IOP Conf. Series: Earth and Environmental Science **189** (2018) 062051 doi:10.1088/1755-1315/189/6/062051

# Analysis on Passenger Flow Changes during Holidays——A Case Study of Beijing-Shanghai High-Speed Railway

**P W Han[1], L Nie[1,2]**

[1] School of Traffic and Transportation, Beijing Jiaotong University, No.3 Shangyuancun, Haidian District, Beijing, China

[2] Corresponding author at: Department of Transportation Management Engineering, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

E-mail address: lnie8509@yahoo.com

**Abstract.** In China, the number of passengers taking the high-speed train will increase dramatically during the short-term holidays. However, only a few researchers have analysed the characteristics of high-speed railway passenger flow change during the holidays and all of them just focused on the total number change, lacking analysis on different OD types. Therefore, in order to provide specific advice for making and optimizing the holiday high-speed train line plan, this paper studied passenger flow change during holidays based on OD classification, taking Beijing-Shanghai High-Speed Railway (BSHR) as an example. Firstly, we classified OD pairs into different groups according to OD level and OD distance, while the OD level is defined by the station level which was obtained by Hierarchical Clustering Method. Secondly, by counting and calculating the increased number of passengers during each holiday for each OD group, we drew the conclusion that the increased passenger flow of BSHR in holidays is mainly from the 2$^{nd}$ level OD, 1$^{st}$ level OD and [0,500)km OD. Due to this result, we suggest to increase the service frequency of 1$^{st}$ and 2$^{nd}$ level OD pairs in holidays. Also, we prefer to add short-distance trains and try to avoid adding too many long-distance trains.

## 1. Introduction

Passenger flow distribution is the decisive factor for high-speed train line planning, and passenger flow fluctuation decides how to adjust the existing train line plan. In recent years, with the rapid development of China's high-speed railway, the attractiveness and competitiveness of railway have been greatly improved in passenger transport market. However, the railway planners have begun to face the problem of tight transport capacity, because more passengers will strain transport resources, i.e. the capacity of trains. In this case, deep understanding and mastering of the distribution characteristics and changing rules of passenger flow have important auxiliary decision-making value for making and optimizing the train line plan, transportation resource allocation and marketing strategy, which will improve the utilization efficiency of transportation resources to some extent.

In China, there are many short-term holidays each year, including five 3-day holidays, i.e. New Year's Day, Ching Ming Festival, Labor Day, Dragon Boat Festival and Mid-Autumn Festival, and one 7-day holiday, i.e. National Day. Obviously, from figure 1, we can see that the short-term holidays in China have shown different passenger flow fluctuation with other days. According to figures, the amount of days among all short-term holidays (including the previous day and the day after a holiday) is 35, and the sum of passenger flow volumes during these periods is over 176 million, accounting for

9.6% and 12.2% of which over the whole year respectively. Especially in the peak day of short-term holidays, the total increased number of passengers will go up to 2.35 million, and the maximum value of passenger flow fluctuation coefficient is 1.76 with respect to workdays. It is evident that such sharp increase of passenger flow will lead to more demands for trains and seats, but the maximum number of trains running on the track is limited and the capacity of one train (set) is also fixed. Therefore, in order to transport more passengers with limited resources, it is urgently necessary to generate the line plan making good use of the trains operated in holidays. This work greatly depends on systematical analysis of the passenger flow variation as well as the distribution of changed volumes.

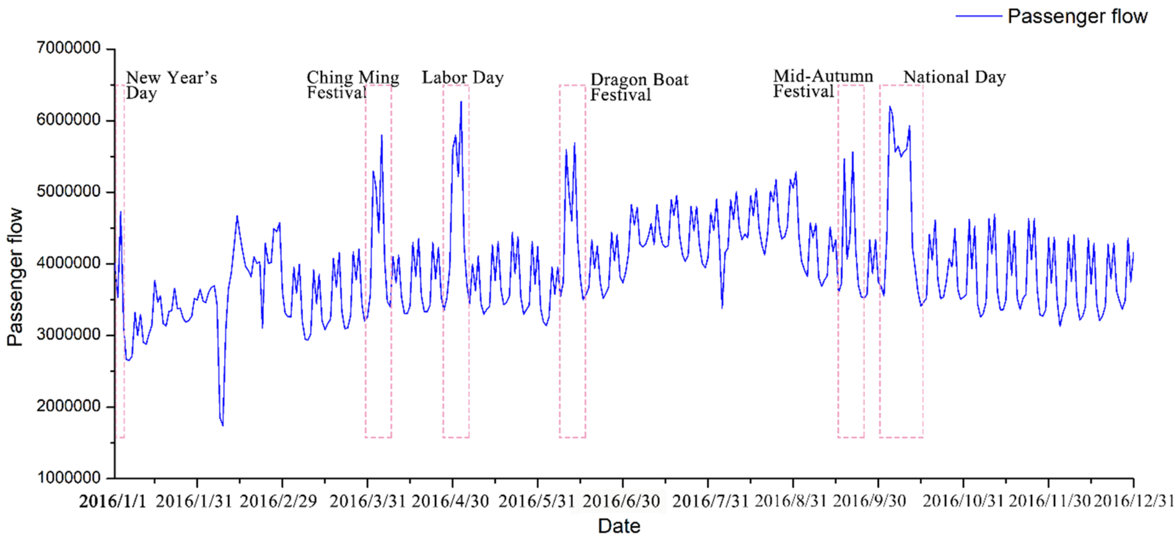

Figure 1. Trends in daily passenger flow of China's high-speed railway in 2016

In other countries, extensive studies have been carried out to explore the trends and spatial and temporal distribution of passenger flow in metro system, but few papers focused on railway. Tanaka M, Sato N and Sakuma et al [1] utilized passenger flow data information to analyse the distribution characteristics and rules of passenger flow for Japanese railway, with the help of data mining. Myojo S [2] and Myojo S [3] analysed the daily flow volume of urban rail system, based on passengers' traveling data stored by Automatic Fare Gate (AFG) in stations, which even includes travellers' detailed route and transfer information. The conclusions can be used for short-term flow forecasting and train rescheduling. For passengers' traveling behaviour, Suh S D, Hyun J and Kim [4] testified that the quality of transport service has an important influence on travellers' behaviour, by the research towards Korean Train Express (KTX). Also, Albalate D and Fageda X [5] studied the same topic for Spanish high-speed railway, using empirical evidence.

However, in China, many researchers have analysed railway passenger flow features with different backgrounds. Feng B, Bao X and Wang Q [6] adopted the Grey Neural Network method to study the trends of flow. Ma Y X and Gao S [7] studied the periodicity, trend and stationarity of short term flow. They concluded that the short-term passenger flow of the railway is generally stable and periodically changes from week to week. Specifically, as for the characteristics of passenger flow for different railway lines, the features of Beijing and Tianjin intercity passenger flow were analysed by combining some related characteristics (such as travel purpose and occupational structure), carried out by Zhang M and Zhang C [8]. Wen C L [9] studied the spatial and temporal distribution characteristics of the high-speed rail passenger flow in Hainan East Ring, and used it for the train stop plan and train scheduling optimization. Lu Y [10] carried out statistical analysis on the passenger flow structure, passenger flow trend and time distribution of Zhengzhou-Xi'an high-speed railway, and further optimized the operation strategy and train operation plan. And the spatial and temporal distribution characteristics of passenger flow in Wuhan-Guangzhou high-speed railway were summarized by

Zhang L [11] and Tong X J [12]. Meanwhile, Lu Y [10], Zhang L [11] and Tong X J [12] all contain the analysis of passenger flow changes in holidays. Lu Y [10] studied the changes of the daily passenger flow of Zhengzhou-Xi'an high-speed rail during the Ching Ming Festival. Zhang L [11] analysed the trend of daily dispatched number of passengers before and after the holiday of some stations of Wuhan-Guangzhou high-speed railway. Tong X J [12] briefly summarized the characteristics of passenger flow during short-term holidays.

In summary, most studies on railway passenger flow characteristics and change laws focus on the temporal and spatial distribution. A few of them involve passenger flow changes in holidays, but they only analyse the trend of total number of passengers during the vacation and do not explore the composition of the changed passenger flow. Such kind of analysis is insufficient to provide specific advice and guidance for making and optimizing train line plan in holidays. In view of this situation, this paper selects Beijing-Shanghai High-Speed Railway (BSHR), which is the busiest line in the Chinese high-speed railway network, and divides related OD pairs into different groups according to OD level, OD distance and OD level with distance respectively. It should be noted that the OD level definition is based on station level, which is obtained by cluster analysis in accordance with the number of passengers boarding and getting off trains at the station. Then we can get the number of passengers increased during holidays for each group. This result will tell us how passenger flow changes in each holiday and will be more detailed and more practical for holiday line plan optimization.

## 2. OD classification

The number of related OD pairs of BSHR is huge and the difference among them is obvious, so it is better to analyze the data for different types of OD pairs. In this case, the article will firstly classify OD pairs from two points of view, i.e., OD level and OD distance. Since each OD pair is made up by two stations, we define the OD level by station level.

### *2.1. Station clustering*

We use the Hierarchical Clustering Method to obtain the station level. Since the station level is related to many factors, including the number of passengers, urban population, urban GDP, etc., in order to simplify the cluster process, we only use the most directly related factor, i.e. the number of passengers, as the clustering index. It can be calculated as follows:

$$x_k = \sum_{i=1}^{S_1} p_{ik} + \sum_{j=1}^{S_2} p_{kj} \tag{1}$$

In formula (1), $x_k$ represents the sum of passengers departing from station $k$ or arriving at station $k$ per day. Correspondingly, $\sum_{i=1}^{S_1} p_{ik}$ and $\sum_{j=1}^{S_2} p_{kj}$ indicate the total number of passengers getting off and on at station $k$ respectively. Specifically, the index $i$ represents the departure station so that $S_1$ is the set of stations that passengers depart from, while the index $j$ indicates the destination and $S_2$ is the set of stations that passengers in station $k$ leave for.

Then we can calculate the inter-class distance based on the clustering index. Firstly, for a class $G$, we set $x_1, x_2, \cdots, x_n$ as the clustering index for each element of $G$. Then the gravity of $G$, which is denoted as $\bar{x}_G$, can be calculated by formula (2).

$$\bar{x}_G = n^{-1} \sum_{k=1}^{n} x_k \tag{2}$$

When $\bar{x}_G$ is obtained, we can get the diameter of $G$, denoted by $D_G$. The formula is as follows:

$$D_G = \sum_{i \in G} (x_i - \overline{x}_G)^2 \tag{3}$$

Finally, the distance between class $G_p$ and $G_q$ can be calculated. In this article, we use the Ward's method to calculate the inter-class distance.

$$D_w^2(p \cdot q) = D_{p+q} - D_p - D_q \tag{4}$$

In this formula, $D_w^2(p \cdot q)$ is the inter-class distance of class $G_p$ and $G_q$ . $D_{p+q}$ , $D_p$ and $D_q$ indicate the diameter of $G_p \cup G_q$ , $G_p$ and $G_q$ respectively, which can be calculated according to formula (3). What's more, before the calculation of diameter for each class, the gravity $\overline{x}_{p+q}$ , $\overline{x}_p$ and $\overline{x}_q$ should be obtained first by formula (2).

Based on the symbol definition and formulas proposed above, we then use the Hierarchical Clustering Method to cluster stations. The procedure is as follows:

Step 1: We divide N stations into N original classes, i.e. each station is an independent class.

Step 2: The inter-class distance is calculated by formulas (2)-(4). Then we combine the closest two classes into a new class. After clustering in this way, the whole sample is clustered into N-1 classes.

Step 3: We calculate the inter-class distance again based on the new classes and choose the closest two classes to form a new class.

Step 4: If the number of classes is 1, i.e. all stations have been clustered into one class, the process is finished. If not, the step 3 will be repeated.

Following the method above, we cluster stations related to BSHR into a final class. Then, we use the scree plot to decide the number of classes to be divided, according to the clustering trend and variance. Therefore, we capture three classes of stations. The 1st class is the set of big stations, which has 40 elements. The 2nd class is for medium stations while the third one represents small stations, having 84 and 184 elements respectively.

### *2.2. OD grouping*

The correspondence between the OD level and the station level is shown in table 1, which can also be regard as the definition of OD levels. As presented, there are three OD levels in this paper. The first level is set for OD pairs linking 1st class of stations while OD in second level connects 1st class of stations with the ones in lower classes. And OD pairs formed by lower classes of stations belong to 3rd OD level.

Table 1. OD level definition based on the origin and destination's station level

| Items | Level Number | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Station level of origin** | 1st | 1st | 1st | 2nd | 3rd | 2nd | 2nd | 3rd | 3rd |
| **Station level of destination** | 1st | 2nd | 3rd | 1st | 1st | 2nd | 3rd | 2nd | 3rd |
| **OD level** | 1st level | 2nd level | | | | 3rd level | | | |

What's more, we also divide OD pairs by their length according to traditional rules. Thus the groups classified by distance are (0, 500] km OD group, (500, 1000] km OD group, (1000, 1500] km OD group and (1500, +∞) km OD group. The structural analysis of the changed passenger flow in holidays will be carried out on the basis of OD level and OD distance classification.

## 3. Passenger flow analysis based on OD classification

The samples to be analyzed are most 3-day holidays and the 7-day holiday from January 1, 2016 to June 30, 2017, including Ching Ming Festival, Labor Day, Dragon Boat Festival, Mid-Autumn Festival and National Day. Considering that passenger flow begins to grow rapidly one day before the holiday and will fall back to the normal level at least one day after the holiday, this article also counts the day before and after the holiday into the scope of vacation period, which is called General Holiday

Period (GHP). Therefore, the average number of daily passengers we considered is the average during the GHP, and the (normal) workday passenger flow used for comparison refers to the passenger flow value on the second day before the GHP. For example, the Ching Ming Festival in 2016 is from April 2 to April 4, so the GHP is from April 1 to April 5 and the workday passenger flow for comparison takes the value on March 30.

To make the results more reliable, we also pre-processed and screened the data within the selected time period before analysis. Firstly, we checked if there is the seasonal adjustment of timetable during the selected period. If so, we deleted the affected holiday from samples. Secondly, we deleted the OD influenced by the stoppage of trains due to weather or other unexpected factors in the scope of dates chosen. Finally, in the case that the Nanjing-Shanghai section has the parallel of Nanjing-Shanghai intercity railway, we removed the passenger flow transported by Nanjing-Shanghai intercity railway from the OD pairs in this section.

### *3.1. Passenger flow increment in each OD level group*

The OD level determines the scale of passenger flow between two cities to a certain extent. The ascending amount of passengers in each level is presented in figure 2. Overall, for BSHR, 50%-60% of the increased passenger flow in short-term holidays comes from the $2^{nd}$ level OD pairs. This proportion in $1^{st}$ level group and $2^{nd}$ level group is 30%-40% and 10%-15% respectively. Therefore, $2^{nd}$ level OD pairs contribute to the major passenger flow growth. For the comparison from holiday to holiday, it also can be concluded that in 3-day holidays, the growth of passenger flow mainly comes from OD pairs connecting big stations with small or medium stations, while in the 7-day holiday, there is relatively more passengers arisen between big stations.

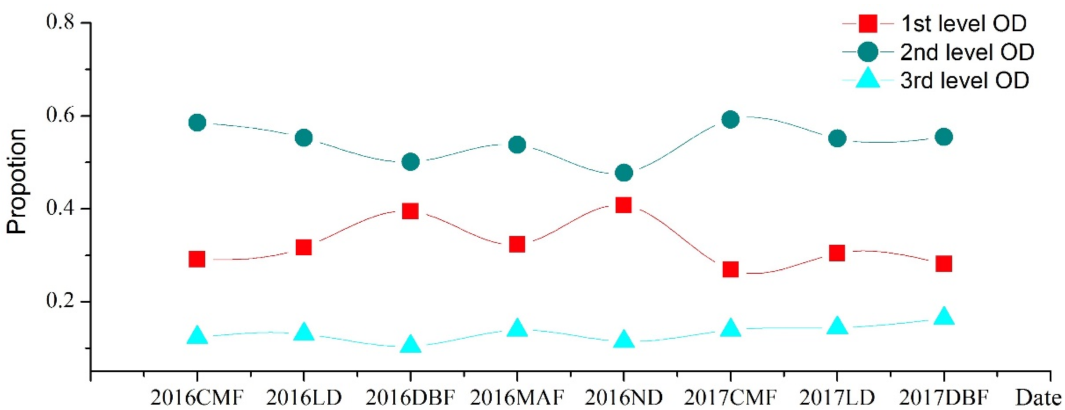


Figure 2. The proportion of the total passenger flow growth in BSHR that each OD level's passenger flow growth accounts for during some holidays in 2016 and 2017

### *3.2. Passenger flow increment in each OD distance group*

As for the growth distribution on all groups divided by distance, it can be clearly obtained from figure 3 that the increased value is mostly contributed by [0,500) km group, with the percentage between 60% and 70% except the National Day, and even five 3-day holiday samples have more than 70% of increased passenger flows in this group. About 20% and 10%-20% of passenger increment comes from [500, 1000) km group and [1000, 1500) km group respectively, totally accounting for 30%-40% of the total number. And there is few passengers arisen in OD pairs longer than 1500km.

Still, as the only 7-day holiday, the National Day shows different features with other 3-day holidays. Specifically, in [0,500) km group, the proportion of total passenger flow increment is obviously lower than it in 3-day holidays. Conversely, in other distance groups, this percentage is higher to different degrees compared to it in 3-day holidays. Therefore, it can be concluded that due to the longer time period of the vacation, the number travelers from medium and long distance during the

National Day has increased more, but in 3-day holidays, short-distance OD pairs account for a higher proportion of passenger flow growth.

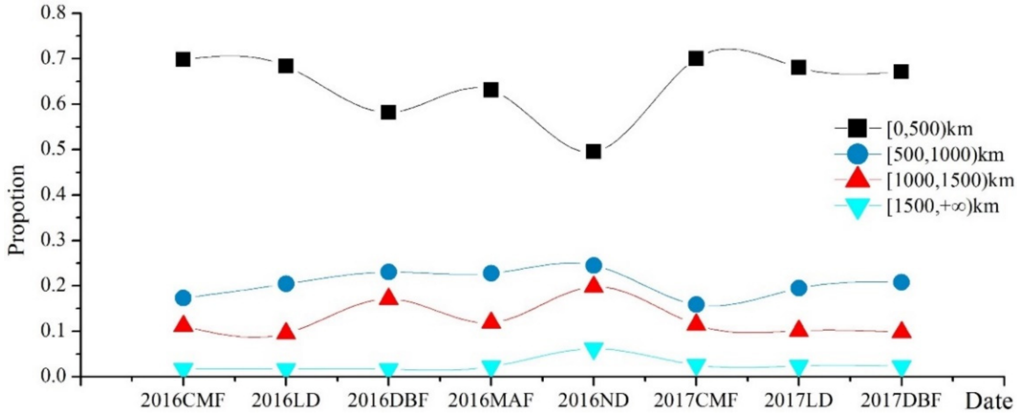


Figure 3. The proportion of the total passenger flow growth in BSHR that each OD distance group's passenger flow growth accounts for during some holidays in 2016 and 2017

*3.3. Passenger flow increment in each OD level-distance group*

Besides conclusions drawn above, we match the OD levels with OD distance groups. In this way, we obtain 12 new sub-categories to explore the detailed structure of increased number of passengers during the holidays. All categories divided by OD level with OD distance are shown in table 2.

Table 2. All OD level-distance groups

| | [0,500)km | [500,1000)km | [1000,1500)km | [1500,+∞)km |
|---|---|---|---|---|
| 1st level | √ | √ | √ | √ |
| 2nd level | √ | √ | √ | √ |
| 3rd level | √ | √ | √ | √ |

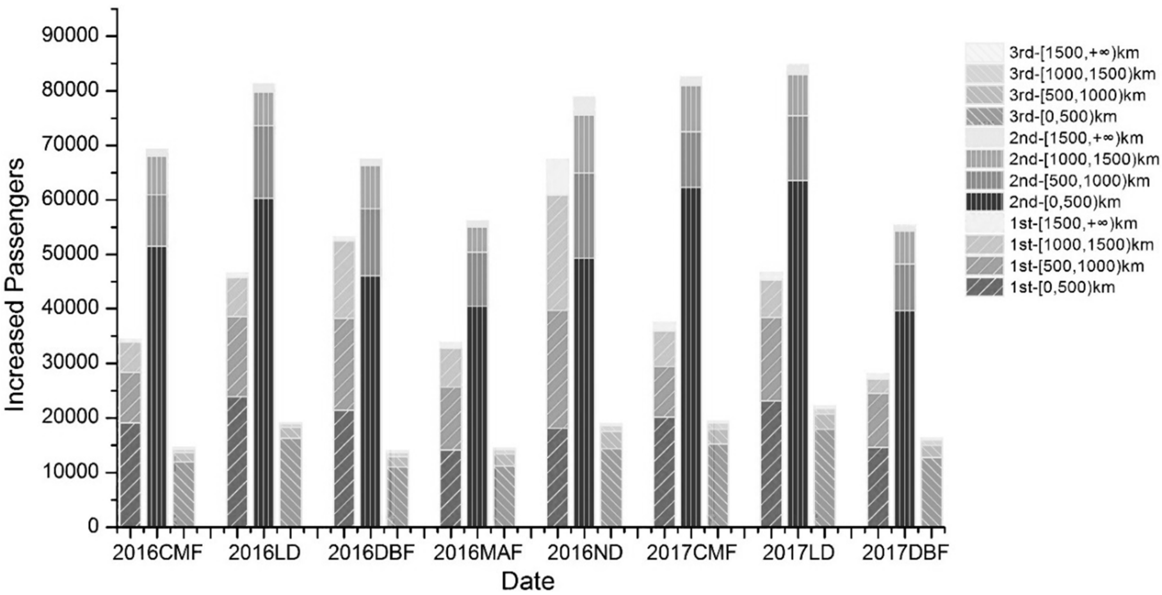


Figure 4. Amounts of increased passenger flow in each OD level-distance group of BSHR during some holidays in 2016 and 2017

With the same method, we can obtain ascending numbers of passengers over all sub-categories, as presented in figure 4. In summary, the category with the largest proportion of passenger flow growth is level 2 - [0,500) km group, followed by level 1 - [0,500) km group, level 3 - [0,500) km group, level 1 - [500, 1000) km group and level 2 - [500, 1000) km group. These categories are the main sources of the passenger flow increment of BSHR during the holidays.

For the difference between 7-day holiday and 3-day holidays, the proportions of increased passenger flow in level 1 - [1000, 1500) km group, level 1 - [1500, +∞) km group, level 2 - [1000, 1500) km group and level 1 - [1500, +∞) km group, i.e. long-distance OD pairs in level 1 and 2, are greater during the 7-day holiday, which indicates that more people choose long-distance travel thanks to the long vacation period. Besides, 1st and 2nd level OD pairs in [500, 1000) km also have a higher proportion of increased amount.

## 4. Conclusion

Taking BSHR as an example, this paper firstly divides the OD pairs into different groups according to OD distance and OD level. The latter division is accomplished with the help of station level classification using Hierarchical Clustering Method. Based on OD classification, we then analyse characteristics of passenger flow changes during the holidays, and draws the following conclusions.

(1) In holidays, the increased passenger flow of the BSHR is mainly from the OD pairs formed by the big station and the small or medium station, followed by the OD pairs between big stations.

(2) In terms of distance, the growth of passenger flow during the holidays mainly comes from short distance OD pairs, i.e. OD pairs with [0,500) km length.

(3) The amount of increased passenger flow between the medium and small stations or with the length more than 500km is relatively lower.

Based on the conclusions above, we can provide some suggestions for the adjustment of the train line plan of BSHR during the holidays. Firstly, we need to add some stops to the existing line plan to increase the service frequency of OD pairs made up by the big station and the medium or small station, as well as OD pairs between big stations. What's more, as for trains added in holidays, we had better pay more attention to short-distance trains and try to avoid adding too many long-distance trains.

As for the future study, there are many aspects that can be extended. First of all, the analysis in this paper is based on the average daily passenger flow during the holidays, and it can be further extended to the daily and even the hourly passenger flow changes. Secondly, due to the geographical location and economic environment of different railways, passenger flow in each railway line shows various features, which may be different with that of BSHR. Besides, we can also explore some other characteristics, such as daily flow density changes on each section, the influence of geography on passenger flow, the influence of urban attributes on passenger flow and so on. These work will enable us to get more precise, personalized and universal passenger flow changing rules during the holidays.

## Acknowledgments

The research carried out in this paper is supported by National Natural Science Foundation of China (Grant Nos. U1434207), the 111 Project of China (No. B18004) and National key research and development plan (No. 0201700).

## References

[1] Tanaka M, Sato N, Sakuma Y, Mafune K and Goto K 2002 A study on the application of data mining technology to the analysis of passenger flow data *RTRI Report* vol 16 no 11 pp 37-42

[2] Myojo S 2006 Daily Estimation of Passenger Flow in Large and Complicated Urban Railway Network Railway *Proc. of the 7th World Congress on Railway Research* Montreal Canada Jun. 4-8 2006

[3] Myojo S 2006 Method to Estimate Passenger Flow Using Stored Ticket Gate Data *Quarterly Report of RTRI* vol 47 no 4 pp 178-181

[4] Suh S D, Hyun J and Kim 2015 Effects of Korean Train Express (KTX) Operation on the

National Transport System *Yul* **5** 175-189
[5] Albalate D and Fageda X 2016 High speed rail and tourism: Empirical evidence from Spain *Transportation Research Part A* **85** 174-185
[6] Feng B, Bao X and Wang Q 2015 Research of railway passenger volume forecast based on grey and neural network *Journal of Railway Science & Engineering* **12(5)** 1227-1231
[7] Ma Y X and Gao S 2010 Analysis on timing rule for railway short time passenger flow *Railway Transport and Economy* **32(2)** 87-90
[8] Zhang M and Zhang C 2010 Analysis of passenger flow structure and travel characteristics of Beijing-Tianjin Intercity High-speed Railway *China Storage & Transport* **6** 88-89
[9] Wen C L 2014 Passenger Flow Characteristics Analysis and Train Operating Organization Optimization for Hainan East High-speed Railway *Master Thesis* Central South University, Changsha, China chapter 2 pp 17-32 (in Chinese)
[10] Lu Y 2013 Study on Passenger Flow Characteristics of Zhengzhou-Xi'an High-Speed Railway and Marketing Strategy *Master Thesis* Southwest Jiaotong University, Chengdu, China chapter 2 pp 7-25 (in Chinese)
[11] Zhang L 2012 Analysis on the Change and Forecast of Passenger Flow of Wuhan-Guangzhou High-speed Railway *Master Thesis* Central South University, Changsha, China chapter 2 pp 15-27 (in Chinese)
[12] Tong X J 2014 Research on Characteristics of Temporal and Spatial Passenger Flow Distribution of Wuhan-Guangzhou Railway Transport Corridor *Master Thesis* Central South University, Changsha, China chapter 2-3 pp 15-44 (in Chinese)