Understanding urban mobility patterns from a spatiotemporal perspective

Citation for published version (APA):

Gan, Z., Yang, M., Feng, T., & Timmermans, H. J. P. (2020). Understanding urban mobility patterns from a spatiotemporal perspective: daily ridership profiles of metro stations. *Transportation*, 47(1), 315-336. <https://doi.org/10.1007/s11116-018-9885-4>

Document license:
TAVERNE

DOI:
[10.1007/s11116-018-9885-4](https://doi.org/10.1007/s11116-018-9885-4)

Document status and date:
Published: 01/02/2020

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Understanding urban mobility patterns from a spatiotemporal perspective: daily ridership profiles of metro stations

Zuoxian Gan^{1,2} · Min Yang^{1,2} · Tao Feng³ · Harry Timmermans^{3,4}

Published online: 14 May 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract Smart card data derived from automatic fare collection (AFC) systems of public transit enable us to study resident movement from a macro perspective. The rhythms of traffic generated by different land uses differ, reflecting differences in human activity patterns. Thus, an understanding of daily ridership and mobility patterns requires an understanding of the relationship between daily ridership patterns and characteristics of stations and their direct environment. Unfortunately, few studies have investigated this relationship. This study aims to propose a framework of identifying urban mobility patterns and urban dynamics from a spatiotemporal perspective and pointing out the linkages between mobility and land cover/land use (LCLU). Relying on 1 month's transactions data from the AFC system of Nanjing metro, the 110 metro stations are classified into 7 clusters named as employment-oriented stations, residential-oriented stations, spatial mismatched stations, etc., each characterized by a distinct ridership pattern (combining boarding and alighting).

✉ Min Yang
yangmin@seu.edu.cn

Zuoxian Gan
zxgan@seu.edu.cn

Tao Feng
T.Feng@tue.nl

Harry Timmermans
H.J.P.Timmermans@tue.nl

¹ School of Transportation, Southeast University, No. 2 Southeast University Road, Nanjing 211189, People's Republic of China

² Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University, No. 2 Southeast University Road, Nanjing 211189, People's Republic of China

³ Urban Planning Groups, Department of Urban Science and System, Eindhoven University of Technology, PO Box 513, 5600 Eindhoven, The Netherlands

⁴ College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, People's Republic of China

A comparison of the peak hourly ridership of the seven clusters is conducted to verify whether the clustering results are reasonable or not. Finally, a multinomial logit model is used to estimate the relationship between characteristics of the local environment and cluster membership. Results show that the classification based on ridership patterns leads to meaningful interpretable clusters and that significant associations exist between local LCLU characteristics, distance to the city center and cluster membership. The analytical framework and findings may be beneficial for improving service efficiency of public transportation and urban planning.

Keywords Urban mobility · Ridership patterns · Smart card data · Station clustering · LCLU

Introduction

Understanding urban mobility patterns, and exploring the relationship between daily travel rhythms and land use are important for city planners and transit authorities alike. These topics are essential for formulating effective policies and supporting scheduling, routing, and urban planning solutions (Kim et al., Kim et al. 2016; Tupper et al. 2016). Urban places are neither static nor homogeneous. The daily dynamics and heterogeneity of an urban area are reflected in the movement of people. As gathering and distribution centers of passengers, public transit (PT) stations and their service areas provide the windows to capture urban human movement and explore the different daily travel patterns in different urban places. In turn, this knowledge serves to understand the structure of large cities from a spatiotemporal perspective.

PT stations facilitate the movement of people and are characterized by diverse mobility patterns, thanks to their two salient features: nodes (nodes of networks) and places (places in the city) (Bertolini 1996). On the one hand, station areas are the key nodes in public transit networks and make it convenient for passengers to travel from one place to another place. On the other hand, station areas do not only provide good accessibility but also afford numerous activity opportunities. People are usually required to tap-in at the starting station and tap out at the destination station when they travel by public transportation. The different number of boardings and alightings of a station during different temporal intervals (profiles) characterize fluctuations in ridership. The different types of fluctuations are perhaps the best evidence to sketch the rhythm of daily urban mobility.

Data from Automatic fare collection systems (AFC) provide an opportunity to obtain a good understanding of urban mobility patterns. Nowadays, AFC are widely used in PT systems (bus, light railway and metro etc.) all over the world. People are usually required to use their smart cards when they get on and get off a PT vehicle (the transaction of tapping-out may not be necessary in some PT systems). Since the transactions are time stamped and geo tagged, individual travel data, including boarding time, alighting time, original station, destination station and fare become available. Therefore, it becomes easy to derive travel trajectories and examine daily regularities across stations. Ridership patterns can be measured within given temporal intervals (Reads et al. 2016).

People moving for different activities to specific urban functional areas result in the dispersion of passengers at different time periods and PT stations. Since heterogeneous ridership patterns of different PT stations may be related to urban structure, especially local land cover/land use (LCLU) (Chen et al. 2009; Reads et al. 2016), attempts to infer and

analyze the linkage between the ridership profiles and the local LCLU seems indispensable in the quest to better understand resident mobility in an urban context. For example, due to the conventional 9-to-5 work mechanism, a metro station surrounded by residential areas may dispatch a lot of people during the morning peak and attract a large number of people during the afternoon peak, whereas the ridership profile of a metro station dominated by commercial and business areas may have a reverse profile.

Although the number of studies concerned with people movement and urban mobility has rapidly increased lately, only a few of these have explored the boarding and alighting profile on a weekday basis (Chen et al. 2009; Roth et al. 2011; Ma et al. 2013; Munizaga et al. 2014; Zhong et al. 2015; Reades et al. 2016; El Mahrsi et al. 2017). To the best of our knowledge, this study presents one of the first attempts of exploring the relationship between local LCLU and metro ridership patterns, while at the same time analyzing the layout of urban functions in a Chinese city. In this study, we focus on identifying the spatiotemporal ridership profiles of metro stations and detecting the relationship between these profiles and local LCLU using smart card data collected for the Nanjing Metro System, China. The objective of this study can be summarized as follows:

1. Extracting several distinct spatiotemporal ridership profiles from continuously collected transaction records. Combining daily boarding and alighting sequences should deepen our understanding of urban mobility patterns and urban form from a space–time perspective. It may give transportation operators a glimpse of what corresponding operational measures should be adopted to satisfy the diurnal citizen movement.
2. Exploring the linkage between ridership patterns and local LCLU to understand heterogeneity in ridership patterns and highlight how urban structure motivates urban mobility during a diurnal cycle. This large-scale analysis provides suggestions for urban planners to detect flawed urban planning and implement urban sustainable development.

The remainder of the paper is organized as follows. The literature review is discussed in the next section, and the study area and dataset are presented in “[Study area and data collection](#)” section. Related approaches are introduced in “[Methodology](#)” section. “[Results](#)” section presents the results of the empirical study in Nanjing. The final conclusions and limitations of this study are discussed in “[Discussion and conclusions](#)” section.

Literature review

Conventional approaches to analyze travel patterns of urban residents are normally based on datasets collected from questionnaires or travel dairies, the sample sizes of which are generally small and not easily updated in real time (Doherty and Miller 2000; Axhausen et al. 2002). These drawbacks limit our ability to understand urban mobility in detail. The emergence of information and communication technologies (ICT) and the rise of big data have advanced research on travel behavior, activity mining, mobility flows and mobility regularities (Batty et al. 2012). Big data with specific time–space information of travelers make it possible to analyze urban mobility from a spatiotemporal perspective. Moreover, the sample size of these datasets also improves the accuracy of the analysis. Big travel data can be divided into GPS devices, taxi trajectories, mobile phone data, smart card data (SCD) and social media data (González et al. 2008; Chen et al. 2009; Roth et al. 2011;

Feng and Timmermans 2013; Ma et al. 2013; Munizaga et al. 2014; Zhong et al. 2015; Xu et al. 2015; Manley et al. 2018; Reades et al. 2016; Goulet-Langlois et al. 2016; Yang et al. 2016; El Mahrssi et al. 2017). In contrast to taxi trajectories and social medial check-in data, the other three types of data have two salient advantages: high spatiotemporal precision and tracking long-term resident movements.

In recent years, many studies on urban mobility are based on the application of SCD. Morency et al. (2006) analyzed the variability in travel behavior of transit users, in which the behavior of regular transit users is associated with the number of boardings per day and boarding time. Based on real-time data derived from the ‘Oyster’ cards in the London subway system, Roth et al. (2011) state that patterns of intraurban mobility are very heterogeneous in terms of volume, but not in terms of distance travelled. Ceapa et al. (2012), also using ‘Oyster’ card data, found that crowding levels at stations are highly regular and overcrowded stations are identified to help passengers change their travel patterns. Some similar work can be found in other cities (see for example Liu et al. 2009; Kusakabe and Asakura 2014; Zhong et al. 2015).

Other studies have given special attention to the clustering of ridership patterns of PT stations. Based on over 15 million tap-in and tap-out data of 296 metro stations in London, Reades et al. (2016) aggregated boarding and alighting activities to 15-min intervals, threw away the day-of-week information and classified the stations into five clusters. Yin et al. (2016) utilized the maximum values, kurtosis and skewness, the equilibrium coefficient and peak hour factors of boardings and alightings to classify the stations into eight clusters. Tan et al. (2017) used principal component analysis to extract the main characteristics of the ridership time-sequences and then applied the *k*-means algorithm to identify seven types of stations in the Guangzhou Metro System. Moreover, they found that diversity and types of metro stations varied as the distance to the city center increased.

As for the linkage between urban mobility patterns and local environment, based on boarding data collected from New York City Transit, Chen et al. (2009) first calculated the values of 19 variables for each station, including morning peak hourly ridership, afternoon peak hourly ridership, morning peak hourly riders/total daily riders, afternoon peak hourly riders/total daily riders and early morning hourly ridership, etc., then divided the New York City subway stations (diurnal ridership patterns) into five clusters (commercial, highly mixed-use, moderately mixed-use, residential and residential with transfer), and finally explored the relationship between the diurnal ridership patterns and local land use. However, this research did not reflect the complete ridership patterns on a day basis because of lacking alighting data. Kim et al. (2016) conducted similar work in Seoul, South Korea. They used ridership counts of selected intervals to classify the subway stations (diurnal ridership patterns) into five clusters and found that the diurnal patterns were associated with land use within the station area of influence.

This study takes this stream of literature one step further. In this study, according to longitudinal observations of ridership, metro stations are clustered into specific categories within distinctive ridership profiles. The classification of metro stations is regarded as a bridge to understand the relationship between the local environment and daily urban mobility patterns.

Study area and data collection

Nanjing, the heart of Nanjing Metropolitan, is the capital of Jiangsu Province and one of most important political, economic, and cultural centers in eastern China, known also by its famous historical and cultural richness. It covers a total of 6587.02 square kilometers

Fig. 1 Study area and Nanjing Metro network

and 11 districts, including 5 core districts (Xuanwu District, Qinhuai District, Gulou District, Jianye District and Yuhuatai District), 3 inner districts (Qixia District, Jiangning District and Pukou District) and 3 remote districts (Liuhe District, Lishui District, Gaochun

District) (Fig. 1). The city represents the typical microcosm of Chinese cities, which are still under rapid population urbanization and urban growth. Its total population increased from 4.7 million in 1985 to 8.27 million in 2016 and the number of urban residents increased to 6.78 million, accounting for 82% increase. Accordingly, the urban built-up area of Nanjing increased from 133.75 square kilometers in 1980 to 960.08 square kilometers in 2012, and 1125.78 square kilometers in 2016. The gross domestic product (GDP) of Nanjing is 1050 billion RMB (about 166.7 billion USD), ranked the 11th in all Chinese cities in 2016 and the real GDP per capital is 127,264 RMB (about 20,187 USD), ranked the 3th, which is the second largest commercial center in the Yangtze Delta.

To respond to the “bus priority” policy emphasized by the Chinese Government and relieve urban traffic pressure, Nanjing government sets “Nanjing Urban Rail Transit Network Planning” in 2002. In 2005, Nanjing built its first urban rail line—metro line 1. By April 2015, six metro lines were completed running through nine districts (two remote districts, Lishui and Gaochun are not served) with 112 stations and the total mileage is 225 km, ranked the fourth in China (after Shanghai, Beijing and Guangzhou) and the eleventh worldwide (before many other megacities such as Hong Kong, Mexico City, Berlin and Washington, DC). The Nanjing Metro System has been a major transportation mode for urban daily commuting. It carries 717 million passengers annually and its share in 2015 was about 34.8% of the passenger volume of all types of public transportation.

The data about metro ridership were collected from the Nanjing Metro Corporation (NMC). The provided metro dataset consists of a full month (April 2015) of usage information. Every trip record includes smart-card ID, start time, end time, line number of boarding, line number of alighting, station ID of boarding, station ID of alighting and trip duration. The total number of trip records was 43 million; the average daily number of boarding and alighting was close to 1.5 million. For the purpose of analysis, the dataset was first aggregated and converted into hourly records (Fig. 2). There were 21 weekdays, 8 weekends and 1 holiday in April 2015. We focused on people’s travel patterns on weekday and the records of 9 non-weekdays were not considered. Considering that people’s travel patterns are replicable on a daily basis,¹ the average daily weekday ridership of each station is used without including the day of week information, which eliminates the possible influence of any extraordinary events (such as a large-scale concert).

In addition to the metro trip dataset, we also used the land use data of the Nanjing region from the Nanjing Urban Planning Bureau (NUPB). This file includes different land use types (administration and public services, commercial and business facilities, industrial, green and open space, residential, others and unconstructed land) based on Chinese national standards—“Code for classification of urban land use and planning standards of development land” (MOHURD 2011). We further calculated the ratio of different land uses utilizing the land use file in ArcGIS 10.3. University and spot profiles which were dummy coded using Baidu Map (whether there is a university or key scenic spot within the metro station influence area).

Methodology

Cluster analysis was used to classify the Nanjing Metro stations according to the different diurnal travel patterns. Next, a multinomial logit regression was estimated to explore the relationship between each distinctive cluster and local land use characteristics. The

¹ We examined the normalized data (see subsection “Variable selection and normalization”) of station boardings and alightings between weekdays, but no significant difference was found.

Card ID	Begin_time	line num. of boarding	In_station ID	Card_type	Line num. of alighting	Out_station ID	End_time	T(seconds)
01911DBC** *****	28APR15_1 8:12:29	4	1	67	4	2	28APR15_1 8:24:07	698
0604006D*** *****	10APR15_1 0:56:35	4	1	51	4	2	10APR15_1 1:06:52	617
060405FC*** *****	18APR15_1 4:57:24	1	6	51	1	9	18APR15_06 :39	555
06049358*** *****	18APR15_1 9:05:34	1	7	51	2	30	18APR15_1 9:26:26	1252
0804453D*** *****	01APR15_1 5:59:44	1	10	80	5	44	01APR15_1 6:29:36	1792
0604CBA7** *****	26APR15_0 8:56:27	3	41	51	1	12	26APR15_0 9:24:31	1684
...	...	...	...	...	...	...	...	...

Day	Station ID	Counts	6:00~7:00	7:00~8:00	8:00~9:00	...	22:00~23:00	23:00~24:00
1	1	Boarding Alighting	114 37	321 158	331 462	...	24 87	2 17
1	2	Boarding Alighting	224 200	857 1123	754 4482	...	237 122	24 50
1	3	Boarding Alighting	245 137	775 1416	680 4952	...	109 108	10 15
...	...	...	...	...	...	...	...	...
2	1	Boarding Alighting	94 18	314 161	314 433	...	31 62	3 12
2	2	Boarding Alighting	193 168	799 1144	668 4543	...	299 121	17 53
...	...	...	...	...	...	...	...	...

Fig. 2 Nanjing Metro data used in the study (April 2015) (the line number of boarding/alighting in the AFC system is not coded according to the actual metro line. For example, the “4” in the column of line num. Boarding/alighting represents Nanjing Metro Line 10 instead of Nanjing Metro Line 4)

following sub-sections will first present the variable selection and normalization, then the cluster analysis and finally the estimation of the multinomial logit model.

Variable selection and normalization

There are two common approaches to define the variables in cluster analysis of time-sequences. One is using the statistical characteristics of ridership profiles as variables, because the distribution of daily ridership varies by station over time and these specific periods are most useful to recognize different daily travel patterns. Morning and afternoon peak hourly ridership, and the ratio of morning and afternoon ridership over daily total etc. are normally utilized (Chen et al. 2009). A similar approximate method is using the eigenvalues of total time-sequence to identify different ridership patterns. The kurtosis/skewness of daily boarding and alighting ridership, the number of maximum ridership of boarding and alighting and the equilibrium coefficient of boarding and alighting were applied in the clustering of Beijing Metro stations (Yin et al. 2016).

The other approach is to directly use the station ridership as candidate variables rather than selecting some distinctive period or eigenvalues in advance (Kim et al. 2016; Tan et al. 2017). Since hourly ridership is one important aspect of travel patterns at each metro station, hourly ridership itself can be used as a variable in the classification of stations. In contrast to the first approach, this approach contains more information.

The second approach was therefore used in this study. Each station has 36 (boarding/alighting $\times$ 18 h intervals) variables for cluster analysis. Because the average daily ridership differs between stations, Z-score normalization was used to convert the hourly counts into hourly indices. Larger positive indices represent greater hourly ridership. The Z-score equation is expressed as follows:

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \quad (1)$$

where X_{ij} is the i th hourly ridership being normalized of station j , μ_j is the arithmetic mean of hourly ridership of station j , and σ is the standard deviation of hourly ridership of station.

Cluster analysis

The k -means cluster analysis, popular for its simplicity and efficiency, was used to identify the clusters of stations according to the ridership patterns (Berkhin 2006). It iteratively divides the dataset into k clusters based on minimization of the distance of each data point to its cluster center. Given a set of observations $(x_1, x_2, \dots, x_n)$, where each observation is a d -dimensional vector (in this study, d is equal to 36, namely, 36 normalized hourly station ridership), the k -means algorithm aims to partition the n observations into k clusters $C = \{C_1, C_2, \dots, C_i, \dots, C_k\}$ to minimize the within-cluster sum of distance variance (SDV). They are calculated as follows.

$$SDV = \sum_i^k \sum_{X \in C_i} ||X - \mu_i||^2 \quad (2)$$

where μ_i is the cluster center of category C_i .

One challenge in k -means cluster analysis is determining the number of clusters. Three most frequently used indices are the DB index (Davies and Bouldin 1979), the Dunn index (Dunn 1973), and the XB index (Xie and Beni 1991). The DB index presents the ratio of the sum of intra-cluster distances to inter-cluster separation and, therefore, the optimal number of cluster is determined by minimizing the value of the index. The Dunn index measures the ratio of the separation (minimum inter-cluster distance) to the compactness (maximum intra-cluster distance). The greater the Dunn index value, the better the clustering is. The XB index aims to quantify the ratio of the total variation within clusters to the minimum distance between cluster centers. The best number of cluster should minimize the value of the index. They are defined as:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j:j \neq i} \left(\frac{e_i + e_j}{D_{ij}} \right) \quad (3)$$

$$Dul = \min_x \left(\frac{\min_{j:j \neq i} \left(\frac{\min_{x \in C_i, y \in C_j} d(x, y)}{\max_k \left(\max_{x, y \in C_k} d(x, y) \right)} \right)}{\max_k \left(\max_{x, y \in C_k} d(x, y) \right)} \right) \quad (4)$$

$$XBI = \frac{\sum_i \sum_{x \in C_i} d^2(x, \mu_i)}{n \times \min_{i,j \neq i} d^2(\mu_i, \mu_j)} \quad (5)$$

where e_i and e_j are average intra-cluster distances of Clusters i and j , and $e_i = \frac{\sum_i \|x_i - u_i\|}{n_i}$; D_{ij} is the inter-cluster distance of Clusters i and j , and $D_{ij} = \|u_i - u_j\|$; n_i is number of observation that belong to cluster i . $d(x, y)$ is the distance between observations x and y . n is the number of observations in the dataset.

Multinomial logit model

After the metro stations were clustered, a multinomial logit model was used to estimate the relationship between local LCLU characteristics and clusters. The probability of a metro station being classified into one cluster is described as follows.

$$P(y_i = j | \mathbf{x}) = \frac{e^{\mathbf{x}_i \beta_j}}{\sum_j e^{\mathbf{x}_i \beta_j}} \quad (6)$$

where $y_i = j$ represents the station i being classified into cluster j , $\mathbf{x}$ is the vector of independent variables including different LCLU characteristics, β_j is the coefficient.

Results

Clustering of metro stations

As described above, first the number of clusters should be decided. Here, we run tests for $k = 3$ to $k = 10$. Results for the DB index, Dunn index and XB index are shown in Fig. 3. Based on these results, we decided to use 7 clusters for further analyses.

Finally, all 110 metro stations² were classified into 7 clusters using the k -means algorithm. Figure 4 shows the spatial distribution of the clustering results. Figures 5 and 6 show the daily ridership patterns of different clusters. Each line in the figure corresponds to a metro station. To highlight the temporal profiles of each cluster, the average ridership and the quantiles 0.25 and 0.75 of the total ridership within each cluster are shown in the figures. Two specific clusters are presented in Fig. 5. The stations in the two clusters are close to special amenities, such as a traffic hub, scenic spot or university campus. The following text will first describe and analyze the two clusters “Traffic hub and scenic spot stations” and “University influence stations”.

² Two metro stations (Lukou Airport Station—the terminal station of Metro Line S1 and Jinniuhu Station—the terminal station of Metro Line S8) were finally excluded due to two reasons. One is that the ridership patterns in the two stations are different from that of other stations, and they are often classified into a stand-alone cluster when they are considered in the cluster analysis. The other is that the SDV within the two stations is higher than the SDV without the two stations of the same number of clusters. For example, with 7 clusters, the SDV within them is 530.88 while the SDV without them is just 424.90.

Fig. 3 Indices calculated for different number of clusters. **a** DB index, **b** Dunn index, **c** XB index

Cluster 1 Traffic hub and scenic spot stations. 6 stations including Nanjing Railway Station, Nanjing South Railway Station and Fuzimiao Station³ are classified into this cluster. Figure 5a shows the changes in ridership profiles in the cluster are smaller than in other clusters. As summarized in Table 1, the morning/afternoon peak hourly boarding/alighting are not the most significant ones among all clusters. The ratio of the morning peak hourly boarding and the morning peak hourly alighting and of the afternoon peak hourly boarding and afternoon peak hourly alighting both approximate 100%. It corresponds to the reality that the temporal distribution of ridership is more balanced and the peak hour factor is not very salient in these stations, compared to other types of stations.

Cluster 2 university influenced stations. 8 stations including NJU (Nanjing University Xianlin Campus Station), CPU (China Pharmaceutical University) Station and SEU (Southeast University Jiulonghu Campus Station) are included in this cluster. These stations share some common characteristics: the new campuses of one or more universities are located near to a metro station. As shown in Fig. 5b and Table 1, the ridership patterns of these stations differ from Cluster 1. In particular, the morning peak hourly boarding and the morning peak hourly alighting differ greatly (157.80%). Their ridership profiles also differ from the other five clusters. Compared to other clusters, this cluster has one more peak in its ridership profile, which relates to boarding around noon.

The remaining clusters shown in Fig. 6 are named “Employment-oriented stations”, “Mixed mainly employment-oriented stations”, “Residential-oriented stations”, “Mixed mainly residential-oriented stations” and “Spatial mismatched stations”, because their ridership patterns show a typical flow characteristic of morning and evening tides. Except for

³ Fuzimiao, also refers to Confucius Temple, located in Qinhuai District (near the city center), is a very famous open AAAA (the highest grade of Chinese scenic spots) scenic spot in Nanjing. With the ancient magnificent architectures, profound culture and superior geographical position, it has been one of the top tourist spots in Nanjing with an average of more than 100 thousand visitors per day.

Fig. 4 Spatial distribution of cluster membership

the cluster “Spatial mismatched stations” which has large boardings and alightings in both morning and afternoon peak hours, other stations within these clusters have a wide gap between the boardings and alightings in morning (and afternoon) peak hours.

Fig. 5 Daily ridership profiles for clusters 1 and 2. The time line charts on the left shows how the hourly normalized ridership of each cluster membership varied across a weekday, whereas the average ridership (also is the cluster center) and quantiles 0.25 and 0.75 of the total ridership is also shown on the right. **a** Cluster 1: traffic hub and scenic spot stations. **b** Cluster 2: university influence stations

Cluster 3 Employment-oriented stations. 5 stations are included. Figure 6a shows the afternoon peak hourly boarding and morning peak alighting are distinctive. The two temporal ridership account for 27.21% of total daily boardings and 30.39% of total daily alightings. However, the morning peak hourly boarding and afternoon peak hourly alighting are both small and account for not more than 13.50% (12.29% of afternoon peak hourly alighting is the smallest among all clusters). Table 1 shows that boardings are significantly less than alightings during weekday morning peak hours (43.83%), while the opposite pattern is observed for weekday afternoon peak hours (221.40%, the highest value in the sixth column). Thus, stations in this cluster present a distinctive ridership pattern—“morning alighting and afternoon boarding”.

Cluster 4 Mixed mainly employment-oriented stations. 22 stations are included in this cluster and their ridership patterns are similar to those of Cluster 3. Similarly, the ratio of morning peak alighting and afternoon peak hourly boardings is very large, while the ratio of morning peak boarding and afternoon peak alightings is very small. Table 1 shows that the ratio of afternoon peak hourly boarding and afternoon peak hourly alighting is 210.16% which is the second highest value.

Cluster 5 Residential-oriented stations. This cluster includes 35 stations, which is the largest of all clusters. It is evident that four peaks are observed. The morning peak boarding and afternoon peak alighting are higher than the other two peak ridership (Fig. 6c). As shown in Table 1, three temporal ridership are higher than those in other clusters: morning peak boarding (32.79%), afternoon peak alighting (26.43%) and the ratio of morning peak boarding over morning peak alighting (341.56%). These indexes indicate that, on weekdays, many people go to other places in the morning and come back to the same place in the afternoon. The ridership patterns of these stations are just opposite to that in Cluster 3, which is “morning boarding and afternoon alighting”.

Cluster 6 Mixed mainly residential-oriented stations. This cluster consists of 11 stations. Its ridership profile is similar to those in Cluster 5. Figure 6c, d show that the difference

Fig. 6 Daily ridership profiles for clusters 3, 4, 5, 6 and 7. **a** Cluster 3: employment-oriented stations. **b** Cluster 4: mixed mainly employment-oriented stations. **c** Cluster 5: residential-oriented stations. **d** Cluster 6: mixed mainly residential-oriented stations. **e** Cluster 7: spatial mismatched stations

Table 1 Comparison of peak hourly ridership of 7 clusters (peak hour as defined in the study is 7:00–9:00 and 17:00 ~ 19:00)

Cluster	Morning peak boarding/daily boarding (%)	Afternoon peak boarding/daily boarding (%)	Morning peak alighting/daily alighting (%)	Afternoon peak alighting/daily alighting (%)	Morning peak boarding/ morning peak alighting (%)	Afternoon peak boarding/ afternoon peak alighting (%)
1 (Traffic hub and scenic spot stations)	13.19	17.67	14.34	18.55	91.98	95.26
2 (University influenced stations)	20.53	16.32	13.01	20.25	157.80	80.59
3 (Employment-oriented stations)	13.32	27.21	30.39	12.29	43.83	221.40
4 (Mixed mainly employment-oriented stations)	13.13	27.72	31.91	13.19	41.15	210.16
5 (Residential-oriented stations)	32.79	12.20	9.60	26.43	341.56	46.16
6 (Mixed mainly residential-oriented stations)	25.21	11.61	13.64	20.21	184.82	57.45
7 (Spatial mismatched stations)	26.24	21.05	22.36	19.98	117.35	105.36

The two largest values in each column are highlighted in italics

Fig. 7 LCLU in Nanjing and the temporal characteristics of dispersion and convergence of each cluster. Key traffic hubs, new university campuses and key scenic spots are labelled in the maps. The values of temporal characteristics are based on the difference between the average normalized lightings and boardings. **a** LCLU in the study area. **b** A larger version for the core city. **c** The temporal characteristics of each cluster

between the middle two ridership peaks (afternoon boarding and morning alighting) and the outer two ridership peaks (morning boarding and afternoon alighting) is not very significant. Statistical data analysis shown in Table 1 also supports this conclusion. The ratio of morning peak boarding and morning hourly alighting is 184.82%, the second biggest value following the corresponding value (341.56%) of Cluster 5. The ratio of afternoon peak boarding and afternoon peak alighting is the second smallest value (57.45%), following the corresponding value (46.16%) of Cluster 5.

Cluster 7 Spatial mismatched stations. 23 stations are included in this cluster. Figure 6e shows that the peaks of the ridership profiles are obvious but their gaps are not as significant as those in Clusters 3, 4, 5 and 6. Table 1 shows that the ratios of the four peak hourly ridership are all more than 19%, but the value in the fifth and sixth columns are not much higher or lower than 100% (117.35 and 105.36% respectively). It indicates that many out-of-station people arrive during the morning peak hours and leave during the afternoon peak hours, while many local residents go out during the morning peak hours and come back during the afternoon peak hours. This ridership pattern reflects a common phenomenon in many cities today - “the separation of workplace and residence”.

Table 2 Statistical characteristics of local environment

LCLU characteristics	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Administrative area (%)	7.78	34.72	2.92	18.69	12.91	8.23	15.38
Commercial and business area (%)	18.39	5.42	5.23	17.44	7.43	4.65	10.96
Industrial area (%)	5.90	9.44	39.55	9.40	10.66	6.99	13.23
Green space (%)	15.51	10.43	1.49	8.89	8.81	3.02	11.01
Residential area (%)	27.65	22.17	5.88	27.84	35.88	28.40	28.34
Others (%)	14.96	3.84	9.72	5.10	6.64	10.92	8.67
Unconstructed (%)	11.46	13.98	35.19	12.82	17.40	37.76	12.41
Traffic hub or scenic spot dummy (%)	83.33	0	0	9.09	5.71	0	17.39
University dummy (%)	0	87.50	0	22.73	22.86	0	13.04
Distance to city center (m)	6035	21388	29,587	7577	15,790	37,039	9668

Relationship between ridership patterns and local environment

In the above sections, stations with similar ridership patterns were classified into clusters using the *k*-means clustering algorithm. Stations close to a university campus or scenic spots were clustered into Clusters 1 and 2, while the other stations in other 5 clusters present the characteristics of employment-oriented, residential-oriented or spatial mismatched development. Figure 7a shows the location of the metro stations and the LCLU in the study area. A local enlarged drawing of the core city is presented in Fig. 7b. Based on the analysis of the difference between average boarding and alighting profiles in hourly intervals of each cluster, Fig. 7c describes the dispersion and convergence of the mean temporal pattern for each cluster. It is based on the differences between the numbers of normalized alightings and boardings. The warm colors represent the convergence, which indicates that the number of tap-ins exceeds the number of tap-outs in the given time interval. The darker the color, the larger the gap. Cool colors represent the dispersion. A characteristic of clusters 3 and 4 is that the major convergence is concentrated from 7:00 to 11:00 and dispersion is concentrated from 17:00 to 18:00, while clusters 5 and 6 shows the reverse trend. Figure 7c also shows that cluster 1 demonstrates convergence during the daytime but it is not very significant, while in cluster 2, dispersion occurred first from 6:00 to 18:00 and convergence subsequently occurred from 18:00 to 24:00. The results in Fig. 7c are in accordance with the plots in Figs. 5 and 6.

Table 2 shows the mean values of the LCLU characteristics of each cluster. On average, Clusters 1 and 2 have the highest proportion of traffic hub, scenic spots and university campuses, respectively.⁴ Cluster 3 (Employment-oriented stations) has the highest proportion of industrial land use, and the smallest proportion of residential land use. Except for Cluster 1, Cluster 4 (Mixed mainly employment-oriented stations) has the highest proportion of

⁴ The distance threshold of the influence area of metro station is identified 800 m (0.5 mile) in the study, which is generally accepted and used as the standard walking distance and the adequate representation of a person's willingness to walk to rail transit (for example: Kuby et al. 2004; Hess and Almeida 2007; Zhao et al. 2013). Moreover, the Ministry of Housing and Urban–Rural Development of People's Republic of China (MOHURD) launched a design guidance of urban rail—"Guidelines for Planning and Design of Urban Rail", and defined a radius of 800 m as the influence area of rail stations (MOHURD 2015).

commercial & business areas (17.44%), which is two times more than for Clusters 5 and 6. The highest proportion of residential land use occurs in Cluster 5 (Residential-oriented stations). Compared to Clusters 3, 4, 5 and 6, the last cluster (spatial mismatched stations) has a more balanced ratio of different land uses.

Because the spatial distribution of clusters shows that the mixed mainly employment-oriented stations are more concentrated in the core city, whereas the residential-oriented stations and mixed mainly residential-oriented stations are typically located out of the city center (see Fig. 4), we also calculated the average distance to the city center for each cluster. As expected, results presented in Table 2 show that the average distance to the city center among these clusters differs widely between the clusters.

The results of the multinomial logit model are listed in Table 3. The log-likelihood value increases from -190.68 for a constant-only model to -92.05 for the final model. Pseudo $R^2=0.51$ indicates a high goodness-of-fit. Thus, it suggests a clear link between the local environment and different ridership patterns. For Cluster 1, the traffic hub or scenic spot dummy is a positively related variable, while it is insignificantly associated with other ridership patterns. As expected, there is a significant relationship between the university dummy, distance to the city center and high-education based stations (Cluster 2). It accords with the phenomenon that remote stations close to new university campuses are clustered in this category, whereas the stations close to the old campuses in core urban areas are not included. The model confirms that a station located in an area with a higher percentage of commercial and business land use is more likely assigned to Clusters 3 and 4 than to Cluster 7. Meanwhile, a negative relationship is found between the percentage of commercial and business land use and residential based stations (Clusters 5 and 6). In addition, the model indicates that a station located in a high residential land use area is most likely to belong to Cluster 5 or 6, and least likely to Cluster 3 and 4. Controlling the variables of LCLU, the significant and positive coefficients of distance to the city center for Clusters 3, 5 and 6 suggest that employment-oriented stations, residential-oriented stations and mixed mainly residential-oriented stations are more likely located in the outskirts of the city. In the contrast, the mixed mainly employment-oriented stations are more likely located in the core urban areas.

Discussion and conclusions

In this study, we proposed a framework for understanding urban mobility patterns from daily ridership profiles of metro stations. All metro stations were allocated to specific clusters based on hourly intervals of both boarding and alighting rather than boarding or alighting alone. According to their distinctive weekday ridership patterns, each cluster was then given a specific and meaningful name. After that, different peak hourly ridership characteristics was compared to test and verify whether the clustering results are reasonable or not. Finally, a multinomial logit model was utilized to figure out if there is significant relationship between distance to city center, LCLU within the influence area of metro station and ridership profiles for helping understand the reasons behind the travel phenomenon.

According to their distinctive weekday ridership patterns, each cluster could be given a specific and meaningful name: traffic hub and scenic spot stations, university influenced stations, employment-oriented stations, mixed mainly employment-oriented stations, residential-oriented stations, mixed mainly residential-oriented stations and spatial mismatched stations. Then, we summarized and compared the peak hourly ridership of 7

Table 3 Regression results of the multinomial logit model (base on Cluster 7: spatial mismatched stations)

Variables	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
	Coeff.	Z	Coeff.	Z	Coeff.	Z	Coeff.	Z	Coeff.	Z	Coeff.	Z
Constant	-4.48	-0.79	-9.69**	-2.32	-10.95*	-1.66	-1.07	-0.49	-3.58	-1.77*	-12.38**	-2.42
Administrative area	-33.00*	-1.70	5.64	0.89	-7.48	-0.33	1.54	0.47	-1.97	-0.52	6.65	0.79
Commercial and business area	18.48*	1.90	-2.33	-0.15	4.46	0.15	10.73*	1.95	-8.72	-1.31	-20.33	-1.13
Industrial area	-13.90	-0.70	2.05	0.33	5.03	0.58	0.10	0.03	-2.53	-0.72	-1.41	-0.19
Residential area	-1.42	-0.19	5.90	0.94	-14.33	-0.69	-1.96	-0.61	8.96***	2.80	8.28	1.23
Traffic hub or scenic spot dummy	7.80*	1.93	-23.29	-0.00	-16.35	-0.00	-1.18	-1.12	-0.17	-0.15	-20.26	-0.00
University dummy	-15.16	-0.01	3.79**	2.02	-11.80	-0.00	1.05	0.96	1.79	1.56	-18.17	-0.00
Distance to city center	0.0009	0.34	0.0003***	2.65	0.0005***	2.65	-0.00002	-0.28	0.0002***	2.62	0.0005***	3.46
LL (0)	-190.68											
LL (β)	92.05											
2[LL (0)- LL (β)]	197.27											
Pseudo R ²	0.51											

*Indicates significance at 0.1 level

**Indicates significance at 0.05 level

***Indicates significance at 0.01 level

clusters to check out our clustering of metro stations. The results show that, among these stations, the employment-oriented stations and the mixed mainly employment-oriented stations have the highest ratios of afternoon peak hourly boarding and morning peak hourly alighting. Meanwhile, the residential-oriented stations and the mixed mainly residential-oriented stations have the highest ratios of afternoon peak hourly alighting and morning peak hourly boarding in weekdays. The four peak hours of spatial mismatched (mixed-use) stations are obvious, but there is no significant difference in their sizes. Overall, the fluctuations of each cluster are quite different on a weekday basis.

The results of the multinomial logit model confirm that the spatial distribution of different types of metro stations reflects the spatial distribution of different urban functions (LCLU). The type of traffic hub and scenic spot stations (Cluster 1) are normally located near to key traffic hubs (Nanjing Railway Station and Nanjing South Railway Station) and scenic spots (e.g. Fuzimiao and Yuhuatai Scenic Areas). Stations close to new university campuses in suburban areas are allocated to Cluster 2. Due to the fact that the data used in the study concerned in weekdays, the classification of stations is mostly related to residence and employment, with 96 out of 110 stations clustered into employment-based, residential-based and mixed-use stations (Clusters 3, 4, 5, 6 and 7). Most mixed mainly employment-oriented stations are concentrated in core urban districts. The high number of job opportunities around these metro stations attracts people for work during the morning peak hours and pushes them back home in the afternoon peak hours. Thus, the morning rush at these stations normally occurs in the boarding while the afternoon rush is reflected in the alighting (Chen et al. 2009).

Employment-oriented stations are located in suburban areas with a high coverage of industrial land use. There are three employment-oriented stations gathered in Liuhe District, which is a province-level economic development zone. Most residential based stations (Clusters 5 and 6) are located in the inner and remote districts. The urbanization process in these districts occurred later than in the core districts. With the rapid urban sprawl of the last two decades, these areas were developed from small towns where urban functions are not good enough. Normally, there is a lack of commerce and business firms. Thus, these stations present the opposite ridership pattern to those of Clusters 3 and 4. The distribution of mixed-use stations is relatively dispersed, but they are more likely located in the core districts. Comparing the land use around these metro stations shows that Cluster 3 has the highest percentage of industrial land use and Cluster 4 has the highest proportion of commercial and business land use, which indicates that there is a significant and positive relationship between the percentage of commercial and business land use and the employment-based stations. While clusters 5 and 6 have a lower ratio of employment-based LCLU but a higher percentage of residential land use. It indicates that the percentage of residential land use is significantly associated with the clustering of residential-based stations, while its association with the clustering of employment-based stations is negative and significant.

Nanjing is still under rapid urban sprawl and population growth, and at the same time its rail system is fast expanded.⁵ The results of the study may have some implications for Nanjing urban design and transit planning. Many stations have a common characteristic: unbalanced usage during peak hours. For example, employment-based stations tend to have much tap-in passengers but few tap-out passengers whereas residential-based stations have

⁵ According to the official information from National Development and Reform Commission of People's Republic of China, there will be 15 metro lines under operation or building in Nanjing by 2020. http://www.sdpc.gov.cn/zcfb/zcfbtz/201505/t20150520_692548.html.

the opposite pattern. The reverse occurs in the afternoon. The implication of this finding may be important for transit operators to work out suitable train schedules to improve service efficiency. Moreover, these results may especially be useful for daily arrangements and services in metro stations, especially the connection between metro and other transportation modes. In terms of medium- and long-term planning, these results can help planners better understanding urban mobility.

Interpreting our findings, we recognize the necessity of some further work. First, the study focused on weekday metro transaction data. Further studies may combine the ridership patterns of weekdays and weekends to further understand urban mobility and the relationship between daily ridership patterns and the local environment. Second, the ridership patterns may also be influenced by the socioeconomic characteristics of the stations' influencing areas, such as age composition, household structure, income and housing floor-space etc. The influence of these variables can be examined in future research. Third, we have to acknowledge that the metro system is a restricted representation of urban mobility although it carries a large volume of passengers in a city and plays an important role in daily travel. Our approach needs to be extended certainly to other traffic modes such as bus, tram etc. to enrich the analysis. The integration of more transportation modes may be considered more accurate to understand urban mobility patterns and the link between mobility patterns and urban land use. Lastly, the conceptual framework proposed in the present paper may be applied in other cities having rail transit systems. Comparative and empirical studies across cities are encouraged in future.

Acknowledgements This research is funded by National Natural Science Foundation of China (Grant Nos. 71771049, 51378120 and 51338003), the Six Talent Peaks Project in Jiangsu Province (Grant No. 2016-JY-003) and China Scholarship Council (Grant No. 201606090149). We are grateful for the valuable suggestions from the associate editor (Professor Mark Horner) and three anonymous reviewers.

Authors' Contribution ZG Literature Search and Review, Data Analysis and Manuscript Writing. MY Content planning and Manuscript Revising. TF Manuscript Revising and Editing. HT Manuscript Revising and Editing.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Axhausen, K.W., Zimmermann, A., Schönfelder, S., Rindsfuser, G., Haupt, T.: Observing the rhythms of daily life: a six-week travel diary. *Transportation* **29**(2), 95–124 (2002)
- Batty, M., Axhausen, K.W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., Portugali, Y.: Smart cities of the future. *Eur. Phys. J. Spec. Top.* **214**(1), 481–518 (2012)
- Berkhin, P.: Survey of clustering data mining techniques. In: Kogan, J., Nicholas, C. (eds.) *Grouping Multidimensional Data*, pp. 25–71. Springer, Heidelberg (2006)
- Bertolini, L.: Nodes and places: complexities of railway station redevelopment. *Eur. Plan. Stud.* **4**(3), 331–345 (1996)
- Ceapa, I., Smith, C., Capra, L.: Avoiding the crowds: understanding tube station congestion patterns from trip data. In: *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, pp. 134–141 (2012)
- Chen, C., Chen, J., Barry, J.: Diurnal pattern of transit ridership: a case study of the New York City subway system. *J. Transp. Geogr.* **17**(3), 176–186 (2009)

- Davies, D.L., Bouldin, D.W.: A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **2**, 224–227 (1979)
- Doherty, S.T., Miller, E.J.: A computerized household activity scheduling survey. *Transportation* **27**(1), 75–97 (2000)
- Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J. Cybernet.* **3**(3), 32–57 (1973)
- El Mahrsi, M.K., Côme, E., Oukhellou, L., Verleysen, M.: Clustering smart card data for urban mobility analysis. *IEEE Trans. Intell. Transp. Syst.* **18**(3), 712–728 (2017)
- Feng, T., Timmermans, H.J.: Transportation mode recognition using GPS and accelerometer data. *Transp. Res. Part C* **37**, 118–130 (2013)
- González, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. *Nature* **453**(7196), 779–782 (2008)
- Goulet-Langlois, G., Koutsopoulos, H.N., Zhao, J.: Inferring patterns in the multi-week activity sequences of public transport users. *Transp. Res. Part C* **64**, 1–16 (2016)
- Hess, D.B., Almeida, T.M.: Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York. *Urban Stud.* **44**(5&6), 1041–1068 (2007)
- Kim, M.K., Kim, S.P., Heo, J., Sohn, H.G.: Ridership patterns at subway stations of Seoul capital area and characteristics of station influence area. *KSCE J. Civil Eng.* **3**, 964–975 (2016)
- Kuby, M., Barranda, A., Upchurch, C.: Factors influencing light-rail station boardings in the United States. *Transp. Res. Part A* **38**(3), 223–247 (2004)
- Kusakabe, T., Asakura, Y.: Behavioural data mining of transit smart card data: a data fusion approach. *Transp. Res. Part C* **46**, 179–191 (2014)
- Liu, L., Hou, A., Biderman, A., Ratti, C., Chen, J.: Understanding individual and collective mobility patterns from smart card records: a case study in Shenzhen. In: 12th International IEEE Conference Intelligent Transportation Systems. ITSC'09, St. Louis, Missouri, United State (2009)
- Ma, X., Wu, Y.J., Wang, Y., Chen, F., Liu, J.: Mining smart card data for transit riders' travel patterns. *Transp. Res. Part C* **36**, 1–12 (2013)
- Manley, E., Zhong, C., Batty, M.: Spatiotemporal variation in travel regularity through transit user profiling. *Transportation* **45**(3), 703–732 (2018)
- Ministry of Housing and Urban-Rural Development of People's Republic of China (MOHURD): Code for classification of urban land use and planning standards of development land. Ministry of Housing and Urban-Rural Development of People's Republic of China, Beijing (2011)
- Ministry of Housing and Urban-Rural Development of People's Republic of China (MOHURD): Guidelines for planning and design of urban rail. Ministry of Housing and Urban-Rural Development of People's Republic of China, Beijing (2015)
- Morency, C., Trépanier, M., Agard, B. (2006) Analysing the variability of transit users behaviour with smart card data. Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE, Toronto, Ontario, Canada (2006)
- Munizaga, M., Devillaine, F., Navarrete, C., Silva, D.: Validating travel behavior estimated from smartcard data. *Transp. Res. Part C* **44**, 70–79 (2014)
- Reades, J., Zhong, C., Manley, E.D., Milton, R., Batty, M.: Finding pearls in London's oysters. *Built Environ.* **42**(3), 365–381 (2016)
- Roth, C., Kang, S.M., Batty, M., Barthélemy, M.: Structure of urban movements: polycentric activity and entangled hierarchical flows. *PLoS ONE* **6**(1), e15923 (2011)
- Tan, Z., Li, S., Li, X., Liu, X., Chen, Y.: Clustering of metro stations in Guangzhou based on passenger flow. *Trop. Geogr.* **37**(1), 102–111 (2017). (in Chinese)
- Tupper, L. L., Matteson, D. S., Handley, J. C.: Mixed data and classification of transit stops. arXiv preprint: [arXiv:1611.04026](https://arxiv.org/abs/1611.04026) (2016)
- Xie, X.L., Beni, G.: A validity measure for fuzzy clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **13**(8), 841–847 (1991)
- Xu, Y., Shaw, S.L., Zhao, Z., Yin, L., Fang, Z., Li, Q.: Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach. *Transportation* **42**(4), 625–646 (2015)
- Yin, Q., Meng, B., Zhang, L.: Classification of subway stations in Beijing based on passenger flow characteristics. *Prog. Geogr.* **35**(1), 126–134 (2016). (in Chinese)
- Yang, X., Zhao, Z., Lu, S.: Exploring spatial-temporal patterns of urban human mobility hotspots. *Sustainability* **8**(7), 674 (2016)
- Zhao, J., Deng, W., Song, Y., Zhu, Y.: What influences Metro station ridership in China? Insights from Nanjing. *Cities* **35**, 114–124 (2013)

Zhong, C., Manley, E., Arisona, S.M., Batty, M., Schmitt, G.: Measuring variability of mobility patterns from multiday smart-card data. *J. Comput. Sci.* **9**, 125–130 (2015)

Zuoxian Gan is a PhD candidate in the School of Transportation, Southeast University. He has been awarded a scholarship under the China Scholarship Council to pursue his study in the Urban Planning Group, Department of Urban Science and System, Eindhoven University of Technology as a joint PhD candidate. His research interests are urban rail transit, travel behavior analysis and transport geography.

Min Yang is Professor of the School of Transportation, Southeast University. He received his PhD degree in transportation engineering from the same university in 2007. His areas of interest include travel demand analysis and behavior modeling, comprehensive urban transportation system planning and integrated traffic hub planning.

Tao Feng is Assistant Professor of the Urban Planning Group of the Eindhoven University of Technology. His areas of interest include travel demand forecasting, travel behavior analysis, big data, housing and mobility decision and land use and transportation. He is the associated editor of journal *Asia Transport Studies*.

Harry Timmermans is Professor of the Urban Planning Group of the Eindhoven University of Technology. He has research interests in modeling decision-making processes and decision support systems in a variety of application domains, including transportation. He is editor of the *Journal of Retailing and Consumer Services*, and serves on the board of several other journals in transportation, geography, urban planning, marketing, artificial intelligence and other disciplines. He is member of the Board of the International Association of Travel Behavior Research (IATBR), and of several scientific committees of the Transportation Research Board.