Integrating shared e-scooters as the feeder to public transit: A comparative analysis of 124 European cities

Citation for the original published paper (version of record):

Li, A., Gao, K., Zhao, P. et al (2024). Integrating shared e-scooters as the feeder to public transit: A comparative analysis of 124
European cities. Transportation Research, Part C: Emerging Technologies, 160.
http://dx.doi.org/10.1016/j.trc.2024.104496

(article starts on next page)

Transportation Research Part C 160 (2024) 104496

Contents lists available at ScienceDirect

Transportation Research Part C

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

Integrating shared e-scooters as the feeder to public transit: A comparative analysis of 124 European cities

Aoyong Li [a,b], Kun Gao [c,*], Pengxiang Zhao [d,e], Kay W. Axhausen [f]

[a] *State Key Lab of Intelligent Transportation System, Beihang University, Beijing, China*
[b] *Hangzhou Innovation Institute, Beihang University, Hangzhou, China*
[c] *Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, Sweden*
[d] *GIS Centre, Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden*
[e] *K2 — The Swedish Knowledge Centre for Public Transport, Lund, Sweden*
[f] *Institute for Transport Planning and Systems (IVT), ETH Zürich, Zürich, Switzerland*

ARTICLE INFO

Keywords:
Shared micro-mobility
Integration with public transit
First-and-last-mile problem
Spatial and temporal patterns
Explainable machine learning

ABSTRACT

E-scooter sharing is a potential feeder to complement public transit for alleviating the first-and-last-mile problem. This study investigates the integration between shared e-scooters and public transit by conducting a comparative analysis in 124 European cities based on vehicle availability data. Results suggest that the integration ratios of e-scooter sharing in different cities show significant variations and range from 5.59% to 51.40% with a mean value of 31.58% and a standard deviation of 8.47%. The temporal patterns of integration ratio for first- and last-mile trips present an opposite trend. An increase in the integration ratio for first-mile trips is related to a decrease in the integration ratio for last mile in the time series. Additionally, these cities can be divided into four clusters according to their temporal variations of the integration ratios by a bottom-up hierarchical clustering method. Meanwhile, we explore the nonlinear effects of city-level factors on the integration ratio using explainable machine learning. Several factors are found to have noticeable and nonlinear influences. For example, the density of public transit stations and a higher ratio of the young are positively associated with the integration ratio to a certain extent. The results potentially support transport planners to collectively optimize and manage e-scooter sharing and public transport to facilitate multi-modal transport systems.

1. Introduction

Shared micro-mobility systems, including electric scooter (e-scooter) sharing and bike sharing, have gained popularity worldwide due to various benefits. Shared e-scooters are usually dockless, they offer greater flexibility and allow for less physical effort due to their smaller size compared with shared bikes (Gössling, 2020; Gao et al., 2021a; Li et al., 2022; Attard, 2022). E-scooter sharing has been commonly considered a feasible solution for the first-and-last-mile problem due to its convenience and suitability for short-distance trips (Zuo et al., 2018; Li et al., 2020; Gao et al., 2021b; Li et al., 2021b). Its potential for seamless integration with public transit systems paves the way for enhanced multi-modal travel. Thus, understanding the usage patterns of e-scooter sharing as a feeder mode to public transit is crucial for informing transportation policy and offering strategic improvements.

In recent years, a strand of studies has investigated the integration of shared micro-mobility systems (e.g., e-scooter and bike sharing) with public transit, such as exploring the relationships between them (Kong et al., 2020; Gao et al., 2023a), analyzing the

* Corresponding author.
E-mail address: gkun@chalmers.se (K. Gao).

https://doi.org/10.1016/j.trc.2024.104496
Received 12 March 2023; Received in revised form 23 January 2024; Accepted 24 January 2024
Available online 31 January 2024
0968-090X/

spatial and temporal patterns of first-and-last mile trips using shared micro-mobility (Fearnley et al., 2020; Guo et al., 2023), the accessibility of shared micro-mobility services (Zuo et al., 2020; Li et al., 2020), the value of shared micro-mobility for solving the first-and-last mile problems (Baek et al., 2021; Luo et al., 2021; Zuniga-Garcia et al., 2022), and the determinants of travel behavior using shared micro-mobility as a feeder to public transit system (Cheng et al., 2022; Liu et al., 2022).

However, these studies predominantly focus on bike-sharing systems, with limited attention to e-scooter sharing. Additionally, they are often constrained by using data from a single city, limiting the generalizability of conclusions. As far as we know, no study has examined the integration of shared e-scooters and public transit in multiple cities and conducted a comparative analysis to reveal their similarity and divergence under various urban contexts. The lack of comprehensive data analytic results from multiple-city comparison obstructs the efficient promotion of the integration of e-scooter sharing and public transit towards multimodal transportation systems. In addition, single-city research is also insufficient to unravel the mechanism behind data analytic results concerning the integration of e-scooter sharing and public transit. Furthermore, a number of built environment (e.g., road network, public transport stops, city size) and socio-demographic factors (e.g., population and employment density) could have substantial impacts on the usage of shared micro-mobility for first-and-last mile problem, namely integration of shared micro-mobility and public transit systems. These factors vary across cities (Fishman, 2016; Kong et al., 2020). Thus, the integration of shared micro-mobility and public transit and the corresponding usage patterns of using shared micro-mobility as a feeder to public transit may vary across different cities, which needs further explorations to provide insights about the underlying mechanism.

In this study, we collected the vehicle availability data from e-scooter sharing systems in more than one hundred European cities. Our aim is to investigate to what extent shared e-scooters are used to integrate with public transit for first- and last-mile problems in a data-driven manner and how the integration is influenced by city-level factors using interpretable machine learning. We aim to address the following three research questions (RQ).

1. **RQ1**: Do the integration ratios of shared e-scooters with public transit present similar or distinct patterns in different cities across space and time?
2. **RQ2**: What factors can potentially explain the variations in the integration ratios of shared e-scooters with public transit among different cities?
3. **RQ3**: What are the quantitative impacts of different city-level features on the integration ratios of shared e-scooters with public transit among different cities?

By combining the e-scooter sharing trip data with public transit systems, we identify the trips of using shared e-scooters as a feeder to public transit by developing an inference method. Based on the massive trip data in 124 cities, the integration ratios of e-scooter sharing with public transit at the city level are quantitatively calculated for each city. Herein, the integration ratio is defined as the percentage of trips using e-scooter sharing as a feeder to public transit over all e-scooter sharing trips. In particular, the spatial and temporal patterns of integration ratios in these cities are analyzed. A comparative analysis is further implemented to understand their similarities and dissimilarities using unsupervised clustering methods. Moreover, efforts have been made to explain the mechanism behind the differences among different cities. Multi-source data are collected to extract and quantify the factors depicting city-level features, and utilized to investigate the impacts of these factors on the integration ratio of e-scooter sharing and public transit. A machine learning approach is employed to measure the relationship between potential factors and the integration ratio. Meanwhile, a data-driven interpretation method is leveraged to explain the black box of machine learning and uncover the nonlinear effects of each factor on the integration ratio. To our knowledge, this is the first study to provide insights and knowledge for the above-mentioned three research questions through comparison analysis of more than 100 cities, which offers unique empirical evidence and support for promoting the integration of shared micro-mobility and public transit.

The remainder of this paper is structured as follows. Section 2 presents a review of relevant literature. Section 3 introduces the collected e-scooter sharing data and study areas, followed by the methodology in Section 4. Afterward, we present the results accompanied by discussions in Section 5. Finally, Section 6 summarizes the findings and potential future work.

2. Literature review

Shared micro-mobility services, particularly suitable for short-distance urban travel, are widely utilized as feeders to public transit in response to first-and-last-mile problem. Many studies have examined the integration of various shared micro-mobility types with public transit, focusing on usage patterns, accessibility evaluation, and behavioral analysis.

2.1. Exploring the relationship between shared micro-mobility and public transit

The first strand focuses on exploring the relationship (e.g., competing and complementing) between shared micro-mobility systems and public transit systems. Most relevant studies employed survey-based methods to analyze the relationship of shared micro-mobility with other transport modes (Shaheen et al., 2013; Fishman et al., 2014; McQueen et al., 2021; Montes et al., 2023). For example, questions such as "Did you use shared e-scooters to access a bus?" and "How often do you use e-scooters to access a bus?" were proposed to examine whether the users utilize the e-scooter sharing for the integration with public transit in surveys (Portland Bureau of Transportation, 2019; Chicago Department of Transportation, 2020). According to survey results in Chicago Department of Transportation (2020), the shared micro-mobility systems may compete with or complement public transit systems. In Chicago, more than 30% of the respondents would use e-scooters to get to or from the stations of public transit such as bus, rail,

Table 1
The summary of numerical findings.

Reference	Time	Area	Main findings
Chicago Department of Transportation (2020)	2020	Chicago	More than 30% of the respondents use shared e-scooters to connect public transits
San Francisco Municipal Transportation Agency's (2019)	2019	San Francisco	Nearly 30% of users use shared e-scooters to connect public transits
6t-bureau de recherche (2019)	April, 2019	French	About 23% of e-scooter trips are combined with other transport modes.
Christoforou et al. (2021)	May to Jun, 2019	Paris	About 37% of e-scooter trips are combined with other transport modes.
Fearnley et al. (2020)	November to December, 2019	Oslo	Nearly 23% of respondents would use e-scooter to connect public transit modes.
Aarhaug et al. (2023)	October to November 2021	Oslo	About 20% of e-scooter trips are combined with other public transit modes.
Transport for London (2019)	June 2021 to November 2022	London	32% of e-scooter trips were integrated with a mode of public transit modes.

or Metro. In 2019, nearly 30% of e-scooter users took the e-scooter as the first and last-mile connection tool, according to the survey made by the San Francisco Municipal Transportation Agency (SFMTA) (San Francisco Municipal Transportation Agency's, 2019). In Europe, researchers implemented a quantitative survey in many French cities, where a total of 4382 responses were collected from users of e-scooter sharing and 21 semi-structured interviews were conducted with Lime users. In this survey, the usage patterns of e-scooter sharing and user profiles were analyzed, which indicated that about 23% of e-scooter trips were combined with other transportation modes in France (6t-bureau de recherche, 2019; Laa and Leth, 2020). Christoforou et al. (2021) focused on Paris and implemented surveys from 7 predefined locations and collected questionnaires from 459 respondents between May and June 2019. The answers covered travel habits, general usage, last trip, and socio-demographic information. Results showed that about 37% of e-scooter trips were combined e-scooters with public transport modes (i.e., buses, metro, Reseau Express Regional, and streetcar). Between November 8th and December 1st 2019, Fearnley et al. (2020) performed a web survey on e-scooter sharing usage in Oslo, where 549 responses were collected. According to the results, about 57% of respondents would use e-scooter sharing in combination with other transport modes. Specifically, nearly 23% of them would use e-scooter sharing as the first- and last-mile for public transport modes. Aarhaug et al. (2023) also took Oslo as the study area and collected 1921 responses in 2021. The study reached a similar conclusion, where approximately 20% e-scooter trips were integrated with public transport modes.

Survey-based studies, while providing empirical and useful insights at the aggregated or average levels, often lack the spatial and temporal details in their data, which were insufficient for high-resolution analysis (Luo et al., 2021; Naroa et al., 2022). With advancements in information and communication technologies, the shared micro-mobility data can be collected passively, allowing more detailed studies of the relationship between shared micro-mobility and public transport systems based on inference approaches (Campbell and Brakewood, 2017; Kong et al., 2020; Zuniga-Garcia et al., 2022; Cheng et al., 2022). A rough idea of the relationship between shared micro-mobility and public transit is based on the distance of the trip origin and destination to nearby public transit stations (Lin et al., 2019; Chen et al., 2022). For example, Cheng et al. (2022) extracted the integration trips of dockless bike-sharing services and urban rail transport systems by building a 50-meter buffer for each rail station entrance. However, the distance to public transit stations is not the only determinant of the relationship between shared micro-mobility and public transit. Consequently, some studies proposed to combine additional information such as bus schedules, surveys, user characteristics, and trip purposes, to identify the integration of shared micro-mobility and public transit (Zuniga-Garcia et al., 2022; Luo et al., 2021). Zuniga-Garcia et al. (2022) proposed a two-stage framework to assess the relationship between e-scooter sharing and bus transit services by isolating the effects of several factors including spatial location, time of the day, weekday, and weather. To some degree, their methods are more reasonable while needing more input information (i.e., more data resources). Taking advantage of these shared micro-mobility data between June 2021 and November 2022, the Transport for London (TfL) analyzed the usage patterns of e-scooter sharing comprehensively and found that 32% of e-scooter trips were integrated with a mode of public transport, in a combination of London Councils and three e-scooter operators (Transport for London, 2019). To improve the readability, we summarize the numerical findings in Table 1.

2.2. The determinants of the integration between shared micro-mobility and public transit

Moreover, some other studies made efforts to investigate the integration of shared micro-mobility and public transport systems by exploring the determinants. The practical aim is to understand the determinants and promote the integration of shared micro-mobility with public transit by improving relevant factors such as infrastructure or built environments. To name a few, Cheng et al. (2022) employed a quantile regression approach to estimate the nonlinear effects of the built environment on the integration of the dockless bike-sharing system and urban rail transport in Nanjing, China. Their results demonstrated that built environment factors had various impacts and magnitudes by selecting a different set of rail stations according to the quantiles of integration demand. For example, the relationship between the length of minor roads and the integrated use shows a strong relationship for the stations at a low quantile, while becoming less significant at medium and high quantiles. Liu et al. (2022) analyzed the impact of passengers' socio-economic attributes and perceived congestion on the route choice behavior in combined travels using bike-sharing systems and metro systems. Baek et al. (2021) regarded e-scooter sharing as a potential alternative for last-mile trips and designed a stated preference (SP) experiment in Seoul to examine the value of travel time of using e-scooter sharing. Their results illustrated that e-scooters can be better integrated with subway stations. Ziedan et al. (2021) studied the influences of e-scooters on transit ridership in Louisville, Kentucky, using fixed effects regression models. They found that e-scooter sharing did not have a significant impact on local bus ridership, but it was unclear if e-scooter sharing could complement public transit as first- and last-mile feeder transport modes. In addition, Liu et al. (2022) utilized smart card data to explore the route choice behavior of bike-metro trips (i.e., using bike sharing as the feeder to metro stations). They employed a multinomial logit model to model the route choice behavior and quantify the impacts of different factors, including passengers' socio-economic attributes and perceived congestion.

2.3. The comparison among multi cities

Relevant studies have shown the significance of the integration of shared micro-mobility and public transport. However, these studies mainly focus on only a certain or small number of cities, ignoring the variety among different cities with different contexts. Little attention has been paid to the comparison of the integration of shared micro-mobility and public transport systems in multiple cities. For instance, Kou and Cai (2019) compared the travel patterns of bike-sharing systems in eight cities, but only the trip distance and duration were analyzed without focusing on the integration of bike-sharing with public transit. Li et al. (2022) conducted a comprehensive comparison of shared e-scooter systems in 30 cities from four perspectives, including temporal usage patterns, statistical characteristics, utilization efficiency, and wasted electricity during idle time. The integration of e-scooter sharing with public transit was not investigated at all. As far as we are concerned, there is no existing study that reveals and compares the integration of e-scooter sharing with public transit in multiple cities, which is important in determining appropriate and tailored planning and management in different urban contexts. Meanwhile, no study has analyzed the quantitative impacts of different city-level features on the integration of e-scooter sharing with public transit across different cities, because such a study requires massive data from multiple cities.

In this study, we address these gaps by employing empirical data on e-scooter sharing from 124 cities to explore how e-scooter sharing systems integrate with public transport. We analyze and compare spatio-temporal integration patterns and utilize multi-source data to assess city-level influences on the patterns. By applying explainable machine learning, we aim to elucidate the non-linear effects of these factors on the integration of e-scooter sharing and public transit, contributing unique insights to this field of research.

3. Data and study area

To achieve this aim, we collected the vehicle availability data from e-scooter sharing systems across 124 cities in 15 European countries for the analysis. These countries include Germany, Sweden, the United Kingdom, Italy, France, Poland, Switzerland, Norway, Austria, Finland, Belgium, the Netherlands, Spain, Hungary, and Slovakia. Significantly, the growing market for shared micro-mobility in European cities makes them ideal subjects for this analysis (O'Brien, 2021).

The original vehicle availability data includes the locations of all idle shared e-scooters in a city, which are collected continuously every 10 s. Based on the raw data, the e-scooter sharing trips are extracted using the method in Zhao et al. (2021). For each recorded trip, the available information consists of trip ID, e-scooter identifier, longitude and latitude of the origin and destination of the recorded trip, and starting and ending timestamps of the trip. Due to potential issues like GPS positioning errors or manual re-balancing operations, trips erroneously recorded were filtered out following the criteria outlined in Li et al. (2022).

Table 2 summarizes the data statistics at the country level. The statistics at the city level of 124 cities are available in the supplementary materials and not presented herein in case of redundancy. The columns of "No. City", "No. E-scooters", "No. trips" represent the number of cities, the number of e-scooters, and the number of trips in all cities in the given country, respectively. It is important to note that the data collection periods across different cities were not the same. "Days (Sum)" in Table 2 denotes the sum of number of days in the data for all cities in a country. "Days (Max)" and "Days (Min)" in Table 2 represent the maximum and minimum number of days in the available data of a city in the country, respectively. The period ranges from three months to more than one year, reflecting the progressive nature of our data collection process. Cities with earlier data collection, such as cities in Switzerland, exhibit longer collection periods, whereas some other cities, particularly in the UK, display shorter data collection duration. "Average daily trips per city" refers to the average daily number of trips per city in a country.

4. Methodology

We first collect the vehicle availability data from e-scooter sharing systems and extract all trips using shared e-scooters in each city. Secondly, we identify the trips that integrate e-scooter sharing and public transit, namely the trips using shared e-scooters as the feeder transport mode for the first-mile and last-mile travel to public transit stations. In this regard, the integration ratio of using e-scooter sharing as the feeder to public transport in each city can be attained. Afterward, the spatial and temporal patterns of the integration ratios in different cities are investigated and compared. Lastly, an explainable machine learning method is employed to explore the influences of city-level factors on the integration. The city-level factors are extracted and fused from multiple data resources.

Table 2
The summary of the data used for analysis.

Country	No. City	No. E-scooters	No. trips	Days (sum) (Sum days in all cities)	Days (max) (Max days in a city)	Days (min) (Min days in a city)	Average daily trips per city
Austria	4	4,675	1,833,427	880	440	146	2,083
Belgium	2	8,528	669,004	258	129	129	2,593
Finland	3	34,226	8,870,938	1,117	407	350	7,942
France	7	26,724	8,728,435	2,046	446	144	4,266
Germany	44	176,527	46,457,741	10,776	497	129	4,311
Hungary	1	3,211	182,676	147	147	147	1,243
Italy	9	13,679	2,215,159	1,946	479	129	1,138
Netherlands	2	1,276	170,147	293	147	146	581
Norway	5	36,905	12,553,214	1,475	494	163	8,511
Poland	9	16,504	2,132,406	1,932	303	143	1,104
Slovakia	1	1,769	136,648	299	299	299	457
Spain	2	2,468	848,751	425	295	130	1,997
Sweden	16	55,450	13,915,142	4,292	433	145	3,242
Switzerland	7	8,845	2,156,108	2,953	507	130	730
United Kingdom	12	29,709	7,140,771	2,557	306	128	2,793
Total	124	420,496	108,010,567	31,396	–	–	3,440

4.1. Identifying e-scooter trips integrating with public transit

As the e-scooter trip data are passively collected, no information can directly reveal whether the user used the shared e-scooter as a feeder to get access to public transit or not, in a specific recorded trip. Several inference methods have been proposed in relevant literature to identify such feeder trips. In this study, a straight-forward but widely adopted method is employed. The approach is based on the distance between the origin and destination of an e-scooter trip and nearby transit stations (Lin et al., 2018; Chen et al., 2022; Cheng et al., 2022). An e-scooter trip is regarded as a first-mile trip (FMT) to public transit if the distance between the end point of the e-scooter trip and the nearest transit station is smaller than a given threshold and the distance between starting point of the e-scooter trip and the nearest transit station is larger than the given threshold. If the distance between the starting point of an e-scooter trip and the nearest transit station is shorter than a given threshold, and the distance from the ending point of the e-scooter trip to its nearest transit station is larger than the given threshold, the e-scooter trip will be treated as the last-mile trip (LMT). The distance threshold is set as 50 meters according to existing studies (Chen et al., 2022; Cheng et al., 2022). Please note that there is no perfect approach to infer whether a trip using e-scooter sharing is integrated with public transit or not. Some other methods may combine more information for identifying the integration of e-scooter sharing and public transit. However, considering that our study covers the e-scooter sharing data of 124 cities, it is hard to collect additional data for all the cities. Therefore, we adopt a practically feasible and well-utilized method from the relevant literature.

Again, we focus on the integration ratio of using e-scooter sharing with public transport systems. The integration ratio refers to the percentage of trips using e-scooter sharing as a feeder to public transit in all trips using e-scooter sharing, reflecting the probability of using e-scooter sharing to integrate with public transit. For a certain period in a city, the ratio of first-mile integration trip (RFMT) and the ratio of last-mile integration trip (RLMT) are defined as the number of FMT and LMT divided by the total number of e-scooter trips.

$$RFMT = \frac{N_{FMT}}{N_T} \tag{1}$$

$$RLMT = \frac{N_{LMT}}{N_T} \tag{2}$$

where N_{FMT} and N_{LMT} denote the number of FMT and LMT, respectively. N_T is the total number of e-scooter trips in the period. The sum of RFMT and RLMT at a given period for a city is defined as the ratio of integration trip (RIT).

4.2. Temporal signature and clustering analysis

We propose a temporal signature to capture the temporal fluctuations (or patterns) of RFMT and RLMT in a city. It is also utilized to measure the similarities and dissimilarities of integration ratio among different cities. Considering the difference in the covered periods of data in different cities, the temporal signature is calculated for each city based on the day of the week and the hour of the day. Specifically, the RFMT and RLMT are first calculated for each hour in a day and in each city. Then, the average value for a specific hour among different weeks is calculated as the element in the temporal signature, which can be expressed as:

$$S_i = [E_{i,1,1}, \dots, E_{i,j,k}, \dots, E_{i,7,24}] \tag{3}$$

where S_i is a 1×168 vector that covers the average RFMT or RLMT in each hour from Monday to Sunday. i is the city number. j is from 1 to 7 representing the day of the week, i.e., Monday to Sunday. k is from 1 to 24, representing the hour in a day. The temporal signatures for RFMT and RLMT for ith city are calculated and denoted as S_i^f and S_i^l, respectively.

Additionally, we grouped these cities into different clusters, which can reflect the similarities and dissimilarities between different groups. To consider both RFMT and RLMT in temporal signatures simultaneously, the RFMT and RLMT for a city are concatenated into a new vector (S_i^{fl}) with 336 elements, where i represents the ith city. To find the similarity of cities in different levels, a bottom-up hierarchical clustering method is adopted. Each city (S_i^{fl}) starts as its own cluster, and then, pairs of clusters are successively

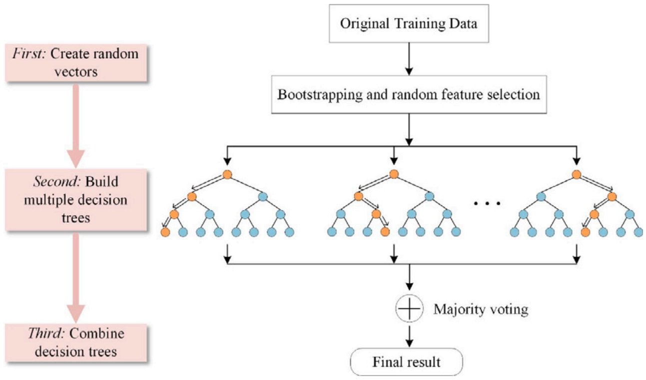


Fig. 1. Illustration of the random forest method.
Source: Cheng et al. (2019).

merged together according to the similarity among clusters. Finally, a hierarchical tree with different levels of similarity is generated, where cities within the lower level cluster are more similar than the cities within the high level cluster. Specifically, suppose s, t represent the cluster in a lower level, and they are clustered into cluster u and v in a higher level. To represent the similarity between different clusters, we adopt the Euclidean distance according to Li et al. (2022), which is defined as:

$$d(s,t) = \sqrt{(s_1 - t_1)^2 + (s_2 - t_2)^2 + \cdots \ldots + (s_{336} - t_{336})^2} \tag{4}$$

Afterward, it is necessary to determine the method calculating the distance between newly generated clusters. The Ward's method is employed to minimize the variances of S_i^{fl} within a group (Ward, 1963). As a incremental, it is formulated as:

$$dis(u,v) = \sqrt{\frac{|u|+|s|}{T}d(v,s)^2 + \frac{|u|+|t|}{T}d(v,t)^2 - \frac{|v|}{T}d(s,t)^2} \tag{5}$$

where s and t are jointed as the new cluster u, v is other unused clusters in the tree, $|*|$ is the number of elements in this vector and $T = |u| + |s| + |t|$.

4.3. Influences of city-level factors on RFMT and RLMT

After analyzing RFMT and RLMT, we employ explainable machine learning to quantitatively reveal the potential associations between the RFMT or RLMT and the factors at the city level. The machine learning method is adopted since it does not make any assumptions on the relationships between city-level factors and dependent variable, and can unravel the impacts of factors in a data-driven way (Gao et al., 2021c).

We use the random forest to model the relations between integration ratios and city-level factors. The dependent variables are RFMT and RLMT calculated in Section 4.1. The explanatory variables at the city level are calculated for each city based on multiple data sources. As a tree-based model, the random forest has been widely applied, which combines multiple individual decision trees by using the bootstrapping method to optimize model fit and prediction. The process of a random forest model can be illustrated in Fig. 1. Different observations are bootstrapped for each tree. Also, the candidate splitting variables in each tree are chosen by a random selection from the full set of explanatory variables. In each single tree, splitting is continued until the tree reaches the maximum depth. After estimating base tree models, the majority voting strategy is employed to compute the result. The final predicted outcome is the average of the predicted values of all trees.

We divide the city-level explanatory factors into five categories, including socio-demographic variables, land use variables, public transit variables, road network variables, and other variables. Considering that a city encompasses various terrains, such as forests, mountains, and rivers, where shared e-scooters are not available or rarely utilized, the municipal boundaries of a city do not accurately represent the real service area of the e-scooter sharing within the city as shared e-scooters mainly operate in central areas of European cities. Hence, we focus on the e-scooter sharing service areas for our analysis in calculating city-level factors, and all relevant factors are calculated within the defined service areas. We estimate the service area of e-scooter sharing in a city using an approach derived from e-scooter trip data in Li et al. (2022). The method mainly contains three steps. (1) For each city, we determine the range of the e-scooter sharing data by assessing the locations of both trip origins and destinations. (2) Using this range, we then generate a grid layer with each cell measuring 100 m. The number of trip origins and destinations within each cell is counted. (3) We eliminate the cells with less than four trip origins and destinations. The remaining cells constitute the e-scooter sharing service area for the given city.

The socio-demographic variables representing the information of the population are calculated based on the dataset acquired from the WorldPop Hub (WorldPop, 2023). The website provides spatial raster datasets depicting the population structure — that is, the count of individuals with different genders and age periods per cell. In each city, we restrict our attention to the dataset within the defined e-scooter sharing service area and the city-level socio-demographic variables are subsequently consolidated by gender and age groups. These computed socio-demographic variables encompass population, population density, male proportion, and ratios of different age groups. To account for the influence of the native, we have also integrated the native-to-foreign resident ratio to represent the migration level. Since there is no information related to the native at the small range, we procure the data from the source (City Population, 2016).

Road network variables, including road densities and the proportions of different levels of roads, are sourced from OpenStreetMap (OSM). Before initiating calculations, we first trim the road network to match the calculated service area. The newly adjusted road networks are then categorized into distinct levels, including major roads, branch roads, and pedestrian-only roads. The term "major road" is designated for thoroughfares predominantly used by automobiles. "Branch roads", on the other hand, refer to roads that cater to micro-mobility solutions, including cycling lanes and residential streets. Lastly, "pedestrian-only roads" are exclusive pathways for pedestrians, typically maintaining a low-speed limit.

Additionally, leveraging public transportation data from OpenStreetMap, we compute the variables relevant to public transit. We calculate the density of public transit stations by dividing the number of stations located within the service area by the total area of the service zone. The public transit stations considered in this analysis encompass bus stops, tram stops, and subway stations. It should be noted that the station in the OpenStreetMap refers to station platform, so two station platforms for opposite directions of a route with the same bus station name are both counted as two stations in this study. Another crucial variable we assess is the number of public transit trips, which is counted by using the General Transit Feed Specification data (GTFS) in different cities. Generally, a city with more public transport trips has a higher public transport accessibility.

Besides, some other variables are calculated as summarized in Table 3. There are some differences in the periods and sample sizes of e-scooter sharing data in different cities. To avoid their impacts on analysis, we consider "Number of trips" and "Number of days" to represent the number of collected trips and the number of days in each city, respectively. Two time-related factors "Weekday" and "Period of the day" are considered. The hours are grouped according to the change in RFMT and RLMT (Fig. 5), i.e., [3:00, 9:00), [9:00, 15:00), [15:00, 21:00), [21:00, 3:00). Additionally, the number of e-scooters will affect the convenience of e-scooter sharing and consequently influence the integration ratio with public transit. So the feature "Number of e-scooters" is also taken into account. The detailed description of city-level factors is summarized in Table 3.

4.4. The interpretation of model results

After modeling the relationship between RFMT (or RLMT) and city-level factors, we leverage a data-driven interpretation technique to reveal the relationship. Partial dependence analysis has been widely used to investigate nonlinear effects in many relevant studies (Wang et al., 2022; Pérez-Fernández and García-Palomares, 2021). However, an assumption of partial dependence analysis is that the features have no correlation with each other, which may be violated in empirical studies. We use an alternative method, accumulated local effect analysis. The accumulated local effect analysis still works even if the used features are correlated (Gao et al., 2023b), and can be formulated as:

$$\begin{aligned} ALE(x_I) &= \int_{\min(x_I)}^{x_I} E_{x_R|x_I}[\frac{\partial \Phi(x_I, x_R)}{\partial x_I}|x_I = k_I]\mathrm{d}(k_I) - c_I \\ &= \int_{\min(x_I)}^{x_I} \int \frac{\partial f(x_I, x_R)}{\partial k_I}\mathrm{P}(x_I|k_I)\mathrm{d}x_I \mathrm{d}z_I - c_I \end{aligned} \quad (6)$$

where $\frac{\partial \phi(x_I, x_R)}{\partial x_I}$ represents the gradient at the point and the local effect of x_I on the dependent variable. The gradient can be replaced by the first-order difference if the utilized model has no gradient. Since the accumulated local effect analysis ignores the outlier values by involving conditional distribution, the estimated local effect of a feature on the dependent variable is not impacted by the correlation. More technical details about accumulated local analysis are available in Molnar (2020).

5. Results and discussion

5.1. The integration ratio of first-and-last-mile trips

The distribution of RIT is shown in Fig. 2. The RIT, namely the proportion of e-scooter sharing being used as the feeder tool from or to public transit stations, shows considerable variation across cities. The value of RIT ranges from 5.59% to 51.40% in different cities. The mean value is 31.58% and the standard deviation is 8.47%. In over 76 cities, the RIT exceeds 30%, indicating that in these cities, more than 30% of e-scooter trips are integrated with public transit systems. However, a few cities have a very low integration ratio. For instance, Corby in the UK has the lowest RIT (5.59%). It suggests that e-scooter sharing is not commonly used for first-and-last-mile connections to public transit in Corby. Notably, two other UK cities, Oxford and Portsmouth, also demonstrate high RIT values, with Oxford exceeding 50%, the highest in our study. This manifests that e-scooter sharing is popular as a feeder to public transit in Oxford and Portsmouth.

To validate our findings, we conducted a comparative analysis of our RIT values with existing studies based on surveys or other approaches. In Paris, our RIT value is about 32.4%, marginally lower than the 37% reported by Christoforou et al. (2021). The RIT value for London in our study is around 35%, slightly exceeding the 32% derived from surveys by Transport for London but very close to it (Transport for London, 2019). A small discrepancy is observed for the results in Oslo, where our results indicate that 26% of e-scooter sharing trips are integrated with public transport, slightly larger than 23% and 20% reported in Fearnley

Table 3
The description of considered factors.

Explanatory variables	Unit	Definitions	Min	Max	Mean
Socio-demographic features					
Service Area	km^2	The service area was estimated by using the e-scooter data	0.443	422.605	45.107
Population		Number of individuals living in the service area	1739	2,123,080	183,247
Population density	$/km^2$	The population within the service area divided by the size of the service area	624.33	17,620.95	3685.14
Male ratio	%	Number of males divided by the population of the given city	46.71%	50.42%	48.74%
Ratio of the young	%	Number of individuals aged 15–29 divided by the population of the given city	12.67%	23.95%	16.89%
Ratio of the adult	%	Number of individuals aged 30–50 divided by the population of the given city	22.27%	33.82%	26.30%
Ratio of the elder	%	Number of individuals aged 50–65 divided by the population of the given city	11.90%	23.74%	19.44%
Other ratio	%	Number of individuals aged smaller than 15 or bigger than 65 divided by the population of the given city	30.33%	41.25%	37.37%
Ratio of the native	%	The percentage of native individuals in a city	50.00%	96.93%	83.03%
Land use variables					
Ratio of commercial	%	The percentage of commercial area in the service area	0.0%	19.3%	4.82%
Ratio of industrial	%	The percentage of industrial area in the service area	0.0%	22.3%	4.9%
Ratio of leisure	%	The percentage of leisure area in the service area	1.1%	66.8%	8.0%
Ratio of residential	%	The percentage of residential area in the service area	2.2%	92.6%	69.4%
Ratio of retail	%	The percentage of retail area in the service area	2.9%	16.1%	2.9%
Road network variables					
The length of road	km	The length of roads within the service area of a city	58.797	47,060.053	552.320
Ratio of major roads	%	The length of major roads divided by the total length of road network in a city	6.9%	40.1%	18.4%
Ratio of branch roads	%	The length of branch roads divided by the total length of road network in a city	15.8%	76.3%	46.7%
Ratio of pedestrian-only roads	%	The length of pedestrian-only roads divided by the total length of road network in a city	7.9%	73.6%	34.9%
Density of major roads	km/km^2	The length of major roads divided by the area of the city	10.588	64.400	23.235
Density of branch roads	km/km^2	The length of branch roads divided by the area of the city	31.557	143.131	59.196
Density of pedestrian-only roads	km/km^2	The length of pedestrian-only roads divided by the area of the city	13.678	124.763	45.815
Public transit variables					
Station density	$/km^2$	The number of public transit stations divided by the area of the city	1.347	69.820	17.250
Number of public transit trip		The number of public transport trips counted by GTFS data	32	271 283	30 111
Other features					
Number of trips		The number of collected trips in the given city	2243	14,478,828	871 052
Number of days		The number of days in the given city	128	507	253
Time features					
Workday		The value is 1 if the trips are in workday, otherwise is 0	0	1	–
Period of the day		The period of the day, which has four values 0 [3:00, 9:00), 1 [9:00, 15:00), 2 [15:00, 21:00), 3 [21:00, 3:00)	0	3	–
Supply					
Number of e-scooters		The number of e-scooters in the given city	264	41,592	3391
Location					
Longitude		The longitude of the city	−5.989	24.938	9.565
Latitude		The latitude of the city	36.718	63.431	51.427

et al. (2020) and Aarhaug et al. (2023), respectively. These comparisons reveal that our RIT values are consistent with those from other sources, underscoring the reliability of our analysis. The observed differences could stem from various factors, including survey methodologies, respondent biases, and distinct definitions of public transit. In summary, this empirical comparative analysis substantiates the credibility of our findings.

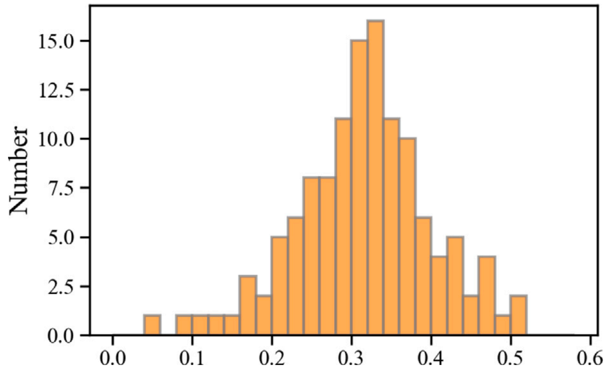


Fig. 2. The distribution of RIT (the ratio of integration trip).

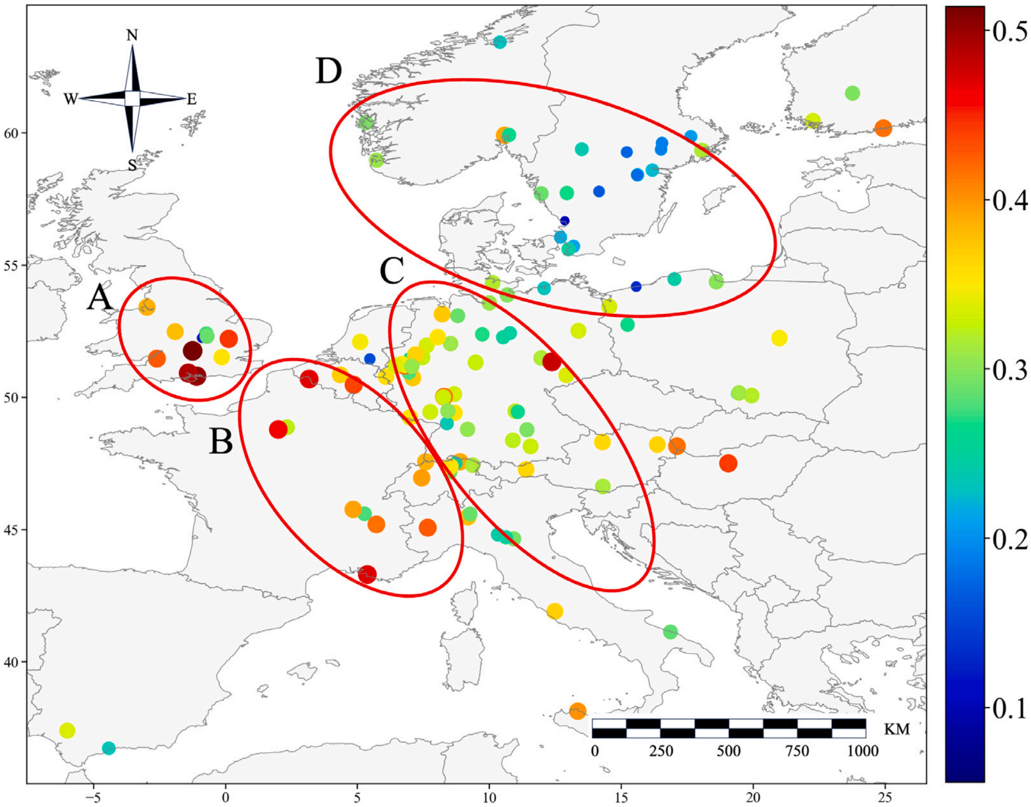


Fig. 3. The spatial distribution of RIT.

5.2. Spatial pattern analysis

The spatial distribution of the integration ratio is depicted in Fig. 3. The results exhibit a clear pattern with higher RIT values in cities located in areas *A* and *B* (mainly in the UK and France), while cities in area *C* (including Germany, Austria, Italy, etc.) tend to have lower RIT. RIT in cities of area *D* is the lowest among all the areas. One potential explanation is that cities in area *D* experience lower temperatures compared to other areas, particularly in winter. Consequently, e-scooter sharing is less preferred as a mode of connection with public transport in colder weather. To quantify this spatial pattern, we calculate the spatial autocorrelation of RIT using Moran's I and obtain a value of 0.248, with significance at the confidence level of 99% ($p < 0.001$). This suggests that the RIT in a city is more similar to that of an adjacent city than that of a distant city, indicating the presence of spatial dependency in the integration ratio. However, it is important to note some exceptions. For instance, Corby in area *A* exhibits the lowest RIT, while Leipzig in area *C* records an RIT of 47.96%, surpassing most other cities in the same area.

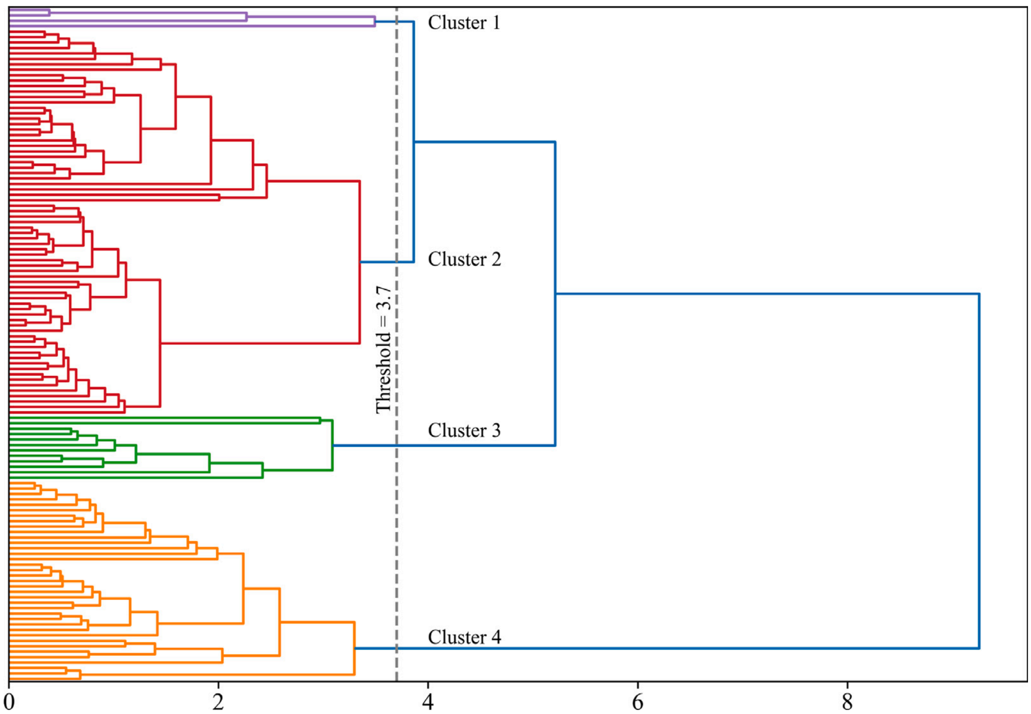


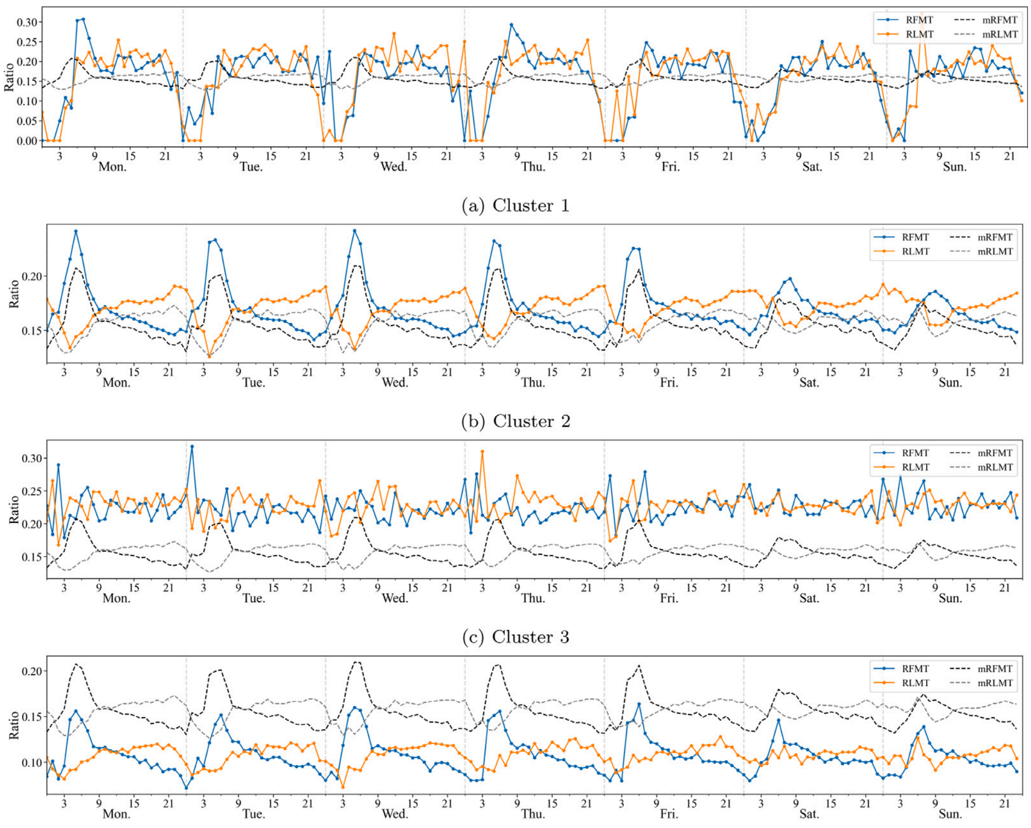
Fig. 4. The hierarchical relationship of the temporal features for the RFMT and RLMT between cities.

5.3. Temporal pattern analysis and clustering

Fig. 4 visualizes the results of hierarchical clustering based on the temporal signature in Section 4.2. The x-axis represents the distance between clusters, which is calculated using the Ward linkage. The y-axis represents the cities, but their names are omitted for readability. A version of the figure including city names is available in the supplementary materials. Adjusting different thresholds for clustering may lead to different clustering results. According to the shape of the clustering hierarchical tree, we use a threshold shown in the gray line in Fig. 4 to divide the cities into four clusters.

In Fig. 5, we further present the average temporal signatures of cities in each cluster on a weekly basis. These clusters differ primarily in the magnitude for average RFMT and average RLMT, the range of RFMT and RLMT, and fluctuation in temporal patterns of RFMT and RLMT. We display the average temporal signatures of all cities in black and gray dash curves in Fig. 5 as benchmarks, which are helpful for identifying the differences in temporal signatures in different clusters. The unique patterns of each cluster can be interpreted as follows:

- **Cluster 1** (Fig. 5(a)): Cluster 1 has the largest variations in RFMT and RLMT values, as compared to other clusters. The RFMT values of the cities in Cluster 1 range from 0 to 30.69% with a standard deviation being 7.42%, while the RLMT values vary from 0 to 31.18% with a standard deviation being 7.57%. Cities, including Bristol, Liverpool, Oxford, and Trzebinia, have zero RFMT and RLMT values at midnight. This is likely ascribed to the fact that e-scooter sharing in these cities is not available at midnight due to regulations. For example, the business operation time of e-scooter sharing in Oxford is between 4 a.m. and 10 p.m (Oxford E-scooter, 2022).
- **Cluster 2** (Fig. 5(b)): Cluster 2 contains 71 cities from Austria, Belgium, Finland, France, Germany, Italy, Norway, Poland, Spain, Sweden, Switzerland, and the United Kingdom. This cluster consists of most cities, while the ranges of RFMT and RLMT in Cluster 2 are lower than those of Cluster 1. The ranges of RFMT and RLMT are 14.16%–24.18% (the mean and standard deviation are 16.90% and 2.15%) and 12.60%–19.24% (the mean and standard deviation are 17.08% and 1.31%), respectively. The smallest RFMT and RLMT values of this cluster are higher than 0, as e-scooter sharing is still operating at midnight in these cities. The cities in the same country may have different regulations for e-scooter sharing. For example, two UK cities, e.g., London and Birmingham, are in Cluster 2, because both two cities have no restrictions and can be used at midnight. In terms of the fluctuation in the temporal patterns (Fig. 5(b)), temporal patterns of RFMT and RLMT in this cluster are distinct from those in Cluster 1. There is one obvious peak for both RFMT and RLMT on each day of the week. The peak of the average RFMT occurs at midnight when the RLMT reaches the lowest point. As for RLMT, the peak in each day is around 11 p.m., where the value of RLMT is small.
- **Cluster 3** (Fig. 5(c)): Cluster 3 contains 12 cities. Compared with Cluster 2 (Fig. 5(b)), the average magnitudes of RFMT and RLMT are higher in Cluster 3. The RFMT ranges from 17.90% to 31.76% and the range of RLMT is from 16.76% to 31.00%,



(a) Cluster 1

(b) Cluster 2

(c) Cluster 3

(d) Cluster 4

Fig. 5. Temporal signature of RFMT and RLMT during one week for each cluster. Y-axis shows the values of RFMT or RLMT. The black and gray dash curves are the average temporal signatures of RFMT and RLMT among all cities, and are used for comparison benchmarks.

where the mean value for RFMT and RLMT are 22.43% and 22.85%, respectively. Similar to Cluster 2 (Fig. 5(b)), the peaks of RFMT and RLMT in Cluster 3 are at the midnight. However, the temporal changes of RFMT and RLMT in Cluster 3 display fluctuating patterns. On the contrary, there is no obvious increasing or decreasing pattern during the daytime, as displayed in Fig. 5(b).

- **Cluster 4** (Fig. 5(d)): This cluster contains 31 cities. The temporal patterns of RFMT and RLMT are similar to those in Clusters 2. However, the magnitudes of RFMT and RLMT are lower than those in Cluster 2, with the RFMT ranging from 7.2% to 16.37% and the RLMT ranging from 7.26% to 12.80%, where the mean value for RFMT and RLMT are 10.73% and 10.80%, respectively. The temporal pattern of the RFMT is similar to that in Cluster 2, with a clearly visible peak during the daytime. According to Fig. 5(d), we can find that the temporal pattern of RFMT is similar to that in Cluster 2 (Fig. 5(b)), where the RFMT shows a decreasing trend after reaching the peak at about 5 a.m.. However, the temporal pattern of RLMT is similar to that in Cluster 3 (Fig. 5(c)), which does not show a clear increasing or decreasing trend.

Although the four clusters show some distinct temporal patterns of RFMT and RLMT, they also present some similarities. Overall, the temporal patterns of RFMT and RLMT have opposite trends. More specifically, an increase in RFMT generally accompanies a decrease in RLMT. The phenomenon is particularly prominent in Cluster 2 (Fig. 5(b)). Cluster 1 is an exception where the values of RFMT in the early morning (around 5 a.m.) tend to be higher than those during the daytime as the e-scooter sharing system closes at midnight. This may be because fewer public transit services are available in the early morning, as many lines of public transport are still not open. Therefore, more people will use e-scooters to go to a farther public transit station with public transit services.

Furthermore, in Fig. 6, we present a summary of city-level variables, comprising two subfigures. Fig. 6(a) provides an overview of socio-demographic and land-use variables, while Fig. 6(b) summarizes infrastructure-related and supply variables. Each line in the sub-figures represents a cluster, and each point of the line represents the mean value of a specific city-level variable for the cities within that cluster. To ensure comparability, we normalize the variable values by dividing them by their respective maximum values.

Analyzing Fig. 6(a), we observe that among the four clusters, only population and population density exhibit noticeable variance. On the other hand, other demographic variables, such as the ratio of the young, and the elderly, and gender distribution, show limited

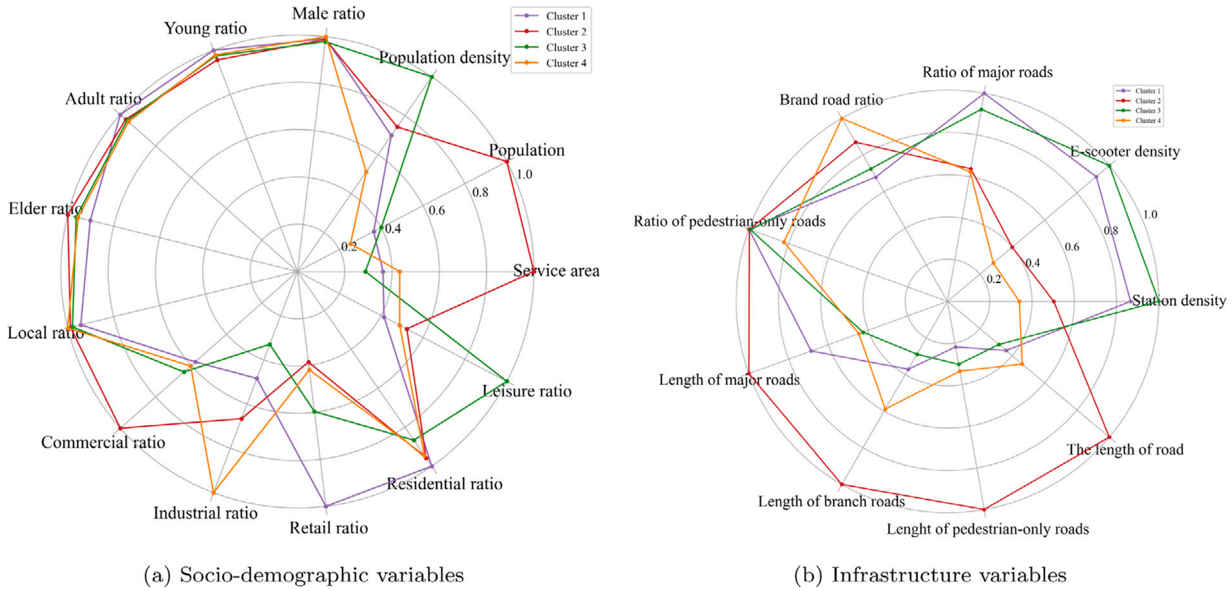


Fig. 6. The summary of city-level variables. The values of a factor are normalized by the maximum value among the cities for the convenience of comparison, so the value of a factor is between 0 and 1.

Table 4
Performance metrics of random forest algorithms.

Performance metrics	RFMT	RLMT
Coefficient of determination (R^2)	0.915	0.941
Root Mean Square Error (RMSE)	1.28%	1.07%
Explained Variances (ER)	0.957	0.970

differences between the clusters. Regarding land-use variables, the Residential ratio does not display significant variation when compared to the other four land use variables. Notably, each cluster achieves the highest value in certain categories. For instance, Cluster 1 attains the maximum value for the Retail ratio, Cluster 2 has the highest Commercial ratio, Cluster 3 is predominantly focused on the Leisure ratio, and Cluster 4 demonstrates the highest Industrial ratio.

Regarding infrastructure, there is a distinct disparity in the ratio of major roads and branches among the clusters. Clusters 1 and 3 are characterized by a higher percentage of major roads, while Clusters 2 and 4 exhibit a greater proportion of branch roads. Notably, there is a high correlation between the length of roads and different types of roads, leading to a shared pattern among them. Particularly, cities within Cluster 2 stand out with the maximum mean value for various road variables, including the length of roads, length of major roads, length of branch roads, and length of pedestrian-only roads. Conversely, the curves of the other three clusters are closely aligned and show minimal differentiation in these road-related variables. Furthermore, it becomes evident that Clusters 1 and 2 boast higher E-scooter density and Station density compared to the other clusters, indicating distinctive mobility infrastructure characteristics in these particular clusters.

Additionally, Fig. 7 presents a spatial representation of the clustering results. While the figure offers valuable insights, its interpretability is somewhat limited. For example, the second cluster is primarily concentrated within Western Europe, such as Germany, Switzerland, Austria, etc. The fourth cluster is mainly located in Northern Europe. However, the boundaries between different clusters are not clear. This is due to the fact that the integration ratios of different cities are impacted by various factors, with location being just one of them. In addition, the uneven distribution of cities across clusters further impacts the visualization's meaning. Cluster 1 encompasses merely 4 cities, whereas Cluster 2 comprises over 70 cities. Hence, it is necessary to invest more effort into delving into the factors impacting the integration ratio in different cities.

5.4. Influences of city-level factors

We model the effects of various city-level factors on RFMT and RLMT using two separate random forest models, where only a part of variables within Table 3 are selected. We adopt these variables by avoiding the correlation between them, and the two final models are built with the highest R^2. The predictive performance of the two models is summarized in Table 4. The results show that the used models have pretty good performance in modeling the effects of city-level factors with a R^2 of around 0.915 for RFMT and 0.941 for RLMT and thus sufficient to reflect the impacts of different factors on the RFMT and RLMT.

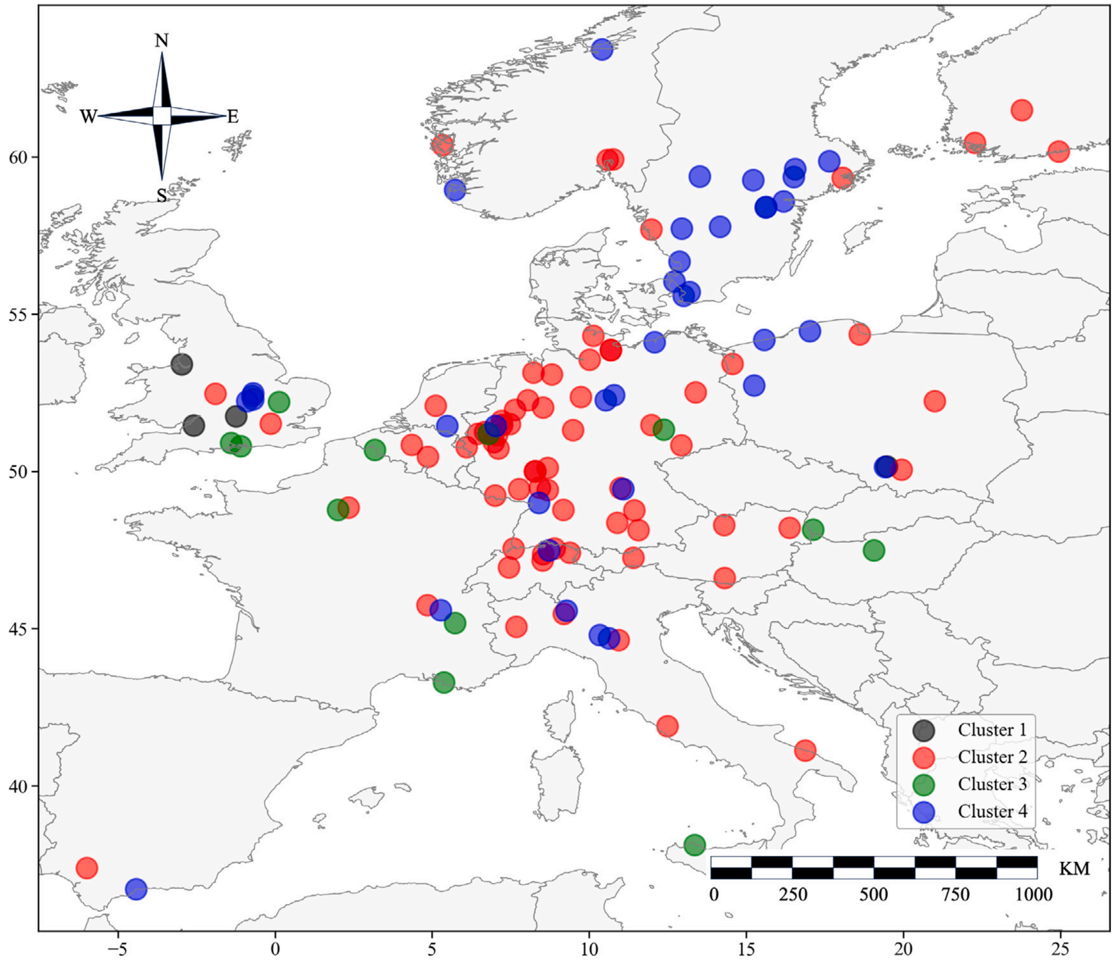


Fig. 7. Spatial representation of clustering results.

Fig. 8 shows the importance of adopted variables that are estimated by using SHapley Additive exPlanations (SHAP) values. SHAP value is based on the game theory and more details can be found in Lundberg and Lee (2017). Figs. 9, 11, 10, and 12 present quantitative influences of selected factors on RFMT and RLMT. In each figure, the x-axis shows the value of a factor and the y-axis is the predicted difference in RFMT or RLMT with a unit of % when the factor changes, i.e., the accumulated local effect value (ALE). The impact of these variables are analyzed and discussed along with the feature importance.

Fig. 9 illustrates the impact of socio-demographic features on RFMT and RLMT. The population density does not show a single consistent pattern. Compared with the RLMT, the population shows a more significant impact on RFMT. A denser population will lead to a higher integration ratio for the first mile. For the RFMT, the integration ratio will increase along with population density. When the population density surpasses 13 thousands per square kilometer, the impact becomes opposite (i.e. decreasing impact), which may be attributed to a special case in a city. The integration ratio approximately increases with a higher "Ratio of the native" in a city, which implies the native people contributes to more usage of e-scooter sharing users for connecting to public transit. Fig. 9(c) shows that a higher "Ratio of the young" is positively related to the integration ratio when the "Ratio of the young" is smaller than 17.5%. This is reasonable and in line with some prior studies as e-scooter sharing has been indicated to be more prevalent among young people (Oeschger et al., 2023; Pazzini et al., 2022; Montes et al., 2023; Brussels, 2019). For example, the young people take up half of all the e-scooter users in Brussels (Brussels, 2019). Montes et al. (2023) found the young preferred to select shared micro-mobility as the connecting tool. However, after the "Ratio of the young" continues increasing after 0.175, it can be observed that the integration ratio decreases with the "Ratio of the young". A stronger impact of "Ratio of the young" on the RLMT than RFMT can be observed. It can be explained from the following perspectives. First, although the young ratio increases, it does not necessarily mean that more young users use e-scooters to connect to public transport. The users may use e-scooters for other trip purposes. Second, the "Ratio of young" may be correlated to some other determinants that are not considered in our analysis, which needs more exploration with more data in future studies. Third, the reason why the impact on RFMT is more pronounced compared with RLMT may be that the young may focus on the first mile more than the last mile. The public transport has scheduled timetables, and the passengers need to catch the public transport on time. But for the last-mile trips, the passengers are generally

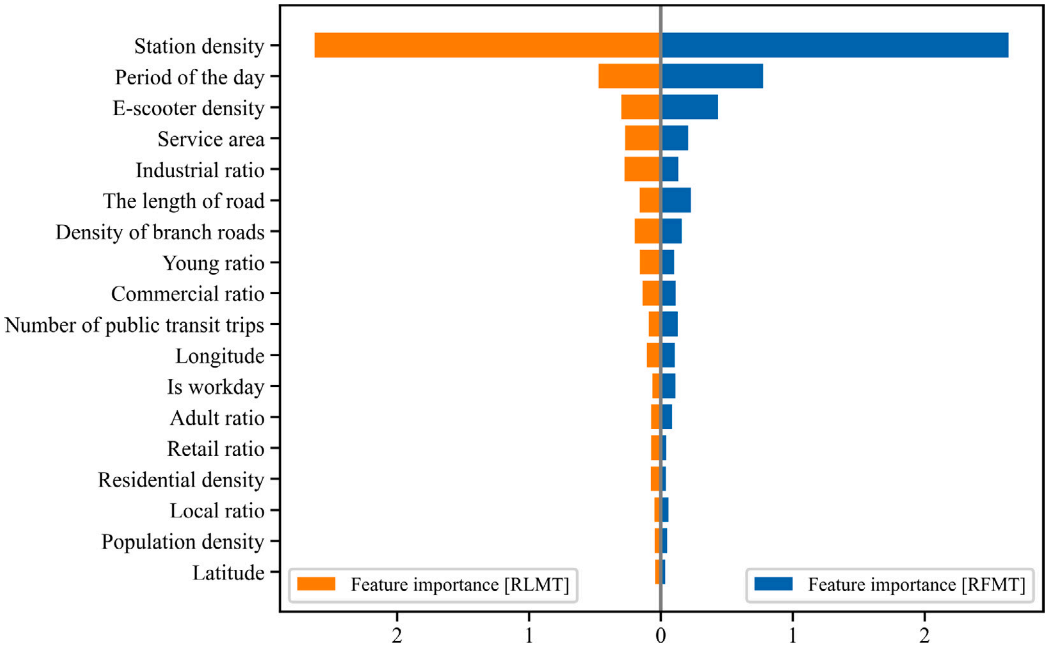
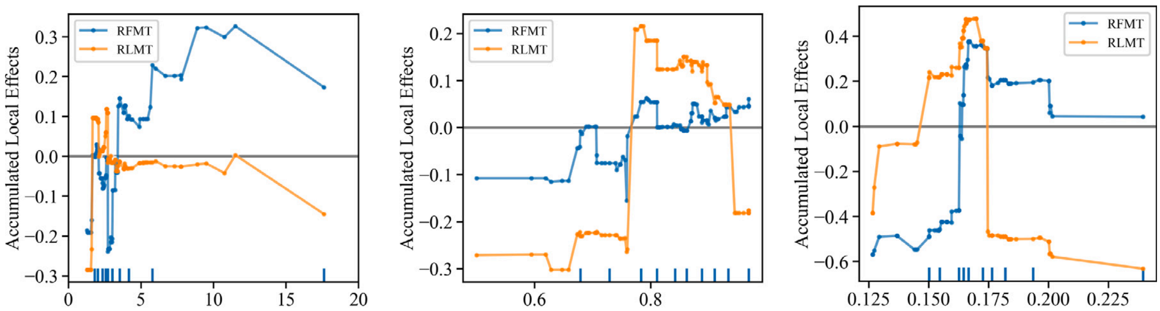


Fig. 8. Feature importance of adopted factors estimated by SHAP.



(a) Population density (10^3/km^2) (b) Ratio of the native (c) Ratio of the young

Fig. 9. The effects of socio-demographic features on the RFMT and RLMT. The unit of Y-axis is %. The unit of X-axis for each figure is within the parentheses of each sub-figure.

more flexible and do not need to consider the schedules of public transit anymore or strict time constraints, resulting in less urgency for using e-scooter sharing for quick or on-time last-mile connections than the first-mile trips.

In terms of land use features, the ratios of commercial, industrial, residential, and retail area are analyzed. According to Fig. 8, the ratio of residential area has the lowest impact on RFMT and RLMT. Thus, we focus on the remaining three land use variables, whose impact on RFMT and RLMT are shown in Fig. 10. Within the three variables, the impacts of "Commercial ratio" and "Retail ratio" on RFMT and RLMT have similar patterns. RFMT and RLMT increase along with the "Commercial ratio" and "Retail ratio", which means a city with more commercial and retail activities has higher integration ratio of e-scooters with public transit. However, the "Industrial ratio" will negatively relate to RFMT and RLMT as demonstrated in Fig. 10(a). A potential explanation is the area with industrial purpose is far away from the city center and generally are not well connected with public transit. Thus, the short distance micro-mobility service for connecting to public transit is not prevalent.

The road network variables and public transit variables are together shown in Fig. 11. As for road network variables, only "Density of branch road" (Fig. 11(a)) shows noticeable impacts, though not as significant as other factors. RFMT and RLMT decrease at the beginning and then keep stable with an increase in "Density of branch roads" (Fig. 11(a)). The potential reason is that more branch roads are facilitators of using e-scooter sharing for various travel purposes (e.g., entertainment), rather than the usage of e-scooter sharing to public transit stations, which indirectly results in decreases in RFMT and RLMT. However, the decreasing trend ends after a certain threshold. "Station density" has obvious effects on RFMT and RLMT (Fig. 11(b)). The larger the density of public transit stations is in a city, the higher the RFMT and RLMT. The findings are reasonable as more public transit stations in a city

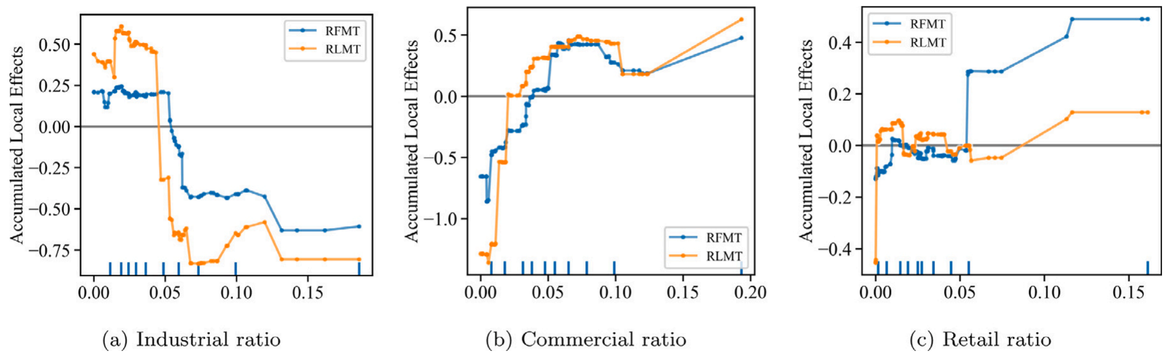


Fig. 10. The effects of land use features on the RFMT and RLMT. The unit of Y-axis is %. The unit of X-axis for each figure is within the parentheses of each sub-figure.

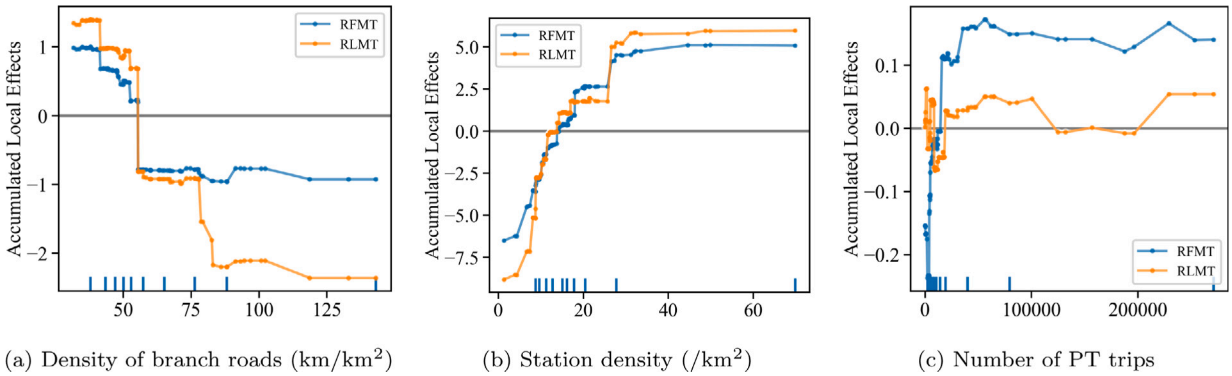


Fig. 11. The effects of Road network and Public transit features on the RFMT and RLMT. The unit of Y-axis is %. The unit of X-axis for each figure is within the parentheses of each sub-figure.

mean higher accessibility to different places using public transit, which is beneficial for increasing the usage of e-scooter sharing as the feeder to public transit. It should be noted that when the station density is over a threshold (about 15), its effects on RFMT and RLMT are not noticeable anymore. Although "Number of PT trips" does not have a significant impact on RFMT and RLMT as the previous two factors, the impact curves still show obvious patterns. The integration ratio will increase along with the number of public transport trips, which means the promotion of the public transit's accessibility can promote the usage of e-scooter sharing.

The relation of "Density of e-scooters" (i.e., the supply) is shown in Fig. 12(a). When "Density of e-scooters" is very small, the values of RFMT and RLMT increase sharply with "Density of e-scooters", implying that increasing the supply can enhance the integration ratio of e-scooter sharing and public transit. However, when "Density of e-scooters" exceeds 250, this variable's impact becomes stable, meaning the supply can satisfy the demand of e-scooters. According to Fig. 12(b), it is found that RFMT and RLMT increase with the size of the service area of e-scooter sharing. However, when the service area is over a threshold (about 100 km^2), it does not show further effects, as displayed in Fig. 12(b). A potential explanation is that the first-and-last mile problem is more serious in larger cities so the e-scooter sharing will be more useful as the integration tool. Figs. 12(c) and 12(d) show the impact of cities' locations on RFMT and RLMT. According to Fig. 12(c), RFMT and RLMT decrease along with Longitude. Although there is an increase after about 15, the increase could be ignored. It means along with the east–west direction, the RFMT and RLMT become smaller. Similarly, the Latitude also has a negative impact on RFMT and RLMT (Fig. 12(d)), meaning the city with high latitude tends to have a smaller RFMT and RLMT. It can be easily found that the findings based on the location variables match with the spatial patterns shown in Fig. 3. Based on Figs. 12(e) and 12(f), the impact both variables on RFMT and RLMT show opposite patterns. Compared with "Is Workday", the "Periods of a day" has more significant impacts on RFMT and RLMT based on Fig. 8. From the first period (3:00–9:00) to the fourth period (21:00–3:00), the value of RFMT reduces, indicating a lower percentage of e-scooter sharing trips as an integration tool. The finding is in line with the temporal tendency in Fig. 5. The value of RFMT in the first period is higher. This is probably because more people commute to their workplaces or schools in the morning peaking hours, where e-scooter sharing is an expedient option to go to public transit stations. In comparison, the value of RLMT in the first period is lower than that in the third and fourth periods. This may be attributed to the increased demand for using e-scooters as egress mode from public transit stations for entertainment or other trips after work and in the evening. The effects of "Workday" on RFMT and RLMT present contrary directions as well (Fig. 12(f)).

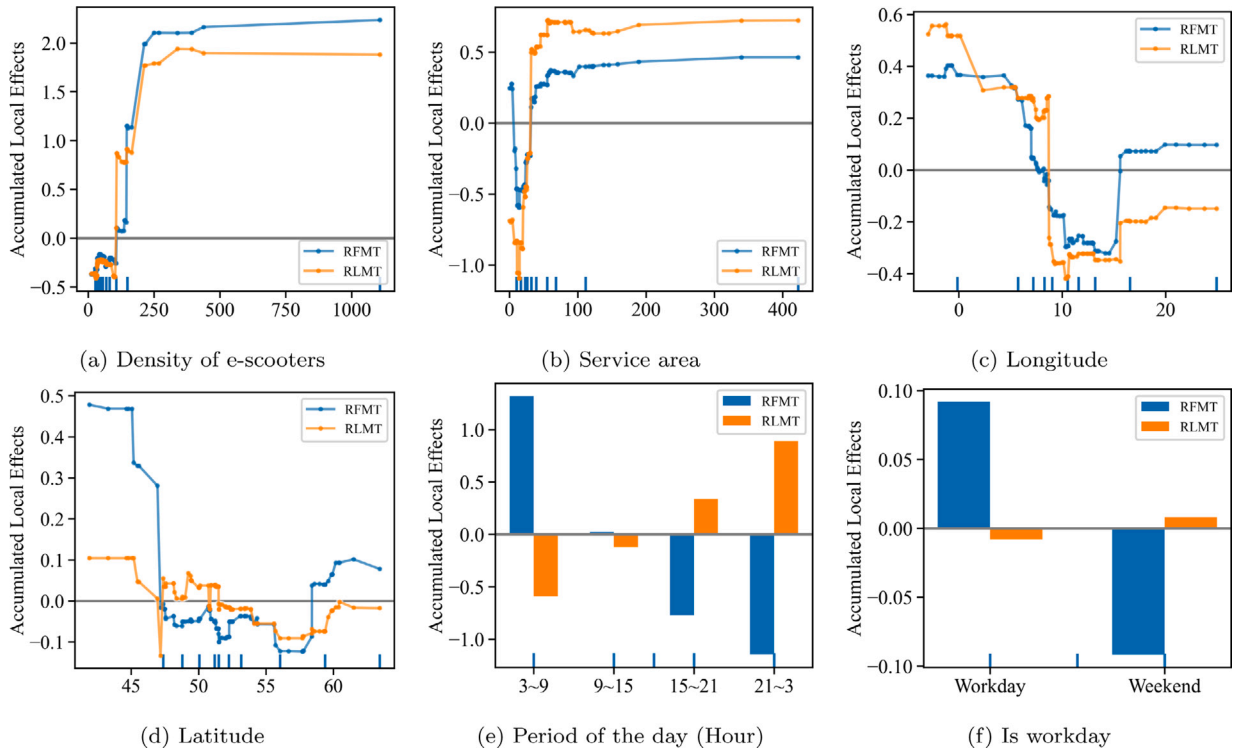


Fig. 12. The effects of other features on the RFMT and RLMT. The unit of Y-axis is %. The unit of X-axis for each figure is within the parentheses of each sub-figure.

6. Conclusion

This study provides a large-scale analysis of how e-scooter sharing is integrated with public transit for the first-and-last-mile problem. Specifically, the spatial and temporal patterns of integration ratios of e-scooter sharing with public transit in 124 European cities are analyzed and compared. Meanwhile, we investigate the non-linear effects of city-level factors on the integration ratio of e-scooter sharing with public transit using machine learning and a model-agnostic interpretation method. The main findings of this study are summarized as follows.

First, the integration ratio of e-scooter sharing with the public transit system shows large variations among different cities. The integration ratios among 124 European cities range from 5.59% to 51.40% with a standard deviation of 8.47%. For instance, in cities like Corby, e-scooter sharing shows poor integration as a feeder mode for public transit. In addition, the integration ratio of e-scooter sharing with public transit presents obvious spatial dependency. The cities with high RIT are mainly located in central Europe and the UK. The cities within Northern Europe generally have a lower integration ratio.

Second, we analyze the temporal patterns of RFMT and RLMT in different cities. Different cities are divided into four clusters using a hierarchical clustering method (Fig. 5). The temporal patterns of RFMT and RLMT show an opposite trend. The increasing tendency of RFMT generally coexists with a decrease in RLMT. More importantly, the temporal patterns in different cities are obviously and significantly divergent in terms of range, fluctuation patterns, and magnitude, which are extensively discussed.

Third, we explore the non-linear impacts of various city-level factors, such as socio-economic, road network, and public transit factors, on RFMT and RLMT values, offering insights into enhancing integration. Several factors have noticeable influences on RFMT and RLMT. For example, public station density is positively associated with both RFMT and RLMT. Compared to “Workday”, “Periods of a day” has a larger impact on the integrated usage of e-scooter sharing and public transit. Higher ratios in the young are positively related to the value of RFMT and RLMT.

Overall, the findings are useful for city planners and policymakers to develop evidence-based policies regarding the implementation of micro-mobility in different city contexts. However, it is important to acknowledge some limitations that need to be addressed. First, we identify the integration relationship between e-scooter sharing and public transit only depending on the spatial relationship, i.e., the Euclidean distance, between trip origin/destination and their nearby public transit stations. Although the method has been implemented in many studies, it is still coarse. Additional information, including trip distance, origin and destination times, timetables of public transit and the distribution of Points of Interest (POIs) around trip endpoints, could be utilized for more accurate integration analysis of shared micro-mobility and public transit (Kong et al., 2020; Luo et al., 2021). In the future, we will focus on an analysis of fewer cities and develop a more robust method to identify the integration trip at the trip level using both passively collected data and actively collected data from questionnaires. Second, there are some potential influencing factors that we ignored due to data limitations such as education level and income. These factors could potentially relate to user behavior and thus the integration of e-scooter sharing with public transit. It is rather interesting to investigate the impacts of these factors if more data resources are available. Another important point to mention is that the relations of factors in the models are statistical correlations, Third, this study does not distinguish the integration of e-scooter sharing with different public transit modes. This is because public transit modes significantly differ across cities. For instance, cities such as Milan and Paris have extensive metro systems, while Zurich does not have metro systems. However, the interaction between e-scooter sharing and different types of public transit may vary. The limitation presents an intriguing research opportunity to concentrate on cities with similar public transit modes and investigate the nuances in their interplay. Moreover, we only focused on how the e-scooter sharing integrated with the public transit system, i.e., the complementary role of e-scooter sharing, while the e-scooter sharing could also compete with existing transport services. For example, the e-scooter sharing, as a shared micro-mobility service, can substitute the public transit system as well (Li et al., 2021a; Ziedan et al., 2021; Guo et al., 2023). Therefore, it will present a more comprehensive picture of the impacts of e-scooter sharing on public transport by considering both the complementary and substitution roles. Unraveling the substitutions of e-scooter sharing to other transport modes in multiple cities is a promising direction for future work.

CRediT authorship contribution statement

Aoyong Li: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Kun Gao:** Conceptualization, Methodology, Formal analysis, Validation, Writing – original draft, Writing – review & editing, Funding acquisition, Data curation. **Pengxiang Zhao:** Writing – original draft, Validation, Investigation, Data curation. **Kay W. Axhausen:** Writing – original draft, Validation, Resources, Formal analysis, Data curation.

Acknowledgments

This research has been supported by the Young Scientists Fund of the National Natural Science Foundation of China (52202389), State Key Laboratory of Resources and Environmental Information System, Area of Advance Transport in Chalmers University of Technology, Sweden (Project No. 2021-0040), e-FAST project (Project No. P2022-00414) funded by Swedish Energy Agency, and a grant administered through K2 - The Swedish Knowledge Centre for Public Transport (grant number 2023001).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.trc.2024.104496.

References

6t-bureau de recherche, 2019. Usages et usagers des trottinettes electriques en free-floating en france. URL https://www.6-t.co/article/trottinettes-freefloating.

Aarhaug, J., Fearnley, N., Johnsson, E., 2023. E-scooters and public transport – Complement or competition? Res. Transp. Econ. (ISSN: 0739-8859) 98, 101279.

Attard, M., 2022. Active travel and sustainable transport. COMM. Transp. Res. 2, 100059, 1, https://www.sciopen.com/article/10.1016/j.commtr.2022.100059.

Baek, K., Lee, H., Chung, J.-H., Kim, J., 2021. Electric scooter sharing: How do people value it as a last-mile transportation mode? Transp. Res. D (ISSN: 1361-9209) 90, 102642.

Brussels, 2019. Distribution of electric scooter users in Brussels (Belgium) in 2019, by age. URL https://www.statista.com/statistics/1169554/distribution-of-e-scooter-users-in-brussels-by-age/.

Campbell, K.B., Brakewood, C., 2017. Sharing riders: How bikesharing impacts bus ridership in new york city. Transp. Res. A (ISSN: 0965-8564) 100, 264–282.

Chen, W., Chen, X., Chen, J., Cheng, L., 2022. What factors influence ridership of station-based bike sharing and free-floating bike sharing at rail transit stations? Int. J. Sustain. Transp. 16 (4), 357–373.

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 Behav. Soc. (ISSN: 2214-367X) 14, 1–10.

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. Transp. Res. A (ISSN: 09658564) 162, 175–187.

Chicago Department of Transportation, 2020. E-scooter pilot evaluation summary. URL https://www.chicago.gov/city/en/depts/cdot/provdrs/future_projects_andconcepts/news/2021/may/cdot-releases-evaluation-of-2020-e-scooter-pilot-program.html.

Christoforou, Z., De Bortoli, A., Gioldasis, C., Seidowsky, R., 2021. Who is using e-scooters and how? Evidence from Paris. Transp. Res. D (ISSN: 13619209) 92, 102708.

City Population, 2016. City population - Population statistics in maps and charts for cities, agglomerations and administrative divisions of all countries of the world. URL https://www.citypopulation.de/.

Fearnley, N., Johnsson, E., Berge, S.H., 2020. Patterns of E-scooter use in combination with public transport. Findings (ISSN: 2652-8800).

Fishman, E., 2016. Bikeshare: A review of recent literature. Transp. Rev. (ISSN: 0144-1647) 36 (1), 92–113, Publisher: Routledge.

Fishman, E., Washington, S., Haworth, N., 2014. Bike share's impact on car use: Evidence from the United States, Great Britain, and Australia. Transp. Res. D (ISSN: 1361-9209) 31, 13–20.

Gao, K., Li, A., Liu, Y., Gil, J., Bie, Y., 2023a. Unraveling the mode substitution of dockless bike-sharing systems and its determinants: A trip level data-driven interpretation. Sustainable Cities Soc. 104820.

Gao, K., Yang, Y., Gil, J., Qu, X., 2023b. Data-driven interpretation on interactive and nonlinear effects of the correlated built environment on shared mobility. J. Transp. Geograp. 110, 103604.

Gao, K., Yang, Y., Li, A., Li, J., Yu, B., 2021a. Quantifying economic benefits from free-floating bike-sharing systems: a trip-level inference approach and city-scale analysis. Transp. Res. A 144, 89–103.

Gao, K., Yang, Y., Li, A., Qu, X., 2021b. Spatial heterogeneity in distance decay of using bike sharing: An empirical large-scale analysis in Shanghai. Transp. Res. D 94, 102814.

Gao, K., Yang, Y., Qu, X., 2021c. Examining nonlinear and interaction effects of multiple determinants on airline travel satisfaction. Transp. Res. D 97, 102957.
Gössling, S., 2020. Integrating e-scooters in urban transportation: Problems, policies, and the prospect of system change. Transp. Res. D (ISSN: 1361-9209) 79, 102230.
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. Transp. Res. D 122, 103880.
Kong, H., Jin, S.T., Sui, D.Z., 2020. Deciphering the relationship between bikesharing and public transit: Modal substitution, integration, and complementation. Transp. Res. D (ISSN: 1361-9209) 85, 102392.
Kou, Z., Cai, H., 2019. Understanding bike sharing travel patterns: An analysis of trip data from eight cities. Physica A (ISSN: 0378-4371) 515, 785–797.
Laa, B., Leth, U., 2020. Survey of E-scooter users in Vienna: Who they are and how they ride. J. Transp. Geograp. 89, 102874.
Li, A., Gao, K., Zhao, P., Qu, X., Axhausen, K.W., 2021a. High-resolution assessment of environmental benefits of dockless bike-sharing systems based on transaction data. J. Clean. Prod. (ISSN: 0959-6526) 296, 126423.
Li, A., Huang, Y., Axhausen, K.W., 2020. An approach to imputing destination activities for inclusion in measures of bicycle accessibility. J. Transp. Geograp. (ISSN: 0966-6923) 82, 102566.
Li, A., Zhao, P., Haitao, H., Mansourian, A., Axhausen, K.W., 2021b. How did micro-mobility change in response to COVID-19 pandemic? A case study based on spatial-temporal-semantic analytics. Comput. Environ. Urban Syst. 90, 101703.
Li, A., Zhao, P., Liu, X., Mansourian, A., Axhausen, K.W., Qu, X., 2022. Comprehensive comparison of e-scooter sharing mobility: Evidence from 30 European cities. Transp. Res. D (ISSN: 1361-9209) 105, 103229.
Lin, D., Zhang, Y., Zhu, R., Meng, L., 2019. The analysis of catchment areas of metro stations using trajectory data generated by dockless shared bikes. Sustainable Cities Soc. (ISSN: 2210-6707) 49, 101598.
Lin, J.-J., Zhao, P., Takada, K., Li, S., Yai, T., Chen, C.-H., 2018. Built environment and public bike usage for metro access: A comparison of neighborhoods in Beijing, Taipei, and Tokyo. Transp. Res. D (ISSN: 1361-9209) 63, 209–221.
Liu, Y., Feng, T., Shi, Z., He, M., 2022. Understanding the route choice behaviour of metro-bikeshare users. Transp. Res. A (ISSN: 0965-8564) 166, 460–475.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), In: Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc..
Luo, H., Zhang, Z., Gkritza, K., Cai, H., 2021. Are shared electric scooters competing with buses? a case study in Indianapolis. Transp. Res. D (ISSN: 1361-9209) 97, 102877.
McQueen, M., Abou-Zeid, G., MacArthur, J., Clifton, K., 2021. Transportation transformation: Is micromobility making a macro impact on sustainability? J. Plan. Lit. 36 (1), 46–61.
Molnar, C., 2020. Interpretable Machine Learning. Lulu. com.
Montes, A., Geržinic, N., Veeneman, W., van Oort, N., Hoogendoorn, S., 2023. Shared micromobility and public transport integration - A mode choice study using stated preference data. Res. Transp. Econ. (ISSN: 0739-8859) 99, 101302, URL https://www.sciencedirect.com/science/article/pii/S0739885923000422.
Naroa, C., Iñigo, M., Luis, A., Kent, L., 2022. On the simulation of shared autonomous micro-mobility. COMM. Transp. Res. (ISSN: 2772-4247) 2, 100065, 4, https://www.sciencedirect.com/science/article/pii/S2772424722000154.
O'Brien, O., 2021. European update: 360,000 e-scooters available across the continent. URL https://zagdaily.com/trends/european-update-360000-e-scooters-available-across-the-continent/.
Oeschger, G., Caulfield, B., Carroll, P., 2023. Investigating the role of micromobility for first- and last-mile connections to public transport. J. Cycl. Micromob. Res. (ISSN: 2950-1059) 1, 100001.
Oxford E-scooter, 2022. E-scooter trial in Oxford. URL https://www.oxfordshire.gov.uk/residents/roads-and-transport/connecting-oxfordshire/active-travel/e-scooter-trial.
Pazzini, M., Cameli, L., Lantieri, C., Vignali, V., Dondi, G., Jonsson, T., 2022. New micromobility means of transport: An analysis of E-scooter users' behaviour in Trondheim. Int. J. Environ. Res. Public Health (ISSN: 1660-4601) 19 (12).
Pérez-Fernández, O., García-Palomares, J.C., 2021. Parking places to moped-style scooter sharing services using GIS location-allocation models and GPS data. ISPRS Int. J. Geo-Inf. 10 (4), 230.
Portland Bureau of Transportation, 2019. 2018 E-scooter findings report. URL https://www.portland.gov/transportation/escooterpdx/2018-e-scooter-findings-report.
San Francisco Municipal Transportation Agency's, 2019. Powered scooter share mid-pilot evaluation. URL https://www.sfmta.com/sites/default/files/reports-and-documents/2019/08/powered_scooter_share_mid-pilot_evaluation_final.pdf.
Shaheen, S.A., Cohen, A.P., Martin, E.W., 2013. Public bikesharing in North America: Early operator understanding and emerging trends. Transp. Res. Rec. (ISSN: 0361-1981) 2387 (1), 83–92, Publisher: SAGE Publications Inc.
Transport for London, 2019. Tfl publishes new data on London's e-scooter rental trial as it celebrates its second anniversary. URL https://modeshift.org.uk/modeshift-news/tfl-publishes-new-data-on-londons-e-scooter-rental-trial-as-it-celebrates-its-second-anniversary.
Wang, Y., Zhan, Z., Mi, Y., Sobhani, A., Zhou, H., 2022. Nonlinear effects of factors on dockless bike-sharing usage considering grid-based spatiotemporal heterogeneity. Transp. Res. D (ISSN: 1361-9209) 104, 103194.
Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. J. Amer. Statist. Assoc. (ISSN: 0162-1459) 58 (301), 236–244.
WorldPop, 2023. World pop.. URL https://hub.worldpop.org/.
Zhao, P., Haitao, H., Li, A., Mansourian, A., 2021. Impact of data processing on deriving micro-mobility patterns from vehicle availability data. Transp. Res. D (ISSN: 1361-9209) 97, 102913.
Ziedan, A., Darling, W., Brakewood, C., Erhardt, G., Watkins, K., 2021. The impacts of shared e-scooters on bus ridership. Transp. Res. A (ISSN: 0965-8564) 153, 20–34.
Zuniga-Garcia, N., Tec, M., Scott, J.G., Machemehl, R.B., 2022. Evaluation of e-scooters as transit last-mile solution. Transp. Res. C (ISSN: 0968090X) 139.
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 (ISSN: 0264-2751) 99, 102614.
Zuo, T., Wei, H., Rohne, A., 2018. Determining transit service coverage by non-motorized accessibility to transit: Case study of applying GPS data in Cincinnati metropolitan area. J. Transp. Geograp. (ISSN: 0966-6923) 67, 1–11.