Identifying urban functional areas and their dynamic changes in Beijing: Using multiyear transit smart card data

Wang, Z., Liu, H., Zhu, Y., Zhang, Y., Basiri, A., Büttner, B., Gao, X. and Cao, M.

This is an accepted manuscript of an article to be published in the Journal of Urban Planning and Development, 147 (2).

The final definitive version is available online:

https://doi.org/10.1061/(ASCE)UP.1943-5444.0000662

UP2883

Identifying Urban Functional Areas and Their Dynamic Changes in Beijing: Using Multiyear Transit Smart Card Data

Authors:

Zijia Wang

Associate Professor, Department of Highway and Railway Engineering, School of Civil and Architectual Engineering, Beijing Jiaotong Univerity, No. 3 Shangyuan Village, Haidian District, Beijing, 100089, PR China. Email: zjwang@bjtu.edu.cn

Haixu Liu

Engineer, Transportation Research Centre, Beijing Urban Construction Design and Development Group Co., Limited, No. 5, Fuchengmen Beidajie, Xicheng District, Beijing, 100032, PR China. Email: liuhaixu@bjucd.com

Yadi Zhu

Research Associate, Department of Highway and Railway Engineering, School of Civil and Architectual Engineering, Beijing Jiaotong Univerity, No. 3 Shangyuan Village, Haidian District, Beijing, 100089, PR China. Email: yadizhu@bjtu.edu.cn

Yuerong Zhang

PhD Candidate, Bartlett School of Planning, University College London, Central House, 14 Upper Woburn Place, London, WC1H 0NN, United Kingdom. Email: yuerong.zhang.14@ucl.ac.uk

Teaching Assistant, Bartlett Centre for Advanced Spatial Analysis, University College London, 90 Tottenham Court Road, London, W1T 4TJ, United Kingdom. Email: yuerong.zhang.14@ucl.ac.uk

Anahid Basiri

Professor, School of Geographical and Earth Sciences, University of Glasgow, Glasgow, G12 8QQ, United Kingdom. Email: ana.basiri@glasgow.ac.uk

Benjamin Büttner

Head of Research Group Accessibility Planning, Department of Civil, Geo and Environmental Engineering, Technical University of Munich, Arcisstr. 21, Munich, 80333, Germany. Email: benjamin.buettner@tum.de

Xing Gao

PhD Candidate, Bartlett School of Planning, University College London, Central House, 14 Upper Woburn Place, London, WC1H 0NN, United Kingdom. Email: xing.gao@ucl.ac.uk

Mengqiu Cao

Senior Lecturer, School of Architecture and Cities, University of Westminster, 35 Marylebone Road, London, NW1 5LS, United Kingdom. Email: m.cao@westminster.ac.uk (corresponding author)

Abstract

A growing number of megacities have been experiencing changes to their landscape due to rapid urbanisation trajectories and travel behaviour dynamics. Therefore, it is of great significance to investigate the distribution and evolution of a city's urban functional areas over different periods of time. Although the smart card automated fare collection system (SCAFC) is already widely used, few studies have used smart card data to infer information about changes in urban functional areas, particularly in developing countries. Thus, this research aims to delineate the dynamic changes that have occurred in urban functional areas based on passengers' travel patterns, using Beijing as a case study. We established a Bayesian framework and applied a Gaussian mixture model (GMM) derived from transit smart card data in order to gain insight into passengers' travel patterns at station level and then identify the dynamic changes in their corresponding urban functional areas. Our results show that Beijing can be clustered into five different functional areas based on the analysis of corresponding transit station functions, namely: multimodal interchange hub and leisure area; residential area; employment area; mixed but mainly residential area; and a mixed residential and employment area. In addition, we found that urban functional areas have experienced slight changes between 2014 and 2017. The findings can be used to inform urban planning strategies designed to tackle urban spatial structure issues, as well as guiding future policy evaluation of urban landscape pattern use.

Keywords

Urban functional areas; Dynamic changes; Urban planning; Travel pattern; Smart card data; Beijing

1. Introduction

Urbanisation leads to rapid growth on a city scale, and a large number of people tend to move to the city seeking a better working and living environment. Urban immigration causes the socio-economic attributes of different regions in a city to change dramatically, and it is therefore necessary for city planners, economists and resource managers to comprehensively understand the distribution of, and changes in, urban functional areas (Pham et al., 2011). However, some traditional urban structure detection methods, such as remote sensing images (Heiden et al., 2012; Van de Voorde, Jacquet, and Canters, 2011), primarily concentrate on the changes in urban physical structure, but these cannot accurately reflect the socio-economic composition of urban areas revealed by urban mobility patterns (Chen et al., 2017). In addition, functional changes in a city happen relatively slowly. Therefore, only examining data for a single year may not precisely reflect the dynamic changes in a city's urban functional areas. Furthermore, the systematic collection of long-term data would require a massive investment of manpower, time and material resources, which would be a significant constraint on conducting the relevant research. With the rapid development of big data, it has increasingly been applied in different fields of urban studies. These studies involve, for example, the use of mobile data (Sagl et al., 2014), social network data (Hasnat et al., 2018), and smart card data (Zhao et al., 2018), and have been validated in multiple cities. To take the smart card data as an example, it consists of a large amount of spatio-temporal information on users' long-term activity, which makes it possible to study cities at the individual level, while the huge volume of data also increases the accuracy of the research. At the same time, these data are by-products of residents' activities, which have low acquisition costs but consist of long-term information. Therefore, methods based on big data can be seen as an effective way to measure the dynamic changes in urban functional areas.

************************Please insert Figure 1 here*****************************

Beijing has a geographical area of 16,808 square kilometres. The total number of usual residents living in Beijing was 21.54 million in 2018. Transport emissions and traffic jams are currently two primary issues in the city (Wang et al., 2015; Cao et al., 2017). In order to alleviate traffic congestion caused mainly by rapid urbanisation and an increase in private car usage, the urban transit system has been dramatically developed to tackle the resulting issues (Jiang et al., 2017). The Beijing transit system comprised 22 lines and 278 stations (all transfer stations are only counted once) by the end of 2017 (Fig. 1) (Beijing Transport Institute, 2018). The total mileage of Beijing transit is predicted to reach 1,000 kilometres, and the annual ridership to reach 4.53 billion, by the end of 2020, according to Beijing's Urban Master Plan (2016-2030) edited by Beijing's Municipal Commission of Planning and Natural Resources. Along with the development of the transit system, use of the smart card automated fare collection system (SCAFC) has become widespread, enabling a large amount of smart card data to be collected. In Beijing, smart cards can be used for different transport modes, such as buses and the metro, although this study primarily focuses on the data relating to travel by metro. The average amount of daily SCAFC data generated exceeds 5 million, consisting of data on more than 2.8 million passengers, which includes trips that started by bus, but involved transferring to the metro. The metro has become one of the most important sustainable transport modes for urban residents, while the large amount of SCAFC data generated from it has revealed urban mobility patterns particularly well (Pelletier et al., 2011; Wang et al., 2018).The aim of this paper is to delineate the dynamic changes that have occurred in urban functional areas, based on passengers' travel patterns, using Beijing as a case study. As urban functional development is a relatively slow process, in order to study the dynamic changes in urban functional areas, this paper also identifies the socio-economic attributes of urban areas for different periods of time by using multi-year smart card data and analyses the evolution of urban functional areas between 2014 and 2017. The paper is organised as follows: the relevant literature is reviewed in section 2; section 3 describes the methods used; section 4 and section 5 present the modelling results and a discussion about passenger travel patterns and the resulting inferences for the corresponding urban functional areas; and the last section draws conclusions.

2. Literature review

The application of smart card data in analysing travel behaviours does not have a long history, largely because the new data sources like smart card data have only recently been available. The large volume of individual level data provides us with a new lens through which to examine the dynamics of human movement (Zhong et al., 2014), and thus a more comprehensive view of urban dynamics. Taking advantage of the disaggregated spatio-temporal information (Gan et al., 2018), studies using smart card data have been divided into various sub-types, such as travel behaviours (Zhao el al., 2017; Kieu et al., 2015), urban structure (Zhong et al., 2014), station hierarchies (Roth et al., 2011; Zhang et al., 2019) and local environment inferences (Chen et al., 2009). However, the ideas underlying these applications are the same, that is to use human movement as a sensor with which to disclose intangible urban patterns.

The fundamental aim of studies that use smart card data is to reveal passengers' travel patterns, including their origin-destinations, journey length, travel frequency etc. Because different travel purposes exhibit various travel patterns, the purpose of trips can be inferred and detected (Zou et al., 2018) by differentiating the regularity and variability of spatiotemporal characteristics. The most intuitive case is that trips relating to work and education usually take place during peak times, while entertainment trips are made during off-peak times (Lee and Hickman, 2014). For example, Alsger and colleagues (2018) proposed the logical inference framework with which to infer the purposes of trips on public transport and deduced five different trip purposes (work, home, education, shopping and recreational) in Brisbane, Queensland. Furthermore, classifying passengers into different clusters derived from their travel patterns can infer their socio-economic attributes (Goulet-Langlois., 2016; Zhu et al., 2018), and enable analysis of potential factors which may affect passengers' travel elasticity (e.g. avoid travelling at peak times) (Halvorsen et al., 2016; Huang et al., 2019).

In addition, to some extent, knowledge about the association between transit passengers' travel patterns and their travel purpose can be extended to reveal the dynamics of the surrounding urban functional areas based on the corresponding transit stations (Alsger et al., 2018). More specifically, the frequencies with which passengers visit transit stations can be used to infer which areas they live or work in (Hasan, 2013). Furthermore, information about regional clustering of job-housing distribution around transit stations can be obtained by analysing high-frequency passengers' individual job-housing distribution (Ma, 2017; Huang et al., 2018). Moreover, transit stations located in a transport hub (i.e. multimodal interchange hub) or entertainment areas are more likely to attract low-frequency passengers, and the regularity of passengers' travel patterns for this type of transit station is weaker compared to commuters' travel patterns.

Station ridership patterns means the time series of ridership entry to and exit from the station.

The regularity of a ridership pattern often changes over time (Zhong et al., 2016; Li et al., 2017). Some studies show that the built environment around transit stations is statistically significantly associated with station ridership patterns (Ma et al., 2018; Taylor et al., 2009; Thompson and Brown, 2006). Similar results have also been found in the case studies of Shenzhen (Gong, 2017), Nanjing (Gan et al., 2020), and Sydney (Blainey, 2013). In the case of Beijing, Zhu et al. (2019) also pointed out that there is a significant relation between station ridership patterns and the built environment during peak times. Meanwhile, Zhong et al. (2014) investigated passenger volume at station entrances and exits to infer the dynamics of the urban functional areas around the corresponding transit stations. Similar results were also obtained by Long and Thill (2015) using combined smart card and household travel survey data to provide a new approach to identifying the dynamics operating in urban functional areas, particularly with regard to jobs-housing relationships in Beijing.

In summary, we can see that smart card data can be used to help analyse travel patterns at both disaggregated and aggregated levels. Passengers' travel patterns can also further reflect the dynamics of urban functional areas, particularly around transit stations. That is to say, the built environment around the transit station shows an association with its ridership pattern; inferences about the urban functional areas can be made by analysing ridership patterns for the corresponding transit stations. Previous empirical studies (e.g., Ma et al., 2017; Alsger et al., 2018; Gan et al., 2020) have shown the validity of these deductive results.

However, most existing literature has two limitations. First, it only considers either an analysis of individual travel behaviour pattern or a station-oriented clustering analysis of ridership patterns when attempting to detect characteristics of stations. Second, most existing literature has focused more on high-frequency passengers. Less attention has been paid to low-frequency passengers, mainly due to a lack of sufficient spatio-temporal information, which may reduce the extent to which it can accurately reflect the dynamics of urban functional areas. Therefore, to bridge these gaps, this paper also contributes to the existing theories in two ways. Firstly, we include both individual travel patterns and station ridership patterns in the analysis, in order to provide planners and policymakers with a more finely-grained picture of station functional areas and their dynamic changes. Secondly, we consider both low-frequency and high-frequency passengers' travel patterns. The particular significance of considering different types of travel patterns is that it improves the accuracy of identifying the dynamics operating in urban functional areas.

3. Methods

3.1. Spatio-temporal travel probability

Each passenger's long-term travel data reflects his/her travel pattern, which is derived from the frequency of the passenger's visits to different transit stations (Hasan, 2013). However, the aforementioned type of research has not taken different time periods into consideration. Building on the aforementioned basic approach, this paper takes into account visiting frequencies during different periods of time for different transit stations, and calculates travel probability under different spatio-temporal circumstances, following Bayesian theory (Zhong et al., 2014; Alsger et al., 2018).

More detailed processes are described below:

(1) Record the long-term travel database of each passenger identified by different smart card numbers based on SCAFC data, which contains all the travel records of the passenger during 5

working days from 2014 to 2017, respectively.

(2) Calculate the number of days on which they used the metro, and the frequencies of entry and exit for different transit stations during different periods for each passenger.

(3) Use the aforementioned statistical data to calculate the probability of visiting frequencies of the station for each passenger during a given period of time.

Taking the calculation of the probability of a passenger entering the station *S* during the time period *T*, given as $P(Entry|S,T)$, as an example, first let:

$$P(Metro|T) = Day_{metro} / Day_{all} \tag{1}$$

Equation (1) shows the probability of a passenger using the metro during the time period *T*.

Where

Day_{all} indicates the number of days of SCAFC data.

$\mathrm{D}ay_{metro}$ is the number of days the passenger used the metro to travel during the time period *T*.

We then select the passenger's travel record for using the metro during the time period *T* to calculate the entry frequency R_o from the station *S*.

$$P(Entry|S,T,Metro) = R_o / R_{all} \tag{2}$$

Equation (2) shows the probability of a passenger entering the station *S* during the time period *T*.

Where

R_o indicates the entry frequency for the station *S*.

R_{all} is the total amount of entry frequencies for all stations.

$$\begin{aligned} &P(Entry|S,T) \\ &= P(Entry|S,T,Metro) \\ &= P(Metro|T) \times P(Entry|S,T,Metro) \end{aligned} \tag{3}$$

Therefore, the probability of a passenger entering the station *S* during the time period *T* can be obtained as shown in equation (3).

Likewise, the probability of a passenger exiting a station during a given time period *T'* can also be calculated following the same steps.

3.2. Gaussian mixture model (GMM)

In recent years, mixture models have been widely applied in the field of SCAFC data mining (Briand et al., 2017; Mohamed et al., 2017). Unlike the traditional clustering method, for instance, based on Euclidean distance, mixture models assume that different indicators follow a specified distribution, and complete the clustering process by analysing multiple mixed distributions. In this paper, we use the Gaussian mixture model (GMM) to complete the cluster process (Reynolds et al., 2000; Zivkovic., 2004).

The underlying principle of the GMM is to fit the data with multiple Gaussian distributions which is shown as follows:

$$X_i \left| Z_{ik} = 1 \sim N(\mu_k, \sigma_k) \right. \tag{4}$$

In formula (4), $Z_{ik} = 1$ means the sample *i* belongs to the cluster *k*, then the sample *i* follows the corresponding Gaussian distribution with the parameter μ_k and σ_k.

When a sample obeys the Mixture Gaussian Distribution, it can be represented by several Gaussian distributions with different parameters, each of which is called component *i* (i=1,2,…, *k*) and is denoted by $N(\mu_k, \sigma_k)$.

We use π_k to represent the probability that sample *i* belongs to component *k*, which means that the sample obeys the Gaussian distribution with the parameter μ_k and σ_k. If we take the sum of all the components $N(\mu_k, \sigma_k)$ and multiply by the probability π_k, we can obtain the probability of sample X_i which is shown in equation (5):

$$X_i \sim \sum_k \pi_k N(\mu_k, \sigma_k) \tag{5}$$

If we multiply the probability of samples i $(i=1,2,\ldots,I)$, where I indicates the total number of samples, we can obtain the likelihood functions $L(X)$ of the total samples as shown in equation (6):

$$L(X) = \prod_I \sum_k \pi_k N(\mu_k, \sigma_k) \tag{6}$$

When the likelihood functions achieve the maximum value, this enables us to obtain the cluster result and the centre of each cluster. The expectation-maximisation algorithm (EM) is used to analyse the model, and the Davies-Bouldin Index (DBI) and Silhouette Coefficient (SC) are used to decide on the number of clusters (Davies & Bouldin, 1979; Rousseeuw, 1987).

4. Data description and parameter selection

4.1. Data description

The dataset in this paper is comprised of Beijing rail transit AFC data from 2014 to 2017, for the same week of each year, and contains more than 0.1 billion travel records and more than 10 million different card holders. The data is divided into five categories, namely: smart card ID (Grant_Card_Code); trip start time (Entry_Time); trip end time (Deal_Time), trip start station (Entry_Station) and trip end station (Exit_Station). As shown in Table 2, the AFC data contains the spatio-temporal information about rail transit passengers.

************************Please insert Table 1 here****************************

4.2. Time period selection

The ridership pattern is roughly the same for the different working days in each of the four years when the passenger flow is measured at 30 minute intervals. As shown in Figure 2, there is a peak in ridership both in the morning and in the evening, while the ridership between the morning and evening peaks remains stable. Therefore, we chose 6:00 to 10:00 for the morning peak period, 10:00 to 16:00 for the off-peak period, and 16:00 to 20:00 for the evening peak period, which correspond to the red, green and blue areas in Figure 2.

************************Please insert Figure 2 here*****************************

4.3. Travel probability division

The travel probability calculated by the method described in section 3.1 is continuous, and it is therefore difficult to obtain a full and accurate understanding of passengers' travel patterns from it. Therefore, the travel probability is divided into three levels, based on two assumptions:

Assumption 1: Most passengers travel by rail transit in the morning and evening periods only once.

Assumption 2: Most passengers have only one Origin-Destination (OD) in the morning and evening periods.

To verify the two assumptions, we calculate the proportion of passengers with different travel times during different periods and the proportion of passengers who visited different stations at different times during each year, and we then calculate and use the average value.

************************Please insert Figure 3 here*****************************

As shown in Figure 3, more than 90% of passengers travelled only once in the morning and evening periods, and more than 75% of passengers used only one entry station and one exit station, indicating that most of the passengers have a stable OD in the morning and evening periods; therefore, the aforementioned two assumptions have been verified. For most of the passengers, the travel probability only relates to the number of travel days based on the two assumptions. Therefore, this paper takes typical passengers who travelled only once and had a stable OD in the morning and evening periods as normal, to determine the passenger travel probability.

In this paper, travel probability is defined as either a low probability (0, 0.4], a mid probability (0.4, 0.7], or a high probability (0.7, 1]. For typical passengers, low probability (0,0.4] means that they travel by rail transit no more than two days a week during that period. This type of travel is mostly for shopping or entertainment (*Goulet-Langlois., 2016*). Mid probability (0.4, 0.7] indicates that the passenger travels on three days a week, and high probability (0.7, 1] indicates that the passenger travels on at least four days a week, most of whom are commuters (*Huang et al., 2018*).

4.4. Passengers' travel patterns

Figure 4 shows the number of passengers with different travel probabilities during different time periods from 2014 to 2017. As can be seen from the figure, there are a large number of low probability passengers travelling during different time periods. These passengers travelled in a more random way and did not exhibit stable travel patterns. However, the ridership pattern measured at 30 minute intervals is relatively stable, as shown in Figure 2, which indicates that although the travel mode choice at the individual level was irregular, the ridership pattern within the network as a whole remained regular.

************************Please insert Figure 4 here*****************************

The number of high probability passengers in the morning period is the largest among the three types of travel probability, indicating that rail transit ridership during the morning period is regular, while the number of low probability passengers also indicates that rail transit provides an important alternative method of travel. During the evening period, the number of low probability passengers is largest, while the number of high probability passengers is lower than during the morning period, which indicates that the regularity of ridership in the evening period is weaker than that in the morning period, suggesting that passengers were more likely to use other modes of travel during the evening period. Low probability passengers form the majority during the off-peak period, which indicates that most passengers only occasionally travel by metro during that time, unlike during the morning and evening period when most passengers are regulars.

5. Urban functional area detection

5.1. Feature selection

Information about the characteristics of a transit station can be obtained from both passenger travel patterns and the station ridership pattern. Therefore, this paper uses two types of indicators to identify the characteristics of a station. Table 2 shows how the station characteristics were selected and identified.

************************Please insert Table 2 here*****************************

Following Geng and Yang (2017), *Entry* and *Exit* represent the total number of passengers entering and exiting the station in three different time periods. The *Interval* covers morning, off-peak and evening time periods. The entering station flow entropy value, given as $Entropy_{Entry}$, and exiting station flow entropy value, given as $Entropy_{Exit}$, are calculated as shown below:

$$Entropy_{Entry} = -\sum_{\text{interval}} (Entry_{\text{interval}} / Entry) * \log_3 (Entry_{\text{interval}} / Entry) \tag{7}$$

$$Entropy_{Exit} = -\sum_{\text{interval}} (Exit_{\text{interval}} / Exit) * \log_3 (Exit_{\text{interval}} / Exit) \tag{8}$$

For the passenger travel pattern indicators, the proportion of high probability passengers and low probability passengers reflects the regularity of passengers visiting each of the stations. The higher the proportion of high probability passengers is, the stronger the ridership regularity of the station. This indicates that the station is more likely to be used for commuting purposes. Conversely, the higher the proportion of low probability passengers is, the weaker the ridership regularity of the station. This infers that the station is more likely to be used for a transport hub (i.e. multimodal interchange hub) and/or an entertainment purpose.

For the station ridership pattern indicators, the proportion of passengers who enter the station either in the morning or evening periods gives an indication of the characteristics of that station. The higher the proportion of passengers entering a station in the morning and evening periods is, the higher the likelihood that the station serves residential passengers, meaning that the station is located in a residential area. However, the station could serve working passengers, which means that it is more likely to be located in an employment area.

The entropy value for entering or exiting the station reflects the distribution of all-day ridership. The smaller the entropy value is, the more likely it is that the station will have an unbalanced distribution of all-day ridership. This indicates that there would be a peak time for ridership each day. In contrast, the larger the entropy value is, the more likely the station is to have a balanced distribution of all-day ridership, meaning that there is no obvious peak time for ridership each day.

5.2. Cluster analysis

We calculated statistics for 11 features of each station for each year and input them into the model. The meanings of the features, denoted as F1 to F11, can be found in Table 2. Because the same station may belong to a different cluster during different years, in order to compare the data, each station for each year is treated as the sample unit in this paper.

As mentioned in Section 3.2, the Davies-Bouldin Index (DBI) and Silhouette Coefficient (SC) were used to decide on the number of clusters and evaluate the cluster performance of the GMM model (Davies & Bouldin, 1979; Rousseeuw, 1987). The smaller the DBI and the greater the SC, the greater the clustering result.

************************Please insert Figure 5 here*****************************

As shown in Figure 5, when the number of clusters is 5, the DBI of the GMM has the smallest value, while the SC of the GMM has the greater value.Therefore, we classified the stations into 5 clusters based on the GMM model. The cluster centres of travel and ridership pattern indicators are shown in Figures 6 and 7, respectively.

************************Please insert Figure 6 here*****************************

************************Please insert Figure 7 here*****************************

Cluster 1：Multimodal interchange hubs and leisure cluster. Cluster 1 is shown by a yellow bar in Figure 6 and Figure 7. In Figure 6, F1 and F2 represent the proportion of low probability passengers in the evening and morning periods, and the F1 and F2 values of Cluster 1 ranked the highest among the five clusters, which indicates that these types of stations have the highest proportion of low probability passengers and the lowest proportion of high probability passengers in the morning and evening period out of the five clusters. F6 denotes the proportion of low probability passengers out of the total passengers within a day, and the value of this cluster is approximately 0.8, which means 80 per cent of the passengers are classified as low probability passengers throughout the day and visit these station irregularly. In Figure 7, F10 and F11 represent the entropy value for entering and exiting a station, both the entry and exit entropy values of stations in Cluster 1 are high, and the exiting station entropy of this cluster is the highest out of the five clusters, which indicates that the distribution of ridership is balanced throughout the day and there is no obvious peak period. Cluster 1 stations include Beijing south railway station (Fig.8 (A)), Beijing west railway station (Fig.8 (B)), Tiananmen east station and Tiananmen west station (Fig.8 (D)), which are typical traffic hubs and scenic areas where tourist attractions are located. Therefore, the stations in Cluster 1 are characterised as multimodal interchange hubs and leisure clusters, and the areas where these stations are located comprise traffic hubs and/or entertainment areas of the city.

Cluster 2：Residential cluster. This cluster is shown as a blue bar in Figure 6 and Figure 7. In Figure 6, F1 and F2 represent the proportion of low probability passengers in the evening and morning period, while F3, F4 and F5 represent the proportion of high probability passengers in the evening period, morning period and throughout the day. The F1 and F2 values of Cluster 2 are low, indicating that these types of stations have a lower proportion of low probability passengers in the morning and evening periods, while the F3 and F4 values of this cluster are high, indicating that these types of stations have a higher proportion of high probability passengers in the morning and evening periods. The F5 value of this cluster is the highest among the five clusters, which means that these types of stations have the highest proportion of high probability passengers in the whole-day period. All of the five features show that passengers who visit these stations follow a regular travel pattern. In Figure 7, F8 and F9 indicate the proportion of passengers entering a station out of the total passengers during evening and morning peak times. The F8 value of Cluster 2 is the lowest, while the F9 value of Cluster 2 is the highest among the five clusters. This means the station ridership pattern of these kinds of stations is dominated by entry-station passengers in the morning, and by exit-station passengers in the evening. Moreover, the passenger flow in and out of these stations varies greatly during the two periods. F10 and F11 represent the entropy values for entering and exiting a station. Stations in this cluster have the lowest F10 and F11 values, indicating that the ridership is concentrated throughout the day. The Cluster 2 stations include Tiantongyuan station, Huilongguan station, and Pingguoyuan station, which are located in typical residential areas. Therefore, the key characteristic of stations in Cluster 2 is that they are residential, and stations in this cluster are located in urban residential areas.

Cluster 3：Employment cluster. This cluster is shown as a light blue bar in Figure 6 and Figure 7. In Figure 6, all seven features of Cluster 3 are approximately equal to those of Cluster 2, which means that passengers visiting stations in Cluster 3 exhibited a regular travel pattern, like those who visited stations in Cluster 2. In Figure 7, the ridership pattern for Cluster 3 stations contrasts with that of Cluster 2 stations, with the former having the highest F8 and the lowest F 9 values, indicating that the ridership patterns of these stations are comprised mainly of exit-station passengers in the morning and entry-station passengers in the evening, while the passenger flow in and out of the stations varies greatly. Both entry-station and exit-station entropy values are greater only than those of Cluster 2. Cluster 3 stations include Zhongguancun station (Fig.8 (E)), and Guomao station (Fig.8 (G)), which are located in typical employment areas. Thus, stations in Cluster 3 are characterised as employment clusters and stations in this cluster are located in urban employment areas.

Cluster 4：Mixed but mainly residential cluster. This cluster is shown by an orange bar in Figure 6 and Figure 7. The proportion of high probability passengers using such stations, which is indicated by F3, F4 and F5, is lower than for stations in Cluster 2 and Cluster 3; however, compared to Cluster 1, Cluster 4 has lower F1, F2, and F6 values and higher F3, F4, and F5 values, which means these stations have more high probability passengers and fewer low probability passengers. To an extent, passengers who visited such stations display a regular travel pattern. However, compared to passengers at stations near employment or residential areas, they have more choice of travel modes, apart from rail transit. From the perspective of station ridership patterns, that of stations in Cluster 4 is similar to Cluster 3, which is characterised as residential. However, the entropy values are at a middling level, suggesting that the ridership concentration distribution was not significant throughout the day. Therefore, the key characteristic of these stations is residential-oriented and stations in this cluster are located in urban mixed but mainly residential areas.

Cluster 5：Mixed employment and residential cluster. This cluster is shown by a gray bar in Figure 6 and Figure 7. In Figure 6, all seven features of Cluster 5 are approximately equal to those of Cluster 4, indicating that the passenger types served by these kinds of stations are similar to those of Cluster 4. In Figure 7, the F8 and F9 values are around 0.5, which means the number of passengers entering and exiting these types of stations is roughly the same during the peak period. At the same time, in Figure 7, the F10 and F11 values are the highest among the five clusters, indicating that the entropy of passengers is large and the passenger flow distribution is relatively average throughout the day. Stations in this cluster serve both working and residential passengers. Therefore, these kinds of stations are classified as mixed residential and employment stations, and hence they are located in mixed employment and residential areas.

************************Please insert Figure 8 here*****************************

5.3. Spatial distribution

The characteristics of stations reflect the function of the city around the station (Gan et al., 2018; Zhao et al., 2018; Zhu et al., 2018). Figure 8 shows the spatial distribution of stations in different clusters. The results enable us to gain greater insight into the evolution of urban functional areas in Beijing between 2014 and 2017.

From 2014 to 2017, the city had a clear circular structure and this has not changed significantly. The core area of the city (also the centre of the rail transit network) is the most scenic area, containing world-famous landmarks such as Tiananmen Square. It also includes transportation hubs such as Beijing West Railway Station and Beijing South Railway Station. There are two typical urban employment areas located in the area between the core area and the fourth ring road (Fig.8): Zhongguancun Technology Park (Fig.8 (E)) and Guomao (central business area of Beijing), Fig.8 (G)). The remaining areas are mainly mixed employment and residential areas adjacent to the two typical employment areas. It is worth noting that mixed employment and residential areas are mainly distributed in the north of Beijing, while the south is mainly residential. The outer ring of the city's fourth ring road is comprised mainly of residential areas, while another typical employment area, called Wangjing (Fig.8(F)), is located in the northeast. There is also an isolated mixed employment and residential area surrounded by residential areas in the southwest, known as Fengtai Technology Park (Fig.8 (H)). Beijing Economic-Technological Development Area (Fig.8 (I)), which is made up of an employment area and two surrounding mixed employment and residential areas, is located in the southeast. These two areas are important employment areas in the south of the city; however, they have not been identified as typical employment areas, like Zhongguancun Technology Park (Fig.8 (E)) and Guomao (Fig.8 (G)), for many years.

According to the spatial distribution of various urban functional areas in Beijing over the years studied, we found that threre is a significant imbalance between jobs and housing in Beijing in general. More jobs are concentrated in the urban central areas, while only a small proportion of jobs are distributed in the outer part of the city. The outer part of the city contains more residential areas. Therefore, this may also lead to long distance commuting and traffic congestion (Zhao and Hu, 2019), and cause air pollution, particularly for people who travel by private vehicles (Cao et al., 2017). To some extent, these results also reflect the combined issue of car dependence and housing affordability (Cao and Hickman, 2018; Dewita et al., 2018, 2020), as well as inferring potential issues associated with transport-related social inequity (Cao, 2019; Cao and Hickman, 2019, 2020; Zhao and Cao, 2020; Zhang et al., 2018). On the other hand, the expansion of jobs from the typical employment area to the surrounding area has relieved traffic congestion in the city. In the near future, it will be necessary to continue to create and extend job opportunities to the outer areas, at least in Beijing. Mixed employment and residential cluster areas, in which mixed employment and residential cluster stations are located, are important in terms of creating more jobs, because these areas already have a relatively good supply of jobs close to residential areas. Thus, encouraging the expansion of jobs within the outer part of the city is an effective way to reduce urban traffic congestion, as well as reducing transport-related social inequity, particularly for the low-income migrants (Zhao and Cao, 2020).

In terms of the residential areas, it is necessary to constantly improve the surrounding services and facilities, such as shopping malls, hospitals, and schools, etc., as this can effectively enhance the living standards of local residents, and can also generate a large number of job opportunities, which can be filled by local residents in order to reduce the travel distance between their workplace and home, and thus in turn reduce traffic congestion.

With regards to transport interchange hubs and tourism business areas, the management of floating populations should be improved. More services and facilities need to be provided in these areas, such as information centres, restaurants, and hotels, etc.

5.4. Evolution process

************************Please insert Figure 9 here*****************************

The evolution of each area’s urban function is shown in Figure 9. For example, the areas that were residential areas in 2014 were mainly still residential in 2015, while a few areas had transformed into mixed but mainly residential areas. The general trend of evolution is that the urban functional areas are in accordance with the order of their spatial distribution. As shown in Figure 9, residential areas (Cluster 2) can only transform into mixed but mainly residential areas (Cluster 4) in four years, and only the mixed but mainly residential areas (Cluster 4) can transform into residential areas (Cluster 2). Mixed employment and residential areas (Cluster 5) are more complicated. On the one hand, they can transform into employment areas (Cluster 3) or mixed but mainly residential areas (Cluster 4). On the other hand, the aforementioned two areas can transform into mixed employment and residential areas.

The aforementioned phenomenon indicates that the evolution of urban functional areas has to follow a process, and this process is longest in relation to the transition from a residential area to an employment area. Therefore, it is difficult to transform a residential area into an employment area in a short time, but mixed employment and residential areas often have a good foundation, making it easier to change the urban function of these areas. Currently, the development of the southern part and the northern part of Beijing is unbalanced. A large number of employment areas are concentrated in the north, while the southern part of the city is comprised mainly of residential areas. In order to achieve a better balance between the north and the south, the development of the southern part of the city should focus on the Fengtai Technology Park and Beijing Economic-Technological Development Area according to the general law of evolution. These two areas both have mixed employment and residential areas, and the Beijing Economic-Technological Development Area already has an employment area. The aim should be to improve transportation, policy, and other factors in theses two areas, so that they will attract more jobs, and effectively change the function of the southern part of the city.

6. Conclusions

This paper identified the characteristics of stations based on SCAFC data, and then detected the spatial distribution of different urban functional areas. Using multi-year data enabled us to arrive at the general law of urban functional areas spatial distribution and dynamics. Advice was given on the further development of Beijing's urban areas.

This research makes a fivefold contribution. First, smart card data have long been used to analyse passenger capacity, and visualise and predict travel behaviour, such as the origin and destination (OD) trajectories. This study extended the aforementioned research to infer urban functional areas based on passengers' travel patterns and ridership patterns at metro stations. Second, different types of unsupervised machine learning approaches/clustering approaches have been employed to assist in finding and increasing the accuracy of the number of clusters. Third, most of the existing research only considers high-frequency passengers, and pays little attention to low-frequency passengers (Ma, 2017; Huang et al., 2018). This paper applied a method for calculating the spatio-temporal travel probability by following Bayesian theory, which measured the travel patterns of low-frequency passengers and high-frequency passengers according to the same rule. Fourth, in this paper, 11 features were selected: features 1 – 7 reflect the travel patterns of passengers who visited the station based on spatio-temporal travel probability; while features 8 – 11 reflect the station ridership patterns. The GMM cluster method was used to identify the characteristics of the station based on the 11 features so that both individual travel patterns and station ridership patterns could be considered. Finally, we identified the function of the urban areas based on the station cluster results. Using multi-year SCAFC data allowed us not only to determine the function of the urban areas across the spatial distribution of each year, but also to chart the evolution process. Through undertaking cluster analysis using the features of individual travel patterns and station ridership patterns, we found that Beijing's functional areas can be divided into five categories, namely: multimodal interchange hub and leisure area; residential area; employment area; mixed but mainly residential area; and a mixed residential and employment area. Residential or mixed but mainly residential areas served by transit stations were primarily distributed in outer Beijing between the fourth ring road and the sixth ring road, whereas mixed residential and employment areas were located in inner Beijing. Meanwhile, urban functional areas experienced slight changes between 2014 and 2017.

The results derived from this paper could be very useful for Beijing's urban planners. According to the research results, the Fengtai Technology Park and Beijing's Economic-Technological Development Area could perhaps provide the key to effectively alleviating the imbalance between the north and the south of the city . These two areas already account for a significant number of jobs, and they would be likely to attract more jobs if transportation links and policy measures were improved, thereby promotingthe development of the southern part of the city and achieving a more equal balance between north and south Beijing. Furthermore,it would provide an incentive for people tomove to the south of the city, thus helping to reduce the pressures on urban land and traffic congestion.

In terms of policy implications, this research would enable urban planners to understand the urban functional area dynamics more accurately and easily. Urban planners could formulate appropriate policies for different functional areas to promote city development in order to improve the living standards of residents, and provide better travel services for floating people and tourists, while reducing traffic congestion. The effects of policies on different areas could also be evaluated by detecting functional areas dynamics after policy implementation.

However, the paper has two limitations. First, observable urban dynamics often take place over a long time span. Thus, the four year time span from 2014 to 2017 used in this research could be seen as a relatively short time window and only small changes were detected, as was apparent from the results shown in Figure 9. We were limited by the data availability, but analysis covering a longer time period of, for example, ten years could be undertaken in a future study when data becomes available. Second, the model that we propose for identifying urban functional area dynamics based on smart card data produces the results that simulate urban functional area dynamics without testing and comparing them to actual changes that occurred during the years between 2014 and 2017. This limitation could be addressed in future research.

Data Availability

The smart card data derived from Beijing Transportation Information Centre are confidential, and will therefore not be made publicly accessible.

Acknowledgements

The authors would like to extend their appreciation to the chief editor and the anonymous reviewers for their valuable comments on the initial draft of this paper. This research is funded by the National Natural Science Foundation of China (Project No. 51808392), the EPSRC (EPSRC Reference: EP/R035148/1), the SCUE Research Fund, and School Funding from the University of Westminster.

References

Alsger, A., Tavassoli, A., Mesbah, M., Ferreira, L., & Hickman, M. (2018). Public transport trip purpose inference using smart card fare data. T*ransportation Research Part C: Emerging Technologies*, 87, 123-137.

Beijing Municipal Commission of Planning and Natural Resources. (2017). Beijing's Urban Master Plan (2016-2030). http://ghgtw.beijing.gov.cn/col/col5096/index.html /Accessed 2 April 2019

Beijing Transport Institute. (2018). *2017 Beijing Transport Development Annual Report*. Available at: http://www.bjtrc.org.cn/List/index/cid/7.html/ (accessed 13 July 2020),

Blainey, S., & Mulley, C. (2013, October). Using Geographically Weighted Regression to forecast rail demand in the Sydney Region. In Australasian Transport Research Forum, Brisbane.

Briand, A. S., Côme, E., Trépanier, M., & Oukhellou, L. (2017). Analyzing year-to-year changes in public transport passenger behaviour using smart card data. *Transportation Research Part C: Emerging Technologies*, 79, 274-289.

Chen, C., Chen, J., & Barry, J. (2009). Diurnal pattern of transit ridership: a case study of the New York City subway system. *Journal of Transport Geography*, 17(3), 176-186.

Chen, Y., Liu, X., Li, X., Liu, X., Yao, Y., Hu, G., ... & Pei, F. (2017). Delineating urban functional areas with building-level social media data: A dynamic time warping (DTW) distance based k-medoids method. *Landscape and Urban Planning*, 160, 48-60.

Cao, M. (2019). *Exploring the Relation between Transport and Social Equity: Empirical Evidence from London and Beijing*. PhD thesis, The Bartlett School of Planning, UCL.

Cao, M., Chen, C-L., & Hickman, R. (2017). Transport emissions in Beijing: A scenario planning approach. *Proceedings of the Institution of Civil Engineers – Transport*, 170(2), 65-75.

Cao, M., & Hickman, R. (2018). Car dependence and housing affordability: An emerging social deprivation issue in London. *Urban Studies*, 55(10), 2088-2105.

Cao, M., & Hickman, R. (2019). Understanding travel and differential capabilities and functionings in Beijing. *Transport Policy*, 83, 46-56.

Cao, M., & Hickman, R. (2020). Transport, Social Equity and Capabilities in East Beijing. In: Chen, C.-L., Pan, H., Shen, Q. and Wang, J. (eds.), *Handbook on Transport and Urban Transformation in China*. Cheltenham: Edward Elgar, 317-333.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. IEEE transactions on pattern analysis and machine intelligence, 2, 224-227.

Dewita, Y., Burke, M., & Yen, B.T.H. (2020). The relationship between transport, housing and urban form: Affordability of transport and housing in Indonesia. *Case Studies on Transport Policy*, 8(1), 252-262.

Dewita, Y., Yen, B.T.H., & Burke, M. (2018). The effect of transport cost on housing affordability: Experiences from the Bandung Metropolitan Area, Indonesia. *Land Use Policy*, 79, 507-519.

Gan, Z., Yang, M., Feng, T., & Timmermans, H. (2020). Understanding urban mobility patterns from a spatiotemporal perspective: daily ridership profiles of metro stations. *Transportation*, 47, 315-336.

Geng, W., & Yang, G. (2017). Partial correlation between spatial and temporal regularities of human mobility. *Scientific reports*, 7(1), 6249.

Gong, Y., Lin, Y., & Duan, Z. (2017). Exploring the spatiotemporal structure of dynamic urban space using metro smart card records. *Computers, Environment and Urban Systems*, 64, 169-183.

Goulet-Langlois, G., Koutsopoulos, H. N., & Zhao, J. (2016). Inferring patterns in the multi-week activity sequences of public transport users. *Transportation Research Part C: Emerging Technologies,* 64, 1-16.

Halvorsen, A., Koutsopoulos, H. N., Lau, S., Au, T., & Zhao, J. (2016). Reducing subway crowding: analysis of an off-peak discount experiment in Hong Kong. *Transportation Research Record,* 2544(1), 38-46.

Hasan, S., Schneider, C. M., Ukkusuri, S. V., & González, M. C. (2013). Spatiotemporal patterns of urban human mobility. *Journal of Statistical Physics*, 151(1-2), 304-318.

Hasnat, M. M., & Hasan, S. (2018). Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. *Transportation Research Part C: Emerging Technologies*, 96, 38-54.

Heiden, U., Heldens, W., Roessner, S., Segl, K., Esch, T., & Mueller, A. (2012). Urban structure type characterization using hyperspectral remote sensing and height information. *Landscape and Urban Planning,* 105(4), 361-375.

Huang, J., Levinson, D., Wang, J., Zhou, J., & Wang, Z. J. (2018). Tracking job and housing dynamics with smartcard data. *Proceedings of the National Academy of Sciences,* 115(50), 12710-12715.

Huang, J., Levinson, D., Wang, J., & Jin, H. (2019). Job-worker spatial dynamics in Beijing: Insights from Smart Card Data. *Cities*, 86, 83-93.

Jiang, H., & Levinson, D. (2017). Accessibility and the evaluation of investments on the Beijing subway. *Journal of Transport and Land Use,* 10(1), 395-408.

Kieu, L. M., Bhaskar, A., & Chung, E. (2015). A modified density-based scanning algorithm with noise for spatial travel pattern analysis from smart card AFC data. *Transportation Research Part C: Emerging Technologies,* 58, 193-207.

Lee, S.G., & Hickman, M. (2014). Trip purpose inference using automated fare collection data. *Public Transport,* 6, 1-20.

Li, Y., Wang, X., Sun, S., Ma, X., & Lu, G. (2017). Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks. *Transportation Research Part C: Emerging Technologies,* 77, 306-328.

Long, Y., & Thill, J. C. (2015). Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing. *Computers, Environment and Urban Systems,* 53, 19-35.

Ma, X., Liu, C., Wen, H., Wang, Y., & Wu, Y. J. (2017). Understanding commuting patterns using transit smart card data. *Journal of Transport Geography,* 58, 135-145.

Ma, X., Zhang, J., Ding, C., & Wang, Y. (2018). A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership. *Computers, Environment and Urban Systems,* 70, 113-124.

Mohamed, K., Côme, E., Oukhellou, L., & Verleysen, M. (2017). Clustering smart card data for urban mobility analysis. *IEEE Transactions on Intelligent Transportation Systems,* 18(3), 712-728.

Pelletier, M. P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. Transportation Research Part C: *Emerging Technologies,* 19(4), 557-568.

Pham, H. M., Yamaguchi, Y., & Bui, T. Q. (2011). A case study on the relation between city planning and urban growth using remote sensing and spatial metrics. *Landscape and Urban Planning,* 100(3), 223-230.

Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2000). Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing,* 10(1-3), 19-41.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.

Sagl, G., Delmelle, E., & Delmelle, E. (2014). Mapping collective human activity in an urban environment based on mobile phone data. *Cartography and Geographic Information Science,* 41(3), 272-285.

Taylor, B. D., Miller, D., Iseki, H., & Fink, C. (2009). Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas. *Transportation Research Part A: Policy and Practice*, 43(1), 60-77.

Thompson, G. L., & Brown, J. R. (2006). Explaining variation in transit ridership in US metropolitan areas between 1990 and 2000: multivariate analysis. *Transportation Research Record,* 1986(1), 172-181.

Van de Voorde, T., Jacquet, W., & Canters, F. (2011). Mapping form and function in urban areas: An approach based on urban metrics and continuous impervious surface data. *Landscape and Urban Planning,* 102(3), 143-155.

Wang, W., Attanucci, J. P., & Wilson, N. H. (2011). Bus passenger origin-destination estimation and related analyses using automated data collection systems. *Journal of Public Transportation,* 14(4), 7.

Wang, Z., Chen, F., & Fujiyama, T. (2015). Carbon emission from urban passenger transportation in Beijing. *Transportation Research Part D: Transport and Environment*, 41, 217–227.

Wang, Z. J., Chen, F., Wang, B., & Huang, J. L. (2018). Passengers' response to transit fare change: an ex post appraisal using smart card data. *Transportation,* 45(5), 1559-1578.

Zhang, M., He, S., & Zhao, P. (2018). Revisiting inequalities in the commuting burden: Institutional constraints and job-housing relationships in Beijing. *Journal of Transport Geography*, 71, 58-71.

Zhang, Y., Marshall, S., & Ed, M. (2019). Network criticality and the node-place-design model: Classifying metro station areas in Greater London. *Journal of Transport Geography,* 79, 102485.

Zhao, J., Qu, Q., Zhang, F., Xu, C., & Liu, S. (2017). Spatio-temporal analysis of passenger travel patterns in massive smart card data. *IEEE Transactions on Intelligent Transportation Systems,* 18(11), 3135-3146.

Zhao, P. & Cao, Y. (2020). Commuting inequity and its determinants in Shanghai: New findings from big-data analytics. *Transport Policy*, 92, 20-37..

Zhao, P. & Hu, H. (2019). Geographical patterns of traffic congestion in growing megacities: Big data analytics from Beijing. *Cities*, 92, 164-174.

Zhao, P., Yang, H., Kong, L., Liu, Y., & Liu, D. (2018). Disintegration of metro and land development in transition China: A dynamic analysis in Beijing. *Transportation Research Part A: Policy and Practice*, 116, 290-307.

Zhong, C., Huang, X., Arisona, S. M., Schmitt, G., & Batty, M. (2014). Inferring building functions from a probabilistic model using public transportation data. *Computers, Environment and Urban Systems,* 48, 124-137.

Zhong, C., Batty, M., Manley, E., Wang, J., Wang, Z., Chen, F., & Schmitt, G. (2016). Variability in regularity: Mining temporal mobility patterns in London, Singapore and Beijing using smart-card data. *PloS one,* 11(2), e0149222.

Zhu, Y., Chen, F., Li, M., & Wang, Z. (2018). Inferring the Economic Attributes of Urban Rail Transit Passengers Based on Individual Mobility Using Multisource Data. *Sustainability,* 10(11), 4178.

Zhu, Y., Chen, F., Wang, Z., & Deng, J. (2019). Spatio-temporal analysis of rail station ridership determinants in the built environment. *Transportation,* 46, 2269-2289.

Zivkovic, Z. (2004). Improved adaptive Gaussian mixture model for background subtraction. *Proceedings of the 17th International Conference on Pattern Recognition*, 2, 28-31.

Zou, Q., Yao, X., Zhao, P., Wei, H., & Ren, H. (2018). Detecting home location and trip purposes for cardholders by mining smart card transaction data in Beijing subway. *Transportation,* 45(3), 919-944.

Table 1.

Examples of AFC data

Grant_Card_Code	Entry_Time	Deal_Time	Entry_Station	Exit_Station
1020	2016/2/29 17:25	2016/2/29 17:36	Hujialou	Qingnianlu
1020	2016/3/2 17:29	2016/3/2 17:42	Hujialou	Qingnianlu
1020	2016/3/3 17:21	2016/3/3 17:30	Hujialou	Qingnianlu
1032	2016/2/29 7:35	2016/2/29 8:01	Jinsong	Huixinxijie
1032	2016/2/29 18:04	2016/2/29 18:28	Taiyanggong	Jinsong
1032	2016/3/1 7:42	2016/3/1 8:07	Jinsong	Huixinxijie
		……		

Table 2.

Feature selection and identification of station characteristics

Scale	Index Name	Meaning	Range
Passenger travel pattern	F1	Proportion of low probability passengers to total passengers at evening peak time	[0,1]
	F2	Proportion of low probability passengers to total passengers at morning peak time	[0,1]
	F3	Proportion of high probability passengers to total passengers at evening peak time	[0,1]
	F4	Proportion of high probability passengers to total passengers at morning peak time	[0,1]
	F5	Proportion of high probability passengers to total passengers within a day	[0,1]
	F6	Proportion of low probability passengers to total passengers within a day	[0,1]
	F7	Proportion of mid probability passengers to total passengers within a day	[0,1]
Station ridership pattern	F8	Proportion of passengers entering station to total passengers at evening peak time	[0,1]
	F9	Proportion of passengers entering station to total passengers at morning peak time	[0,1]
	F10	The entropy value for entering station	[0,1]
	F11	The entropy value for exiting station	[0,1]

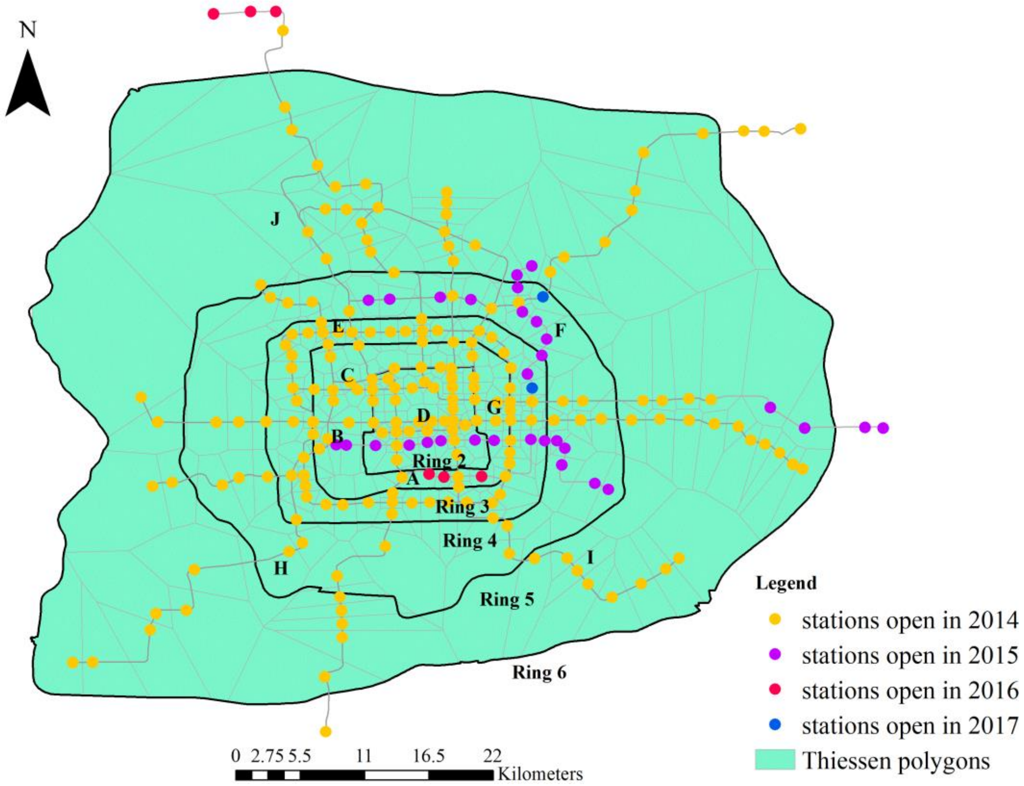


Fig. 1. Metro stations and lines in Beijing (2014-2017)

(Please note that T2\T3 terminal stations are not included in the map)

(A: Beijing south railway station; B: Beijing west railway station; C: Beijing zoo; D: Tiananmen square; E: Zhongguancun technology park; F: Wangjing; G: Guomao; H: Fengtai technology park; I: Beijing economic-technological development area; J: Xierqi)

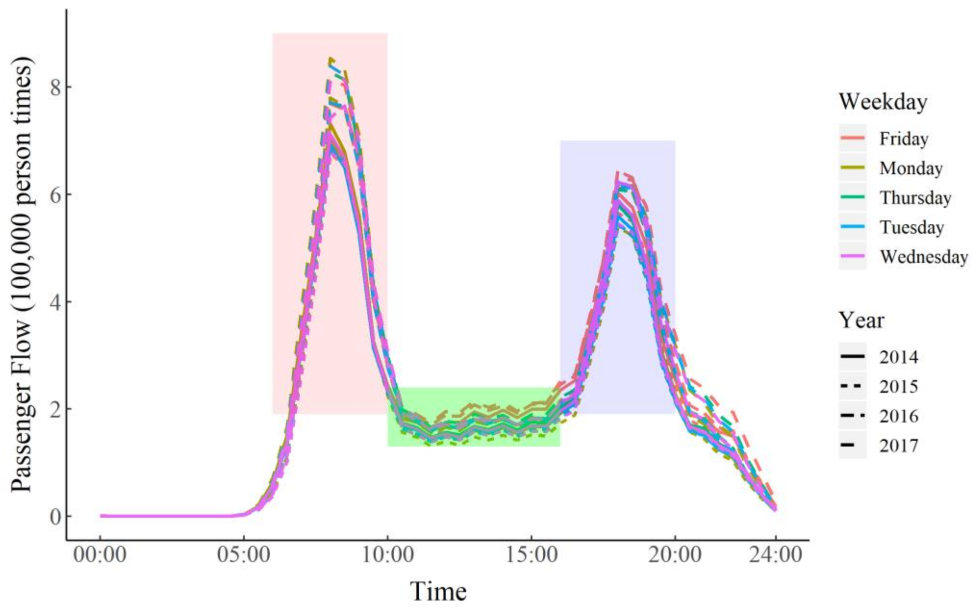


Fig. 2. Distribution of ridership for weekdays in each of the 4 year

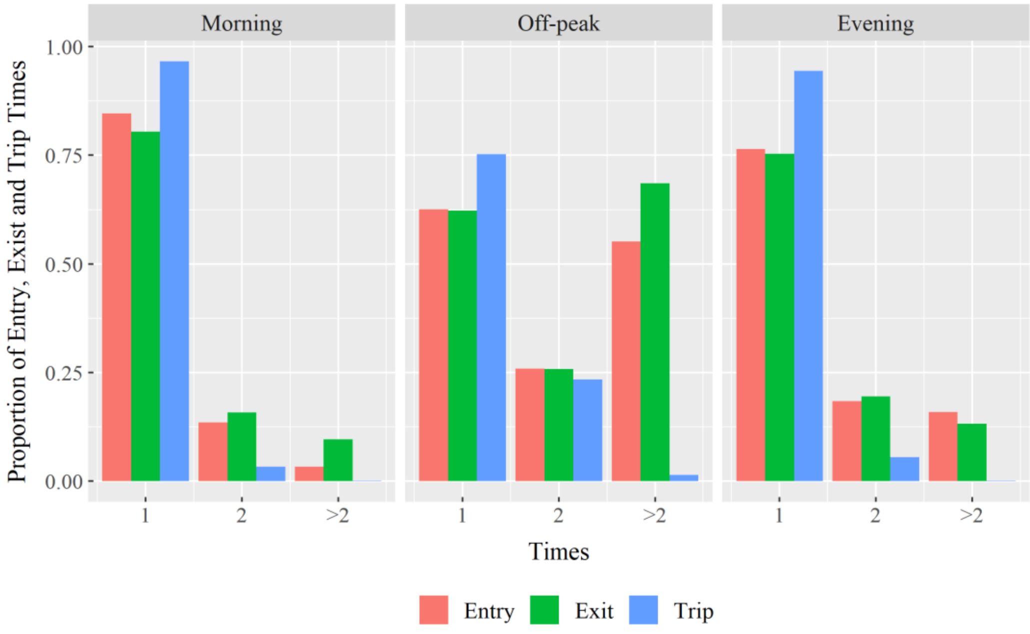


Fig. 3. Travel times and number of stations visited during different time periods

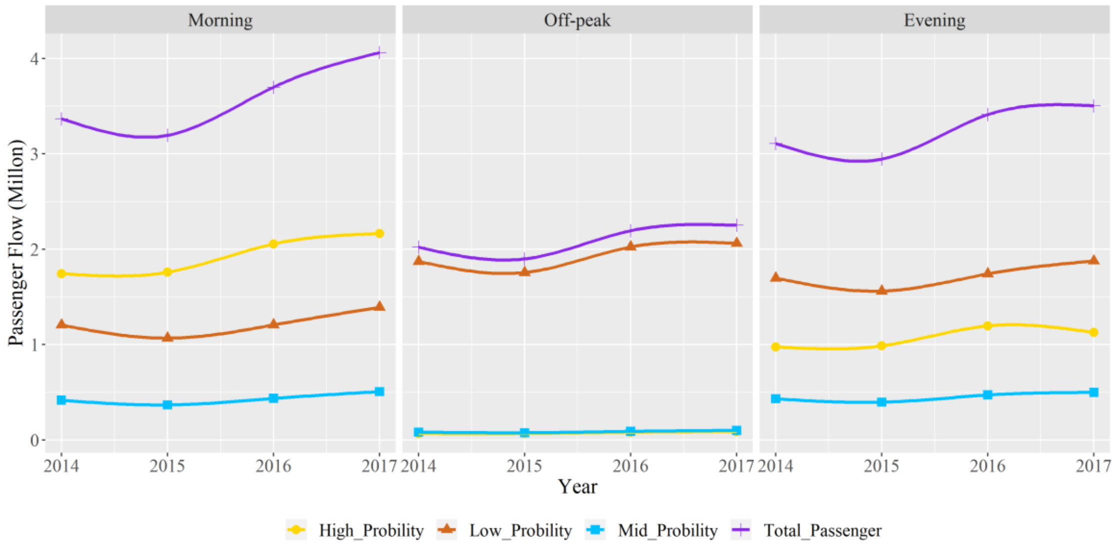


Fig. 4. Ridership travel probabilities during different periods

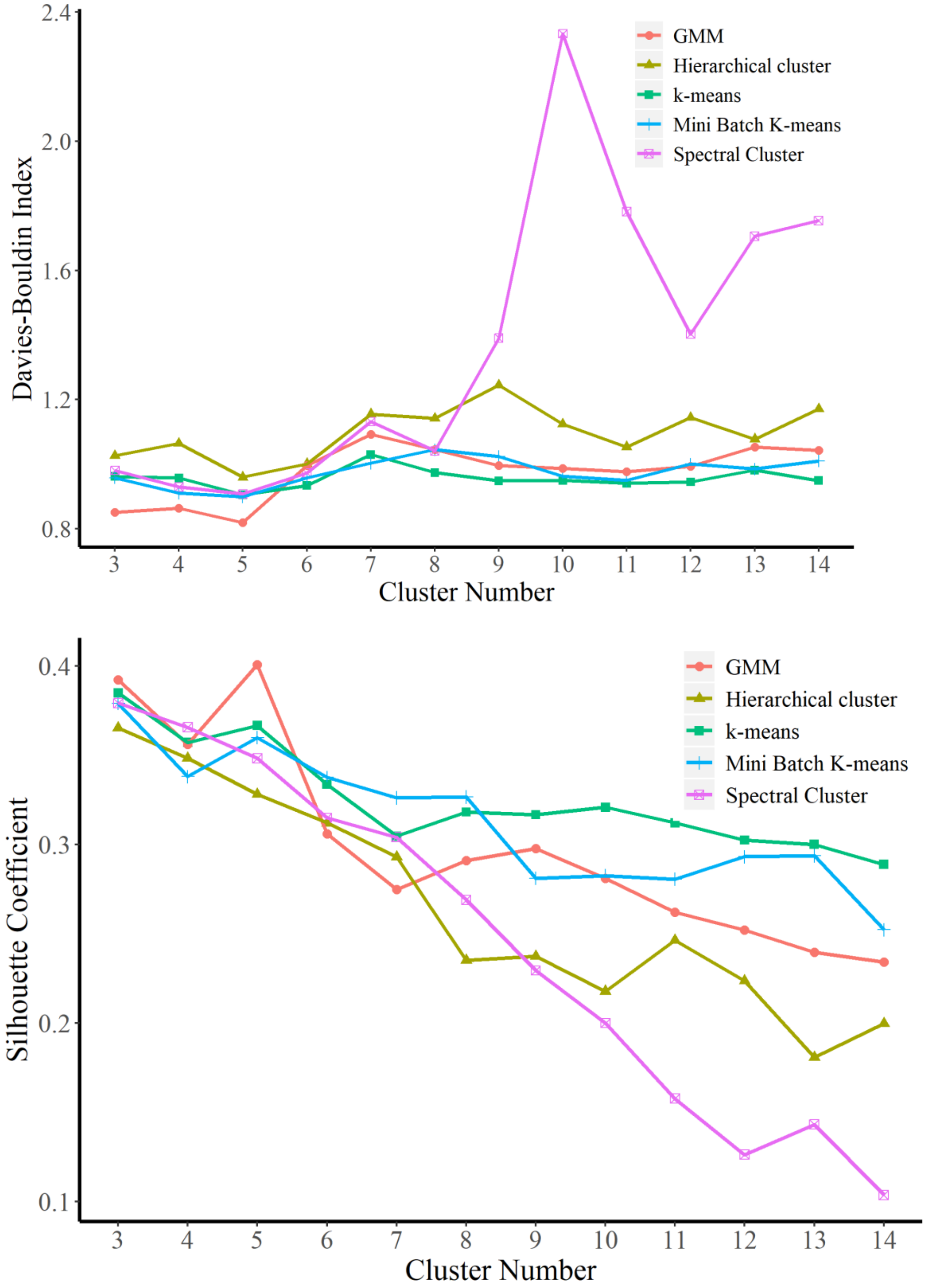


Fig. 5. DBI and SC for different numbers of clusters and different models

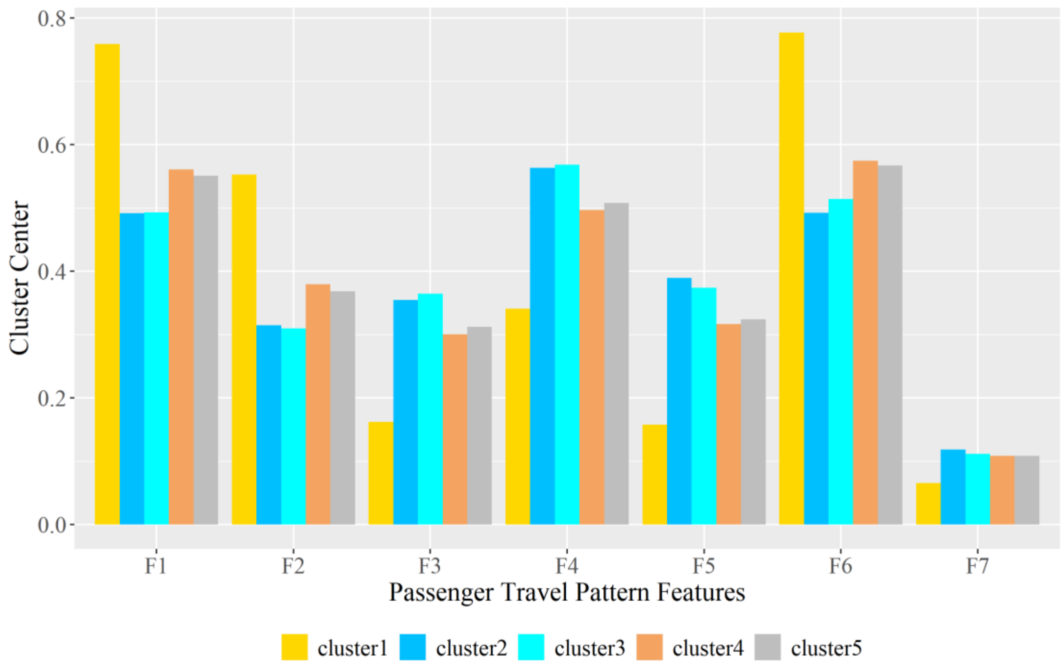


Fig. 6. Travel pattern indicators of each cluster centre

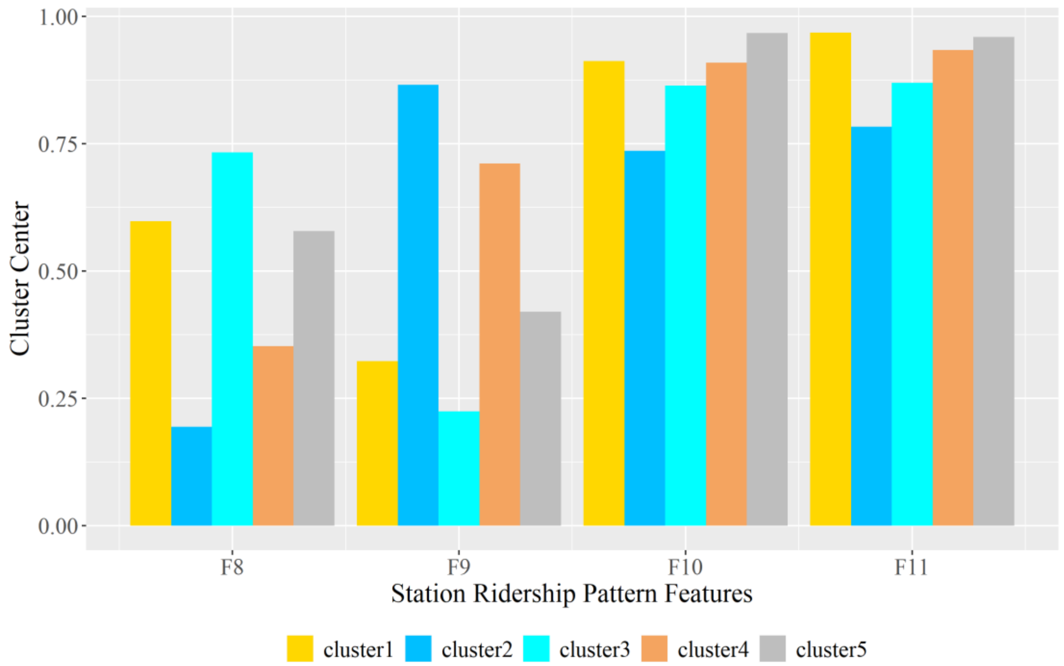
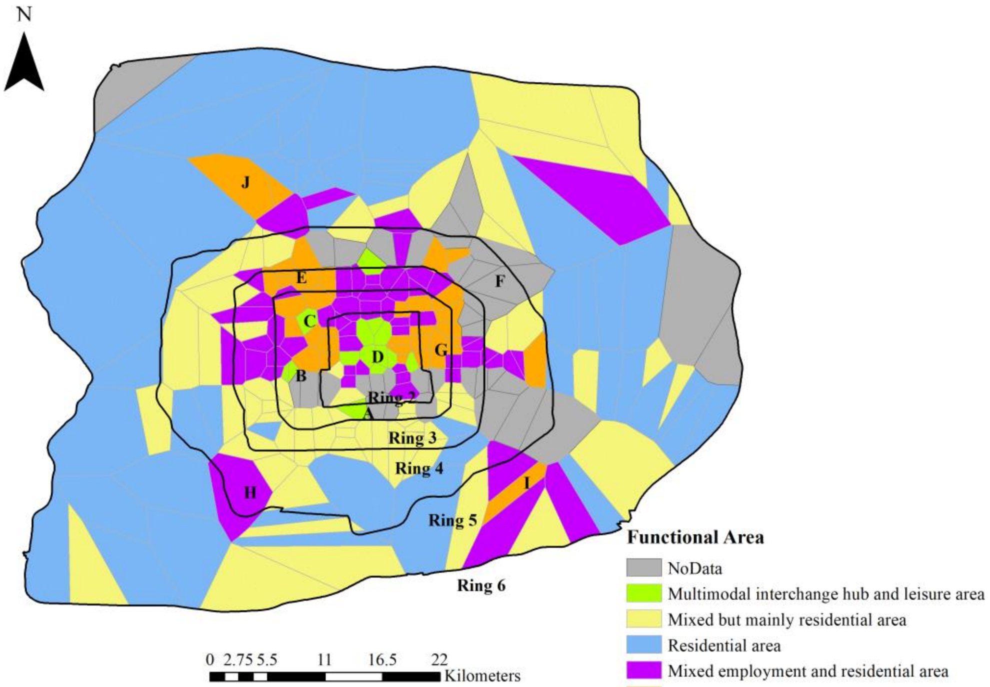
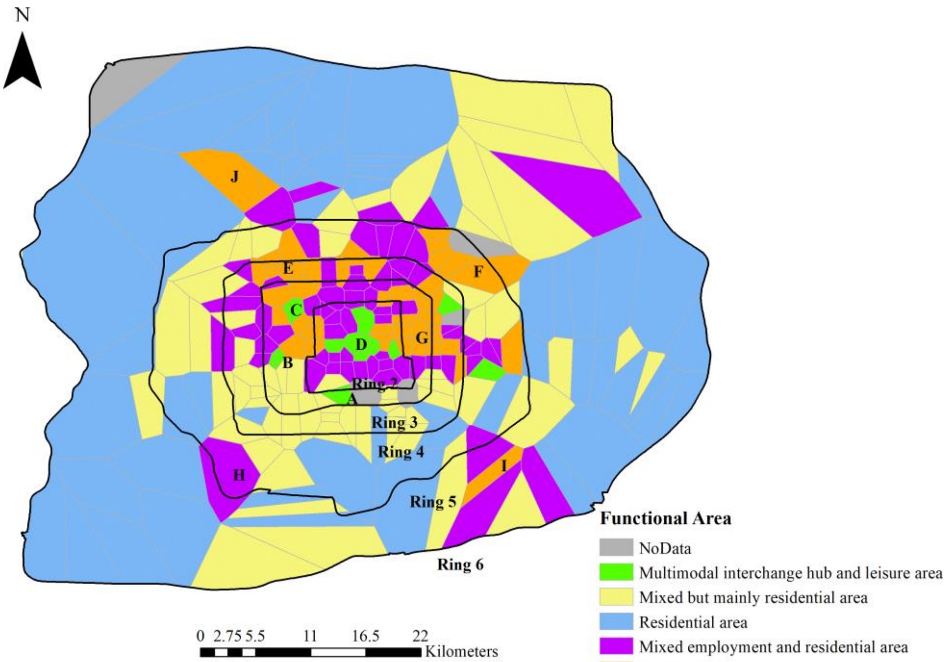


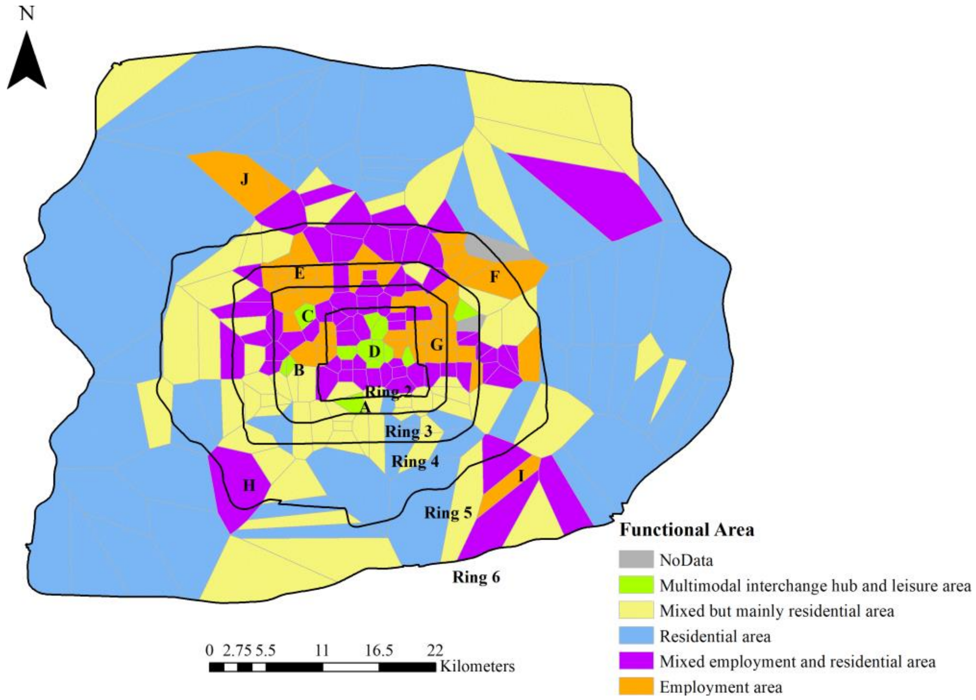
Fig. 7. Ridership pattern indicators of each cluster centre


N
J
E
F
C
D
G
B
Ring 2
A
Ring 3
Ring 4
H
I
Ring 5
Ring 6
Functional Area
NoData
Multimodal interchange hub and leisure area
Mixed but mainly residential area
Residential area
Mixed employment and residential area
Employment area
0 2.75 5.5 11 16.5 22
Kilometers

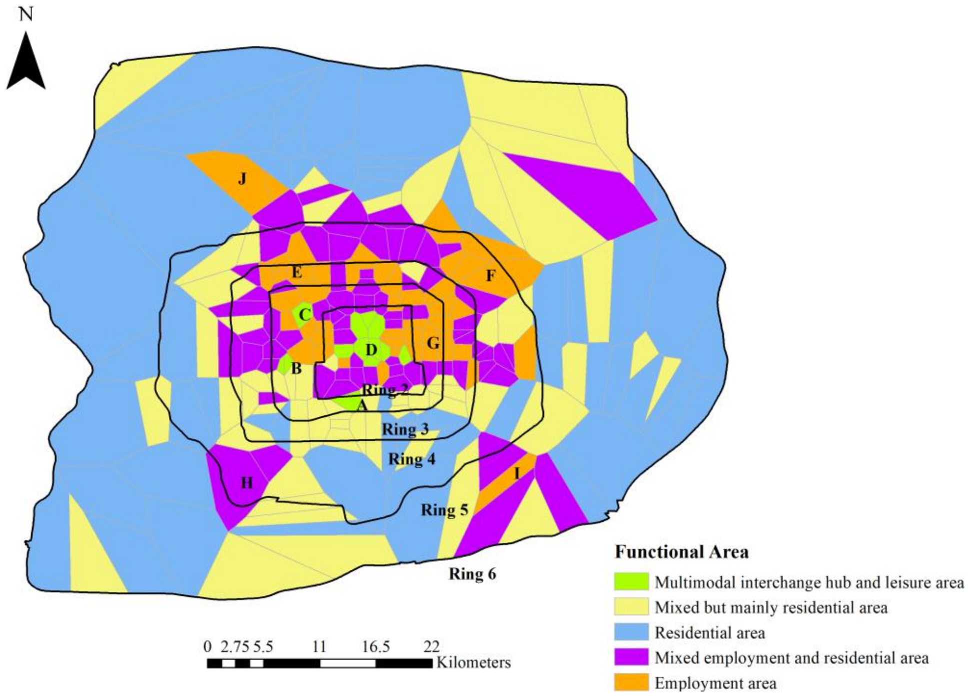
(a)


N
J
E
F
C
D
G
B
Ring 2
A
Ring 3
Ring 4
H
I
Ring 5
Ring 6
Functional Area
NoData
Multimodal interchange hub and leisure area
Mixed but mainly residential area
Residential area
Mixed employment and residential area
Employment area
0 2.75 5.5 11 16.5 22
Kilometers

(b)



(c)



(d)

Fig. 8. Spatial distribution of different clusters

(a:2014, b:2015, c:2016, d:2017)

(A: Beijing South Railway Station; B: Beijing West Railway Station; C: Beijing Zoo; D: Tiananmen Square; E: Zhongguancun Technology Park; F: Wangjing; G: Guomao; H: Fengtai Technology Park; I: Beijing Economic-Technological Development Area; J: Xierqi)

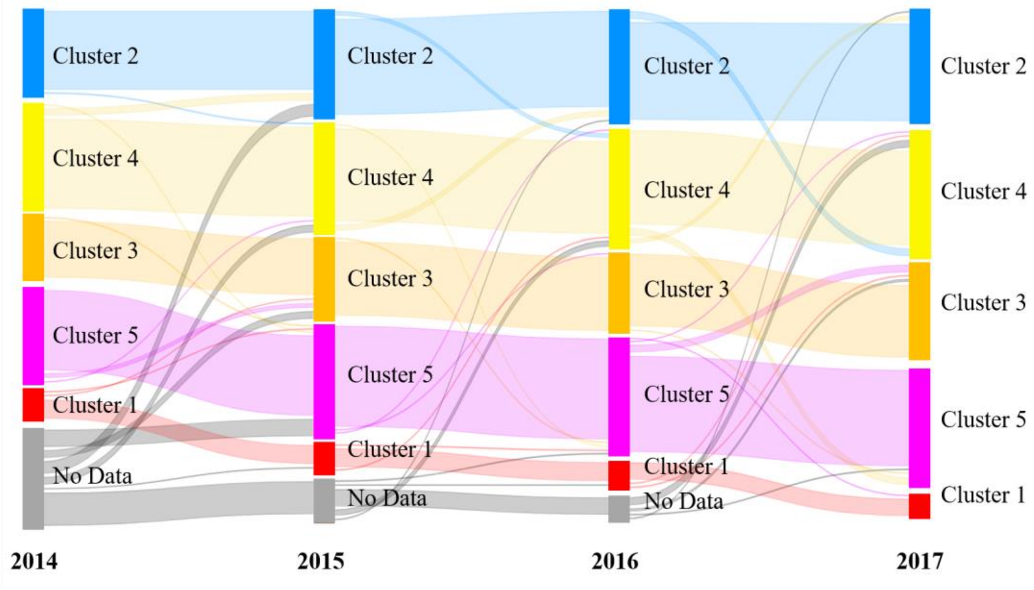


Fig. 9. The evolution process of different clusters of stations

(Cluster1: Multimodal interchange hub and leisure Area, Cluster 2: Residential area, Cluster 3: Employment area, Cluster 4: Mixed but mainly residential area, Cluster 5: Mixed residential and employment area, No Data: Stations not open yet)

752