

Medium-term forecast of daily passenger volume of high speed railway based on DLP-WNN

Medium-term
forecast of HSR
passenger
volume

121

Tangjian Wei

*School of Transportation Engineering, East China Jiao Tong University,
Nanchang, China and
Institute for Transport Studies, University of Leeds, Leeds, UK*

Xingqi Yang

School of Economics and Management, Beihang University, Beijing, China, and

Guangming Xu and Feng Shi

*School of Traffic and Transportation Engineering, Central South University,
Changsha, China*

Received 27 January 2023
Revised 1 February 2023
Accepted 1 February 2023

Abstract

Purpose – This paper aims to propose a medium-term forecast model for the daily passenger volume of High Speed Railway (HSR) systems to predict the daily the Origin-Destination (OD) daily volume for multiple consecutive days (e.g. 120 days).

Design/methodology/approach – By analyzing the characteristics of the historical data on daily passenger volume of HSR systems, the date and holiday labels were designed with determined value ranges. In accordance to the autoregressive characteristics of the daily passenger volume of HSR, the Double Layer Parallel Wavelet Neural Network (DLP-WNN) model suitable for the medium-term (about 120 d) forecast of the daily passenger volume of HSR was established. The DLP-WNN model obtains the daily forecast result by weighed summation of the daily output values of the two subnets. Subnet 1 reflects the overall trend of daily passenger volumes in the recent period, and subnet 2 the daily fluctuation of the daily passenger volume to ensure the accuracy of medium-term forecast.

Findings – According to the example application, in which the DLP-WNN model was used for the medium-term forecast of the daily passenger volumes for 120 days for typical O-D pairs at 4 different distances, the average absolute percentage error is 7%-12%, obviously lower than the results measured by the Back Propagation (BP) neural network, the ELM (extreme learning machine), the ELMAN neural network, the GRNN (generalized regression neural network) and the VMD-GA-BP. The DLP-WNN model was verified to be suitable for the medium-term forecast of the daily passenger volume of HSR.

Originality/value – This study proposed a Double Layer Parallel structure forecast model for medium-term daily passenger volume (about 120 days) of HSR systems by using the date and holiday labels and Wavelet Neural Network. The predict results are important input data for supporting the line planning, scheduling and other decisions in operation and management in HSR systems.

Keywords High speed railway, Passenger flow forecast, Daily passenger volume, Medium-term forecast, Wavelet neural network

Paper type Research paper

© Tangjian Wei, Xingqi Yang, Guangming Xu and Feng Shi. Published in *Railway Sciences*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

The research was supported by the National Natural Science Foundation of China (Grant Nos. 72171236 and 71701216), the National Key R&D Program of China (Grant No. 2020YFB1600400), the China Scholarship Council (202008360277), the Key Science and Technology Research Program of the Educational Department of Jiangxi Province (Grant No. GJJ200605), and the Natural Science Foundation of Hunan Province (Grant No. 2020JJ5783).



1. Introduction and literature review

Passenger flow volume forecast of High Speed Railway (HSR) can provide important data support for scientific and reasonable decision-making by the operation management department. In recent years, passenger flow volume forecast has also been a research focus for scholars and railway operation management departments. In a global scope, [Tsai, Lee, and Wei \(2009\)](#) designed a multi-time unit neural network and a parallel integrated neural network to forecast the passenger flow of HSR in respect of the short-term demand forecast of railway passenger transport. [Wei and Chen \(2012\)](#) used both the empirical mode decomposition and the back propagation neural network to design a hybrid forecast method for short-term metro passenger volume forecast. [Jiang, Zhang, and Chen \(2014\)](#) developed a hybrid model for short-term passenger flow forecast by using both the ensemble empirical mode decomposition and the gray support vector machine to forecast the short-distance, medium-distance and long-distance passenger flows of HSR. [Börjesson \(2014\)](#) forecasted the long-distance passenger flow of HSR by using the elastic coefficient method and the actual data of HSR and air transport. In China, [Wang, Wang, Jia, and Qin \(2004\)](#), [Wang, Liu, Shan, and Zhu \(2010\)](#) and [Liu, Tian, Li, and Zhang \(2016\)](#) analyzed and extracted historical feature data, and improved and established forecast models based on the BP neural network, respectively. [Wang et al. \(2004\)](#) established a monthly seasonal adjustment model based on the Autoregressive Integrated Moving Average (ARIMA) model to forecast the monthly passenger transport volume of railway during the Spring Festival according to the fact that the Spring Festival holiday is not fixed in the Gregorian calendar every year. [Yang and Hou \(2013\)](#) used the wavelet analysis technology and the least square support vector machine to forecast the short-term passenger volume of rail transit. [Yao, Zhou, and Zhang \(2018\)](#) proposed a method for forecasting the real-time inbound and outbound passenger flow volumes at the initial stage of the opening of a new station of urban rail transit based on the improved k -nearest-neighbor nonparametric regression method. [Teng and Li \(2020\)](#) proposed a method for forecasting the passenger transport volume of intercity railway based on the Particle Swarm Optimization (PSO)-Long Short Term Memory (LSTM) combined forecast model in view of the date attributes and weather factors. [Liang, Xu, and Liu \(2020\)](#) considered the temporal-spatial interaction relationship of the passenger flow of urban rail transit and proposed a short-term passenger flow forecast model of urban rail transit that integrates the cyclic door control unit and the graph convolutional neural network. The common feature of the above research is that only the short-term passenger volume was forecasted.

Short-term passenger volume forecast is based on the historical data on real passenger volumes of several periods in the past, and a certain method is used to forecast the passenger volume of the next period or the passenger volumes of a very limited number of periods; that is, the definitive historical data of several periods shall be input to forecast the passenger volume of one period/the passenger volumes of several periods. Depending on the specific time unit, the forecast period can be measured in time units such as year, month, day, hour, minute or even second, but the number of forecast periods is very limited regardless of the time unit.

However, in the actual transport process of HSR, in addition to the short-term passenger volume forecast, passenger volume forecasts with a wider range of forecast periods (days) are also of important application value. For example, currently, the railway ticket pre-sale period is usually 30 days; that is, the train operation schedule from now to the next 30 days has been determined. It means that when preparing the new transportation plan, at least the fluctuation of passenger volumes in the next 30 days shall be considered. For railways in China, the passenger train operation diagram is generally adjusted at relatively fixed time points such as the beginning and end of the Spring Festival travel rush and the summer passenger rush every year; that is, the “service” time of a train operation schedule is about 3 months. During this period, because of the influence of working days, weekends, holidays and related factors, daily passenger volume of HSR will fluctuate day by day. In order to

make the train operation schedules such as the new passenger train operation plan and the operation diagram meet the market demand, it is necessary to forecast the daily passenger volume during daily service time in advance. Therefore, the medium-term forecast of daily passenger volume with a wider range of forecast periods (days) is particularly necessary.

Unlike long-term or short-term forecasts, the medium-term forecast has its own characteristics and difficulties. Compared with the long-term forecast usually expressed in year, since the factors such as population, total Gross Domestic Product (GDP) and land use between ODs (Origin-Destination) of HSR are unlikely to change significantly in the medium-term time frame. During the medium-term forecast, the above macro factor indicators cannot be used to deduce the change of the passenger volume with reference to the long-term passenger volume forecast method. Compared with the short-term forecast expressed in day, the number of forecast periods (days) in the medium-term forecast is far more than that in the short-term forecast. With the progress of the forecast periods (days), it is inevitable to continue to forecast the subsequent daily passenger volume based on the forecasted daily passenger volume, which will significantly reduce the subsequent forecast accuracy due to the accumulation of errors. This is one of the key problems to be solved in this paper.

This paper proposes a medium-term forecast method for the daily passenger volume of HSR with a forecast period of 120 days, which provides data support for the formulation of the train operation schedule by the HSR operation management department. The data on the historical daily passenger volume between O-Ds of HSR is extracted from 12306 for the characteristic analysis to further extract the characteristic factors. The date and holiday labels are designed based on the characteristic factors, and a DLP-WNN forecast model which can perform the medium-term forecast on daily passenger volume of HSR is established in combination with the autoregressive characteristics of daily passenger volume of HSR. The forecast accuracy of the test model is analyzed by examples.

2. Extraction of characteristics of daily passenger volume

The Beijing-Shanghai O-D pair of Beijing-Shanghai HSR was selected, and all daily passenger volumes from Beijing to Shanghai from January 1, 2014 to December 31, 2016 were extracted from 12306 (the national ticketing system in China) for analyzing the characteristics of the daily passenger volumes from perspectives of non-holidays and holidays.

2.1 Characteristics of daily passenger volumes during non-holidays

The daily passenger volumes of Beijing-Shanghai HSR from 2014 to 2016 are shown in Figure 1. According to Figure 1, on an annual basis, except fluctuations on individual days, the annual fluctuations of daily passenger volumes of the three years are roughly the same. On a monthly basis, the daily passenger volumes in the same months in the three years are relatively consistent in terms of change in trend; the daily passenger volumes in January and

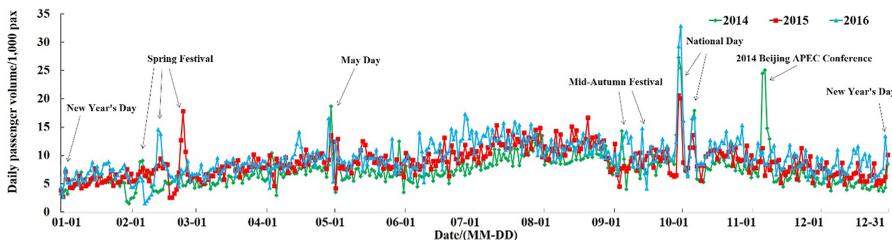


Figure 1.
Daily passenger
volumes of Beijing-
Shanghai HSR from
2014 to 2016

March in each year feature an obvious increasing trend, while the fluctuations in February are relatively large; both July and August are stable at a high quantity, and September has a relatively obvious downward trend compared with August. There is a slight further downward trend in November and December, respectively. On a daily basis, for the same day in each of the three years (for example, the 200th day of each year), the daily passenger volume has a gradual increasing trend with the progress of the year, which indicates that the daily passenger volume fluctuates on an annual basis. It can be concluded that the date attributes such as year, month and day have a periodic impact on daily passenger volume.

Furthermore, the characteristics of daily passenger volumes in each week were analyzed. In China, there are no other statutory holidays in July and August every year, so the rule for the daily passenger volumes of each week in this period is easy to reflect. The daily passenger volumes from Beijing to Shanghai in each week from July to August 2016 were selected for analysis. The change curves for daily passenger volumes from Monday to Sunday each week are shown in Figure 2. It can be seen from the figure that the fluctuation trends of daily passenger volumes in all weeks are very close; the daily passenger volumes in Monday and Tuesday are relatively low; the daily passenger volume in Wednesday features a slight increasing trend, that in Thursday decreases, that in Friday increases possibly because of the approaching weekend, that in Saturday is relatively low, and that in Sunday increases again. It can be concluded that the weekly information also has a periodic impact on the daily passenger volume of HSR.

2.2 Design of date label

To sum up, the four date factors, namely year, month, day and week, have significant impacts on daily passenger volume of HSR, and they will work together to affect the change in daily passenger volumes within a certain number of days. It is difficult to fully represent this impact with simple linear relationships; in other words, it is not easy to fully represent the impact by means of formulas; however, in the medium-term forecast on daily passenger volume of HSR, for any day in the forecast period, the four date factors - year, month, day and week, are defined and can be obtained in advance. Therefore, the four date labels can be set to mark any day, reflecting the dynamic periodic change characteristics of daily passenger volumes in terms of year, month, day and week. See Table 1 for the specific value range of each label. For example, 2014-05-01 is the 121st day of 2014 and Thursday, so the corresponding values of “Year”, “Month”, “Day” and “Week” labels are 2014, 5, 121 and 4, respectively.

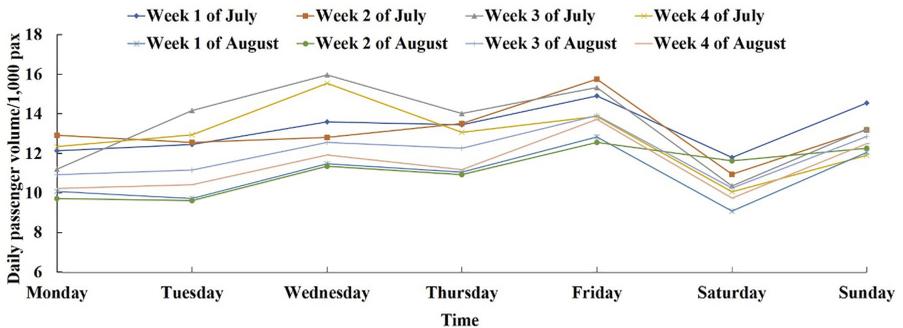


Figure 2. Daily passenger volumes of Beijing-Shanghai HSR in each week from July to August 2016

Values of these 4 date labels can be directly obtained for any day, so the date label is used as definitive data and input into the forecast model. Learning and training are carried out in combination with the historical data of daily passenger volumes for the subsequent forecasts.

2.3 Characteristics of daily passenger volumes during holidays

In addition to the daily dynamic periodic fluctuations, holidays also have a great impact on passenger volume of HSR. It can be seen from Figure 1 that almost all time with large fluctuations in the daily passenger volume falls on holidays. Therefore, the characteristics of daily passenger volumes during holidays will be further analyzed.

First of all, the two-day weekend holiday has a strong periodic impact on passenger volumes of HSR. It can be seen from Figure 2 that it affects not only the passenger volume on holiday, but also the daily passenger volume on Friday. At the same time, the daily passenger volumes are also significantly different between the two days.

In addition to regular weekends, there are also statutory holidays such as the New Year's Day, the Spring Festival, the Tomb-Sweeping Day, the Dragon Boat Festival, the May Day and the Mid-Autumn Festival, and there may be a continuous long holiday taking working days off. See Table 2 for specific statistics on these holidays. Since the daily passenger volumes before and after the holiday will also be affected and fluctuate, the daily passenger volumes of two days before and one day after the holiday are selected for analysis.

For 3-day holidays, for example, the May Day and the Mid-Autumn Festival, the daily passenger volumes of Beijing-Shanghai HSR from 2014 to 2016 are shown in Figures 3 and 4, respectively. The Mid-Autumn Festival in 2015 is only a two-day holiday, which is inconsistent with the holiday in 2014 or 2016; therefore, it is not analyzed together.

From Figures 3 and 4, the fluctuation trends of the daily passenger volumes during May Day holidays in all years are roughly the same; that is, the daily passenger volume in Day 1 before the holiday is at the peak, that in Day 1 of the holiday starts to decrease, that in Day 2 of the holiday is at the lowest point, that in Day 3 of the holiday rises again, and that decreases again in Day 1 after the holiday. The fluctuation trends of the daily passenger volumes during Mid-Autumn Festival holidays in all years are also roughly the same. The change trends of the first five days are very close to those of the May Day, but the difference lies in the last day – the daily passenger volume does not decrease but continues to increase. This indicates that although the day numbers are the same, the impacts of different holidays on daily passenger volume vary.

The daily passenger volumes of Beijing-Shanghai HSR during the Spring Festival and the National Day from 2014 to 2016 are shown in Figures 5 and 6, respectively. For the same holiday, the fluctuation trends of daily passenger volumes of HSR in all years are very close. For the same year, there is a significant difference in the fluctuation of daily passenger volumes of HSR during the two holidays. For the Spring Festival, daily passenger volume of HSR will gradually decrease when the Festival is approaching; the volume on the first day of the Spring Festival is extremely low, and then it gradually rises; the volume is at the peak level on the seventh day and decreases slightly on the first day after the holiday. However, for the National Day, the daily passenger volume begins to rise sharply one day before the

Name	Value
Year	2014, 2015, 2016
Month	1, 2, 3, ..., 12
Day	1, 2, 3, ..., 366
Week	1, 2, 3, ..., 7

Table 1.
Value range of
date label

Holiday type	Number of days/d	Specific time in the Gregorian calendar	Taking working days off or not	Number of days after taking working days off/d
New Year's Day	1	January 1	Yes	3
Spring Festival	3	Variable (Chinese New Year's Eve, the first and second days of the first lunar month; or the first, second and third days of the first lunar month)	Yes	7
Tomb-Sweeping Day	1	April 4 or 5 (the day of the solar term Qingming on lunar calendar)	Yes	3
May Day	1	May 1	Yes	3
Dragon Boat Festival	1	Variable (the 5th day of the 5th lunar month)	Yes	3
Mid-Autumn Festival	1	Variable (the 15th day of the 8th lunar month)	Yes	3
National Day	3	October 1 – October 3	Yes	7

Table 2.
Statistics of statutory holidays in China

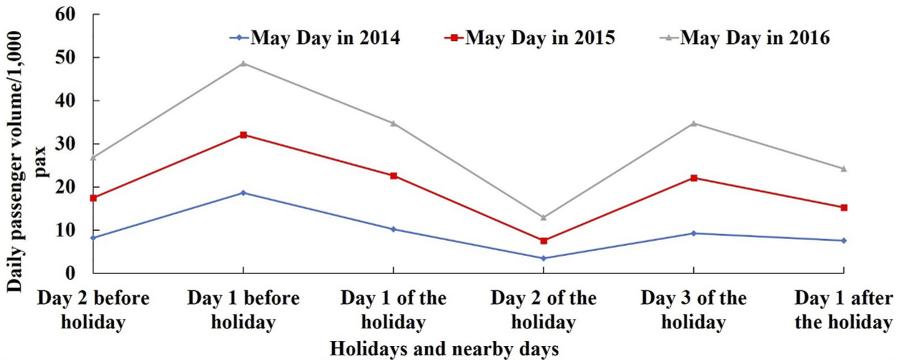


Figure 3.
Daily passenger volumes of Beijing-Shanghai HSR during May Day Holidays in 2014-2016

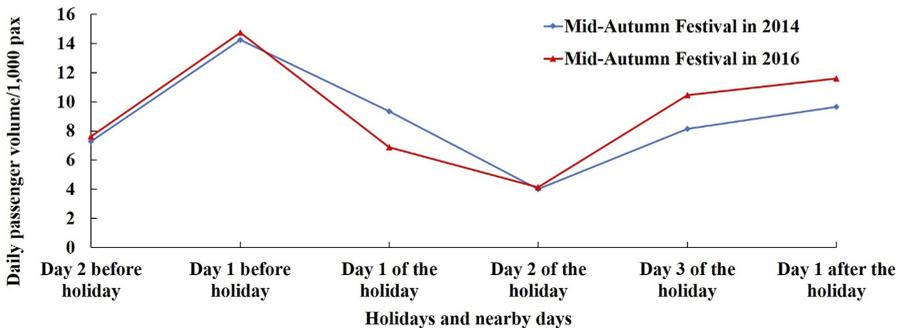


Figure 4.
Daily passenger volumes of Beijing-Shanghai HSR during Mid-Autumn Festival Holidays in 2014-2016

holiday, that on the first day of the holiday is at the peak level, and then that decreases; the daily passenger volumes on the third and fourth days of the holiday are at the lowest point,

and then experience a small peak in the last two days of the holiday. This indicates that although both the Spring Festival and the National Day are seven-day holidays, the impacts of different holidays on daily passenger volume of HSR are completely different.

According to the above analyses, it can be concluded that the impacts of holidays on daily passenger volume of HSR have the following characteristics.

- (1) The impacts on daily passenger volume of HSR are obviously different in case of different day numbers;
- (2) The impacts of the same holiday on daily passenger volume are relatively similar in different years in case of the same day numbers;
- (3) The impacts of holiday types on daily passenger volume vary in case of the same day numbers;
- (4) During the same holiday, daily passenger volumes of all days are obviously different;
- (5) Holiday not only affects daily passenger volume of each day thereof, but also may have a great impact on the daily passenger volume two days before and one day after the holiday.

2.4 Design of holiday label

In the forecast of the daily passenger volumes during holidays, since the four traditional lunar festivals, namely the Spring Festival, the Tomb-Sweeping Day, the Dragon Boat Festival and the Mid-Autumn Festival, are not fixed in the Gregorian calendar every year, and the dates for taking working days off are different every year, it is difficult to periodically capture the

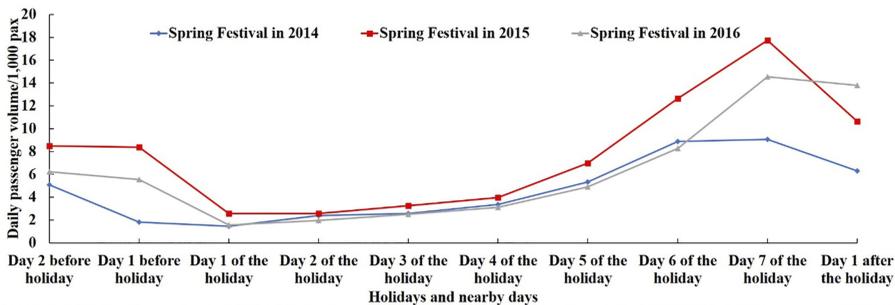


Figure 5.
Daily passenger
volumes of Beijing-
Shanghai HSR during
Spring Festival
Holidays in 2014-2016

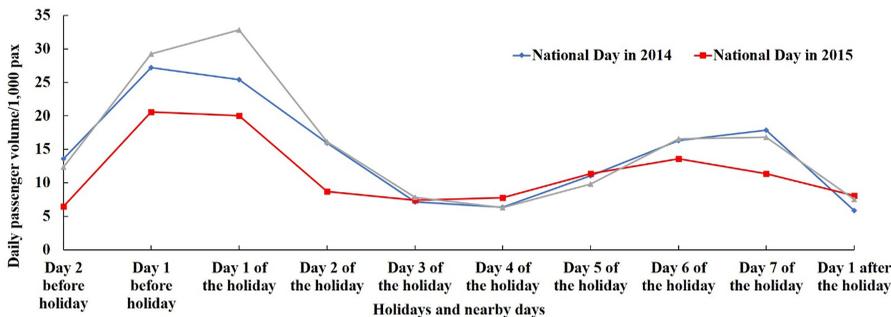


Figure 6.
Daily passenger
volumes of Beijing-
Shanghai HSR during
National Day Holidays
in 2014-2016

impact on daily passenger volume directly from historical passenger flow data and date information, which will affect the accuracy of the forecast. However, in the actual medium-term forecast for 120 days, it can be known in advance whether each day of the forecast period falls on holiday, as well as the types of holidays and day numbers. Therefore, in this paper, it is planned to improve the accuracy of the passenger flow forecast by setting the holiday labels.

The above characteristics (1)–(4) are all the impacts of holidays on their own daily passenger volumes. Therefore, three holiday labels, namely “Day Numbers”, “Holiday Type Corresponding to the Day Numbers” and “Specific Day of the Holiday”, are designed to mark every day. Characteristic (5) reflects the impacts of holidays on the daily passenger volumes on immediately adjacent non-holiday days. Therefore, the label “Impact of Holidays on Nearby Days” is designed to identify the degrees of impacts of various holidays on the daily passenger volumes of one to two days before and after those holidays.

If the daily passenger volume of HSR on any t th day obtained from historical ticket sales data is recorded as $\hat{y}(t)$, its passenger flow change rate β_t can be calculated by the following Equation (1).

$$\beta_t = \frac{\hat{y}(t) - \hat{y}(t-1)}{\hat{y}(t-1)} \quad (1)$$

If the time range of a certain holiday is $[t_1, t_2]$, the change rate β_{t_1-1} of the daily passenger volume $\hat{y}(t_1-1)$ of (t_1-1) d (1 day before the holiday) relative to that $\hat{y}(t_1-2)$ of (t_1-2) d can be obtained by substituting $t = t_1 - 1$ into Equation (1).

The threshold value for the given change rate of the daily passenger volume is assumed to be α . If the absolute value of the change rate is $|\beta_{t_1-1}| \geq \alpha$, it is recognized that the holiday $[t_1, t_2]$ has an impact on the daily passenger volume of (t_1-1) d (one day before the holiday), and the impact degree is β_{t_1-1} ; otherwise, it is considered that the holiday has no impact on the passenger volumes of the days before it, and let $\beta_{t_1-1} = 0$.

In case of $|\beta_{t_1-1}| \geq \alpha$, (i.e. the holiday $[t_1, t_2]$ has an impact on the daily passenger volume of one day before it), let $t = t_1 - 2$, and substitute it into Equation (1) to continue to calculate the change rate β_{t_1-2} of the daily passenger volume of (t_1-2) d. If $|\beta_{t_1-2}| \geq \alpha$, the degree of impact of the holiday on the passenger volumes of two days before it is identified as β_{t_1-2} ; otherwise, let $\beta_{t_1-2} = 0$.

Similarly, by substituting $t = t_2 + 1$ into Equation (1), we can calculate the change rate β_{t_2+1} of the passenger volume of one day after the holiday $[t_1, t_2]$ (i.e. (t_2+1) d). If $|\beta_{t_2+1}| \geq \alpha$, the degree of impact of the holiday $[t_1, t_2]$ on the daily passenger volume of one day after it is identified as β_{t_2+1} ; otherwise, let $\beta_{t_2+1} = 0$.

At this point, the change rates of daily passenger volume of two days before and one day after all holidays can be calculated, and the label value of “Impact of Holidays on Nearby Days” of the corresponding day can be assigned as the passenger flow change rate β . For other days not in the vicinity of holidays, let the label value of “Impact of Holidays on Nearby Days” be 0.

Therefore, the four holiday labels – “Day Numbers”, “Holiday Type Corresponding to the Day Numbers”, “Specific Day of the Holiday” and “Impact of Holidays on Nearby Days”, are set according to the characteristics of the impact of holidays on the daily passenger volume; these labels and the corresponding value ranges are shown in Table 3. The value rule of the “Holiday Type Corresponding to the Day Numbers” label is as follows: in case of the same day numbers, natural integer labels are given in order of holidays. For example, the New Year’s Day, the Tomb-Sweeping Day, the May Day, the Dragon Boat Festival and the Mid-Autumn Festival are all three-day holidays, so the corresponding values of the “Holiday Type

Holiday type	Day numbers/d	Label name and corresponding value			Impact of holidays on nearby days
		Holiday type corresponding to the day numbers	Specific day of the holiday		
Non-holidays	0	0	0	β or 0	
All 1-day holidays	1	1	1	0	
Weekends and other 2-day holidays	2	1	1, 2	0	
New Year's Day	3	1	1, 2, 3	0	
Tomb-Sweeping Day	3	2	1, 2, 3	0	
May Day	3	3	1, 2, 3	0	
Dragon Boat Festival	3	4	1, 2, 3	0	
Mid-Autumn Festival	3	5	1, 2, 3	0	
National Day	7	1	1, 2, 3, 4, 5, 6, 7	0	
Spring Festival	7	2	1, 2, 3, 4, 5, 6, 7	0	

Table 3.
Value range of holiday label

Corresponding to the “Day Numbers” label are 1, 2, 3, 4 and 5, respectively. In addition, in some years, the New Year’s Day or the Dragon Boat Festival falls on a one-day holiday, which is classified into the same type. Similarly, the Mid-Autumn Festival may fall on a two-day holiday, so the two-day holiday and the weekend are classified into the same type.

According to Table 3, holidays can be marked for any day in each year. For example, May 1 is the first day of the three-day holiday, the May Day falls into Type 3 in the category of the three-day holiday, and the first day falls on holiday and is not a nearby day, so the values of “Day Numbers”, “Holiday Type Corresponding to the Day Numbers”, “Specific Day of the Holiday” and “Impact of Holidays on Nearby Days” labels of the day are 3, 3, 1 and 0, respectively. In the medium-term forecast of the daily passenger volume, holiday labels of each day in the forecast period can be calibrated in advance and input into the forecast model as definitive data.

3. Forecast model and method

3.1 Forecast model

Daily passenger volume of HSR changes dynamically with the change of date (Luo, 2020). On the one hand, it is clear that the daily passenger volume of the last few days will have an impact on the daily passenger volume(s) of the coming day(s), that is, the daily passenger volume features autoregression to some extent; on the other hand, the date and holiday attributes of each day of the forecast period will also have an impact on daily passenger volume.

In order to reflect the impacts of the above two aspects at the same time, the DLP-WNN model was designed for the medium-term forecast of daily passenger volume of HSR. The model includes two parallel neural networks, namely subnet 1 and subnet 2, as shown in Figure 7. ζ_{Bias} is a smaller constant term; f is the neuronal transfer function. In this paper, the Morlet mother wavelet basis function (Zhang & Yang, 2015) is used, and its functional expression is shown in Equation (2). Each ellipse represents 1 neuron; $n_{1,1}$, $n_{2,1}$ and $n_{3,1}$ are respectively the number of neurons in layers 1-3 of subnet 1, while $n_{1,2}$, $n_{2,2}$ and $n_{3,2}$ are respectively the number of neurons in layers 1-3 of subnet 2. The dotted line connecting arrows between the neurons in each layer represent the weighted summation between them,

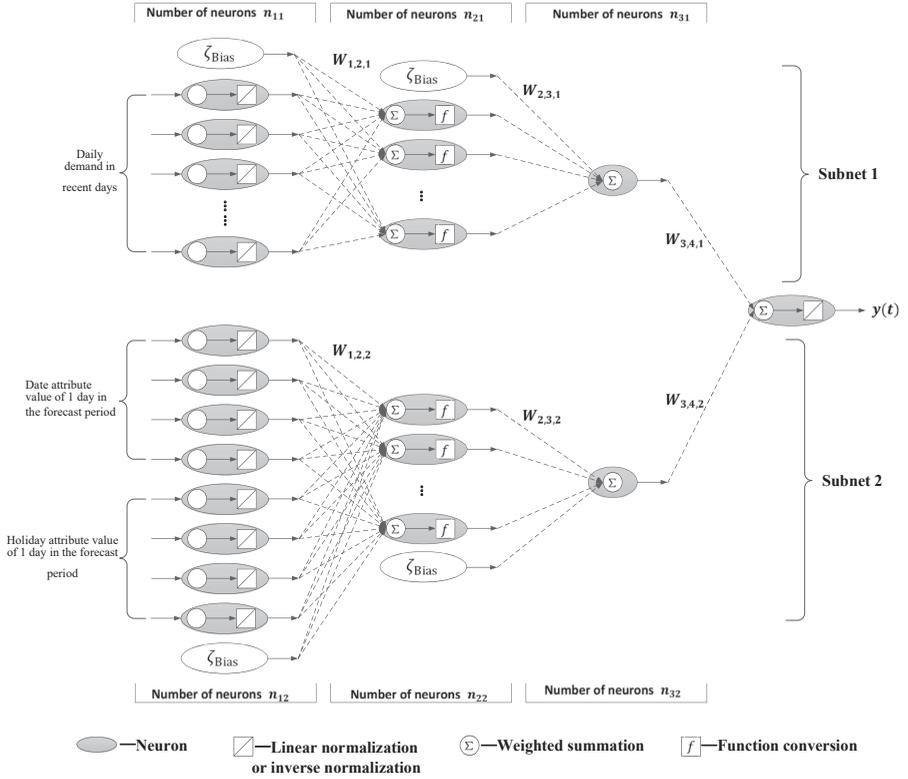


Figure 7. DLP-WNN model for medium-term forecast of daily passenger volume

and each dotted line corresponds to one weight parameter; for example, $w_{1,2,1}$ represents the weight parameter matrix of neurons in layers 1-2 in subnet 1, and the rest are similar.

$$f(x) = \cos(1.75x)e^{-\frac{x^2}{2}} \quad (2)$$

where x is the neuron input data.

Subnet 1 reflects the autoregressive characteristics of the daily passenger volume, that is, the passenger volume on the day of the forecast is affected by the daily passenger volumes of the previous days. Here, by referring to the processing methods of Tsai *et al.* (2009), the data is input by moving the fixed data window, that is, with the progress of the date, the daily passenger volume data of the latest $n_{1,1} - 1$ day is fixedly input every time in layer 1 of subnet 1. Subnet 2 reflects the impact of the date attribute and holiday attribute of the forecast day on the daily passenger volume. The input layer of layer 1 consists of four date label values, four holiday label values and one ζ_{Bias} item corresponding to the day. The output (i.e. the forecasted value of the one-day passenger volume) of the whole network is obtained by weighted summation of the output values of the two subnets. The medium-term forecast result of the daily passenger volume can be obtained by continuously forecasting for 120 days with the aid of this model.

The forecast process of the daily passenger volume in the medium term is designed according to the DLP-WNN model in Figure 7, as shown in Figure 8. The input value of

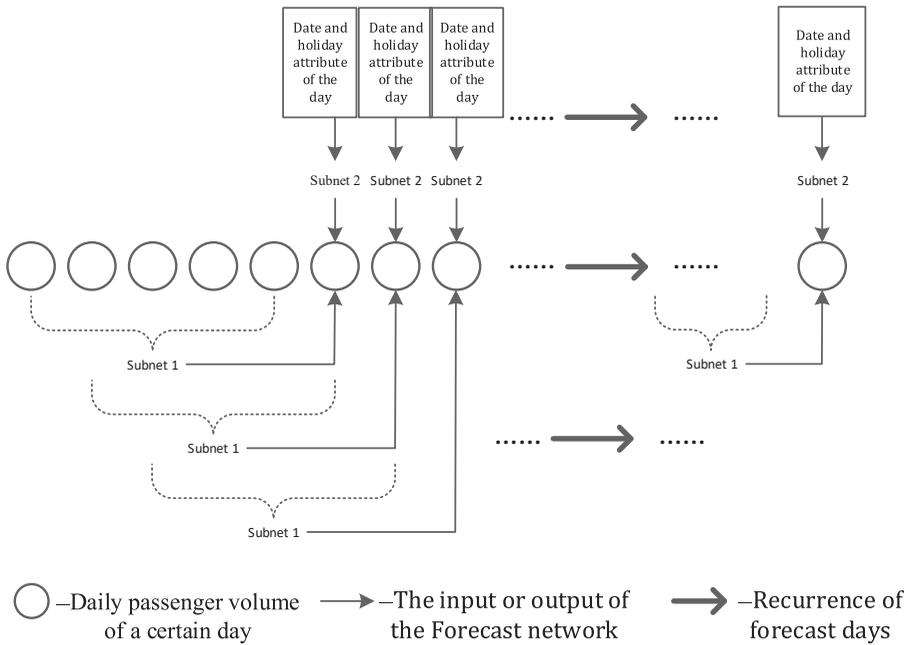


Figure 8. Medium-term forecast process of daily passenger volume based on DLP-WNN model

subnet 1 is the passenger volumes of few days before the forecast day, and on the first day of forecast, the input value is the historical daily passenger volume data; as the forecast days progress, the subsequent input values are gradually updated to the forecast results in the previous days; therefore, the input of subnet 1 can be regarded as forecast data. If the forecast is continued with forecast data, the error accumulation will occur, which will affect the forecast accuracy. This is also a common problem when the short-term forecast method is applied to the medium-term forecast. In order to overcome this error accumulation, two subnets are designed in this paper, and the output of subnet 1 is corrected by subnet 2. Since the input values of subnet 2 are the time attribute label value and the holiday attribute label value of each forecast day, these attribute label values are defined and can be calibrated in advance, being the definitive input data. After learning the historical passenger flow data, the definitive data comprehensively reflects the periodic law of the regular daily passenger volume (time attribute) and the sudden change characteristics of holidays (holiday attribute); therefore, the obtained forecast results can be used to correct the error generated by subnet 1. The forecast results of each day can be obtained by weighted summation of the daily output values of the two subnets. The results can not only continue to reflect the trend of the daily passenger volume, but also reveal the difference of the passenger volumes among different days, especially on holidays, which is helpful to ensure the forecast accuracy.

3.2 Model principle and process

(1) Forward forecast of network

In the DLP-WNN model, the two subnets are of wavelet neural network structure and designed with the wavelet basis function as the transfer function of the middle layer, and

perform the error back propagation while carrying out the forward propagation of signal based on the design idea of the BP neural network. In order to simplify the expression, the forecast principle and operation process are illustrated by taking subnet 1 as an example.

If, in subnet 1, the input data of all neurons in the input layer of layer 1 is linearly normalized to $x_{1,1}(i), i = 1, 2, \dots, n_{1,1}$, the data of all neurons in layer 1 is weighted to obtain the neurons $h_{2,1}(j), j = 1, 2, \dots, n_{2,1}$ in layer 2. The calculation formula is as follows:

$$h_{2,1}(j) = \frac{\left(\sum_{i=1}^{n_{1,1}} w_{1,2,1}(i,j)x_{1,1}(i) \right) - b_j}{a_j} \quad (3)$$

where $h_{2,1}(j)$ is the value of the j -th node in layer 2 of subnet 1; $w_{1,2,1}(i,j)$ is the connection weight between the i -th neuron of layer 1 and the j -th neuron of layer 2 in subnet 1; b_j is the translation factor of the wavelet basis function; a_j is the expansion factor of the wavelet basis function.

All neurons $h_{2,1}(j)$ in layer 2 of subnet 1 obtain the output values $g_{2,1}(j)$ after being subject to the transfer by the wavelet basis function, and the calculation formula is as follows:

$$g_{2,1}(j) = f(h_{2,1}(j)) \quad (4)$$

Then, the calculation formula for all neurons $h_{3,1}(k), k = 1, 2, \dots, n_{3,1}$ in layer 3 of subnet 1 is as follows:

$$h_{3,1}(k) = \sum_{j=1}^{n_{2,1}} w_{2,3,1}(j,k)g_{2,1}(j) = \sum_{j=1}^{n_{2,1}} w_{2,3,1}(j,k)f\left(\frac{\left(\sum_{i=1}^{n_{1,1}} w_{1,2,1}(i,j)x_{1,1}(i)\right) - b_j}{a_j}\right) \quad (5)$$

where $h_{3,1}(k)$ is the value of the k th node in layer 3 of subnet 1; $w_{2,3,1}(j,k)$ is the connection weight between the j th neuron of layer 2 and the k th neuron of layer 3 in subnet 1.

Then, for the entire DLP-WNN network, the forecast values of subnet 1 and subnet 2 for each day are subject to weighted summation to obtain the forecast outputs for each day. For the forecast output $y(t)$ of the daily passenger volume on any t -th day, the calculation formula is as follows:

$$y(t) = \sum_{k=1}^{n_{3,1}} w_{3,4,1}(k)h_{3,1}(k) + \sum_{k=1}^{n_{3,2}} w_{3,4,2}(k)h_{3,2}(k) \quad (6)$$

(2) Network weight correction

In addition to the forward forecast for the whole network, the weight correction of the network needs to be carried out by using the error back propagation. In the correction process, the gradient descent method is used to correct the network weight, so that the network output results continuously approach the desired output.

If the actual daily passenger volume of the t th day is recorded as $\hat{y}(t)$, the calculation formula of the forecast accuracy error E is as follows:

Medium-term
forecast of HSR
passenger
volume

$$E = \frac{1}{2} \sum_{t=1}^M (\hat{y}(t) - y(t))^2 \quad (7)$$

where M is the total days of data training, d.

Therefore, the calculation formula of network weight correction is as follows:

$$w_{3,4,1}^{\mu+1}(k) = w_{3,4,1}^{\mu}(k) - \eta \frac{\partial E}{\partial w_{3,4,1}^{\mu}(k)} \quad (8)$$

where μ is the training times, Nr.

$$w_{2,3,1}^{\mu+1}(j, k) = w_{2,3,1}^{\mu}(j, k) - \eta \frac{\partial E}{\partial w_{2,3,1}^{\mu}(j, k)} \quad (9)$$

$$w_{1,2,1}^{\mu+1}(i, j) = w_{1,2,1}^{\mu}(i, j) - \eta \frac{\partial E}{\partial w_{1,2,1}^{\mu}(i, j)} \quad (10)$$

where η is the learning rate.

The partial derivative of the calculation error E for the weight of each layer is as follows:

$$\frac{\partial E}{\partial w_{3,4,1}^{\mu}(k)} = \frac{\partial E}{\partial y} \frac{\partial y}{\partial w_{3,4,1}^{\mu}(k)} = - \sum_t (\hat{y}(t) - y(t)) h_{3,1}(k) \quad (11)$$

$$\frac{\partial E}{\partial w_{2,3,1}^{\mu}(j, k)} = \frac{\partial E}{\partial y} \frac{\partial y}{\partial h_{3,1}(k)} \frac{\partial h_{3,1}(k)}{\partial w_{2,3,1}^{\mu}(j, k)} = - \sum_t (\hat{y}(t) - y(t)) w_{3,4,1}(k) g_{2,1}(j) \quad (12)$$

$$\begin{aligned} \frac{\partial E}{\partial w_{1,2,1}^{\mu}(i, j)} &= \frac{\partial E}{\partial y} \left(\sum_{k=1}^{n_{3,1}} \frac{\partial y}{\partial h_{3,1}(k)} \frac{\partial h_{3,1}(k)}{\partial g_{2,1}(j)} \times \frac{\partial g_{2,1}(j)}{\partial h_{2,1}(j)} \frac{\partial h_{2,1}(j)}{\partial w_{1,2,1}^{\mu}(i, j)} \right) \\ &= - \sum_t (\hat{y}(t) - y(t)) \left[\sum_{k=1}^{n_{3,1}} w_{3,4,1} \times (k) w_{2,3,1}(j, k) f'(h_{2,1}(j)) \frac{x_{1,1}(i)}{a_j} \right] \end{aligned} \quad (13)$$

where

$$\begin{aligned} f'(x) &= -1.75 \sin(1.75x) e^{-\frac{x^2}{2}} - \\ & x \cos(1.75x) e^{-\frac{x^2}{2}} \end{aligned}$$

where $f'(h_{2,1}(j))$ is the derivative of the Morlet mother wavelet basis function.

After being trained, the network can be used to forecast for $N = 120$ consecutive days, and then the average absolute percentage error M_{MAPE} and the root mean square error R_{RMSE} are used to evaluate the forecast error of N d. The calculation formulas are as follows:

$$M_{MAPE} = \frac{1}{N} \sum_{t=1}^N \frac{|y(t) - \hat{y}(t)|}{y(t)} \tag{14}$$

$$R_{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (y(t) - \hat{y}(t))^2} \tag{15}$$

4. Example analysis

4.1 Data input

Beijing-Shanghai HSR has a total length of 1,318 km, with 24 stations along the line. Typical O-D pairs at short, medium, medium-long and long distances were selected for forecast analysis as in [Table 4](#).

For the O-D pairs in respect of the above four types of distance, the HSR ticketing data from January 1, 2014 to December 31, 2016 was analyzed. The daily passenger volume data from January 1, 2014 to September 2, 2016 was used as the training set, and the DLP-WNN model was adopted for training; the daily passenger volumes of 120 days (from September 3, 2016 to December 31, 2016) were used as the test set, and the errors between the forecast results and the actual values were analyzed and compared to evaluate the rationality and effectiveness of the forecast method proposed in this paper.

4.2 Forecast effect

The medium-term forecast results of daily passenger volumes for typical O-D pairs at 4 different distances are shown in [Figure 9](#). According to the figure, except very few points, in the continuous forecast of 120 days, the errors between the forecasted daily passenger volumes and the actual values of the four O-D pairs are relatively small in general. The mean absolute percentage error M_{MAPE} and the root mean square error R_{RMSE} of the compiled forecast result are shown in [Table 5](#). It can be seen from the error statistics in the table that the values of M_{MAPE} of the medium-term forecast for 120 days on daily passenger volumes of the four O-D pairs are between 7% and 12%. [Figure 4](#) and [Table 5](#) show that the DLP-WNN model results in high accuracy and positive effects for the medium-term forecast for 120 days on daily passenger volumes.

To further test the forecast effect, the DLP-WNN forecast model was compared with other methods. Forecast methods such as the BP neural network, the ELM (extreme learning machine), the ELMAN neural network, the GRNN (generalized regression neural network) and the VMD-GA-BP (variational mode decomposition-genetic algorithm-BP neural network) ([Shi, Yang, Hu, Xu, & Wu, 2019](#)) were used to perform the medium-term forecast for 120 days on daily passenger volumes of typical O-D pairs at the above four types of distance, respectively. The effect comparison is shown in [Figure 10](#).

Table 4. Selection of typical O-D pairs at different distances

O-D pair type (distance)	Typical O-D pair	Actual length/km
Short distance (about 200 km)	Beijing-Cangzhou	210
Medium distance (about 500 km)	Beijing-Qufu	535
Medium-long distance (about 1,000 km)	Beijing-Nanjing	1,023
Long distance (over 1,300 km)	Beijing-Shanghai	1,318

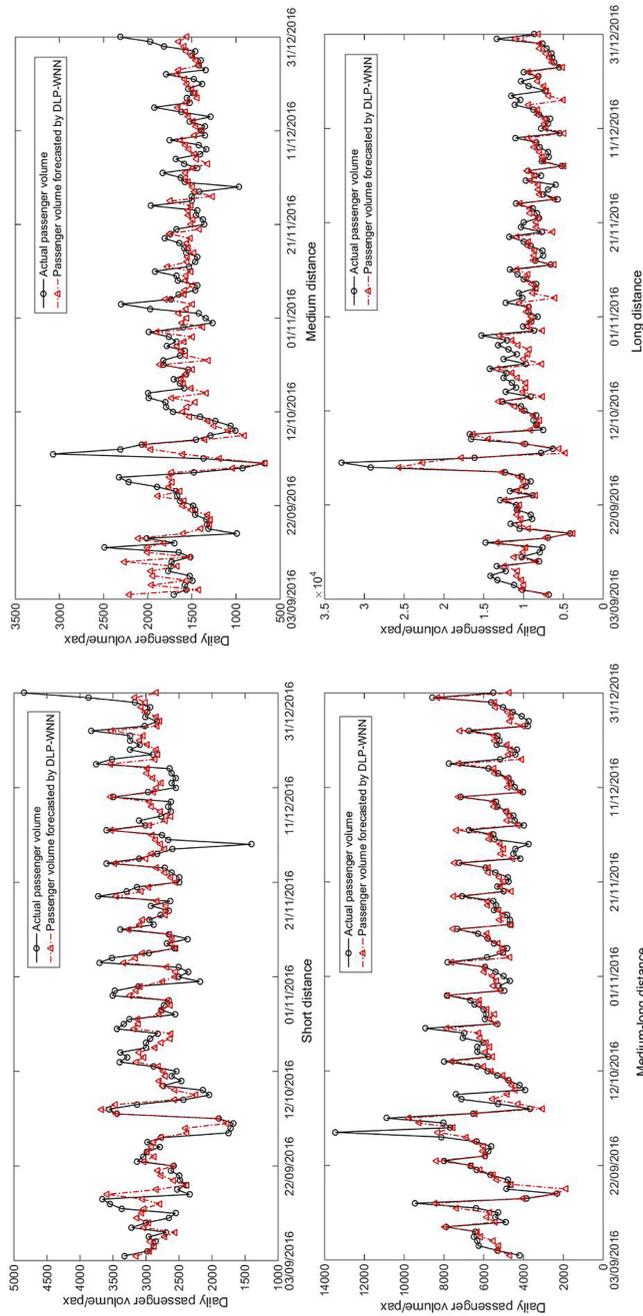


Figure 9.
Medium-term forecast
results for typical O-D
pairs at four distances

From Figure 10 that, the DLP-WNN forecast method has the best medium-term forecast effect for typical O-D pairs at the four types of distance. For the VMD-GA-BP method, it works well in the first few days of the forecast period; however, with the increase in forecast days, it results in the largest forecast deviation. This indicates that this method is only suitable for the short-term forecast, not the medium-term forecast. The mean absolute percentage errors M_{MAPE} and the root mean square errors R_{RMSE} of the results of the medium-term forecast for 120 days on daily passenger volumes for the four O-D pairs with different forecast methods are shown in Tables 6 and 7, respectively. It can also be seen that the DLP-WNN forecast method has the smallest error.

The above analysis shows that the DLP-WNN model is suitable for the medium-term forecast of the daily passenger volume of HSR.

Table 5.
Analysis on medium-term forecast errors of different types of O-D pairs

O-D pair type	$M_{MAPE}/\%$	R_{RMSE}/pax
Short distance	8.48	355.4
Medium distance	11.65	282.1
Medium-long distance	7.96	779.5
Long distance	10.94	1724.6

Table 6.
Comparison of values of M_{MAPE} for medium-term forecast for 120 days on daily passenger volumes with different forecast methods

O-D pair type	M_{MAPE} value/ $\%$					
	DLP-WNN	BP	ELM	ELMAN	GRNN	VMD-GA-BP
Short distance	8.48	10.98	14.08	13.58	13.51	146.60
Medium distance	11.65	13.77	15.75	16.05	16.55	367.15
Medium-long distance	7.96	14.82	18.32	17.39	14.57	129.23
Long distance	10.94	18.06	20.46	21.77	27.73	200.13

Table 7.
Comparison of values of R_{RMSE} for medium-term forecast for 120 days on daily passenger volumes with different forecast methods

O-D pair type	R_{RMSE} value/pax					
	DLP-WNN	BP	ELM	ELMAN	GRNN	VMD-GA-BP
Short distance	355.4	438.9	503.2	488.3	515.1	4207.4
Medium distance	282.1	303.3	350.4	350.4	353.6	5812.7
Medium-long distance	779.5	1432.0	1591.1	1593.8	1309.0	7779.5
Long distance	1724.6	2642.5	3411.9	3487.7	4630.2	20431.4

5. Conclusions

Based on all the daily passenger volume data from Beijing to Shanghai from 2014 to 2016, this paper first analyzed the characteristics of daily passenger volume of HSR during non-holidays and holidays, identifying the dynamic periodic characteristics of daily passenger volume changes and the impact characteristics of holiday attributes respectively. Based on those characteristics, the date and holiday labels were designed correspondingly. In accordance with the autoregressive characteristics of daily passenger volume of HSR, the Double Layer Parallel Wavelet Neural Network (DLP-WNN) model suitable for the medium-term forecast of the daily passenger volume of HSR was established. The specific forecast output of the daily passenger volume was obtained by weighted summation of the output

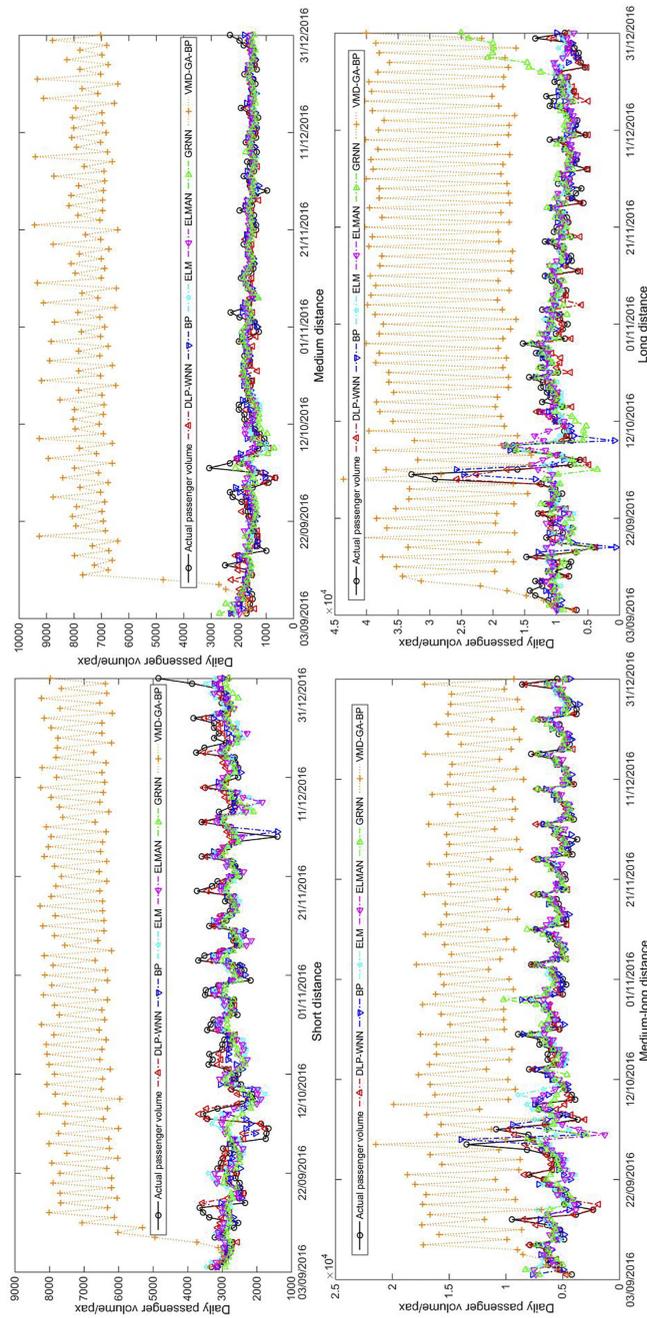


Figure 10.
Effect comparison of
medium-term forecast
for 120 days on daily
passenger volumes for
four O-D pairs with
different forecast
methods

values of two parallel wavelet neural networks. Subnet 1 reflects the impact of daily passenger volumes of a few days before the forecast day and the passenger volume of the forecast day. Subnet 2 reflects the impact of the time attribute and holiday attribute on the forecast day on the passenger volume of the day. The typical O-D pairs at short, medium, medium-long and long distances on Beijing-Shanghai HSR were taken as examples, the DLP-WNN model was used to perform the medium-term forecast for 120 days on daily passenger volumes, and the forecast errors were analyzed and compared in respect of other five forecast methods, reflecting the rationality and effectiveness of the medium-term forecast method proposed in this paper.

It should be noted that the forecast model established in this paper performs the feature learning and forecast based on historical passenger volume data, and its implicit precondition is that the past situation will remain the same in the future; that is, under the premise of no drastic changes in the HSR transportation market, the medium-term forecast method proposed in this paper has beneficial effect. However, in case of extreme situations such as epidemic outbreaks or passenger flow control, the forecast effect will be affected when the historical data does not contain the corresponding type characteristics for the time being. Of course, if the characteristics of passenger flow during the epidemic period can be further analyzed, the corresponding appropriate labels would be designed to establish a forecast model with reference to the ideas of this paper. This will not be discussed in detail here for the time being and may be included in prospective research.

References

- Böerjesson, M. (2014). Forecasting demand for high speed rail. *Transportation Research Part A: Policy and Practice*, 70, 81–92.
- Jiang, X. S., Zhang, L., & Chen, X. (2014). Short-term forecasting of high speed rail demand: A hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China. *Transportation Research Part C: Emerging Technologies*, 44, 110–127.
- Liang, Q., Xu, X., & Liu, L. (2020). Data-driven short-term passenger flow prediction model for urban rail transit. *China Railway Science*, 41(4), 153–162. (in Chinese).
- Liu, H., Tian, H., Li, Y., & Zhang, L. (2016). Study on performance comparison of wind speed hybrid high-precision one-step predicting models along railways. *Journal of the China Railway Society*, 38(8), 41–49. (in Chinese).
- Luo, M. (2020). Research and application of passenger flow analysis for railway passenger transport big data. *China Railway*, 4, 105–110. (in Chinese).
- Shi, F., Yang, X., Hu, X., Xu, G., & Wu, R. (2019). A VMD-GA-BP method for predicting non-holiday passenger flow of high speed railway based on data replacement correction. *China Railway Science*, 40(3), 129–136. (in Chinese).
- Teng, J., & Li, J. (2020). Short-term forecast method for intercity railway passenger flow considering date attributes and weather factors. *China Railway Science*, 41(5), 136–144. (in Chinese).
- Tsai, T. H., Lee, C. K., & Wei, C. H. (2009). Neural network based temporal feature models for short-term railway passenger demand forecasting. *Expert Systems with Applications*, 36(2), 3728–3736.
- Wang, Y., Wang, Z., Jia, L., & Qin, Y. (2004). Study on prediction method of data mining of the passenger traffic volume of railways and its application. *Journal of the China Railway Society*, 26(5), 1–7. (in Chinese).
- Wang, J., Liu, C., Shan, X., & Zhu, J. (2010). Prediction of the railway passenger traffic volume based on Bi-level orthogonalization neural network model. *China Railway Science*, 31(3), 126–132. (in Chinese).

- Wei, Y., & Chen, M. C. (2012). Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C: Emerging Technologies*, 21(1), 148–162.
- Yang, J., & Hou, Z. (2013). A wavelet analysis based LS-SVM rail transit passenger flow prediction method. *China Railway Science*, 34(3), 122–127. (in Chinese).
- Yao, E., Zhou, W., & Zhang, Y. (2018). Real-time forecast of entrance and exit passenger flow for newly opened station of urban rail transit at initial stage. *China Railway Science*, 39(2), 119–127. (in Chinese).
- Zhang, W., & Yang, J. (2015). Dynamic demand forecast of maritime emergency response resources based on wavelet neural network. *Operations Research and Management Science*, 24(4), 198–205. (in Chinese).

Further reading

- Wang, Z., & Wang, Q. (2013). Seasonal adjustment model of China railway monthly passenger traffic volume based on spring festival factors. *Journal of the China Railway Society*, 35(7), 9–13, (in Chinese).

Corresponding author

Guangming Xu can be contacted at: xuguangming@csu.edu.cn