

Research Article

A Hybrid Short-Term Forecasting Model of Passenger Flow on High-Speed Rail considering the Impact of Train Service Frequency

Qingying Lai, Jun Liu, Yongji Luo, and Minshu Ma

State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Jun Liu; jliu@bjtu.edu.cn

Received 8 August 2017; Revised 25 October 2017; Accepted 4 December 2017; Published 20 December 2017

Academic Editor: Ivan Giorgio

Copyright © 2017 Qingying Lai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Short-term forecasting of OD (origin to destination) passenger flow on high-speed rail (HSR) is one of the critical tasks in rail traffic management. This paper proposes a hybrid model to explore the impact of the train service frequency (TSF) of the HSR on the passenger flow. The model is composed of two parts. One is the Holt-Winters model, which takes advantage of time series characteristics of passenger flow. The other part considers the changes of TSF for the OD in different time during a day. The two models are integrated by the minimum absolute value method to generate the final hybrid model. The operational data of Beijing-Shanghai high-speed railway from 2012 to 2016 are used to verify the effectiveness of the model. In addition to the forecasting ability, with a definite formation, the proposed model can be further used to forecast the effects of the TSF.

1. Introduction

As of September 10, 2016, the operating length of China's HSR has exceeded 20,000 km (the HSR in China refers to the train services that have an average speed of 200 km/h or higher. The HSR network consists of upgraded conventional railways and newly built HSR lines). The HSR provides a new choice of transport for some people who used to travel by airline or the highway. The income of the HSR mainly comes from two parts of the passenger flow, one is the newly generated HSR passenger traffic, and the other one is the passenger who transforms from other modes of transport. Therefore, the accurate short-term forecasting of OD passenger flow on HSR is significantly important because (i) it provides the basis for the planning and enhancement of the railway network; (ii) it is the fundament for the investment and construction of HSR; (iii) it affects the revenue management, technical specifications, operation mode, and facility improvement on HSR line [1].

In the past decades, a lot of attention has been paid to the short-term forecasting. These models can be generally divided into three categories: time series models, causal models, and hybrid models.

Firstly, time series models are the functions in which the traffic flow is modeled by the observed values. In general, time series models mainly include autoregressive integrated moving average (ARIMA) [2–4] and exponential smoothing method [5–9]. The ARIMA is a linear combination of time-lagged variables and error terms. It has been widely used in the forecast since the 1970s because it has good performance in modeling linear and stationary time series. The exponential smoothing method includes one exponential smoothing method, quadratic exponential smoothing method, and triple exponential smoothing. The triple exponential smoothing method is also known as the Holt-Winters method, which was first suggested by Holt's student, Winters, in 1960 after reading a signal processing book from the 1940s on exponential smoothing [10], and the model takes into account seasonal changes as well as trends. For ARIMA model and Holt-Winters method, there has not been a consensus as to whether one dominates the other. In different cases, the most appropriate method may be different [11–14]. However, Holt-Winters method may have some advantages in forecasting long-term series because it can use recent data into forecast to update model parameters and improve the accuracy [11].

Second, causal models are that the traffic flow is modeled as a function of some exogenous factors or endogenous factors. The model which considers the effects of multi-influencing factors can greatly enhance the flexibility of the forecasting model. By describing the relationship between transportation capacity, passenger volume, and quantity demanded, Luo et al. [15] construct the passenger volume forecasting model of HSR line, but their work concentrates on the passenger volume of the whole line; thus, it cannot be used to forecast the passenger volume in OD pairs. For the operator, the knowledge of the passenger volume of each OD pair is sometimes more important than a total volume of the railway line. Du et al. [16] considered the influencing factors of HSR passenger volume, but it cannot be applied to the forecast of the DPF due to the limited influence of some factors because some factors cannot be statistics every day. Wardman [17] proposed a forecasting model for railway traffic flow using external factors including GDP, variations in times, and fuel costs. Moreover, some scholars have also conducted other related studies [18–20]. Although the multifactor model is considered to better deal with a timely manner when the external environment changes, the description of the periodic variation in the predicted data is less than the time series model.

Third, hybrid models are the well-established and well-tested approach to improving the forecasting accuracy [21]. For instance, Zeng et al. proposed a hybrid model that combines multilayer artificial neural network and ARIMA [22], and Xu et al. combined genetic algorithms and a grey SVM in the forecasting model [23]. Wei and Chen [24] forecasted the short-term metro passenger flow with a hybrid method, in which empirical mode decomposition (EMD) is done first to extract the features of flow pattern and secondly the neural network is used to forecast the passenger flow based on these extracted features. In other research domains, several studies (such as those of [21, 25, 26]) obtained the result that hybrid forecasts have generally been shown to outperform the forecasts from the single prediction model.

In this paper, we construct a hybrid forecasting model to forecast the daily passenger flow (DPF) by taking the advantages of the time series model and the causal model considering the impact of the TSF. The remainder of this article is organized as follows. In Section 2, features of the TSF on HSR are introduced and the temporal features of OD passenger flow including trend feature, the day of the week feature, and month-of-year feature are discussed, respectively. In Section 3, the Holt-Winters time series model and the linear regression model considering the TSF is constructed, and the hybrid model is built by using the minimum absolute value method. The accuracy of the model is verified by the historical operation data onto the Beijing-Shanghai HSR in Section 4. Finally, Section 5 gives the conclusion of this paper.

2. Daily Passenger Flow Analysis

In order to have a prior knowledge of the DPF on the HSR, the features of the TSF and the temporal features of OD passenger will be introduced here. For different OD markets

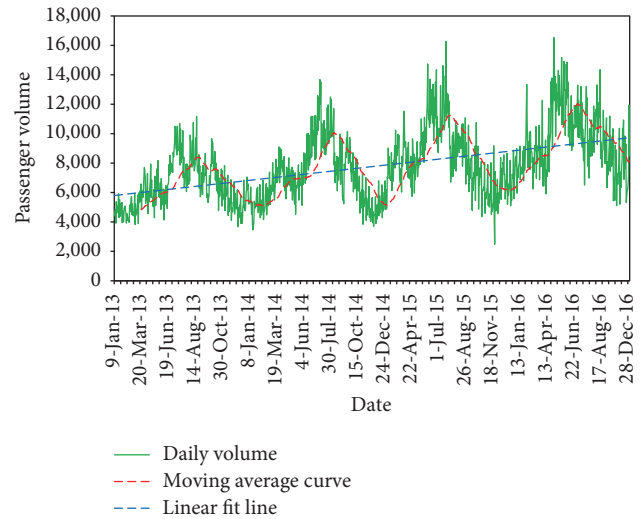


FIGURE 1: OD's (Beijing South Station to Shanghai Hongqiao Station) daily passenger flow.

on the HSR line, the patterns of their DPF might be different. However, there is no doubt that the DPF of any particular OD market creates some regularities over time. In this section, we focus on the OD market of Beijing South Station to Shanghai Hongqiao Station on Beijing-Shanghai HSR line. This HSR line is studied because it is a representative line in China's HSR network. In the first subsection, we will analyze the time series characteristics of the DPF of this OD, including the periodic changes of DPF in weeks, months, and years. In addition, the second subsection illustrates that the train service frequency (TSF) also puts obvious influences on the OD's DPF.

2.1. Time Series Characteristics. Figure 1 shows the flow changes of Beijing South Station to the Shanghai Hongqiao Station from January 2013 to December 2016. In order to examine the long-term trend of the OD's DPF, we calculate a linear regression of the passenger flows of time. The result is demonstrated by the blue dash line, which is similar to the upward trend of HSRs in other countries since the commercial operation. It is believed that the DPF will keep this growing tendency in the following years.

Then, we use the moving average method to generate the red dash line. It is found that the whole line can be divided into four similar parts, and each part corresponds to one year. Hence, a yearly regular pattern of the change of DPF exists.

Moreover, in Figure 2, we compared the DPF of three different weeks, and there has similar change trend in each week. In a cycle of the week, the passenger flow of Monday, Tuesday, Wednesday, and Thursday is relatively lower. On Friday, the flow reaches the highest point as many business people would go back home and tourists go for travel before the weekend. Saturday is another low point, and Sunday is the return date of those business people and tourists traveling on Friday.

From the above analysis, we can find that there exist the yearly, monthly, and weekly periodic characteristics of the

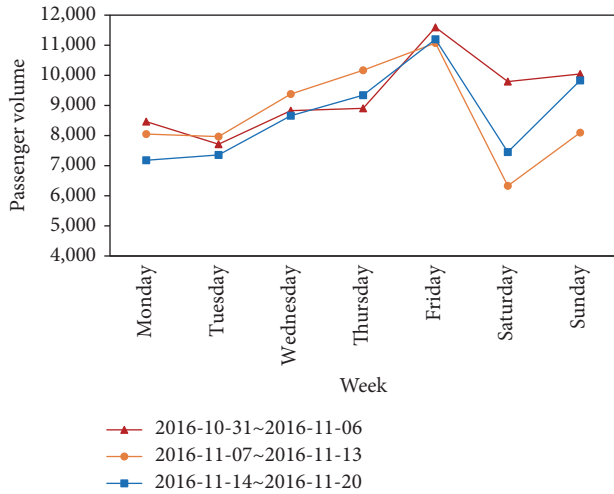


FIGURE 2: The DPF from Beijing South Station to Shanghai Hongqiao Station in different weeks.

OD's DPF. In order to describe the time characteristics of passenger flow changes, this time series data will be modeled using the Holt-Winters model in the next section.

2.2. Analysis of Train Service Frequency Impact. China Railway Corporation (CR) modifies the train timetable at least twice per year, typically in January and July. In each modification, three types of new timetable are generated, namely, normal timetable, weekend timetable, and peak timetable, which are applied during Monday and Thursday, during Friday to Sunday, and in the festival (i.e., the National Day and the spring festival) respectively. Generally, the number of trains in different types of timetable is different. The peak timetable is the most saturated one, followed by the weekend timetable, which is still more saturated compared to the normal one.

As an illustration, Figure 3 shows the DPF and TSF over time; the period in Figure 3 is one day. From the figure, we can see that passenger flow changes when the TSF will show the same trend of change (increase or decrease), which shows the relevance of passenger changes and TSF changes exists. However, when the TSF remains unchanged, the passenger flow is still indicating a trend, pointing out that although TSF can affect the passenger, the cyclical changes in the passenger flow are still the main factor.

From above, the change of TSF may influence the daily passenger flow. As Figure 3, although the daily OD's TSFs are different, the change rate of daily TSF is not obvious. In order to show the changes of OD's TSF in the different time interval during a day, Figure 4 shows the OD's (Beijing South Station to Shanghai Hongqiao Station) TSF from March 27, 2013, to January 14, 2015, on every Wednesday. The display period is from 7 a.m. to 2 p.m., y -axis for each interval is 10 minutes, and x -axis for each interval represents the date. The yellow cell represents the TSF of this OD twice during this period, the green cell represents once, and the white cell represents zero. From Figure 4, we can find that the OD's TSF changes significantly in the different time interval during a day. The

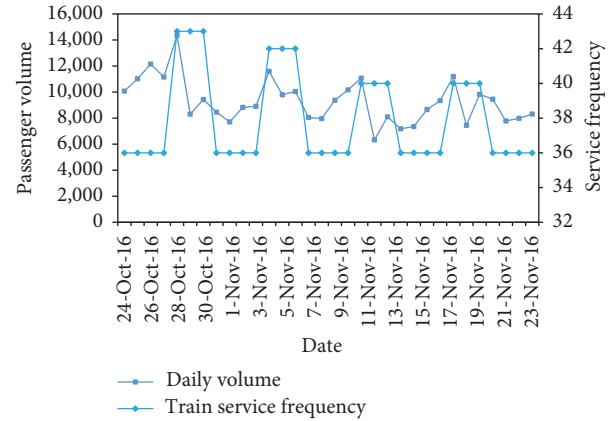


FIGURE 3: Analysis of the correlation between the TSF and DPF in the OD (Beijing South Station to Shanghai Hongqiao Station).

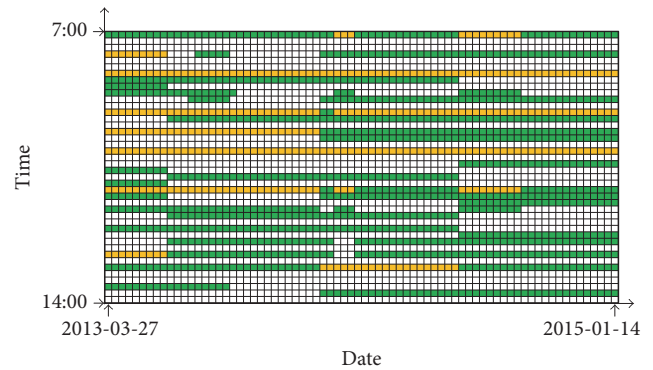


FIGURE 4: Service frequency analysis of the OD (Beijing South Station to Shanghai Hongqiao Station) in different periods.

passenger not only will choose which day to travel but also will choose when to travel, so the change of OD's TSF in different time interval will affect the choice of the passenger. If some of the passengers could not buy the train ticket which they originally want to buy, they would choose another mode of travel or choose another train in different time.

From Figure 4, we can find that when we forecast the passenger flow, we should consider the effect of the TSF. When considering the impact of TSF on the DPF, you cannot directly use the daily OD's TSF but should consider the TSF in the different time interval during a day. Therefore, in the next section, we will build a linear regression model of the DPF. Due to the substantial impacts of TSF on the DPF, the model will take into account not only the timing characteristics as the input variables, but also the TSF in different time intervals during a day.

3. Hybrid Forecasting Model

In this section, we build the Holt-Winters time series model based on the historical DPF's data. Meanwhile, the linear regression model which considers TSF is proposed. The minimum absolute value method is used to combine the two models to form a hybrid model to complement the advantages of the two.

3.1. Holt-Winters Method. The Holt-Winters model consists of three smoothing formulas that reflect the long-term trend of the data, the incremental trend, and the seasonal variation. The prediction formula is used to predict the extrapolation, which applies to situations where demand data exhibits trend and seasonal cycle characteristics.

3.1.1. Holt-Winters Method. Trend feature is represented by a single variable t , which means the length of time. The multiplicative smoothing formulas are as follows:

$$S_t = \alpha \left(\frac{Y_t}{I_{t-L}} \right) + (1 - \alpha) (S_{t-1} + T_{t-1}), \quad (1)$$

$$T_t = \beta (S_t - S_{t-1}) + (1 - \beta) T_{t-1}, \quad (2)$$

$$I_t = \gamma \left(\frac{Y_t}{S_t} \right) + (1 - \gamma) I_{t-L}, \quad (3)$$

where $Y = (Y_1, Y_2, \dots, Y_t, \dots)$ are the historical DPF; L is the length of the seasonal cycle; let $\theta = (\alpha, \beta, \gamma)$ represent the smoothing parameters. We denote the smoothing value, the trend value, and the period coefficient at the time t by S_t , T_t , and I_t , respectively.

And the predicted value at time t of the value n periods ahead is given by

$$v_h(t+n) = \hat{Y}_{t+n} = (S_{t-1} + nT_{t-1}) I_{t-L+n}. \quad (4)$$

3.1.2. Choosing Initial Value. To initiate the updating procedure, we must choose the starting values for the smoothing value, the trend value, and the period coefficient. At present, there are many methods for the selection of initial values. Here, we use a more general way to select the initial value. The equation is as follows:

$$S_0 = \frac{\sum_{t=1}^{cL} Y_t}{cL},$$

$$T_0 = 0,$$

$$I_j = \frac{\sum_{t=1}^{cL} G(j, t \bmod L) Y_t}{\sum_{t=1}^{cL} G(j, t \bmod L)}, \quad (5)$$

$$G(j, t \bmod L) = \begin{cases} 1 & j = t \bmod L \\ 0 & j \neq t \bmod L. \end{cases}$$

We denote the number of cycles c to initialize. “mod” is a symbol which means take the remainder.

3.1.3. Determine the Smoothing Parameters. To obtain more accurate smoothing parameters $\theta = (\alpha, \beta, \gamma)$, we need to ensure that the predicted value and the actual value of the minimum error by

$$E_h(\theta) = \min \sum_{t=cL}^{n-1} \sum_{m=1}^{M_{\max}} (\hat{Y}_{t+m} - Y_t)^2, \quad (6)$$

$$M_{\max} = \max(n-1, t + M_p).$$

And M_p is the number of days to forecast. However, with the increase of time series, the complexity of the above solution will increase; to solve the above problems quickly, this paper adopts the genetic algorithm to get the appropriate smoothing coefficient.

3.2. A Linear Regression Model considering Capability. From the analysis in Section 2.2, we can see that the TSF also changes over time, and there exists the relationship between the DPF and the TSF. It is found that different period of TSF has a different effect on the DPF, so we divide a day into different time periods and consider the variables of temporal features to build the regression model.

3.2.1. Variables of Temporal Features. The temporal features considered in the regression model include trend feature, day-of-week feature, and month-of-year feature. Trend feature is represented by a single variable t like the Holt-Winters model described above. Similar to Lee et al. [27], dummy variables are used to represent day-of-week, month-of-year, and year-of-period features. For the day-of-week feature, there are six dummy variables: w_i ($i = 1 \sim 6$). If the date is the i th day of the week, w_i is 1; otherwise, it is 0. Similarly, there are eleven dummy variables to reflect the different month of the year, m_j ($j = 1 \sim 11$). The number of dummy variables of year-of-period is determined by the number of years to be used, marked as y . It should be noted that, for all these dummy variables, there is an alternative (last day of the week and last month of the year) which is set as the null alternative.

3.2.2. Variables of TSF. As the size of time interval changes, not only the number of variables used will be different, but also the final prediction accuracy will be different for the model. So we assume that the time interval variable is q . We divide the day into different time segments according to the time interval and assume that the number of time segments is h , and $\mathbf{s} = (s_1, s_2, \dots, s_h)$ indicates the TSF of OD in different time segments.

3.2.3. The Linear Regression Model Which Considers TSF. Next, we constructed a linear regression model to describe the relationship between TSF and the DPF, so the DPF can be expressed as $v = v(\mathbf{y}, \mathbf{m}, \mathbf{w}, \mathbf{s})$. Here, α_0 is a constant and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_c)$ is the coefficients of the variable $\{x_i\} \subset \{\mathbf{y}, \mathbf{m}, \mathbf{w}, \mathbf{s}\}$, $1 \leq c \leq C$, and C is the number of variables. $\{x_i^t\}$ is a collection of all designed variables including different temporal features in time t and OD's TSF under different time segments like (7), and ε is the random error:

$$v_1(t) = v(\mathbf{y}, \mathbf{m}, \mathbf{w}, \mathbf{f})$$

$$= \alpha_0 + \alpha_1 x_1^t + \alpha_2 x_2^t + \dots + \alpha_c x_c^t + \varepsilon. \quad (7)$$

The variables x_i contain the sequential variables and discrete variables, and the magnitude of some variables is different. So the size of the partial regression coefficient cannot directly explain the magnitude of the linear effect of the dependent variable. Therefore, it is necessary to standardize the variables and the dependent variables in advance.

3.3. Hybrid Forecasting Model. Both of the above methods can be used to forecast the DPF, but they have some insignificance. The Holt-Winters model takes advantage of time series characteristics of passenger flow, ignoring the impact of different periods of departure frequency. The other method considers the change of OD's TSF in different times, but there are some shortcomings in the time characteristics of the impact of the method. It is natural for a decision maker to consider time-varying forecast combination schemes to avoid the disadvantages of those two methods. This paper uses the minimum absolute method [28] to combine the two approaches to take full advantage of the two methods.

In the hybrid forecasting, the essential step is to determine the weight coefficient of the hybrid forecasting model, so as to achieve the purpose of synthesizing the information of different forecasting methods to improve the accuracy of prediction. The traditional method of combination forecasting is equalization prediction method, least squares method, and so on. However, these methods have several disadvantages, but the minimum absolute method has excellent characteristics to overcome these problems.

From the above, $v_h(t)$ and $v_l(t)$ are predictions of the actual DPF $v_r(t)$; $v_c(t)$ is the predicted value of the hybrid model. So $v_c(t) = k_h v_h(t) + k_l v_l(t)$; k_h and k_l are the weight values of the two methods, $k_1 + k_2 = 0$.

The minimum absolute value method was based on the absolute value error as the objective function of the minimum; the mathematical model is as follows:

$$\min E_c = \sum_i |\hat{\varepsilon}_i|. \quad (8)$$

The constraint satisfies the following formula:

$$AK + \hat{\varepsilon} = Y, \quad (9)$$

where

$$A = \begin{bmatrix} v_h(1) & v_h(2) & \cdots & v_h(t_p) \\ v_l(1) & v_l(2) & \cdots & v_l(t_p) \end{bmatrix}^T, \quad (10)$$

$$\hat{\varepsilon} = [\hat{\varepsilon}_1 \quad \hat{\varepsilon}_2 \quad \cdots \quad \hat{\varepsilon}_{t_p}]^T,$$

$$K = [k_h, k_l]^T,$$

$$Y = [f_r(1) \quad f_r(2) \quad \cdots \quad f_r(t_p)]^T.$$

We assume $\hat{\varepsilon}_i = \hat{\varepsilon}_i^+ - \hat{\varepsilon}_i^-$, $\hat{\varepsilon}_i^+ \geq 0$, $\hat{\varepsilon}_i^- \geq 0$. Thus, the above objective function can be converted into the following equation:

$$\min E_c = \sum_i (\hat{\varepsilon}_i^+ + \hat{\varepsilon}_i^-). \quad (11)$$

So we can transform (11) into a typical linear programming problem, which uses the simplex method to calculate the weight value.

4. Case Analysis

4.1. Experiment Design. In data preparation, daily data of the Beijing South Railway Station to Shanghai Hongqiao Railway Station's passenger flow and TSF from July 1, 2012, to the end of 2016 was collected in our experiments. The official operation time of the Beijing-Shanghai HSR was in July 2011. However, since the passenger flow was still in the breeding stage and the passenger flow characteristics were not visible, this article abandoned the passenger flow data from July 2011 to June 2012. Moreover, the DPF in the holiday has a significantly different pattern from that in ordinary days, among the over two years' daily data, points that are during China's legal holidays (e.g., Spring Festival, Labor Day, and National Day) are removed from the regression. That is to say, only ordinary days are considered in training and testing experiments. For example, when evaluating the forecasting ability of the period from October 16 to November 15, 2013, daily data from July 1, 2012, to October 15, 2016, are used for training.

In China, HSR in the working day (except Friday) use the daily timetable, on Friday and the weekend use the weekend timetable, and in the holiday use the peak timetable. Different types of timetable are designed to accommodate different passenger demand, so in a day at various intervals and the same OD's TSF, there will be some differences. Typically, the daily TSF of the three timetables for the same OD has a quantitative relationship, $f_d \leq f_w \leq f_p$, where f_d , f_w , and f_p are the OD train TSF of daily timetable, weekend timetable, and peak timetable. Thus, when calculating the weight of different methods, this article based on the impact of the various types of timetable will be calculated separately.

According to the data used, starting from Beijing to Shanghai trips starting time ranges from 6 a.m. to 19 p.m. Therefore, this article uses the 6:00 a.m. as the starting point and 17:00 p.m. for the endpoint, according to a particular time interval which will be divided into different time segments. For comparison of several prediction methods, we use three methods to predict the traffic after 30 days, respectively, and compare it with actual traffic.

The performance measures adopted in our research are the widely used mean absolute percent error (MAPE, M) in (12) and variance of absolute percentage error (VAPE, V) in (13). The former one measures the average forecasting accuracy, and the latter one measures the stability of prediction:

$$M = \frac{1}{n_f} \sum_{i=1}^{n_f} \left| v_i - \hat{v}_i \right| \times 100\%, \quad (12)$$

$$V = \frac{\sum_{i=1}^{n_f} (v_i - \hat{v}_i / v_i - M)^2}{n_f}, \quad (13)$$

where n_f denotes the number of days predicted; in this article, we forecast the number of days for 30 days, so $n_f = 30$.

4.2. Hybrid Forecasting Model Analysis. MAPE results of different models for 25 testing experiments were shown in Figure 5 and Table 1, and the time interval selected is 45 minutes. It is found that hybrid model proposed in this work

TABLE 1: MAPE values (%) of different models for 25 testing experiments.

Testing experiment	Hybrid	Holt-Winters	Linear
1	18.1	18.1	20.4
2	17.3	17.0	20.5
3	12.2	12.5	20.4
4	15.7	21.9	21.4
5	12.6	16.8	22.3
6	12.1	13.6	23.5
7	12.3	15.9	23.3
8	20.4	32.9	23.1
9	15.0	23.0	22.4
10	14.7	19.4	15.5
11	19.4	27.2	15.6
12	16.2	22.7	15.0
13	11.2	13.6	14.8
14	11.0	12.6	14.7
15	11.7	11.9	14.5
16	10.6	12.2	14.6
17	10.1	12.7	14.8
18	15.3	23.8	15.2
19	22.8	34.6	15.8
20	12.5	17.1	17.0
21	10.3	12.7	17.4
22	9.7	12.1	17.3
23	16.2	24.3	17.2
24	21.1	42.3	37.9
25	12.2	22.4	39.0
Average	14.4	19.7	19.7

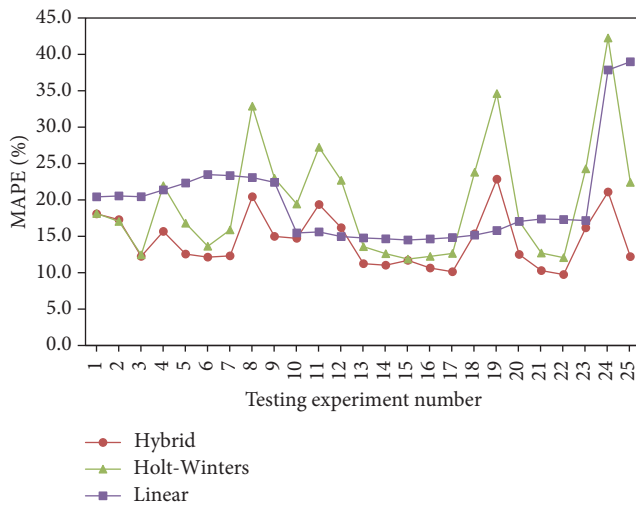


FIGURE 5: MAPE of different models for 25 testing experiments.

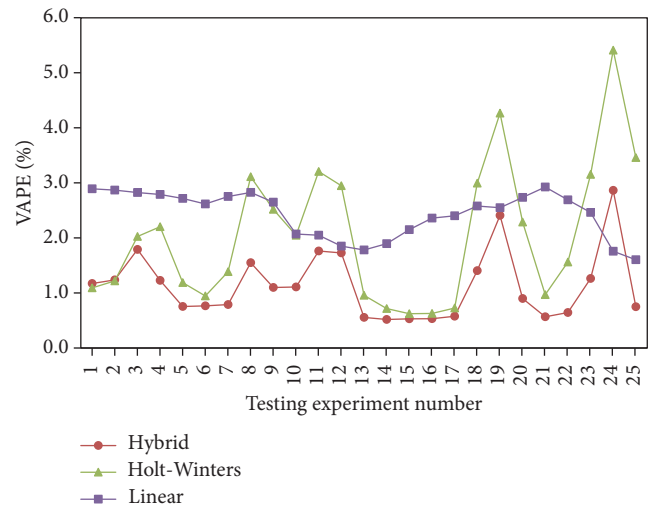


FIGURE 6: VAPE of different models for 25 testing experiments.

performs the best of accuracy (14.4% in average) among these tested models, and the prediction accuracy of the other two methods is approximately the same. For most of the experiments, MAPE of hybrid model is lower than that of the other two approaches. Even for those experiments (11, 19) in which the hybrid model has a worse accuracy than Holt-Winters model, their difference is not significant.

VAPE results are demonstrated in Figure 6 and Table 2. The results show that the proposed hybrid model has a higher stability (1.1% in average) than others and is followed by Holt-Winters, Linear.

From the above comparisons, it is found that hybrid model proposed in this work can perform better in both forecasting accuracy and stability. Besides, since our model

TABLE 2: VAPE values (%) of different models for 25 testing experiments.

Testing experiment	Hybrid	Holt-Winters	Linear
1	1.2	1.1	2.9
2	1.2	1.2	2.9
3	1.8	2.0	2.8
4	1.2	2.2	2.8
5	0.8	1.2	2.7
6	0.8	0.9	2.6
7	0.8	1.4	2.8
8	1.5	3.1	2.8
9	1.1	2.5	2.6
10	1.1	2.0	2.1
11	1.8	3.2	2.0
12	1.7	2.9	1.9
13	0.6	1.0	1.8
14	0.5	0.7	1.9
15	0.5	0.6	2.1
16	0.5	0.6	2.4
17	0.6	0.7	2.4
18	1.4	3.0	2.6
19	2.4	4.3	2.5
20	0.9	2.3	2.7
21	0.6	1.0	2.9
22	0.6	1.6	2.7
23	1.3	3.2	2.5
24	2.9	5.4	1.8
25	0.7	3.5	1.6
Average	1.1	2.1	2.4

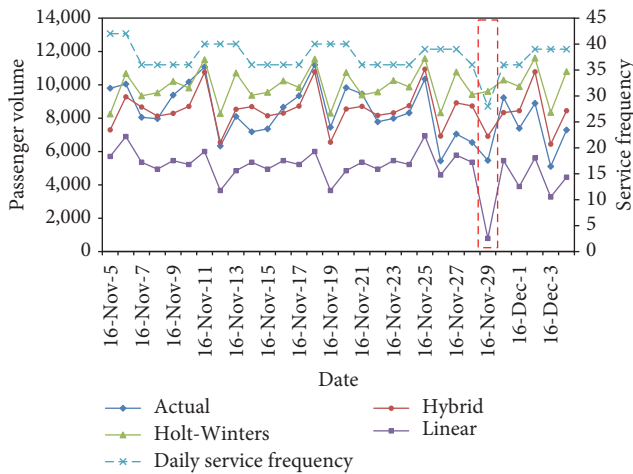


FIGURE 7: Analysis of daily passenger flow by different methods (the period of November 5, 2016, to December 4, 2016).

has established a definite relationship between passenger time characteristics and TSF, it also gives a way to forecast the real DPF. Taking experiment 25 (corresponding to the period of November 5, 2016, to December 4, 2016) as an example, the real DPF, together with the forecasted DPF by three different ways, are illustrated in Figure 7.

From Figure 7, we can find that the forecast value of Holt-Winters model is always higher than the actual value, the estimated value of linear model is always lower than the real value, and this leads to the most of the prediction of the hybrid model between the two, making it closer to the actual value. From the red dotted rectangle frame out of the part, we can find two phenomena. One is where the Holt-Winters only consider the timing characteristics of the DPF, so when the OD's TSF becomes larger, its degree of deviation will increase; the other one is that the linear model overreacts to frequency changes, and also it makes the deviation of the predicted value increase. But the hybrid model is that the use of the difference between the two changes makes the forecast close to the actual DPF.

4.3. *The Impact of Different Time Intervals.* From the model, we can see that the selection of the interval will affect the accuracy of the linear regression model, which in turn affects the accuracy of the hybrid model prediction. 17 different time intervals were selected to explore the influence of different time intervals on the model. At the same time, to enhance the reliability of the results, we will increase the number of samples at each time interval from the 25 to 60.

The MAPE of different time intervals is plotted as a boxplot. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles,

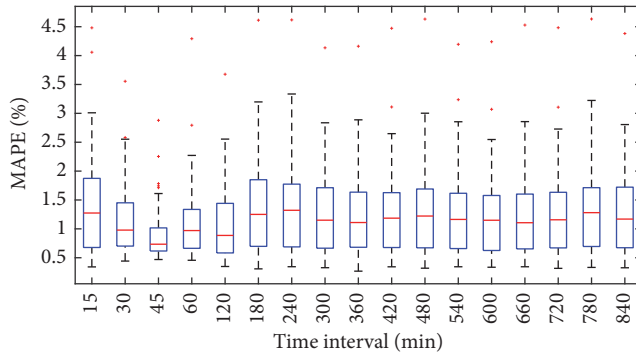


FIGURE 8: MAPE at different time intervals.

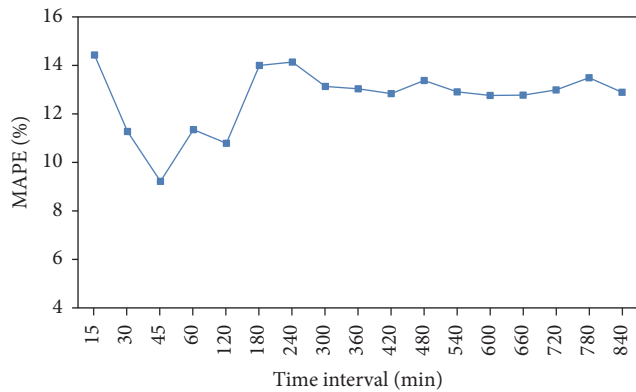


FIGURE 9: Multisample average MAPE at different time intervals.

the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

From Figure 8, we can see that when the interval time is 45 minutes, the 25th and 75th percentiles of MAPE are lower than other interval times. Moreover, its box length also is the shortest. Figure 9 shows the average MAPE for different samples at different time intervals. We can see that the mean error is the smallest 45 minutes, followed by 120 minutes. Therefore, it is shown that the hybrid model with the 45-minute interval is the best in the accuracy and stability of the prediction. Although it has outliers, the overall size of its outliers is still smaller than the outliers of other time intervals.

From Figure 9, we can see that from 15-minute to 180-minute time interval, the predicted MAPE is similar to the parabola. For more than 180-minute time interval, the model prediction MAPE remains unchanged.

5. Conclusions

This paper constructs a hybrid forecasting model, and the model not only makes full use of the time series model for the more accurate prediction of the periodic data but also utilizes the effects of TSF. From the example, we can see that the hybrid model is more accurate and more stable than the single model. Compared with other models, the data sources used are less and easier to obtain. Moreover, the model in this paper can be employed to provide the basic information for the adjustment of the train timetable.

However, in this paper, we consider the TSF in different time during a day but never take the rated transportation capacity into account. Other factors such as the strategy of seat allocation are not considered for simplicity. The refined representation of transportation capacity of HSR line is necessary for the future work.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by China Railway Corporation Technology Research and Development Plan Project (2016X005-E) and National Key R&D Program of China (2016YFB1200600).

References

- [1] X.-S. Jiang, L. Zhang, and M. X. Chen, "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*, vol. 44, pp. 110–127, 2014.
- [2] E. P. George and G. Jenkins, "Time series forecasting and control," *Chapman & Hall/crc Boca Raton*, vol. 2, no. 3, pp. 131–133, 1970.
- [3] C. K. Moorthy and B. G. Ratcliffe, "Short term traffic forecasting using time series methods," *Transportation Planning and Technology*, vol. 12, no. 1, pp. 45–56, 1988.
- [4] S. Lee and D. B. Fambro, "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting," *Transportation Research Record*, vol. 1678, pp. 179–188, 1999.
- [5] W. R. Christiaanse, "Short-term load forecasting using general exponential smoothing," *IEEE Transactions on Power Apparatus and Systems*, vol. 90, no. 2, pp. 900–911, 1971.
- [6] C. Chatfield, "The Holt-Winters Forecasting Procedure," *Journal of Applied Statistics*, vol. 27, no. 3, p. 264, 1978.
- [7] C. Chatfield and M. Yar, "Holt-Winters Forecasting: Some Practical Issues," *The American Statistician*, vol. 37, no. 2, p. 129, 1988.
- [8] P. S. Kalekar, "Time-series Forecasting using Holt-Winters Exponential Smoothing," *Kanwal Rekhi School of Information Technology*, 2004.
- [9] J. D. Bermúdez, J. V. Segura, and E. Vercher, "Holt-Winters forecasting: an alternative formulation applied to UK air passenger data," *Journal of Applied Statistics*, vol. 34, no. 9, pp. 1075–1090, 2007.
- [10] P. R. Winters, "Forecasting sales by exponentially weighted moving averages," *Management Science*, vol. 6, pp. 324–342, 1960.
- [11] Y. H. Guo, X. P. Shi, and X. D. Zhang, "A study of short term forecasting of the railway freight volume in China using ARIMA and Holt-Winters models," in *Proceedings of the International Conference on Supply Chain Management and Information Systems. IEEE*, pp. 1–6, 2011.

- [12] C. P. Da Veiga, C. R. P. Da Veiga, A. Catapan, U. Tortato, and W. V. Da Silva, "Demand forecasting in food retail: A comparison between the Holt-Winters and ARIMA models," *Wseas Transactions on Business and Economics*, vol. 11, no. 1, pp. 608–614, 2014.
- [13] D. Puthran, H. C. Shivaprasad, and K. S. K. Kumar, "Comparing SARIMA and Holt-Winters' forecasting accuracy with respect to Indian motorcycle industry," *Transactions on Engineering and Sciences*, 2014.
- [14] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, Bowker-Saur, London, UK, 2014.
- [15] Y.-J. Luo, J. Liu, X. Sun, and Q.-Y. Lai, "Regression model for daily passenger volume of high-speed railway line under capacity constraint," *Journal of Central South University*, vol. 22, no. 9, pp. 3666–3676, 2015.
- [16] X. Du, R. Shang, and Z. Gao, "High-speed railway passenger flow volume prediction model based on factor-response simulation method," *Information Technology Journal*, vol. 10, no. 9, pp. 1761–1766, 2011.
- [17] M. Wardman, "Demand for rail travel and the effects of external factors," *Transportation Research Part E: Logistics and Transportation Review*, vol. 42, no. 3, pp. 129–148, 2006.
- [18] R. Law and N. Au, "A neural network model to forecast Japanese demand for travel to Hong Kong," *Tourism Management*, vol. 20, no. 1, pp. 89–97, 1999.
- [19] R. Batley, J. Dargay, and M. Wardman, "The impact of lateness and reliability on passenger rail demand," *Transportation Research Part E: Logistics and Transportation Review*, vol. 47, no. 1, pp. 61–72, 2011.
- [20] J. Chi and J. Baek, "A dynamic demand analysis of the United States air-passenger service," *Transportation Research Part E: Logistics and Transportation Review*, vol. 48, no. 4, pp. 755–761, 2012.
- [21] R. R. Andrawis, A. F. Atiya, and H. El-Shishiny, "Combination of long term and short term forecasts, with application to tourism demand forecasting," *International Journal of Forecasting*, vol. 27, no. 3, pp. 870–886, 2011.
- [22] D. Zeng, J. Xu, J. Gu, L. Liu, and G. Xu, "Short term traffic flow prediction using hybrid ARIMA and ANN models," in *Proceedings of the 2008 Workshop on Power Electronics and Intelligent Transportation System, PEITS 2008*, pp. 621–625, Guangzhou, China, August 2008.
- [23] S. Xu, H. Zhao, and X. Lv, "A grey SVM based model for patent application filings forecasting," in *Proceedings of the 2008 IEEE International Conference on Fuzzy Systems, FUZZ 2008*, pp. 225–230, Hong Kong, China, June 2008.
- [24] Y. Wei and M. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 148–162, 2012.
- [25] S. Makridakis and M. Hibon, "The M3-competition: Results, conclusions and implications," *International Journal of Forecasting*, vol. 16, no. 4, pp. 451–476, 2000.
- [26] J. H. Stock and M. W. Watson, "Combination forecasts of output growth in a seven-country data set," *Journal of Forecasting*, vol. 23, no. 6, pp. 405–430, 2004.
- [27] C. k. Lee, T. D. Lin, and C. H. Lin, "Pattern analysis on the booking curve of an inter-city railway," *Journal of the Eastern Asia Society for Transportation Studies*, vol. 6, pp. 303–317, 2005.
- [28] Y. Zhang, M. A. Chuan-Sheng, and K. Wei, "The study of estimation of weight coefficient in combination forecast models-the application of the least absolute value model," *Journal of Transportation Systems Engineering and Information Technology*, vol. 6, no. 4, pp. 125–129, 2006.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

