Bi et al., 2022), random forest(L. Cheng et al., 2019), and Xtreme Gradient Boosting(Ji et al., 2022). These Studies have demonstrated that shared bike trips are influenced by socioeconomic factors, spatial locations, and urban built environment (Bi et al., 2022; Ding et al., 2019; Eren and Uz, 2020; Fu et al., 2023).

For shared e-bike trips, the trip distances of shared e-bikes also follow a log-normal distribution (Li et al., 2024). Compared to shared bike trips, the trip distances of shared e-bikes are slightly longer(Guidon et al., 2019; Reck et al., 2021), but many shared e-bike trips still concentrate within 2 km(Li et al., 2024). These finding suggests that shared e-bike trips are primarily intended for short-distance urban trips, with many trips being used to address the first/last mile problem in public transportation (Choi et al., 2023; Liu et al., 2023; Zhu et al., 2024). Regarding driving factors, Existing Studies show that shared e-bike trips are also influenced by socioeconomic factors, spatial locations, and urban built environment(Liu et al., 2023; Yang et al., 2022; Y. Yu et al., 2022; Zhou et al., 2022). For instance, socioeconomic factors impact the acceptance and frequency of shared e-bike usage, spatial locations influence usage patterns and demand intensity, and the urban built environment affects the availability and convenience of shared e-bikes.

In summary, considerable research has been conducted on (e−)bike trips. However, several gaps remain in the existing research. First, existing studies predominantly focus on shared (e−)bike trips rather than private e-bike trips. To protect personal privacy, we typically collect sparse trajectories of private e-bikes using monitoring stations, rather than the dense trajectories obtained through GPS technology. This sparsity presents challenges in accurately extracting trip behavior from these trajectories. Second, existing studies use classical machine learning approaches to capture the nonlinear relationship between (e−) bike trips and driving factors. These approaches often assume sample independence and identical distribution, neglecting the spatial correlation and heterogeneity inherent in (e−)bike trips. To address these gaps, we propose a method for extracting trip behavior from sparse trajectories and establish a spatial random forest approach to capture the nonlinear relationship between private e-bike trips and driving factors. Additionally, we analyze the mobility patterns of private e-bike trips and reveal the driving mechanisms using interpretable data-driven techniques.

3. Study area and data sources

3.1. Study area

Wuhan is situated in the center of China, in the eastern part of Hubei Province, and along the middle reaches of the Yangtze River. Since 2010, Wuhan's resident population has been on a continuous upward trend, growing from 9.7 million in 2010 to 13 million in 2023. At present, Wuhan has emerged a mega city in the central region of China and a core city in the Yangtze River Economic Belt. As shown in Fig. 1, the area within the Third Ring Road is the heart of Wuhan. Despite covering only 6% of Wuhan's total area, this region accommodates 50% of the city's population. This study focuses on the spatiotemporal patterns of

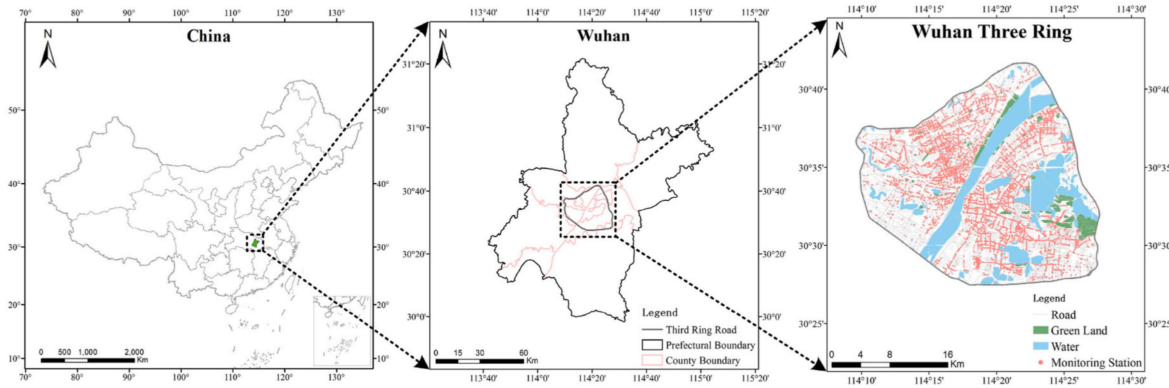


Fig. 1. Sketch map of the study area.

private e-bike trips within the Third Ring Road. Following to the methodology of Fu et al. (2023), we divided the study area into a regular grid of 200m × 200 m cells.

3.2. Data sources and data preprocessing

3.2.1. Data sources

The dataset of this study mainly consists of POI data, house price data, population data, Gaode congestion index data, and private e-bike trajectory data. The POI data, house price data, population data, and Gaode congestion index data are mainly used to construct the driving factors of urban environment, traffic condition, and socio-economics (discussed in Section 3.2), while the private e-bike trajectory data is mainly used to extract the trip behavior. In this subsection, we mainly describe the collection method and field information of private e-bike trajectories.

The trajectories of private e-bikes are collected through monitoring stations rather than GPS devices. When a private e-bike moves near a monitoring station, the real-time location is uploaded to the database. Over 8000 monitoring stations were deployed within the study area, as shown in Fig. 1. The dataset used in this study was collected from December 28, 2020, to January 10, 2021, encompassing over 1 million private e-bike trajectories. Table 1 shows the field information of a single private e-bike trajectory, where *ID* is the unique identifier of the private e-bikes. By sorting the trajectory points in time, a complete trajectory can be obtained. Compared to the GPS continuous tracking and positioning technology used for shared e-bikes, the trajectory points of private e-bikes are only generated when they are near monitoring stations. This method safeguards personal privacy but also results in significantly sparser trajectories.

3.2.2. Data preprocessing

To support this study, we preprocessed spatiotemporal datasets as follows.

(1) We have unified the coordinate systems of POI data, housing price data, congestion data, population data, and trajectory data into the WGS84 coordinate system.

Table 1
Sample of single private e-bike trajectory.

ID	Date Time	Latitude	Longitude
125172	2020/12/28 07:16:24	30.4****	114.4****
125172	2020/12/28 07:18:14	30.4****	114.4****
125172	2020/12/28 07:19:03	30.3****	114.5****
……	……	……	……
125172	2021/01/08 20:14:07	30.4****	114.5****

(2) There are significant differences in human activity patterns during weekdays, weekends, and holidays (S. Zhang et al., 2019; Y. B. Zhang et al., 2024). In this study, we focused on extracting and analyzing e-bike trips during weekdays, removing trajectory data collected on weekends and holidays.

(3) We filtered invalid private e-bike trajectory data, such as trajectory data with too few points or no movement. Additionally, since we focus on private e-bike trips within the Third Ring Road, we removed trajectory data outside this area. After filtering the data, we obtained more than 300,000 private e-bike trajectories.

4. Methodology

As shown in Fig. 2, the proposed framework is divided into three parts: extracting e-bike trips based on sparse trajectories, modeling e-bike trips based on spatial random forest, and revealing e-bike trip mechanisms based on explainable approach (discussed in sections 4.1-4.3). First, we propose a two-step extraction method for trip behavior, which identifies the staying areas in trajectory data and extracts the trip behavior of e-bikes based on the temporal relationship of the staying areas. Second, we establish a spatial random forest method to capture the nonlinear relationship between e-bike trips and driving factors. Finally, we reveal the driving mechanism of e-bike trips and explore the impact of various factors on e-bike trips via the interpretable SHAP method.

4.1. Extracting E-bike trip based on sparse trajectories

The extraction of e-bike trips is a critical step in the proposed framework. Since private e-bikes is only tracked near monitoring stations, the trajectory points are sparse compared to those of shared e-bikes. Additionally, unlike shared e-bikes, there is no order data to assist in extracting the trip behavior of private e-bikes. Given the above difference between shared and private e-bikes, we propose a two-step method to extract the trip behavior from private e-bike trajectories. Specifically, we first identify the staying areas in the trajectory data, and then generate the e-bike trips based on the temporal relationships between these staying areas.

As shown in Fig. 3, the staying areas in the trajectory are primarily divided into two categories: the determined staying area and the potential staying area. When the e-bike is parked near the monitoring station, the real-time trajectory point will be uploaded to the database at specific intervals. These points cluster near the monitoring station, forming determined staying areas. If the e-bike is not parked near a monitoring station, the real-time trajectory points will not be recorded in the database. When the trajectory point is recorded again, the first monitored area might be where the e-bike was parked, which we refer to as a potential staying area.

For potential staying areas, we extract the first or last point of the

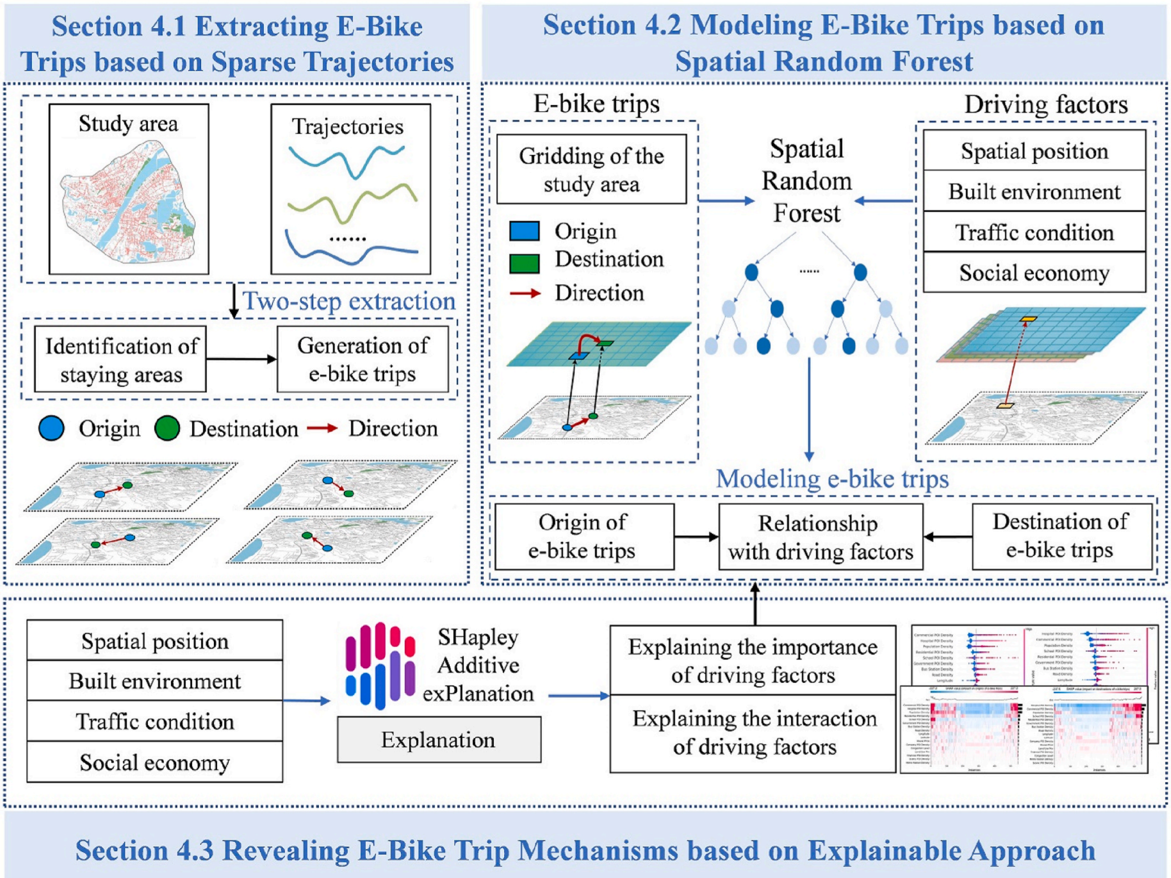


Fig. 2. Research framework for analyzing spatiotemporal patterns of e-bike trips.

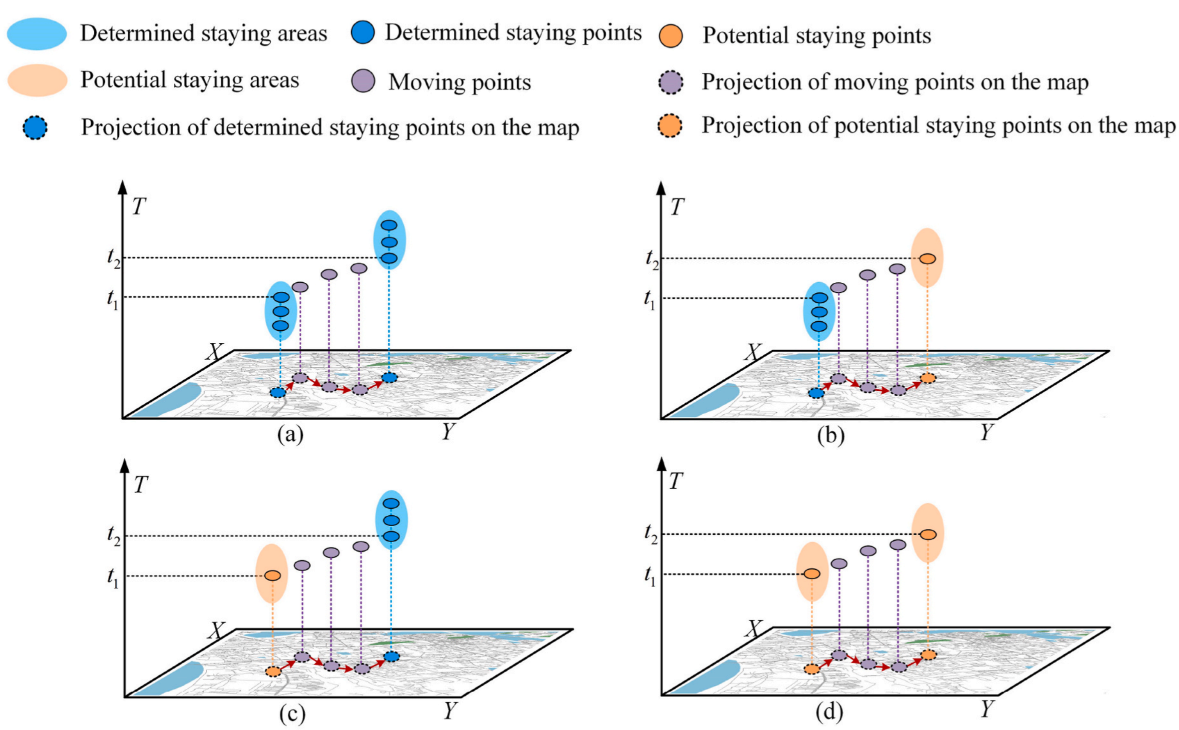


Fig. 3. Extraction of e-bike trips: (a) from determined staying area to determined staying area, (b) from determined staying area to potential staying area, (c) from potential staying area to determined staying area, and (d) from potential staying area to potential staying area.

trajectory fragment. For determined staying areas, inspired by the hierarchical agglomerative clustering algorithm, we extract the center point of the cluster in the trajectory fragment (details are in Appendix A). After identifying the staying area in the trajectory, the e-bike trip can be generated based on the temporal relationship between clusters, including from determined staying area to determined staying area (Fig. 3(a)), from determined staying area to potential staying area (Fig. 3(b)), from potential staying area to determined staying area (Fig. 3(c)), and from potential staying area to potential staying area (Fig. 3(d)).

4.2. Modeling E-bike trips based on spatial random forest

Based on the extraction of e-bike trips, we further model e-bike trips. The random forest is a classical machine learning technique known for its powerful nonlinear fitting abilities, successfully applied across various fields (Valipour Shokouhi et al., 2024; B. Zhang et al., 2023). However, most existing studies simply apply the classical random forest without accounting for the specific characteristics of trip data. Specifically, the classical random forest does not account for the spatial correlation and heterogeneity of trip data, instead assuming that trip data are sample-independently and identically distributed. Therefore, we propose a spatial random forest (SRT) to address the above limitations.

As shown in Fig. 4, the proposed SRF incorporates not only the driving factors of the target grid but also those of its spatial neighbors. To enhance the model's capacity in capturing spatial correlation, we introduced spatial neighborhoods into the classical random forest. Furthermore, we explicitly integrated the spatial position of the target grid into the SRF to improve its capability in capturing spatial heterogeneity. Table 2 outlines the 17 driving factors utilized in this study, categorized into spatial position, built environment, traffic condition, and socioeconomics. As the proposed SRF introduces the spatial neighborhood, we need to fuse the driving factors of the target grid itself with those of the spatial neighbors. The specific fusion method is shown in formulas (1), (2), (3), and (4).

$$\widehat{f}^{i}_{sp} = f^{i}_{sp} \tag{1}$$

$$\widehat{f}^{i}_{be} = \sum_{j \in \Omega_i} f^{j}_{be} \tag{2}$$

Table 2
Dependent and independent variables in spatial random forests.

Input/Output	Variables/Driving Factor	Variable Description (within each grid)
E-bike Trips	Origin*	Total amount of origins
	Destination*	Total amount of destinations
Spatial Position	Longitude+	Longitude of center coordinates
	Latitude+	Latitude of center coordinates
Built Environment	Road density+	Total road length
	Land Use Mix+	Ratio of different land-use types
	Commercial POI Density*	Number of Commercial POIs
	School POI Density*	Number of School POIs
	Company POI Density*	Number of Company POIs
	Hospital POI Density*	Number of Hospital POIs
	Government POI Density*	Number of Government POIs
	Bus Station Density*	Number of Bus Stations
	Metro Station Density*	Number of Metro Stations
	Financial POI Density*	Number of Financial POIs
	Residential POI Density*	Number of Residential POIs
	Scenic POI Density*	Number of Scenic POIs
Traffic condition	Congestion Level+	Congestion duration calculated from Gaode congestion index
Social economy	Population Density*	Population calculated through WorldPop data
	House Price+	Average house prices calculated from Fangtianxia data

+ represents continuous variables, and * represents categorical variables.

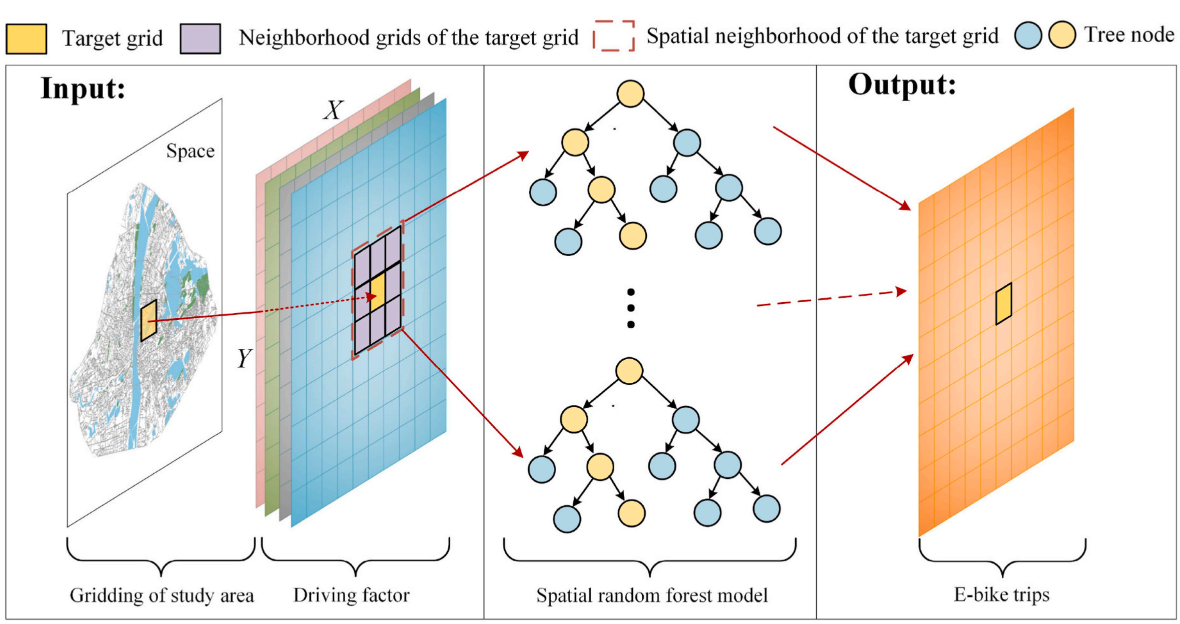


Fig. 4. Definition of spatial random forest.

$$\widehat{f}^{i}_{tc} = \max_{j\in\Omega_i} f^{j}_{te} \tag{3}$$

$$\widehat{f}^{i}_{se} = \frac{\sum_{j\in\Omega_i} \frac{1}{d_{ij}^2} f^{j}_{be}}{\sum_{j\in\Omega_i} \frac{1}{d_{ij}^2}} \tag{4}$$

where $\widehat{f}^{i}_{sp}$, $\widehat{f}^{i}_{be}$, $\widehat{f}^{i}_{tc}$, and $\widehat{f}^{i}_{se}$ represent the fused spatial position, built environment, traffic condition, and socio-economic factors for the i th grid, respectively; Ω_i represents the spatial neighborhood of the i th grid; $f^{j}_{sp}, f^{j}_{be}, f^{j}_{tc}$, and f^{j}_{se} represent the spatial position, built environment, traffic condition, and socio-economic factors of spatial neighbor grid, respectively; d_{ij} represents the spatial distance between the target grid and the spatial neighbor grid. Based on Formulas (1), (2), (3), and (4), the fused spatial position corresponds to the central coordinates of the target grid. The fused built environment is determined by aggregating the count of POIs within the spatial neighborhood. The fused traffic condition is represented by the highest congestion level observed within the spatial neighborhood. The fused socio-economic is calculated as the average house price and population within the spatial neighborhood.

4.3. Revealing E-bike trip mechanisms based on explainable approach

Although the proposed SRF establishes non-linear relationships between seventeen driving factors and e-bike trips, it is still unclear how these factors affect the model's outputs. The SHAP method, a tool for interpreting the output of machine learning models (Lundberg and Lee, 2017), helps understand why the model makes decisions by quantifying the contribution of each factor to the model output. Currently, the SHAP method has been successfully applied to interpret black-box machine learning models across various domains (Ji et al., 2022; Parsa et al., 2020). Therefore, the SHAP method is employed in this study to elucidate the driving mechanism of e-bike trips.

The principle of SHAP method is based on the Shapley value in game theory, elucidating the impact of each driving factor on the model output by computing their respective Shapley values. Specifically, the mathematical model of SHAP is shown in Formula (5).

$$SRF\left(\left\{\widehat{f}^{i}_{1},\ldots,\widehat{f}^{i}_{j},\ldots\widehat{f}^{i}_{17}\right\}\right) = \frac{\sum_{i=1}^{N_t} SRF\left(\left\{\widehat{f}^{i}_{1},\ldots,\widehat{f}^{i}_{j},\ldots\widehat{f}^{i}_{17}\right\}\right)}{N_t} + \sum_{j=1}^{17} SHAP\left(\widehat{f}^{i}_{j}\right) \tag{5}$$

where SRF represents the spatial random forest; $\frac{\sum_{i=1}^{N_t} SRF\left(\left\{\widehat{f}^{i}_{1},\ldots\widehat{f}^{i}_{j},\ldots\widehat{f}^{i}_{17}\right\}\right)}{N_t}$ denotes the average of model outputs for all grids in the test sample; N_t denotes the total number of grids in the test sample; $SHAP\left(\widehat{f}^{i}_{j}\right)$ denotes the Shapley value of the j th driving factor in the i th grid. When $SHAP\left(\widehat{f}^{i}_{j}\right)$ is greater than 0, the j th driving factor of i th grid has a positive effect on e-bike trips. When $SHAP\left(\widehat{f}^{i}_{j}\right)$ is less than 0, the j th driving factor of the i th grid has a negative effect on e-bike trips.

5. Experimental results and analysis

5.1. Spatiotemporal characterization of E-bike trips

In this section, we first analyze the statistical distributions of trip distance and time, and then analyze the spatiotemporal characteristics of the origin and destination of e-bike trips.

Fig. 5(a) and (b) display the statistical distributions of trip distance and time, respectively. The results indicate that the number of e-bike trips first increases and then decreases as the trip distance increases, whereas the number of e-bike trips continuously decreases as the trip time increases. We further fit the statistical distribution of trip distance and time, as shown in Table 3. The results reveal that trip distances follow a clear lognormal distribution, with an Adj. R-Square of 0.99. The fitting result aligns with the characteristics of human trips, where shorter distances are often covered by walking or cycling rather than using e-bikes. In contrast, trip times exhibit a distinct Hill distribution, with an Adj. R-Square of 0.95. In addition, compared to shared (e−)bike trips, private e-bike trips are not commonly employed to address the first/last mile problem in public transportation and are more frequently used for daily commuting needs. For example, 65.6% of private e-bike trips cover distances greater than 1 km, and 68.5% of private e-bike trips last longer than 5 min.

Fig. 6(a) illustrates the time distribution of e-bike trip origins and destinations. The results indicate that the private e-bike trips exhibit a clear bimodal distribution, with the morning peak from 7:30 to 9:20 and the evening peak from 16:00 to 18:45. This results further demonstrate

Table 3
Fitting results of statistical distributions.

Distribution	Trip distance		Trip time	
	R-Square	Adj. R-Square	R-Square	Adj. R-Square
Normal	0.90649	0.90506	0.60716	0.60115
Lognormal	**0.99438**	**0.99429**	0.74208	0.73811
Pulse	0.06327	0.03913	0.85170	0.84788
Giddings	0.97421	0.97382	0.54006	0.53298
Hill	0.91214	0.91079	**0.95319**	**0.95305**

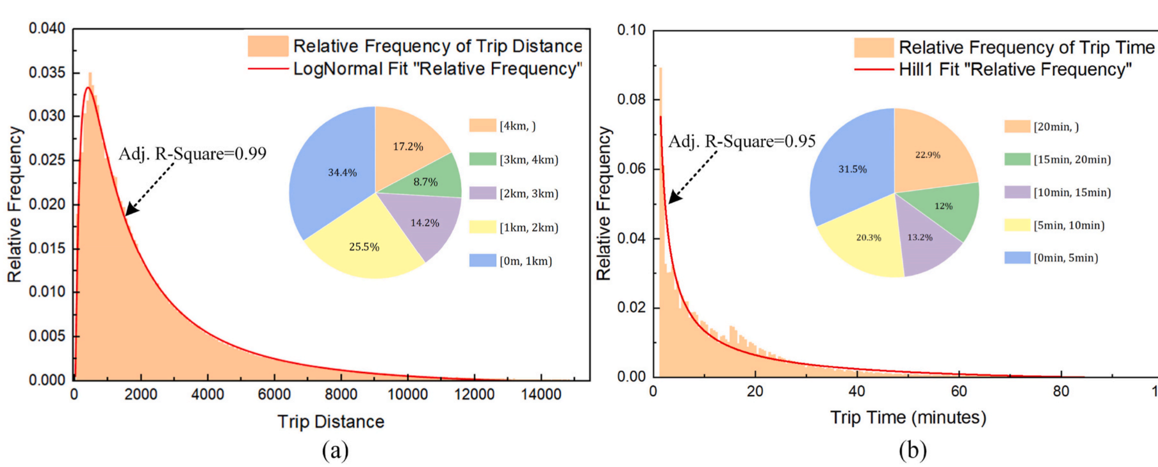


Fig. 5. Statistical distributions of e-bike trips: (a) statistical distributions of trip distance, and (b) statistical distributions of trip time.

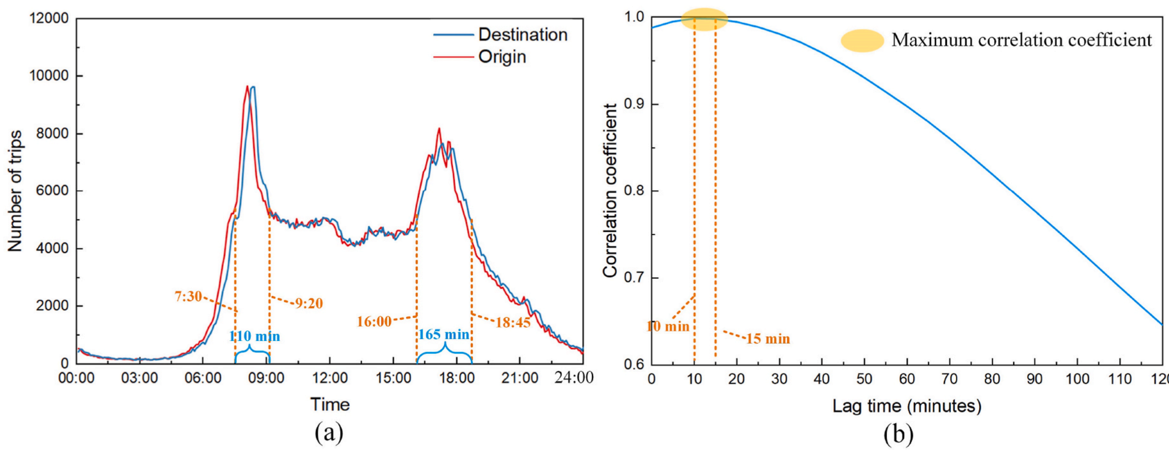


Fig. 6. Temporal distribution of e-bike trips: (a) number of origins and destinations over time, and (b) cross-correlation between the number of origins and destinations.

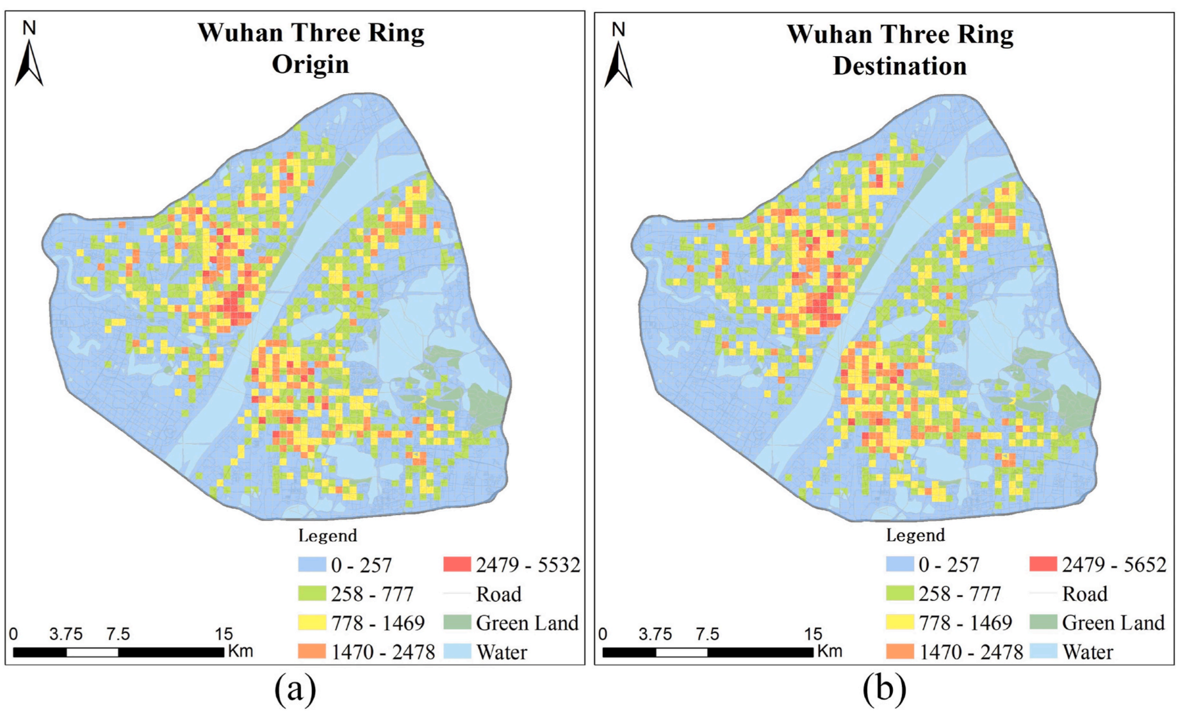


Fig. 7. Spatial distribution of e-bike trips: (a) spatial distribution of origins, and (b) spatial distribution of destinations.

that private e-bikes serve daily commuting needs. Additionally, we further quantified the lag effect between the origin curve and the destination curve using the cross-correlation function, as shown in Fig. 6 (b). The results show that the correlation coefficients of the two curves are highest when the lag time is 10 or 15 min, indicating that most e-bike trips are completed within this time frame.

Fig. 7(a)–(b) display the spatial distribution of the e-bike trip origins and destinations. The results reveal that e-bike trip hotspots are primarily concentrated in the downtown area. This concentration can be attributed to two main factors. First, downtown areas typically host numerous commercial, cultural, and recreational facilities, generating a high demand for short to medium-distance trips (He et al., 2019). Second, downtown areas often experience significant traffic congestion, while e-bikes provide a faster and more flexible travel option on city roads (Rérat, 2021). Additionally, the analysis shows that e-bike trips exhibit clear spatial heterogeneity across the study area, underscoring the necessity of the proposed SRF.

5.2. Driving mechanism of E-bike trips

In this section, we first analyze the fitting accuracy of the proposed SRF. Subsequently, we employ the interpretable SHAP method to unveil the impact mechanism of driving factors, including spatial position, built environment, traffic condition, and socioeconomic factors.

Table 4
Fitting results of spatial random forest and baselines.

Models	Origin		Destination	
	R-Square	Adj. R-Square	R-Square	Adj. R-Square
Linear Regression	0.3659	0.3417	0.3486	0.3237
Decision Tree	0.1926	0.1826	0.1865	0.1764
XGBoost	0.4964	0.4870	0.4639	0.4539
Random Forest	0.5543	0.5460	0.5273	0.5184
Spatial Random Forest	**0.5913**	**0.5836**	**0.5713**	**0.5633**

5.2.1. Modeling accuracy analysis

The Linear regression, Decision Tree, XGBoost, and Random Forest were used as baselines to analyze the advantages of the proposed SRF in terms of fitting accuracy, as shown in Table 4. The results indicate that the fitting accuracy of the XGBoost and Random Forest is significantly better than that of the Linear Regression, suggesting a nonlinear relationship between e-bike trips and the driving factors. Furthermore, the proposed SRF model outperformed the Random Forest model in fitting accuracy, highlighting the benefit of incorporating spatial factors into the conventional Random Forest model. Additional analysis was conducted to examine the stability of fitting accuracy between the proposed SRF and the baselines, as shown in Fig. 8. The results suggest that the proposed SRF model demonstrates high fitting accuracy and exhibits strong stability, affirming its advantages in modeling e-bike trips.

5.2.2. Driving mechanism analysis

Based on the modeling of e-bike trips, Fig. 9 illustrates the relative importance of driving factors to e-bike trip origins. In this visualization, the colors represent the Shapley value of the corresponding driving factor, with a longer black bar on the right indicating a greater importance of the driving factor. The results reveal that the primary impact on e-bike trip origins stems from built environment factors, including *Commercial POI Density, Hospital POI Density, Residential POI Density, School POI Density, Government POI Density,* and *Bus Station Density*. Besides the built environment, the secondary impact on e-bike trip origins comes from spatial position and socio-economic factors, such as *Population Density, Latitude, Longitude,* and *House Prices*. Finally, the factors that have the lowest impact on e-bike trip origins are traffic condition factors, such as *Congestion Level*. Given the high consistency in the relative importance of driving factors for origins and destinations, the relative importance of these factors for destinations is provided in Appendix B, as depicted in Fig. S2.

Building upon identifying the relative importance of driving factors, Fig. 10 further illustrates the impact direction of driving factors on e-bike trip origins. In this visualization, the color of each data point represents the value of the corresponding driving factor, while the position of the data point represents the Shapley value of that factor. Moreover, the positive or negative sign of the Shapley value indicates the direction of the impact direction on e-bike trip origins. For built environment factors, the number of e-bike trips tends to increase with the rise in POI density. Notably, compared to shared (e−)bike trips, transportation stations—particularly metro station density—have a relatively minor impact on private e-bike trips. This discrepancy may be due to fewer people using private e-bikes for connecting transportation purposes. The perceived risk of theft, given the high value of private e-bikes, might deter individuals from leaving them at subway stations or bus stops.

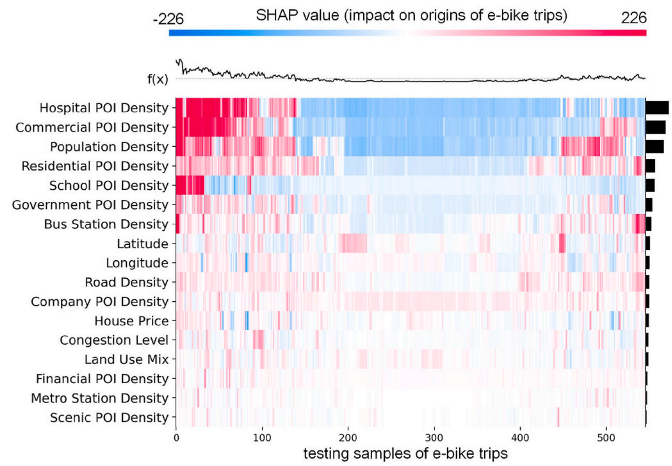


Fig. 9. Relative importance of driving factors to e-bike trip origins.

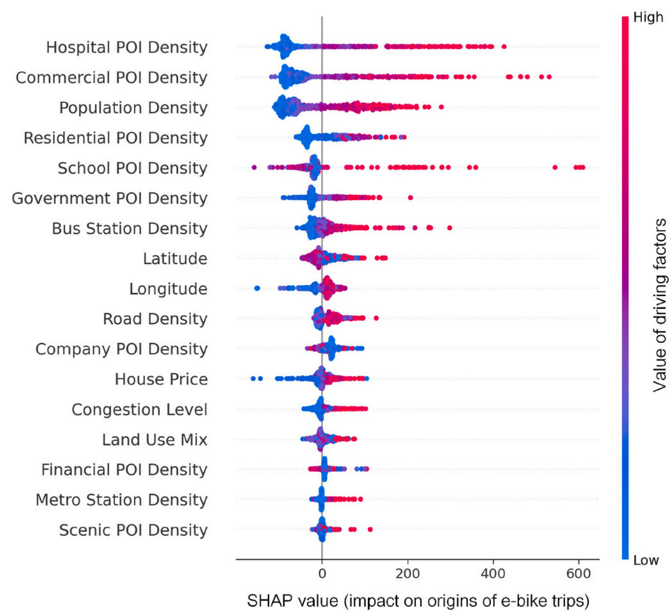


Fig. 10. Impact direction of driving factors on e-bike trip origins.

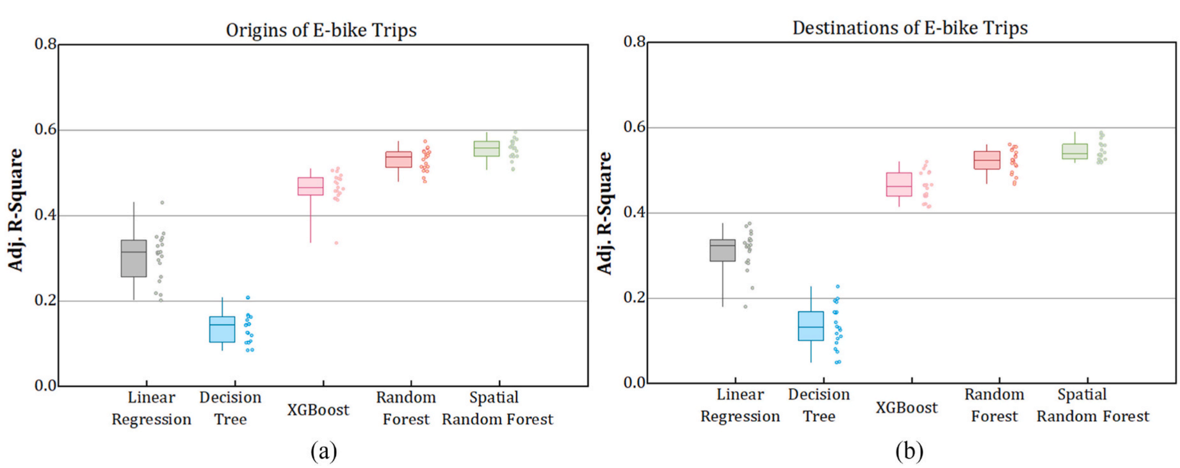


Fig. 8. Stability of spatial random forest and baselines: (a) origin, and (b) destination.

Additionally, the results suggest areas with severe traffic congestion or high housing prices tend to have more private e-bike trips. The reasons for the above results are as follows. For areas with severe congestion, private e-bikes offer a more convenient mode of transportation. For areas with high housing prices, the presence of well-developed commercial and medical facilities increases the demand for short-distance travel among residents. Similar to relative importance of driving factors, the impact direction of driving factors for destinations is provided in Appendix B, as depicted in Fig. S3.

In addition to depicting the impact direction of driving factors on e-bike trips, we further explore whether there is an interaction impact between driving factors, as shown in Supplementary Fig. S4. In this visualization, the greater the dispersion of points along the horizontal axis, the more significant the interaction between two driving factors. The results suggest that interactions among driving factors are primarily concentrated in the built environment. Taking *Commercial POI Density* and *School POI Density* as examples, Fig. 11 show the impact direction of their interaction on e-bike trips. For individual driving factor, the number of e-bikes trips will increase with the increase of *Commercial POI Density* or *School POI Density*, as shown in Fig. 11(a) or Fig. 11(c). Due to the interaction between these factors, the SHAP value exhibits a trend of initially decreasing and then increasing, as illustrated in Fig. 11(b). These results indicate that a higher number of e-bike trips is observed only when both *Commercial POI Density* and *School POI Density* are high.

6. Discussions and conclusions

Understanding human mobility and trip demand through e-bike trajectories is crucial for urban planning, environmental enhancement, and sustainable development. However, existing studies predominantly center on shared (e−)bike trips, with limited research on private e-bike trips. To fill this gap, we established a novel analysis framework based on private e-bike trajectories in Wuhan, China, and revealed the their spatiotemporal patterns using interpretable data-driven techniques.

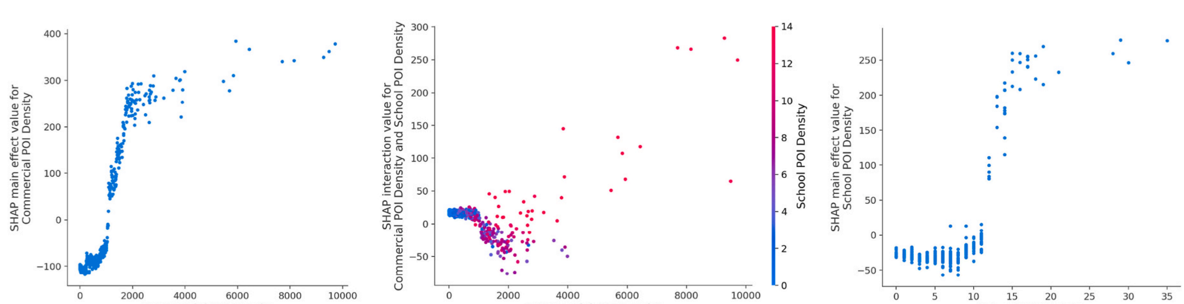
In terms of methodology, we proposed a trip analysis framework based on sparse trajectories of private e-bikes. First, we propose a two-step method, identifying staying areas and generating e-bike trips, to extract trip behavior from private e-bike sparse trajectories. Second, we establish a spatial random forest method to capture the nonlinear relationship between private e-bike trips and driving factors. Finally, we use the interpretable SHAP method to reveal the driving mechanisms of e-bike trips and explore the impact of various factors on these trips. Compared to frameworks designed for shared (e−)bike trajectories (Ji et al., 2022; Tang et al., 2024), our framework has three advantages: (1) It directly extracts trip behavior from sparse trajectories collected by monitoring stations, enriching existing studies of trip behavior extraction; (2) It accounts for spatial correlation and heterogeneity, achieving superior fitting accuracy, particularly in cases involving nonlinear relationships (L. Cheng et al., 2019; B. Zhang et al., 2023; 2024); and (3) It is versatile and can be applied not only to the spatiotemporal analysis of private e-bike trips but also to other modes of transportation, such as motor vehicles and buses.

In terms of empirical insights, we revealed the mobility patterns and driving mechanism of private e-bike trips via extensive real-world trajectories, enriching our understanding of private e-bike trips. First, the statistical distribution of trip distance and trip time for private e-bike trips is similar to that of shared (e−)bike trips(Li et al., 2024). For example, trip distances follow a lognormal distribution, with an Adj. R-Square of 0.99, while trip times exhibit a Hill distribution, with an Adj. R-Square of 0.95. Second, while shared (e−)bikes are crucial in solving the first/last mile problem in public transportation (van Kuijk et al., 2022; Yen et al., 2023; Zuo et al., 2020), the same cannot be said for private e-bikes. For example, 90% of shared bike trips are within 1.3 km(Xu et al., 2023), whereas over 65% of private e-bike trips cover distances greater than 1 km or have trip times exceeding 5 min. These findings suggest that private e-bike trips are rarely used to address the first/last mile problem in public transportation and are more commonly utilized for daily commuting needs. Third, the density of POIs such as *Hospital POI Density*, *Commercial POI Density*, *Residential POI Density, School POI Density*, and *Transportation Station Density* are significantly positively correlated with private e-bike trips. However, compared to shared (e−)bike trips (Y. Guo et al., 2021; Zhu et al., 2024), the *Transportation Station Density,* particularly *Metro Station Density,* ranks lower in importance for private e-bike trips. These results suggest that relatively few people use private e-bikes for connected trips. Finally, private e-bike trips are also positively correlated with *Congestion Level* and *House Price*, meaning that areas with severe traffic congestion or high housing prices tend to have more private e-bike trips.

There are still several limitations in this study. First, the e-bike trip extraction method is not effective in distinguishing trip purposes, such as trips for school or work. Second, in the driving mechanism analysis, we focused solely on urban-related factors such as spatial position, built environment, traffic conditions, and socioeconomic, while neglecting natural factors like weather. Finally, the analysis in this study was limited to weekdays, overlooking variations on holidays and weekends. In future work, we aim to enhance the extraction method for private e-bike trips, incorporate additional driving factors, and investigate spatiotemporal patterns across various time periods.

Funding

This project was supported by National Key Research and Development Program of China [Grant No. 2022YFB3904102]; China National Postdoctoral Support Program for Innovative Talents [Grant No. BX20230360]; China Postdoctoral Science Foundation [Grant No. 2023M743454]; National Natural Science Foundation of China [Grant Nos. 42101423 and 42371470]; Open Fund of National Engineering Research Center of Geographic Information System, China University of Geosciences, Wuhan 430074, China [Grant No. 2023KFJJ09]; Open Fund of State Key Laboratory of Information Engineering in Surveying,



(a) (b) (c)

Fig. 11. Interaction of driving factors: (a) Impact of *Commercial POI Density* on origins of e-bike trips, (b) Impact of interactions on origins of e-bike trips, and (c) Impact of *School POI Density* on origins of e-bike trips.

Mapping and Remote Sensing, Wuhan University [Grant No. 23I03]; Innovation Project of LREIS [Grant No. 08R8A092YA].

Data availability statement

The data supporting the main findings of this study are available at link on https://doi.org/10.6084/m9.figshare.25424899.

CRediT authorship contribution statement

Peixiao Wang: Writing – review & editing, Writing – original draft, Software, Funding acquisition, Conceptualization. **Hengcai Zhang:** Writing – review & editing, Writing – original draft, Project administration, Funding acquisition. **Shifen Cheng:** Writing – review & editing, Funding acquisition. **Feng Lu:** Writing – review & editing, Project administration. **Tong Zhang:** Writing – review & editing. **Zeqiang Chen:** Writing – review & editing.

Declaration of competing interest

The authors declare no conflicts of interest.

Acknowledgments

The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclepro.2024.143444.

References

Bi, H., Ye, Z., Zhu, H., 2022. Examining the nonlinear impacts of built environment on ridesourcing usage: focus on the critical urban sub-regions. J. Clean. Prod. 350, 131314 https://doi.org/10.1016/j.jclepro.2022.131314.

Bieliński, T., Kwapisz, A., Ważna, A., 2021. Electric bike-sharing services mode substitution for driving, public transit, and cycling. Transport. Res. Transport Environ. 96, 102883 https://doi.org/10.1016/j.trd.2021.102883.

Cheng, L., Chen, X., De Vos, J., Lai, X., Witlox, F., 2019. Applying a random forest method approach to model travel mode choice behavior. Travel Behaviour and Society 14, 1–10. https://doi.org/10.1016/j.tbs.2018.09.002.

Cheng, S., Zhang, B., Peng, P., Lu, F., 2023. Health and economic benefits of heavy-duty diesel truck emission control policies in Beijing. Environ. Int. 179, 108152 https://doi.org/10.1016/j.envint.2023.108152.

Choi, S.E., Kim, J., Seo, D., 2023. Travel patterns of free-floating e-bike-sharing users before and during COVID-19 pandemic. Cities 132, 104065. https://doi.org/10.1016/j.cities.2022.104065.

Consulting, IResearch, Lu, Master, 2023. White Paper on China's Two Wheeled Electric Vehicle Industry in 2023.

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. Transport. Res. Transport Environ. 75, 286–296. https://doi.org/10.1016/j.trd.2019.09.005.

Eren, E., Uz, V.E., 2020. A review on bike-sharing: the factors affecting bike-sharing demand. Sustain. Cities Soc. 54, 101882 https://doi.org/10.1016/j.scs.2019.101882.

Fu, C., Huang, Z., Scheuer, B., Lin, J., Zhang, Y., 2023. Integration of dockless bike-sharing and metro: prediction and explanation at origin-destination level. Sustain. Cities Soc. 99, 104906 https://doi.org/10.1016/j.scs.2023.104906.

Fukushige, T., Fitch, D.T., Handy, S., 2021. Factors influencing dock-less E-bike-share mode substitution: evidence from Sacramento, California. Transport. Res. Transport Environ. 99, 102990 https://doi.org/10.1016/j.trd.2021.102990.

Fyhri, A., Fearnley, N., 2015. Effects of e-bikes on bicycle use and mode share. Transport. Res. Transport Environ. 36, 45–52. https://doi.org/10.1016/j.trd.2015.02.005.

Guidon, S., Becker, H., Dediu, H., Axhausen, K.W., 2019. Electric bicycle-sharing: a new competitor in the urban transportation market? An empirical analysis of transaction data. Transport. Res. Rec. 2673 (4), 15–26. https://doi.org/10.1177/0361198119836762.

Guidon, S., Reck, D.J., Axhausen, K., 2020. Expanding a(n) (electric) bicycle-sharing system to a new city: prediction of demand with spatial regression and random forests. J. Transport Geogr. 84, 102692 https://doi.org/10.1016/j.jtrangeo.2020.102692.

Guo, Y., Yang, L., Lu, Y., Zhao, R., 2021. Dockless bike-sharing as a feeder mode of metro commute? The role of the feeder-related built environment: analytical framework and empirical evidence. Sustain. Cities Soc. 65, 102594 https://doi.org/10.1016/j.scs.2020.102594.

Guo, Z., Liu, J., Zhao, P., Li, A., Liu, X., 2023. Spatiotemporal heterogeneity of the shared e-scooter–public transport relationships in Stockholm and Helsinki. Transport. Res. Transport Environ. 122, 103880 https://doi.org/10.1016/j.trd.2023.103880.

He, S., Su, Y., Shahtahmassebi, A.R., Huang, L., Zhou, M., Gan, M., Deng, J., Zhao, G., Wang, K., 2019. Assessing and mapping cultural ecosystem services supply, demand and flow of farmlands in the Hangzhou metropolitan area, China. Sci. Total Environ. 692, 756–768. https://doi.org/10.1016/j.scitotenv.2019.07.160.

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. J. Transport Geogr. 103, 103414 https://doi.org/10.1016/j.jtrangeo.2022.103414.

Li, Q., Zhang, E., Luca, D., Fuerst, F., 2024. The travel pattern difference in dockless micro-mobility: shared e-bikes versus shared bikes. Transport. Res. Transport Environ. 130, 104179 https://doi.org/10.1016/j.trd.2024.104179.

Liu, S., Zhang, F., Ji, Y., Ma, X., Liu, Y., Li, S., Zhou, X., 2023. Understanding spatial-temporal travel demand of private and shared e-bikes as a feeder mode of metro stations. J. Clean. Prod. 398, 136602 https://doi.org/10.1016/j.jclepro.2023.136602.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30. In: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

McCaffery, C., Zhu, H., Tang, T., Li, C., Karavalakis, G., Cao, S., Oshinuga, A., Burnette, A., Johnson, K.C., Durbin, T.D., 2021. Real-world NOx emissions from heavy-duty diesel, natural gas, and diesel hybrid electric vehicles of different vocations on California roadways. Sci. Total Environ. 784, 147224 https://doi.org/10.1016/j.scitotenv.2021.147224.

McKenzie, G., 2020. Urban mobility in the sharing economy: a spatiotemporal comparison of shared mobility services. Comput. Environ. Urban Syst. 79, 101418 https://doi.org/10.1016/j.compenvurbsys.2019.101418.

McQueen, M., MacArthur, J., Cherry, C., 2020. The E-Bike Potential: estimating regional e-bike impacts on greenhouse gas emissions. Transport. Res. Transport Environ. 87, 102482 https://doi.org/10.1016/j.trd.2020.102482.

Parsa, A.B., Movahedi, A., Taghipour, H., Derrible, S., Mohammadian, A., Kouros, 2020. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. Accid. Anal. Prev. 136, 105405 https://doi.org/10.1016/j.aap.2019.105405.

Plazier, P.A., Weitkamp, G., Van Den Berg, A.E., 2017. "Cycling was never so easy!" An analysis of e-bike commuters' motives, travel behaviour and experiences using GPS-tracking and interviews. J. Transport Geogr. 65, 25–34. https://doi.org/10.1016/j.jtrangeo.2017.09.017.

Reck, D.J., Haitao, H., Guidon, S., Axhausen, K.W., 2021. Explaining shared micromobility usage, competition and mode choice by modelling empirical data from Zurich, Switzerland. Transport. Res. C Emerg. Technol. 124, 102947 https://doi.org/10.1016/j.trc.2020.102947.

Rérat, P., 2021. The rise of the e-bike: towards an extension of the practice of cycling? Mobilities 16 (3), 423–439. https://doi.org/10.1080/17450101.2021.1897236.

Rich, J., Jensen, A.F., Pilegaard, N., Hallberg, M., 2021. Cost-benefit of bicycle infrastructure with e-bikes and cycle superhighways. Case Studies on Transport Policy 9 (2), 608–615. https://doi.org/10.1016/j.cstp.2021.02.015.

Tang, J.H.C.G., Huang, Y., Zhu, Y., Yang, X., Zhuge, C., 2024. The association between travel demand of docked bike-sharing and the built environment: evidence from seven US cities. Sustain. Cities Soc. 105325 https://doi.org/10.1016/j.scs.2024.105325.

Valipour Shokouhi, B., de Hoogh, K., Gehrig, R., Eeftens, M., 2024. Estimation of historical daily airborne pollen concentrations across Switzerland using a spatio temporal random forest model. Sci. Total Environ. 906, 167286 https://doi.org/10.1016/j.scitotenv.2023.167286.

van Kuijk, R.J., de Almeida Correia, G.H., van Oort, N., van Arem, B., 2022. Preferences for first and last mile shared mobility between stops and activity locations: a case study of local public transport users in Utrecht, The Netherlands. Transport. Res. Pol. Pract. 166, 285–306. https://doi.org/10.1016/j.tra.2022.10.008.

Venkadavarahan, M., Joji, M.S., Marisamynathan, S., 2023. Development of spatial econometric models for estimating the bicycle sharing trip activity. Sustain. Cities Soc. 98, 104861 https://doi.org/10.1016/j.scs.2023.104861.

Xu, X., Wang, J., Poslad, S., Rui, X., Zhang, G., Fan, Y., 2023. Exploring intra-urban human mobility and daily activity patterns from the lens of dockless bike-sharing: a case study of Beijing, China. Int. J. Appl. Earth Obs. Geoinf. 122, 103442 https://doi.org/10.1016/j.jag.2023.103442.

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. J. Transport Geogr. 104, 103453 https://doi.org/10.1016/j.jtrangeo.2022.103453.

Yen, B.T.H., Mulley, C., Yeh, C.-J., 2023. How public shared bike can assist first and last mile accessibility: a case study of the MRT system in Taipei City, Taiwan. J. Transport Geogr. 108, 103569 https://doi.org/10.1016/j.jtrangeo.2023.103569.

Yu, S., Liu, G., Yin, C., 2021. Understanding spatial-temporal travel demand of free-floating bike sharing connecting with metro stations. Sustain. Cities Soc. 74, 103162 https://doi.org/10.1016/j.scs.2021.103162.

Yu, Y., Jiang, Y., Qiu, N., Guo, H., Han, X., Guo, Y., 2022. Exploring built environment factors on e-bike travel behavior in urban China: a case study of Jinan. Front. Public Health 10. https://www.frontiersin.org/articles/10.3389/fpubh.2022.1013421.

Zhang, S., Liu, X., Tang, J., Cheng, S., Wang, Y., 2019. Urban spatial structure and travel patterns: analysis of workday and holiday travel using inhomogeneous Poisson point process models. Comput. Environ. Urban Syst. 73, 68–84. https://doi.org/10.1016/j.compenvurbsys.2018.08.005.

Zhang, B., Cheng, S., Zhao, Y., Lu, F., 2023a. Inferring intercity freeway truck volume from the perspective of the potential destination city attractiveness. Sustain. Cities Soc. 98, 104834 https://doi.org/10.1016/j.scs.2023.104834.

Zhang, Z., Krishnakumari, P., Schulte, F., van Oort, N., 2023b. Improving the service of E-bike sharing by demand pattern analysis: a data-driven approach. Res. Transport. Econ. 101, 101340 https://doi.org/10.1016/j.retrec.2023.101340.

Zhang, B., Cheng, S., Wang, P., Lu, F., 2024a. Inferring freeway traffic volume with spatial interaction enhanced betweenness centrality. Int. J. Appl. Earth Obs. Geoinf. 129, 103818 https://doi.org/10.1016/j.jag.2024.103818.

Zhang, Y., Manley, E., Martens, K., Batty, M., 2024b. A metro smart card data-based analysis of group travel behaviour in Shanghai, China. J. Transport Geogr. 114, 103764 https://doi.org/10.1016/j.jtrangeo.2023.103764.

Zhou, X., Ji, Y., Yuan, Y., Zhang, F., An, Q., 2022. Spatiotemporal characteristics analysis of commuting by shared electric bike: a case study of Ningbo, China. J. Clean. Prod. 362, 132337 https://doi.org/10.1016/j.jclepro.2022.132337.

Zhou, X., Dong, Q., Huang, Z., Yin, G., Zhou, G., Liu, Y., 2023. The spatially varying effects of built environment characteristics on the integrated usage of dockless bike-sharing and public transport. Sustain. Cities Soc. 89, 104348 https://doi.org/10.1016/j.scs.2022.104348.

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

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