

Short-Term High-Speed Rail Passenger Flow Prediction by Integrating EEMD with Multivariate G-SVM

Yujie Yuan^{a,b,d}, Xiushan Jiang^{a*}, Pei Zhang^c, Chun Sing Lai^{d,*}

^a School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

^b School of Air Traffic Management, Civil Aviation University of China, Tianjin, 300300, China

^c Academy of Armored Force Engineering, Beijing 100072, China

^d Brunel Interdisciplinary Power Systems Research Centre, Department of Electronic and Electrical Engineering, Brunel University London, Kingston Lane, London, UB8 3PH, UK

* Corresponding author: xshjiang@bjtu.edu.cn; chunsing.lai@brunel.ac.uk

ABSTRACT

Short-term prediction of high-speed rail (HSR) passenger flow provides a daily ridership estimation for the near future, which is critical to HSR planning and operational decision making. This paper proposes a new methodology that integrates ensemble empirical mode decomposition with multivariate support vector machines (EEMD-MSVM). There are four steps in this hybrid forecasting approach: (i) utilizing the Kruskal's stress index to determine the correlation strength of passenger flow predictions among HSR multi-stations; (ii) decompose the fluctuation of passenger flows between each HSR stations, using EEMD to generate the intrinsic mode functions (IMFs) and a trend term; (iii) predict the IMF for each correlated station pair using MSVM; (iv) reconstruct the refined IMF components to predict daily multivariate HSR passenger flows. The proposed EEMD-MSVM approach is demonstrated with multiple origin-destination pairs along the Wuhan-Guangzhou HSR in China. Results from various origin-destination pairs, show that the EEMD-MSVM approach outperforms the existing ensemble empirical mode decomposition with grey support vector machine approach (EEMD-GSVM). With the multivariate approach, the mean absolute percentage error in demand prediction for different origin-destination pairs is reduced by 13.9%, 1.2%, 1.0%, 2.0%, and 2.7%, while the mean absolute deviation is reduced by 78.8, 38.0, 4.4, 4.6, and 3.9, respectively. Such increase in short-term demand prediction accuracy can significantly improve HSR service planning, operations, and revenue management in the real world.

Keywords: High-speed rail, passenger flow short-term forecasting, ensemble empirical mode decomposition, multivariate support vector machine

1. Introduction

Short-term prediction of high-speed rail (HSR) passenger flow supplies a daily ridership estimation of the near future (e.g., next day, next week), which exerts a significant effect on HSR planning and operational decision making. Accurate short-term demand forecasting contributes to successful rail revenue management and balances travel demand and supply via proactive measures.

Short-term passenger flow prediction for single rail section has been studied for decades. In the literature, various categories of prediction models (Hong, 2022) have been adopted such as time series models (Harvey, 1990), traffic simulation models (Boxill and Yu, 2000), nonparametric regression models (Nguyen-Tuong et al., 2008), and neural networks (Poirazi et al., 2003, Teresa, 2023). Many results have been obtained for prediction of HSR passenger flow. For example, Ahmed (1979) and Chuwang (2022) proposed the Box-Jenkins time series model to forecast HSR passenger flow. Yet, that research had not adequately taken influence factors (Chen, 2022) into account. The multiple regression model showed good predictions, however, which were characterized by relying on sufficient historical data (Butkevičius et al., 2004, Abramović et al., 2015).

The autoregressive integrated moving average (ARIMA) model has been widely applied in forecasting short-term traffic quantities such as traffic flow rate, travel time, speed, and time occupancy (Hamed et al., 1995; Nijkamp et al., 1996; Lee and Fambro, 1999; Williams, 2003; Kumar, 2015; Shahriari, 2020; Aljuaydi, 2022; Brenner, 2022; Kumar, 2023). With the characteristics of seasonality and trends in traffic data, several studies applied the seasonal ARIMA to predict traffic flow (Williams, 2001; Williams and Hoel, 2003) and international air passenger flow (Faraway and Chatfield, 1998; Lim and McAleer, 2002; Yan, 2012; Vaishnav, 2018; Aijaz, 2020; Bi, 2022; Sibrak, 2022). ARIMA offered good, robust results in modeling linear and stationary time series. However, applications of ARIMA or seasonal ARIMA models were limited by the linear assumption of time-lagged variables; therefore, they might not capture the structure of non-linear relationships (Zhang et al., 1998).

Neural networks have made remarkable achievements in the field of HSR passenger flow forecasting. They are frequently adopted as the modeling approach because they possess the characteristics of adaptability, nonlinearity, and arbitrary function mapping capability (Zhang et al., 1998). Essentially, neural networks can deal with complex, non-linear problems without a priori knowledge regarding the relationships between input and output variables. Tsai et al. (2005) developed short-term forecasting models using a dynamic neural network. A combination forecasting model of the neural network and logit model was established by Reggiani et al. (2014); Himanen. (2019) and El Yaagoubi. (2022). The results showed that the combination achieved a better prediction accuracy, but the calculation of weights needed to be further optimized. A variety of forecasting models,

based on e.g., multiple linear regression, multiple nonlinear regression, autoregressive moving, and coupling wavelet neural network, were used to predict passenger flow; the test results showed that the wavelet neural network performed the best (Adamowski et al., 2012).

In addition to the prediction models for single time series, multivariate traffic flow prediction has attracted researchers' interest. Whittaker et al. (1997) analyzed the mutual interaction of traffic flow and average speed of different locations in a real-world road network and established a multi-state space traffic flow prediction model using the Kalman filter (Chen, 2023). The ARIMA model was applied to predict traffic flow at multiple locations. Stathopoulos and Karlaftis (2001) carried out the spectrum analysis and cross-spectrum analysis, indicating that there was a correlation between traffic flow observed at different locations and that it changed with distance and time of day. Stathopoulos and Karlaftis (2003) analyzed changes in the traffic flow of five different locations using the multivariable state space method. Kamarianakis and Prastacos (2003) applied the space-time autoregressive integrated moving average (STARIMA) to multivariate traffic flow prediction. Vlahogianni and Karlaftis (2004) proposed a multivariate method for road traffic prediction using neural networks, considering the spatial variation of traffic data. Wang and Papageorgious (2005) proposed the extended Kalman filter combined with a macroscopic traffic flow model for real-time traffic flow forecasting on motorways. Chen (2012) studied the influence on traffic prediction of the retrieval daily trend. Numerous studies have been conducted in this field more recently (e.g., Vlahogianni et al., 2014; Zhang et al., 2014; Cai et al., 2016; Jenelius and Koutsopoulos, 2018).

Ensemble empirical mode decomposition (EEMD) has gained prominence in traffic prediction, with studies showcasing its versatility and effectiveness in enhancing forecast accuracy (Huang, 2015). Combining EEMD with advanced machine learning techniques (Bokde et al., 2020), such as Multivariate Gated Recurrent Units and Long Short-Term Memory Neural Networks, has demonstrated promising results, significantly improving the capture of complex temporal patterns in traffic data (Hui et al., 2020). EEMD's preprocessing capabilities provide valuable feature representations, thereby contributing to the overall success of traffic prediction models (Zheng et al., 2019). EEMD has emerged as a valuable tool in traffic prediction, enhancing the accuracy of short-term forecasts and improving our understanding of complex traffic patterns (Pandey et al., 2020). These studies demonstrate the versatility of EEMD as a preprocessing step in conjunction with various prediction models, including neural networks and traditional forecasting techniques. EEMD's ability to capture underlying patterns in traffic data makes it a promising approach for optimizing traffic management and decision-making in transportation systems.

Jiang et al. (2014) proposed a hybrid short-term HSR passenger flow forecasting approach by combining the EEMD and grey support vector machine (GSVM), which

performed well for the univariate short-term prediction, e.g., long-distance/middle-distance/short-distance HSR origin-destination (OD) pairs. However, the computational complexity is high when simultaneously conducting short-term forecasting of HSR passenger flow for multiple stations or OD pairs. It may not satisfy real-time operational requirements. For a massive HSR network, if station correlations are not fully considered in advance—that is, HSR passenger flows are regarded as independent univariate time series—the prediction accuracy can be easily denied.

Previous literature primarily focused on single-section short-term predictions, leaving a gap in comprehensive predictions for multivariate high-speed rail passenger flows. Recognizing this gap, we draw inspiration from traffic flow prediction theories in road networks, acknowledging the similarities and differences between high-speed railways and roads. By bridging these concepts, we aim to extend the predictive capabilities from single sections to multivariate sections, opening new research avenues.

To bridge the gap in multivariate short-term forecasting of high-speed rail passenger flows, this paper extends the existing hybrid approach (EEMD-GSVM) to enable multivariate forecasting for HSR passenger flows with station correlations. The achievement of prediction results of enhancement is better than simultaneous univariate predictions. By using multivariate time series analysis, stations of the HSR network can be classified into several groups according to their correlations. Then multivariate short-term forecasting can be conducted for highly correlated stations of the same group. In this way, the accuracy of real-time forecasting for HSR passenger flows of correlated stations can be improved. We propose the integrated ensemble empirical mode decomposition with multivariate support vector machines (EEMD-MSVM) model for forecasting short-term HSR passenger flows of multiple OD pairs, which extends the existing EEMD-GSVM short-term forecasting model for a univariate OD pair. Our work contributes significantly to the field in several ways:

- This paper introduces the concept of multivariate HSR passenger flow forecasting, breaking away from the traditionally limited single-section predictions. This conceptual advancement marks a pivotal shift in the field's research direction.
- Leveraging enhanced empirical mode decomposition (EEMD) and integrating EEMD with GSVM, we propose a multivariate short-term forecasting model, EEMD-MSVM. Compared to existing single-section models, the novel model demonstrates superior accuracy. It effectively mitigates the adverse effects of various factors influencing high-speed railway passenger traffic forecasts, significantly improving forecast precision. This advancement is crucial for optimizing ticket allocation and train scheduling.
- Emphasizing the importance of selecting appropriate sections for multivariate passenger flow forecasting. Employing AI data processing software, SPSS, we perform correlation analyses for high-speed rail traffic at each station. This strategy

provides a foundation for effectively applying the model.

- Through empirical studies utilizing Wuhan-Guangzhou HSR passenger data, we validate the effectiveness of our EEMD-MSVM approach. By comparing it with EEMD-GSVM, we demonstrate enhanced accuracy and error reduction. Our data analysis reveals that passenger flow fluctuations result from the superposition of oscillations across multiple sections, validating the strong correlation between these sections.

Section 2 of this paper provides the framework for multivariate HSR passenger flow forecasting. Section 3 describes the data collection and case study. The modeling and results are presented in Section 3 and compared with the existing EEMD-GSVM approach. Section 4 offers conclusions.

2. Methodology

2.1 Framework

A typical rail line consists of many stations, and any two stations (origination station and destination station) form an OD. Different OD passenger flows, especially between the same origination station or the same destination station, influence each other. This paper proposes a multi-variable short-term passenger flow prediction model considering the influence of each OD called the EEMD-MSVM model. Taking one OD as one variation, the methodological framework of the EEMD-MSVM approach can be summarized through four key steps, highlighted in Fig. 1.

Step 1: explore the correlation of multivariate HSR passenger flows of different OD based on archived data, to select m HSR sections with strong correlation.

Step 2: decompose empirical modes of the historical passenger flows for every HSR station using EEMD to generate a number of intrinsic mode functions (IMFs) and a trend term (Res).

Step 3: combine the same IMFs of all of the m ODs. Then, the multivariate SVM model is employed to train the combined time series and predict each IMF for every correlated OD; apply trend extrapolation to predict the trend term of each OD.

Step 4: reconstruct the forecasted 1IMF components and 1Res to get passenger flows forecasting of OD1, the same method for obtaining the others. Then, the model is used to predict the passenger flow for next day.

The detailed procedures of the aforementioned steps are depicted from Section 2.2 to Section 2.6.

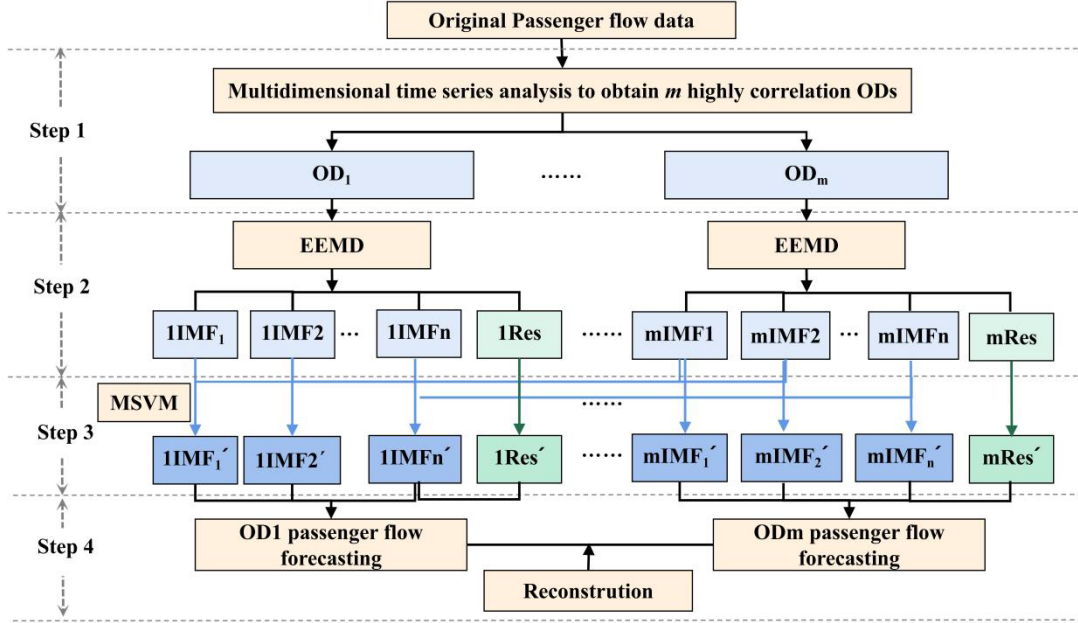


Fig. 1. The methodological framework of the EEMD-MSVM

2.2 Analyzing Correlation with Multidimensional Scaling

Multidimensional scaling (MDS) (Cox et al., 2000.) is a technique that visualizes the level of similarity. It is a form of non-linear dimensionality reduction. The MDS algorithm aims to place each object in N-dimensional space so that the between-object distances are preserved as much as possible. A list of variables is shown in Appendix A.

The purpose of the multidimensional scaling analysis of HSR passenger flow is to determine the correlations of M HSR OD and the quantitative representation of M OD in low-dimensional space, which is perceptual mapping. Then, group the M OD to select m OD with a strong correlation.

First, calculate similarity coefficient matrix C of the M OD, transform the matrix C into generalized distance matrix D , as shown in formula (1), and arrange the non-diagonal elements of the matrix D in ascending order, as shown in formula (2). Then, select the reasonable dimension r to solve the distance matrix; the M OD corresponds to the M point of the Euclidean space and the distance of the points i and j in the r dimensional space can be expressed by formula (3). It also should meet the order $\hat{d}_{i_1j_1} \leq \hat{d}_{i_2j_2} \leq \dots \leq \hat{d}_{i_nj_n}$. Adjusting the position of the M point gradually can make the space distance \hat{d}_{ij} and the generalized distance matching degree enhance unceasingly. $STRESS$ and S_{stress} are used to measure the matching degree. $STRESS$ is a widely used index to measure the degree of matching, which was proposed by Kruskal, so it is called K pressure index (Kruskal, 1964). For each r , we can find a minimum index of pressure-fitting structure. When the minimum pressure index is lower than the given pressure index for the first time, r is the required

dimension. Finally, we use the graphical representation to solve the results.

$$d_{ij} = (c_{ii} + c_{jj} - 2c_{ij})^{1/2} \quad (1)$$

$$d_{i_1j_1} \leq d_{i_2j_2} \leq \dots \leq d_{i_nj_n}, \quad n = \frac{1}{2}M(M-1), \quad i_l < j_l, l = 1, 2, \dots, n \quad (2)$$

$$\hat{d}_{ij} = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + \dots + (X_{ir} - X_{jr})^2} \quad (3)$$

2.3 EEMD for Multivariate Passenger Flow

Let $q_i(t)$ be the original HSR passenger flows of OD i ($i = 1, 2, \dots, m$) in day t . The original passenger flows of OD I during time period T_i is denoted as \mathbf{q}_{i,T_i} :

$$\mathbf{q}_{i,T_i} = [q_i(1), q_i(2), \dots, q_i(t)]^T, \quad i = 1, 2, \dots, m \quad (4)$$

EEMD method is applied to the original HSR passenger flows of OD i during the time period T_i ; more details about EEMD are available in Jiang et al. (2014). The resulting data set includes n IMF components and one trend item denoted by date set C_i :

$$C_i = \left\{ \begin{array}{cccc} q_{i,IMF1}(1), & q_{i,IMF2}(1), & \dots, & q_{i,IMFn}(1), & q_{i,Res}(1) \\ q_{i,IMF1}(2), & q_{i,IMF2}(2), & \dots, & q_{i,IMFn}(2), & q_{i,Res}(2) \\ & & & \vdots & \\ q_{i,IMF1}(t-1), & q_{i,IMF2}(t-1), & \dots, & q_{i,IMFn}(t-1), & q_{i,Res}(t-1) \\ q_{i,IMF1}(t), & q_{i,IMF2}(t), & \dots, & q_{i,IMFn}(t), & q_{i,Res}(t) \end{array} \right\}, \quad i = 1, 2, \dots, m \quad (5)$$

As shown in Fig. 1, the IMF j of OD i is defined as $iIMFj = [q_{i,IMFj}(1), \dots, q_{i,IMFj}(t)]^T$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$, and the trend term of OD i is represented by $iRes = [q_{i,Res}(1), \dots, q_{i,Res}(t)]^T$, $i = 1, 2, \dots, m$.

2.4 Multivariate Passenger Flow Forecasting Model Based MSVM

2.4.1. Divide the train set and test set

When the same IMFs of all of the m OD are combined, divide the train set and test set and select the appropriate kernel function; then the multivariate SVM model is employed to train and predict the combined time series.

The same IMFs are combined as $IMFi$, given by

$$IMFi = \{1IMFi, 2IMFi, 3IMFi, \dots, (m-1)IMFi, mIMFi\}, \quad i = 1, 2, \dots, n \quad (6)$$

The train data set is selected as $IMFi_{train}$, given by

$$IMFi_{train} = \left\{ \begin{array}{l} q_{1,IMFi}(1), \quad q_{2,IMFi}(1), \quad \dots, \quad q_{m,IMFi}(1) \\ q_{1,IMFi}(2), \quad q_{2,IMFi}(2), \quad \dots, \quad q_{m,IMFi}(2) \\ \vdots \\ q_{1,IMFi}(t-a), \quad q_{2,IMFi}(t-a), \quad \dots, \quad q_{m,IMFi}(t-a) \end{array} \right\}, \quad i = 1, 2, \dots, n \quad (7)$$

The test data set is selected as $IMFi_{test}$, given by

$$IMFi_{test} = \left\{ \begin{array}{l} q_{1,IMFi}(t-a+1), \quad q_{2,IMFi}(t-a+1), \quad \dots, \quad q_{m,IMFi}(t-a+1) \\ q_{1,IMFi}(t-a+2), \quad q_{2,IMFi}(t-a+2), \quad \dots, \quad q_{m,IMFi}(t-a+2) \\ \vdots \\ q_{1,IMFi}(t), \quad q_{2,IMFi}(t), \quad \dots, \quad q_{m,IMFi}(t) \end{array} \right\}, \quad i = 1, 2, \dots, n \quad (8)$$

where a is the number of elements in the test set.

After the learning, training, and testing of MSVM, the predicted results of $IMFi$ can final be obtained as follows:

$$IMFi' = \left\{ \begin{array}{l} q'_{1,IMFi}(1), \quad q'_{2,IMFi}(1), \quad \dots, \quad q'_{m,IMFi}(1) \\ q'_{1,IMFi}(2), \quad q'_{2,IMFi}(2), \quad \dots, \quad q'_{m,IMFi}(2) \\ \vdots \\ q'_{1,IMFi}(t), \quad q'_{2,IMFi}(t), \quad \dots, \quad q'_{m,IMFi}(t) \\ \vdots \\ q'_{1,IMFi}(t+f), \quad q'_{2,IMFi}(t+f), \quad \dots, \quad q'_{m,IMFi}(t+f) \end{array} \right\}, \quad i = 1, 2, \dots, n \quad (9)$$

where f is the prediction step ahead.

2.4.2. Kernel function

A kernel function can transform a linear inseparable input space into a linear separable high-dimensional feature space (Joachims, 1998). The performance of SVM depends on the choice of kernel functions. Since the Gaussian radial basis function has a fast learning process with few parameters, this paper chooses it as the kernel function of SVM, as shown in formula (10):

$$k(\mathbf{x}_i, \mathbf{x}) = e^{\left(-\frac{\|\mathbf{x}_i - \mathbf{x}\|^2}{\sigma^2} \right)} \quad (10)$$

where σ is the kernel parameter, $\sigma \in [2^{-10}, 2^{15}]$.

2.4.3. Trend extrapolation predicts the trend term

Trend extrapolation is a forecasting method to determine the predicted value by extrapolating the trend of change from the historical time series of the predicted variables

to the future. Trend extrapolation is usually used to predict the development rule of objects with gradual change rather than jumping change, and it can find a suitable function to reflect the change trend of predicted objects.

The result of EEMD indicates that the change of trend term is smooth and there is no jump fluctuation. Moreover, since this paper is about a short-term prediction of passenger flow, it is very suitable to adopt the trend extrapolation for the prediction of trend term. The fitting function is shown as follows:

$$q'_{i,Res}(t) = F_i(t) \quad (11)$$

After trend extrapolation prediction, the prediction results of trend term matrix are shown as follows:

$$Res' = \left\{ \begin{array}{ccc} q'_{1,Res}(1), & \dots, & q'_{m,Res}(1) \\ \vdots & & \vdots \\ q'_{1,Res}(t), & \dots, & q'_{m,Res}(t) \\ \vdots & & \vdots \\ q'_{1,Res}(t+f), \dots, & & q'_{m,Res}(t+f) \end{array} \right\} = \left\{ \begin{array}{ccc} F_1(1), & \dots, & F_2(1) \\ \vdots & & \vdots \\ F_1(t), & \dots, & F_2(t) \\ \vdots & & \vdots \\ F_1(t+f), \dots, & & F_2(t+f) \end{array} \right\} \quad (12)$$

where Res'_{ie} are the prediction results of trend term matrix, $q'_{i,Res}(t)$ is the prediction value of $q_{i,Res}(t)$.

2.5 IMF and Res Reconstruction for Passenger Flows

The reconstruction process is the inverse process of EEMD. Combined with the characteristics of EEMD, the final prediction results of each OD are obtained by recombining $iIMF_j$ and trend terms. Forecasting results of passenger flow of OD i within a period of T_{t+f} , are given by

$$\mathbf{q}'_{i,T_{t+f}} = [q'_i(1), \dots, q'_i(t), \dots, q'_i(t+f)]^T, \quad i = 1, 2, \dots, m \quad (13)$$

$$q'_i(t) = \sum_{j=1}^n q'_{i,IMF_j}(t) + q'_{i,Res}(t), \quad i = 1, 2, \dots, m \quad (14)$$

where $\mathbf{q}'_{i,T_{t+f}}$ is forecasting vector of OD i for period T_{t+f} , $q'_i(t)$ is the passenger flows prediction result of OD i ($i = 1, 2, \dots, m$) in day t .

By combining all the forecasting results $\mathbf{q}'_{i,T_{t+f}}$ of the m OD within time period T_{t+f} , the final prediction results matrix $\mathbf{Q}'_{T_{t+f}}$ can be obtained. It is shown as follows:

$$\mathbf{Q}'_{T+f} = [\mathbf{q}'_{1,T+f}, \mathbf{q}'_{2,T+f}, \dots, \mathbf{q}'_{m,T+f}]^T = \begin{bmatrix} q'_1(1) & L & q'_1(t) & L & q'_1(t+f) \\ q'_2(1) & L & q'_2(t) & L & q'_2(t+f) \\ M & O & M & O & M \\ q'_m(1) & L & q'_m(t) & L & q'_m(t+f) \end{bmatrix} \quad (15)$$

2.6 Particle Swarm Optimization MSVM Parameters

The particle swarm optimization (PSO) algorithm is applied to optimize the MSVM parameters. The specific algorithm process flow diagram is shown in Fig. 2.

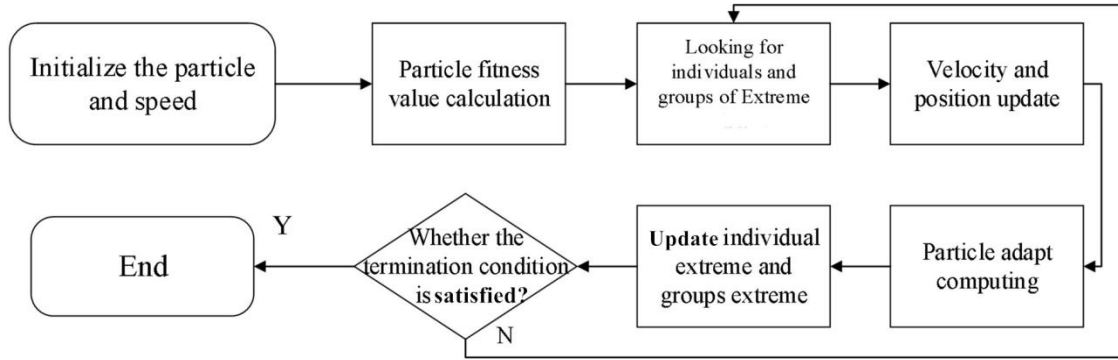
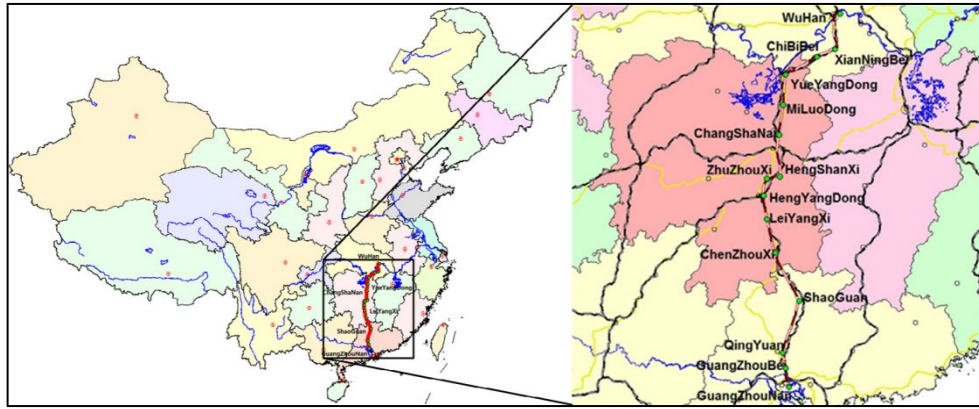


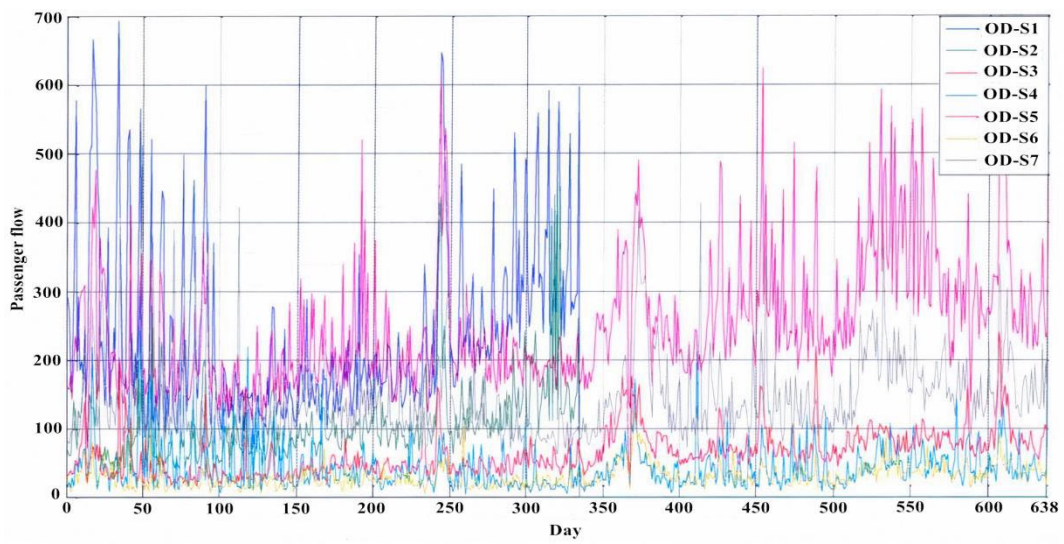
Fig. 2. The specific algorithm process flow of optimizing the MSVM parameters by PSO

3. Data and case study

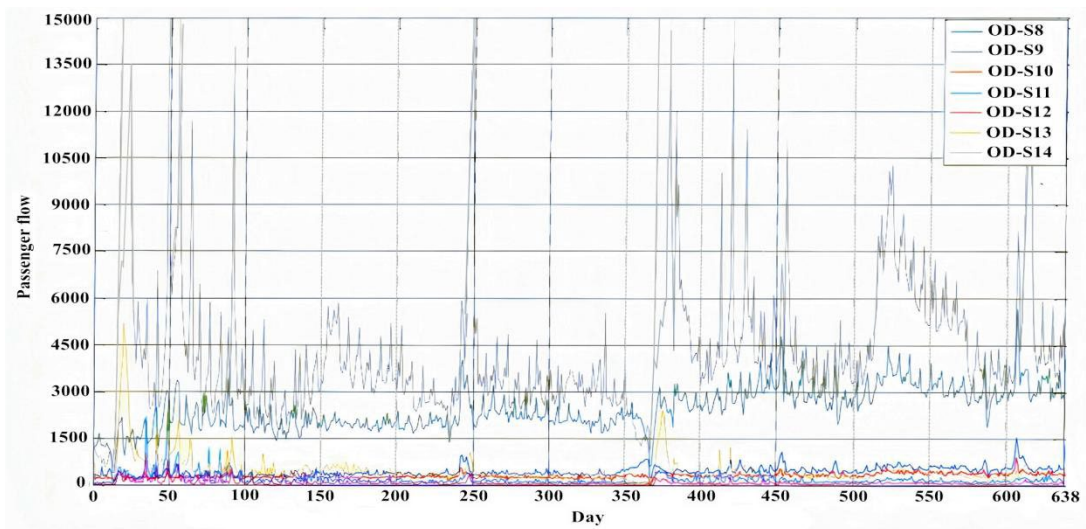
To compare with the earlier study, this paper also takes the same Wuhan-Guangzhou HSR and data, collected from January 1, 2010 to October 31, 2011, to analyze the short-term demand of Wuhan-Guangzhou high-speed railway along with 15 stops, namely Wuhan, Xianning, Chibi, Yueyang, Milluo, Changsha, Zhuzhou, Hengshan, Hengyang, Leiyang, Chenzhou, Shaoguan, Qingyuan, Guangzhoubei, and Guangzhounan. There are 14 OD pairs, starting from Wuhan along the Wuhan-Guangzhou HSR line, as shown in Table 1. In the Wuhan-Guangzhou high-speed railway, the fluctuation of the passenger flow data of 14 sections from Wuhan to other stops is shown in Fig. 3(b) and Fig. 3(c). With the continuous superposition of customers, the passenger flow of 14 OD pairs shows an exponential growth.



(a)



(b)



(c)

Fig. 3. Daily passenger flows of Wuhan-Guangzhou HSR Sections

Table 1

14 OD Pairs starting from Wuhan along the Wuhan-Guangzhou HSR line.

Number	Code	OD Pair
1	S1	Wuhan-Xianning
2	S2	Wuhan-Chibi
3	S3	Wuhan-Yueyang
4	S4	Wuhan-Miluo
5	S5	Wuhan-Changsha
6	S6	Wuhan-Zhuzhou
7	S7	Wuhan-Hengshan
8	S8	Wuhan-Hengyang
9	S9	Wuhan-Leiyang
10	S10	Wuhan-Chenzhou
11	S11	Wuhan-Shaoguan
12	S12	Wuhan-Qingyuan
13	S13	Wuhan-GuangzhouBei
14	S14	Wuhan-GuangzhouNan

3.1 Correlation Analysis Results

Based on the original passenger flow data of 14 OD of the Wuhan-Guangzhou high-speed railway, this paper analyzes the correlation between the various OD using the multidimensional scaling method. SPSS calculates the correlation coefficient between any two data in the n-section passenger flow. The correlation coefficient indicates the mutuality intensity between the cross-section passenger flow data. The larger the coefficient, the stronger the correlation, the smaller the contrary. The correlation coefficient matrix consisting of these correlation coefficients is similar to the original data matrix, and the multidimensional scaling method is used to analyze the correlation coefficient matrix. After the calculation, SPSS gets five outputs: the original distance matrix, the iterative process of the classical solution and the pressure index value, the coordinates of the construction points in the two-dimensional space, and the distance matrix of the optimal scale. The output results are analyzed as follows. Table 2 shows the output of the original distance matrix and Table 3 illustrates the iterative process of the output of the classical solution and pressure index values.

In Table 3, Young's pressure index value is 0.03456, and K pressure index is 0.05983, which are less than 0.10 and $RSQ=0.98569$; this illustrates the good effect of the model. The optimal scaling distance matrix is shown in Table 4.

Table 2

The original distance matrix.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0.0													
2	25.5	0.0												
3	31.3	37.2	0.0											
4	28.1	27.6	30.2	0.0										
5	46.3	52.9	26.3	50.4	0.0									
6	41.4	47.1	21.0	42.2	17.9	0.0								
7	38.7	37.7	50.9	35.9	68.4	61.7	0.0							
8	29.9	33.9	19.5	27.0	34.5	27.7	44.3	0.0						
9	32.6	31.0	36.8	23.4	55.1	46.5	33.9	30.0	0.0					
10	32.8	33.4	32.2	29.0	45.7	37.2	39.6	25.5	26.7	0.0				
11	45.4	42.5	59.9	41.7	75.8	68.9	30.9	51.7	37.8	41.0	0.0			
12	45.4	43.0	61.2	42.5	77.3	70.0	31.8	52.8	37.8	41.2	16.0	0.0		
13	45.7	42.6	60.8	42.9	76.8	69.8	33.2	51.9	37.8	42.0	26.1	21.9	0.0	
14	43.1	37.0	55.4	37.2	71.7	64.4	32.5	46.7	32.3	36.8	28.2	25.1	17.3	0.0

Table 3

Classical solutions of the iterative process and pressure index value output.

Iteration history for two-dimensional solutions (in square distances)		
Young's S-stress formula (1) is used		
Iteration	S-stress	Improvement
1	0.0542	
2	0.0413	0.0128
3	0.0384	0.0029
4	0.0367	0.0017
5	0.0355	0.0012
6	0.0346	0.0010

Iterations stopped because S-stress improvement is less than 0.001

Stress and squared correlation (RSQ) in distances

RSQ values are the proportion of variance of the scaled data (disparities) is accounted for by their corresponding distances.

Stress values are Kruskal's stress formula (1).

For matrix

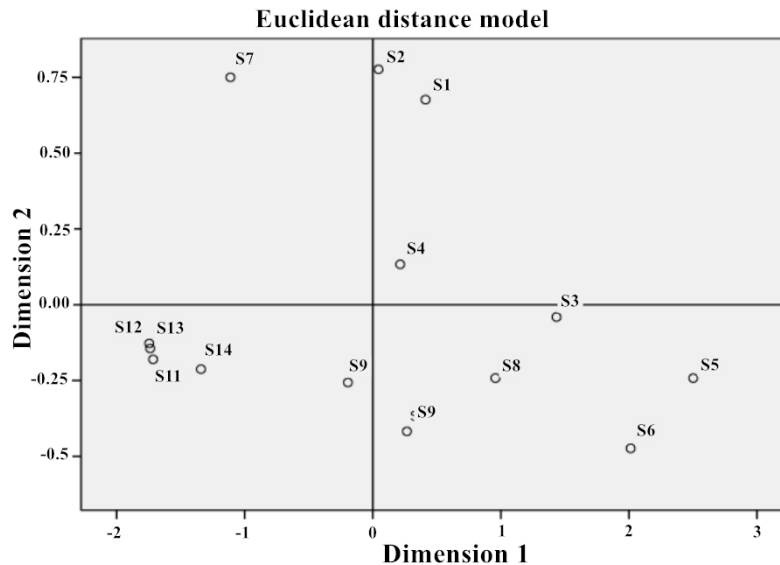
Stress = 0.0598 RSQ = 0.9857

Table 4

The optimal scaling distance matrix.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0.00													
2	0.50	0.00												
3	1.13	1.57	0.00											
4	0.75	0.75	1.13	0.00										
5	2.28	2.69	0.75	2.33	0.00									
6	1.98	2.33	0.50	1.98	0.50	0.00								
7	1.57	1.57	2.67	1.50	3.74	3.36	0.00							
8	1.07	1.37	0.50	0.75	1.50	0.75	2.28	0.00						
9	1.13	1.13	1.57	0.50	2.70	2.28	1.36	1.13	0.00					
10	1.13	1.22	1.13	0.75	2.28	1.57	1.81	0.51	0.75	0.00				
11	2.28	1.99	3.15	1.98	4.22	3.74	1.13	2.67	1.57	1.98	0.00			
12	2.28	1.99	3.18	1.98	4.25	3.78	1.13	2.69	1.57	1.98	0.06	0.00		
13	2.28	1.99	3.17	1.99	4.24	3.77	1.13	2.69	1.57	1.98	0.51	0.50	0.00	
14	1.99	1.57	2.78	1.57	3.84	3.37	1.13	2.30	1.13	1.57	0.75	0.50	0.40	0.00

The passenger flow data of two-dimensional fitting composition under the coordinate figure perception and Euclidean distance linear fitting scatterplot of 14 OD of Wuhan to other stations are shown in Fig. 4 and Fig. 5 respectively.

**Fig. 4.** Section passenger data coordinate perceptual maps

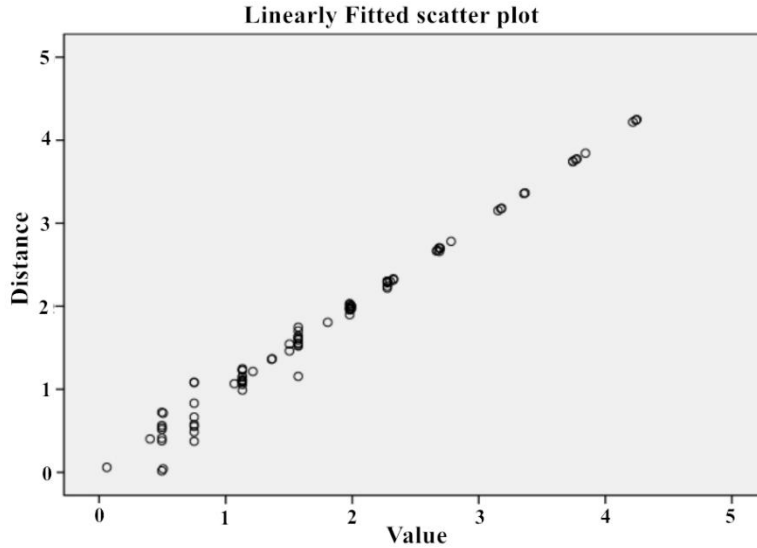


Fig. 5. Section passenger flow under the Euclidean distance of linear scatterplot

The linear fitting scatters diagram shows that the fitting scale of the Euclidean distance adopted by this paper is very consistent with the original distance matrix. Therefore, the use of Euclidean distance is valid.

The section can intuitively show the graph of the data coordinates of the passenger flow data. The OD of S1, S2, and S4 are in the first quadrant, the OD of S7 is in the second quadrant, the sections of S9, S11, S12, S13, and S14 belong to the third quadrant, and the sections of S3, S5, S6, S8, and S10 are located in the fourth quadrant. According to the algorithms of multidimensional scaling, the correlation of points in the same quadrant or short distance is the biggest. So, according to the related degree of the original 14 sections, they are divided into four groups.: The first group includes three sections, S1, S2, and S4; the second group involves one section, S7; the third group contains five sections, S9, S11, S12, S13, and S14; the fourth group comprises five sections, S3, S5, S6, S8, and S10.

In this paper, the fourth group is selected to carry out multivariate passenger flow prediction; the subjects include S3, S5, S6, S8, and S10 five cross-section under test.

3.2 Multivariate of Short-Term Passenger Flow Forecasting

Using the EEMD mode decomposition technique for sections of the passenger demand mode, the short-term traffic flow of multiple ODs is analyzed using MATLAB programming. The sections-EEMD decomposition result includes eight components and one trend item in each section. In the following analysis, we predict the passenger flow of Wuhan to other stations as the obvious example, through the single-section prediction and the multivariate prediction. To make the fluctuations relationship clear between these sections, this paper compared them for each component, with the trend terms and the fluctuation of the original sequence by EEMD decomposed. Because the image is large,

this paper just selects the original passenger flow chart, the comparison charts of IMF1, IMF2 and Res of each section, as shown in Fig. 6- 9.

Although the wave strength of each section is not completely consistent, the law of fluctuations is completely consistent at any time frame, indicating that the five ODs—Wuhan-Yueyang, Wuhan-Changsha, Wuhan-Zhuzhou, Wuhan-Hengyang, Wuhan-Chenzhou—of the passenger flow has a certain correlation. High-frequency component IMF1 is realized by the characteristics of the oscillation of a three-day period or so; high-frequency component IMF2 is realized by the characteristics of the oscillation of a seven-day period or so; and they depict the details of sequence variation high oscillation, in the period of spring, holidays and summer vacation travel, reflecting the high status of these periods of passenger transport. Figures 6 and 7 show the similar volatility characteristics of each section on a three- and seven-day period. IMF component IMF3 and IMF4 realized the oscillation characteristics of a 15-day and 40-day period that are national legal holiday passenger flow fluctuations (effect of a cycle of a 7-day holiday for the daily passenger flow is 5 days before the holiday to the holiday 3 days, making a total of 15 days; the effect of the Spring Festival holiday period of the daily passenger flow is 15 days before the holiday to the festival 25 days, a total of 40 days), reflecting the difference in mechanism of the peak value emergence during holidays and the spring/summer holiday; this is consistent with the daily fluctuations of the law, and to some extent, it represents the fluctuation of the tourist flow. The low-frequency component IMF5 is realized by the unstable period, mainly concentrated in about 66 days and 74 days. The component IMF6 is realized by the shock characteristics of a 167-day period, the component IMF7 is realized by the shock characteristics of a 222-day period, and the component of IMF8 realized the cycle is not stable, no concentration. Because the influence factors of the fluctuation of the passenger flow are complex and an unfavorable observation, IMF5 component-IMF8 component cannot accurately describe what factors lead to the dynamic passenger flow, but they still have the energy of the original sequence. The components of IMF3-IMF8 are not shown in the text.

From Fig. 6 to Fig. 9 the original passenger flow and the volatility of the remaining items in the comparison chart can be seen. The fluctuation characteristics of each section are the same for roughly the same period, so it can be proved that the passenger flow of the six sections namely Wuhan-Yueyang, Wuhan-Changsha, Wuhan-Zhuzhou, Wuhan-Hengyang, and Wuhan-Chenzhou has a strong correlation.

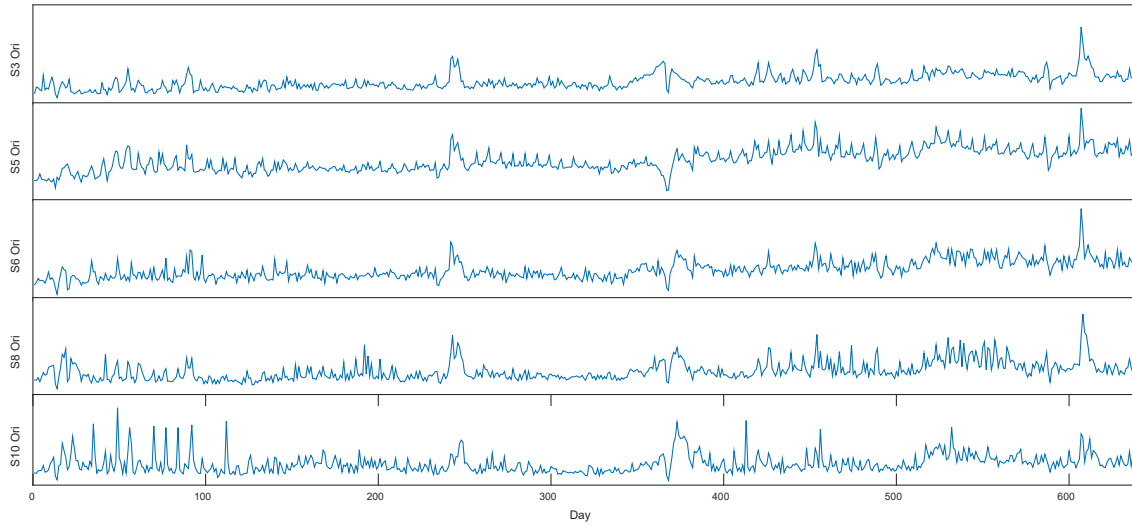


Fig. 6. Comparison chart of the original passenger flow fluctuations in each OD

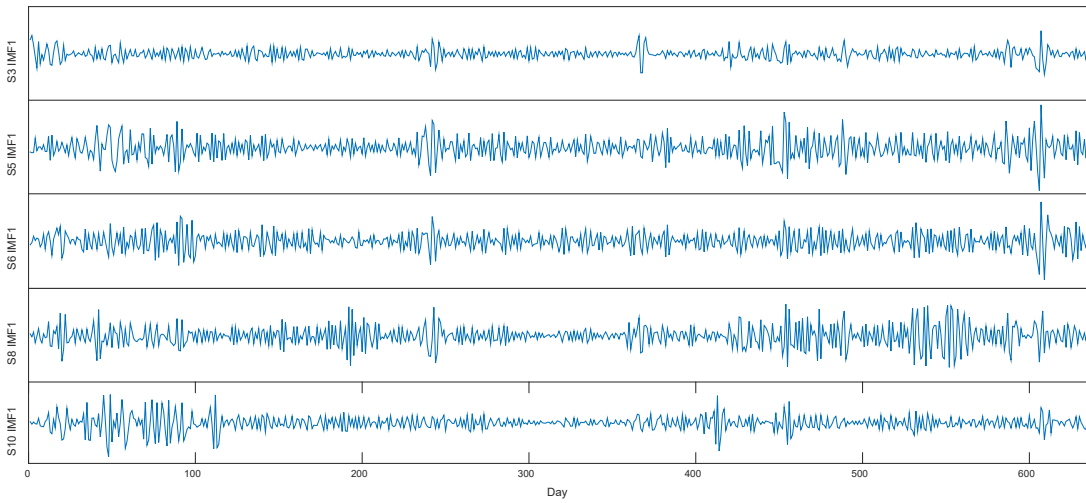


Fig. 7. The comparison chart of fluctuation of IMF1 component in each OD

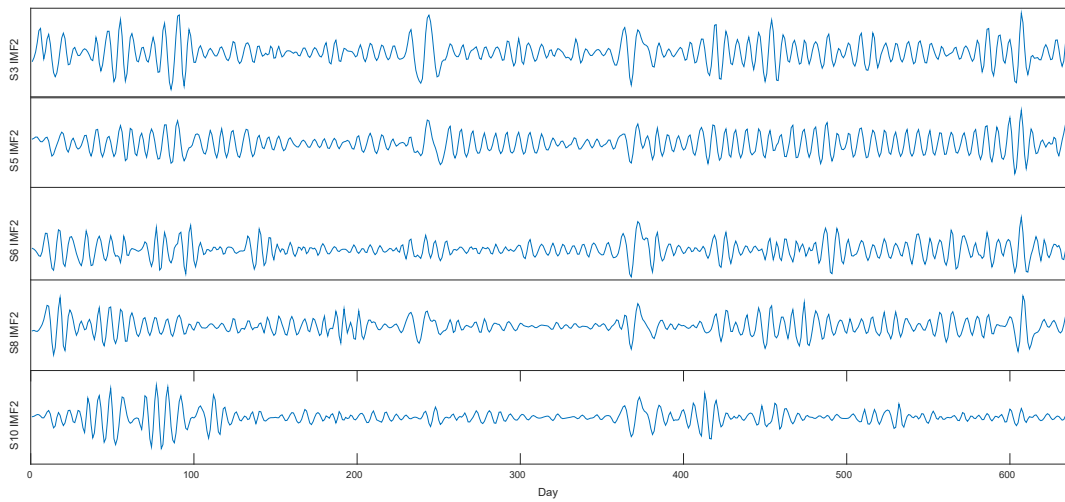


Fig. 8. The comparison chart of fluctuation of IMF2 component in each OD

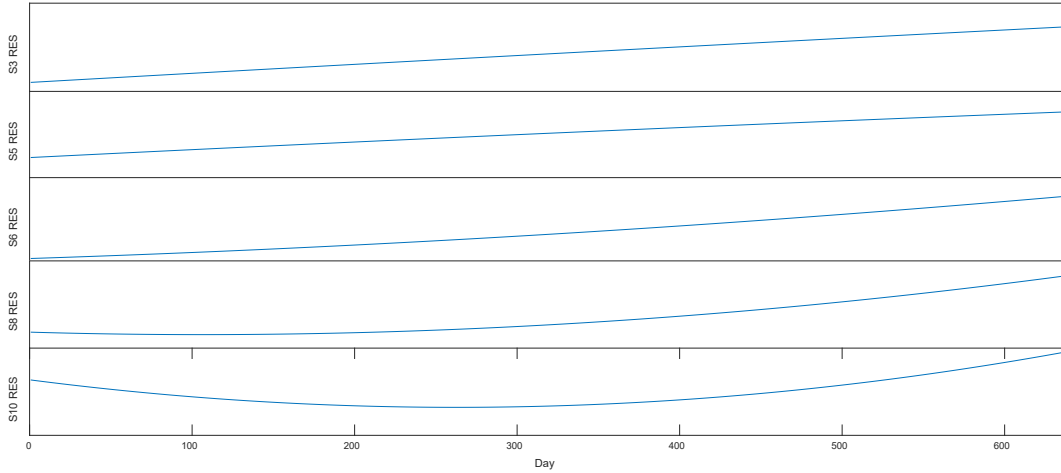


Fig. 9. Comparison of RES remainder fluctuation of each OD

Based on the amalgamated prediction of the constituents and trend components of various observed sections, novel constituents and trend elements for each targeted section are ascertained. Subsequently, gray accumulation-generation procedures are applied to the fresh intrinsic mode function (IMF) constituents and trend components of the respective sections in question. Building upon this foundation, the ultimate prediction error criterion, denoted as the Final Prediction Error (FPE) criterion, is utilized to ascertain the embedding dimension. In this context, the embedding dimension signifies the count of conditional attributes that are pertinent to the support vector machine, as elucidated in Fig. 10. Notably, the optimal setting for the embedding dimension, signifying the number of conditionally attributed factors, is found to be 4 when minimizing the ultimate prediction error, concurrently for both the training and test datasets.

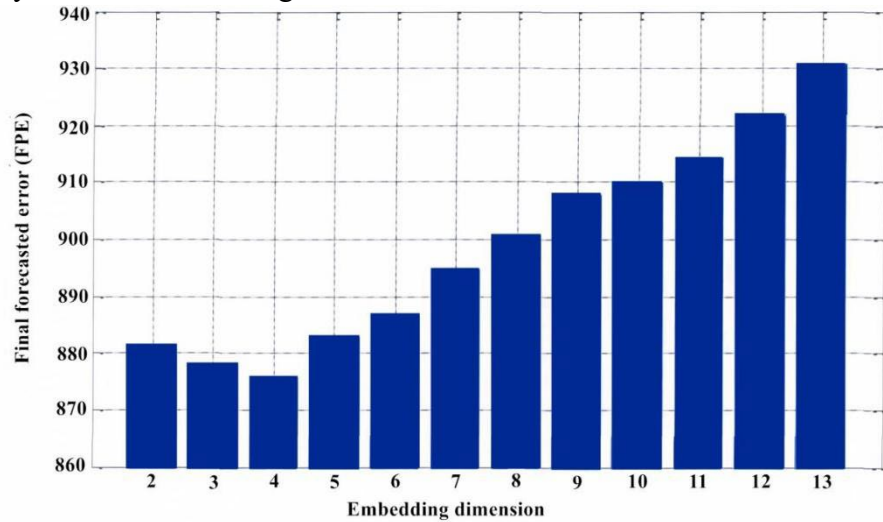


Fig. 10. Final forecasted error standard of FPE

3.3 Parameter Tuning

We adopt the normalized processing to deal with the data of the training set and test set in order to reduce the running time and also use the particle swarm optimization to optimize the model parameters and select the kernel function parameters in the MSVM model. Then, we compute the model parameter and the error with the help of MATLAB. The calculation results of the parameters and errors of each component are shown in Table 5.

The output of the mean square error (MSE) and the Pearson correlation coefficient (r^2) in Table 5 show that the MSVM model has achieved good prediction results and the reliability is strong.

Table 5
The model parameters and error of each component.

IMF		IMF1	IMF2	IMF3	IMF4
Minimum embedding dimension		3	3	3	3
Kernel function-parameters	C	0.1000	19.0434	14.0679	48.0526
	γ	3.6617	0.1000	0.1000	0.8030
Training set	r^2	0.3350	0.9336	0.9975	0.9996
	MSE	0.0080	0.0012	3.8777E-5	5.3373E-6
Test set	r^2	0.2991	0.9429	0.9973	0.9678
	MSE	0.0135	0.0017	9.9371E-4	0.0029
IMF		IMF5	IMF6	IMF7	IMF8
Minimum embedding dimension		3	3	3	3
Kernel function-parameters	C	23.2770	12.4272	9.2349	3.8149
	γ	0.1000	35.0760	0.1000	0.1000
Training set	r^2	0.9999	0.9995	0.9999	0.9998
	MSE	3.5384E-6	2.9779E-5	1.8937E-5	2.4752E-5
Test set	r^2	0.9998	0.9998	0.9994	0.9991
	MSE	1.9836E-5	3.5559E-5	7.4356E-5	1.9830E-4

3.4 EEMD-MSVM Forecast Results

Based on the proposed model, we predict the passenger flow of Wuhan to five stations with high correlation strength (Wuhan-Yueyang, Wuhan-Changsha, Wuhan-Zhuzhou, Wuhan-Hengyang, and Wuhan-Chenzhou). In order to quantitatively study the effect of multivariate forecasting, the sections of passenger were forecasted in a single section and

compared with the proposed method (EEMD-GSVM and EEMD-MSVM). From January 1, 2010, to October 31, 2011, the section of the actual value, passenger flow prediction of the single cross-section and multivariate prediction are needed. By comparing each section passenger flow of raw data, the results of single-section and multivariate passenger flow forecast passenger flow prediction can be obtained.

Compared to single-section passenger flow predictive value, the predictive value of multivariate passenger flow is closer to the actual passenger flow. While compared to single-section passenger flow predictive value, the predictive value of multivariate passenger flow in the same section is closer to the actual passenger flow. The multivariate passenger flows forecasting has higher prediction precision and is more accurate than single sectional passenger flow.

To offer a more intuitive analysis section prediction with a single section of accurate prediction, the prediction error results can be compared to show the merits of the two kinds of forecasting methods. The comparison between results of prediction error indicators uses mean absolute percentage error and mean absolute deviation. Mean absolute percentage error (referred to as MAPE) demonstrates the actual deviation with the actual and estimated values of the absolute value of the actual percentage of the mean values. The average absolute deviation (MAD) demonstrates the actual and estimated values of the mean absolute value of the actual deviation.

Table 6

A comparison of indicators of two types of prediction error method.

	OD pair	MAPE	MAD
Univariate time series forecasting (EEMD-GSVM)	S3	0.2020	118.9446
	S5	0.0578	184.7582
	S6	0.0848	34.9331
	S8	0.1344	42.2542
	S10	0.1207	19.3013
Multivariate time series forecasting (EEMD-MSVM)	S3	0.0627	40.1327
	S5	0.0462	146.7732
	S6	0.0746	30.5758
	S8	0.1145	37.6130
	S10	0.0939	15.3953

Table 6 shows a comparison of indicators of two types of prediction error method. As seen in Table 6, section prediction method has a higher precision than single prediction methods of the area. From MAPE indicators, the multivariate based on short-term

passenger flow prediction value based on the EEMD-MSVM was reduced by 13.94%, 1.16%, 1.01%, 1.98% and 2.67%, compared to the EEMD-SVM single-section short-term prediction value in the sections of S3, S5, S6, S8, and S10 respectively. From MAD (Bokde, N.D. 2020) indicators, the multivariate based on short-term passenger flow prediction value based on the EEMD-MSVM was reduced by 78.81, 37.985, 4.36, 4.6412, and 3.91, compared to the EEMD-SVM single-section short-term prediction value in the sections of S3, S5, S6, S8, and S10 respectively. Based on the analysis results of two types of error index, the prediction accuracy of the multivariate forecasting method compared with single section forecasting method is higher.

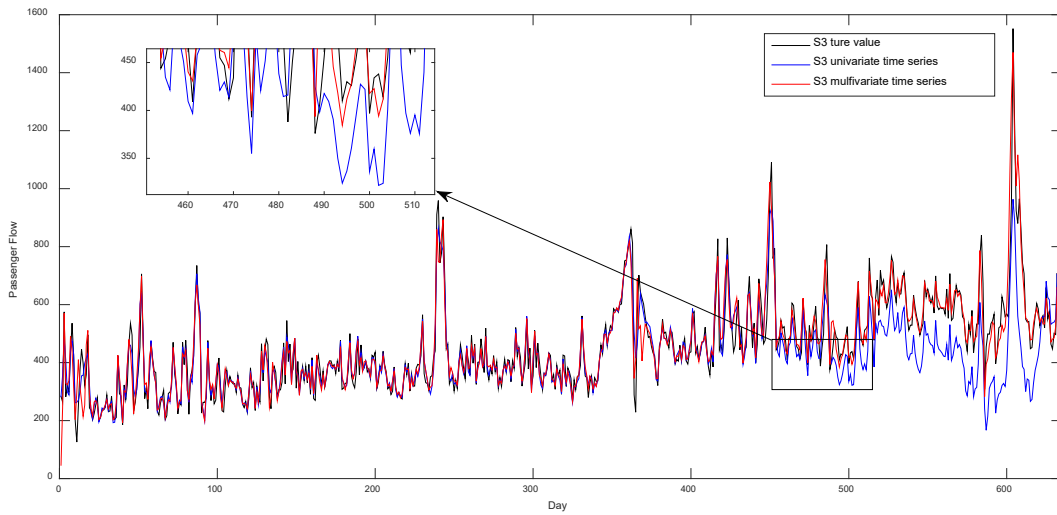


Fig. 11. The comparison of predicted results of S3 section

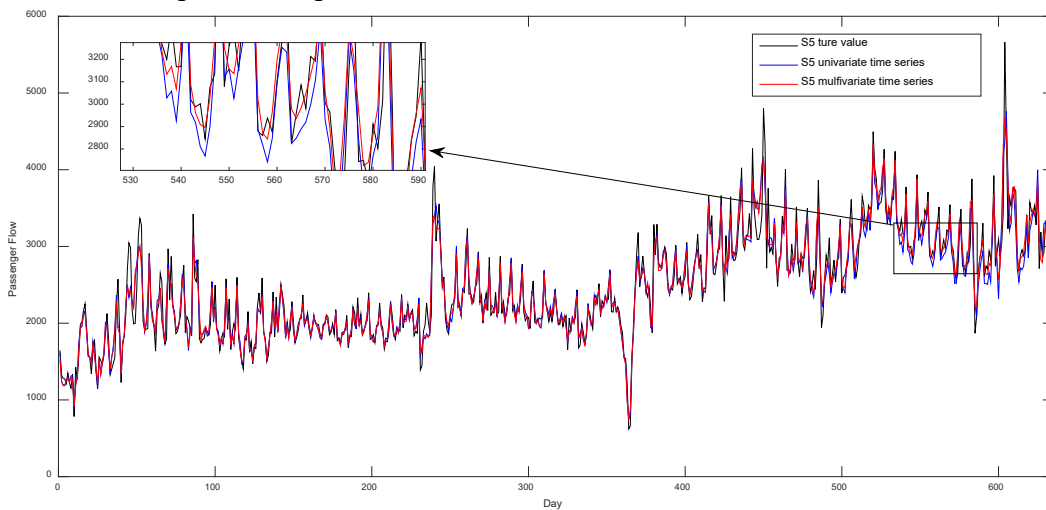


Fig. 12. The comparison of predicted results of S5 section

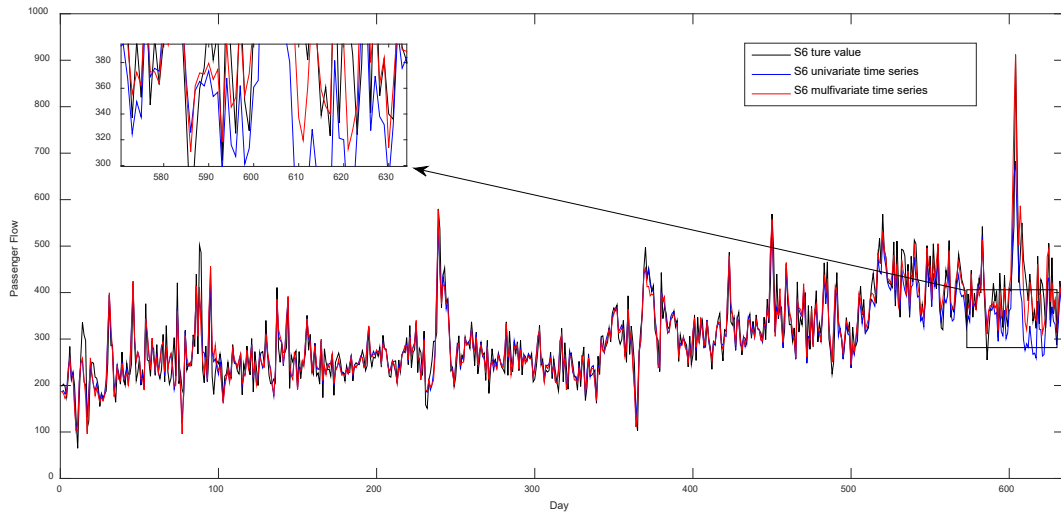


Fig. 13. The comparison of predicted results of S6 section

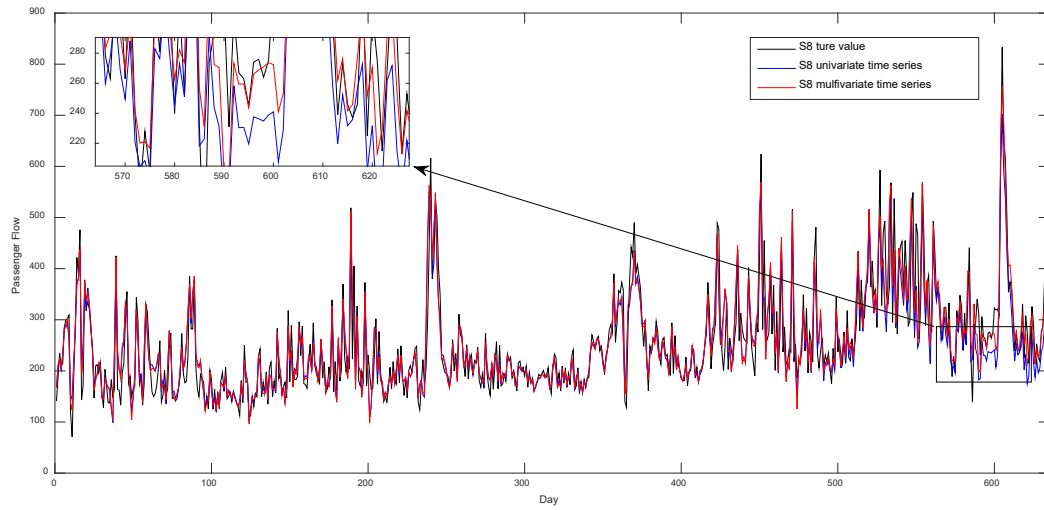


Fig. 14. The comparison of predicted results of S8 section

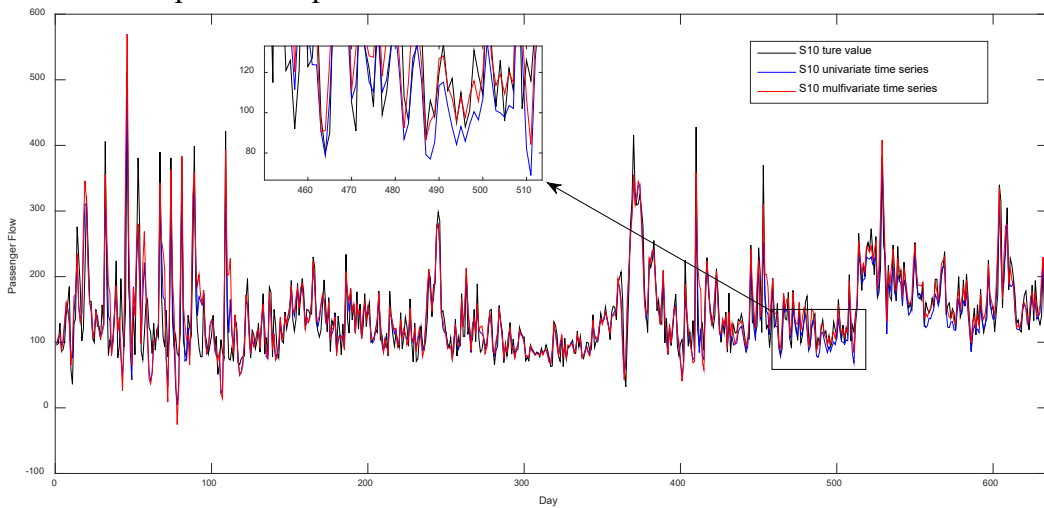


Fig. 15. The comparison of predicted results of S10 section

Fig. 11-Fig. 15 show that in the same day, the prediction results of the multivariate short-term flow prediction model are closer to the actual passenger flow of the day than the predicted results of the univariate short-term passenger flow prediction model, and the prediction of the multivariate short-term passenger flow in the same section is closer to the actual passenger flow of the day than the single-sectional passenger flow forecast. Therefore, the short-term flow prediction accuracy of multivariate is more accurate than the short-term flow prediction accuracy.

4. Conclusion

Drawing from a comprehensive review of the current state of short-term passenger flow prediction in high-speed railways, this paper delves into the theoretical underpinnings of multi-section short-term passenger flow prediction in high-speed railways. It meticulously examines the spatiotemporal dynamics and flux patterns of passenger flow, along with the interplay between different sections of passenger flow. Employing multidimensional scaling techniques, we analyze the multidimensional time series data of passenger flow, elucidating the intricate relationships among high-speed railway passenger flow variables.

Building upon this groundwork, we define the scope of our modeling efforts. Leveraging the principles of multi-section passenger flow prediction in high-speed railways, we devise a novel prediction model based on the EEMD-MSVM methodology. To validate the model's efficacy, we conduct empirical analyses using passenger flow data sourced from the Wuhan-Guangzhou high-speed railway.

In the future, our research agenda will encompass an investigation into the interactions between regional and comprehensive high-speed railway passenger flows. We will dissect the correlation patterns of passenger flow within specific sections, considering the broader context of the entire railway network. Subsequently, we will harness this understanding to refine our passenger flow prediction capabilities.

ACKNOWLEDGEMENTS

This research is supported by the Funds of the National Natural Science Foundation of China (U2034208) and the Key Technologies of digital management and optimization of freight train marshalling plan (N2021X021), Beijing Jiaotong University, China.

Appendix A. Abbreviations

Table A.1

A list of variables.

Variables	Description
d_{ij}	The distance between i and j in the generalized distance matrix
c_{ii}	The similarity coefficient between i and j
X_i	The point corresponding to the variable i in the Euclidean distance matrix constructed on the basis of the generalized distance matrix, $X_i = (X_{i1}, X_{i2}, \dots, X_{ir})$
\hat{d}_{ij}	The distance between i and j in the r -dimensional space.
$q_i(t)$	The original HSR passenger flows of OD i ($i=1,2,\dots,m$) in day t
$iIMFj$	The IMF j of OD i is defined as $iIMFj = [q_{i,IMFj}(1), \dots, q_{i,IMFj}(t)]^T$, $i=1,2,\dots,m, j=1,2,\dots,n$
$iRes$	The trend term of OD i is represented by $iRes = [q_{i,Res}(1), \dots, q_{i,Res}(t)]^T$, $i=1,2,\dots,m$.
$k(\mathbf{x}_i, \mathbf{x})$	The radial basis kernel function, where the vector parameter \mathbf{x} represents the sample information in the prediction model.
σ	Kernel function parameters.
Res'_{ie}	The prediction results of trend term matrix
$q'_{i,Res}(t)$	The prediction value of $q_{i,Res}(t)$.
$\mathbf{q}'_{i,T_{t+f}}$	The forecasting vector of OD i for period T_{t+f}
$q'_i(t)$	The passenger flows prediction result of OD i ($i=1,2,\dots,m$) in day t .
$\mathbf{Q}'_{T_{t+f}}$	The final prediction results matrix

References

- Abramović, B. 2015. Analysis of the mobility of railway passenger transport in small urban areas. *WIT Transactions on The Built Environment*, 146, 665-674, doi: 10.2495/UT150541
- Adamowski, J., Chan, F.H., Prasher, S.O., Ozga-Zielinski, B., Sliusarieva, A., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural

- network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research* 48(1), 273-279. <https://doi.org/10.1029/2010WR009945>
- Ahmed, M.S., Cook, A.R., 1979. Analysis of freeway traffic time-series data by using Box-Jenkins techniques. *Transportation Research Record: Journal of the Transportation Research Board* 722, 1-9.
- Aijaz, I., & Agarwal, P. 2020. A study on time series forecasting using hybridization of time series models and neural networks. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, 13(5), 827-832, <https://doi.org/10.2174/1573401315666190619112842>
- Aljuaydi, F., Wiwatanapataphee, B., & Wu, Y. H. (2022). Multivariate machine learning-based prediction models of freeway traffic flow under non-recurrent events. *Alexandria Engineering Journal*, <https://doi.org/10.1016/j.aej.2022.10.015>.
- Bokde, N.D.; Yaseen, Z.M.; Andersen, G.B. 2020 ForecastTB—An R Package as a Test-Bench for Time Series Forecasting—Application of Wind Speed and Solar Radiation Modeling. *Energies* 2020, 13, 2578. <https://doi.org/10.3390/en13102578>
- Bokde, N., Andrés F., Nadhir A., Siyu T., and Zaher M. Y., 2020. The Hybridization of Ensemble Empirical Mode Decomposition with Forecasting Models: Application of Short-Term Wind Speed and Power Modeling. *Energies* 13, no. 7: 1666. <https://doi.org/10.3390/en13071666>
- Boxill SA, Yu L. 2000. An evaluation of traffic simulation models for supporting ITS development. Centre for Transportation Training and Research, Texas Southern University, Houston, TX; 2000.
- Butkevičius, J., Mazūra, M., Ivankovas, V., Mazūra, S., 2004. Analysis and forecast of the dynamics of passenger transportation by public land transport. *Transport* 19(1), 3-8. doi: 10.1080/16484142.2004.9637944.
- Bi, J. W., Han, T. Y., & Li, H. 2022. International tourism demand forecasting with machine learning models: The power of the number of lagged inputs. *Tourism Economics*, 28(3), 621-645, DOI: 10.1177/1354816620976954.
- Brenner, A., Wu, M., & Amin, S. 2022. Interpretable Machine Learning Models for Modal Split Prediction in Transportation Systems. *arXiv preprint arXiv:2203.14191*, <https://doi.org/10.48550/arXiv.2203.14191>
- Cai, P., Wang, Y., Lu, G., Chen, P., Ding, C., Sun, J., 2016. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. *Transportation Research Part C: Emerging Technologies* 62, 21-34. <https://doi.org/10.1016/j.trc.2015.11.002>
- Chen, C., Wang, Y., Li, L., Hu, J., Zhang, Z., 2012. The retrieval of intra-day trend and its influence on traffic prediction. *Transportation Research Part C: Emerging Technologies* 22, 103-118. <https://doi.org/10.1016/j.trc.2011.12.006>
- Chen, X., Wang, Z., Hua, Q., & Shang, W. (2023). AI-Empowered Speed Extraction via Port-Like Videos for Vehicular Trajectory Analysis. *IEEE Transactions on Intelligent Transportation Systems*. 24 (4), 4541 - 4552. doi: 10.1109/TITS.2022.3167650.

- Chen, X., Liu, S., Liu, R., Wu, H., Han, B., & Zhao, J., (2022). Quantifying Arctic oil spilling event risk by integrating an analytic network process and a fuzzy comprehensive evaluation model. *Ocean & Coastal Management*, 228, 106326. <https://doi.org/10.1016/j.ocecoaman.2022.106326>.
- Chuwang, D. D., & Chen, W. (2022). Forecasting Daily and Weekly Passenger Demand for Urban Rail Transit Stations Based on a Time Series Model Approach. *Forecasting*, 4(4), 904-924.
- Cox, T.F., and Cox, M.A.A., 2000. *Multidimensional scaling, Second Edition*. CRC press.
- El Yaagoubi, A., Ferjani, A., Essaghir, Y., Sheikahmadi, F., Abourraja, M. N., Boukachour, J., Baron, M-L., Duvallet C. & Khodadad-Saryazdi, A. (2022). A logistic model for a French intermodal rail/road freight transportation system. *Transportation Research Part E: Logistics and Transportation Review*, 164, 102819, <https://doi.org/10.1016/j.tre.2022.102819>.
- Faraway, J., Chatfield, C., 1998. Time series forecasting with neural networks: A comparative study using the airline data. *Journal of the Royal Statistical Society: Series C: Applied Statistics* 47(2), 231-250. <https://doi.org/10.1111/1467-9876.00109>
- Harvey, A.C., 1990. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
- Hamed, M.M., Al-Masaeid, H.R., Said, Z.M.B., 1995. Short-term prediction of traffic volume in urban arterials. *Journal of Transportation Engineering* 121(3), 249-254. [https://doi.org/10.1061/\(ASCE\)0733-947X\(1995\)121:3\(249\)](https://doi.org/10.1061/(ASCE)0733-947X(1995)121:3(249))
- Himanen, V., Nijkamp, P., & Reggiani, A. (Eds.). 2019. *Neural networks in transport applications*. Routledge.
- Hong Y, Yuanxun C, Guohui L, 2022. A new traffic flow prediction model based on cosine similarity variational mode decomposition, extreme learning machine and iterative error compensation strategy, *Engineering Applications of Artificial Intelligence*, Volume 115, 105234, ISSN 0952-1976, <https://doi.org/10.1016/j.engappai.2022.105234>.
- Huang, N. E. . 2015. *An Adaptive Approach for Nonlinear and Nonstationary Data Analysis*. ISBN: 978-981-4579-92-6
- Hui H., Jianfeng Z., Tao L., 2020. A Comparative Study of VMD-Based Hybrid Forecasting Model for Nonstationary Daily Streamflow Time Series, *Complexity*, vol. 2020, Article ID 4064851, 21 pages, 2020. <https://doi.org/10.1155/2020/4064851>
- Jenelius, E., Koutsopoulos, H.N., 2018. Urban Network Travel Time Prediction Based on a Probabilistic Principal Component Analysis Model of Probe Data. *IEEE Transactions on Intelligent Transportation Systems* 19, 436-445. <https://doi.org/10.1109/TITS.2017.2703652>
- Jiang, X., 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. <https://doi.org/10.1016/j.trc.2014.03.016>.
- Joachims, T., 1998. Making large-scale SVM learning practical. In: Schölkopf, B., Burges, C. and Smola, A.(eds). *Advances in kernel methods-support vector learning*. MIT Press, Cambridge, MA, pp. 42–56. doi :10.17877/DE290R-14262

- Kruskal, J. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1), 1-27. 10.1007/BF02289565
- Kruskal, J. B. 1964. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29.
- Kumar, S. V., & Vanajakshi, L. 2015. Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *European Transport Research Review*, 7(3), 1-9, <https://doi.org/10.1007/s12544-015-0170-8>
- Kumar, K. 2023. Deep Bi-LSTM Neural Network for Short-Term Traffic Flow Prediction Under Heterogeneous Traffic Conditions. In *Recent Advances in Transportation Systems Engineering and Management* (pp. 597-611). Springer, Singapore.
- Kamarianakis, Y., Prastacos, P., 2003. Forecasting traffic flow conditions in an urban network: comparison of multivariate and univariate approaches. *Transportation Research Record: Journal of the Transportation Research Board* 1857, 74-84. doi:10.3141/1857-09. <https://dx.doi.org/10.3141/1857-09>
- Lee, S., Fambro, D., 1999. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. *Transportation Research Record: Journal of the Transportation Research Board* 1678, 179-188. doi:10.3141/1678-22. <https://dx.doi.org/10.3141/1678-22>
- Lim, C., McAleer, M., 2002. Time series forecasts of international travel demand for Australia. *Tourism Management* 23(1), 389-396. [https://doi.org/10.1016/S0261-5177\(01\)00098-X](https://doi.org/10.1016/S0261-5177(01)00098-X)
- Nguyen-Tuong, D., Seeger, M., Peters, J., 2008. Computed torque control with nonparametric regression models. In: *Proceedings of the American Control Conference*, pp. 212-217. <https://doi.org/10.1109/ACC.2008.4586493>
- Nijkamp, P., Reggiani, A., Tritapepe, T., 1996. Modelling inter-urban transport flows in Italy: A comparison between neural network analysis and logit analysis. *Transportation Research Part C: Emerging Technologies* 4(6), 323-338. [https://doi.org/10.1016/S0968-090X\(96\)00017-4](https://doi.org/10.1016/S0968-090X(96)00017-4)
- Pandey, P., Bokde, N. D. , Dongre, S. , & Gupta, R. . 2020. Hybrid models for water demand forecasting. *Journal of Water Resources Planning and Management*, 147(2). 10.1061/(ASCE)WR.1943-5452.0001331
- Poirazi, P., Brannon, T., Mel, B.W., 2003. Pyramidal neuron as two-layer neural network. *Neuron* 37(6), 989-999. [https://doi.org/10.1016/S0896-6273\(03\)00149-1](https://doi.org/10.1016/S0896-6273(03)00149-1)
- Reggiani, A., Nijkamp, P., Tsang, W.F., 2014. European freight transport analysis using neural networks and logit models. *Tinbergen Institute Discussion Papers*, 97-032/3.
- Sibruk, L., & Zakutynskiy, I. 2022. Recurrent Neural Networks for Time Series Forecasting. Choosing the best Architecture for Passenger Traffic Data. *Electronics and Control Systems*, 2(72), 38-44, 10.48550/arXiv.1901.00069
- Shahriari, S., Ghasri, M., Sisson, S. A., & Rashidi, T. 2020. Ensemble of ARIMA: combining parametric and bootstrapping technique for traffic flow prediction. *Transportmetrica A: Transport Science*, 16(3), 1552-1573, <https://doi.org/10.1080/23249935.2020.1764662>

- Stathopoulos, A., Karlaftis, M.G., 2001. Spectral and cross-spectral analysis of urban traffic flows. In: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, pp. 820-825. <https://doi.org/10.1109/ITSC.2001.948766>
- Stathopoulos, A., Karlaftis, M.G., 2003. A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies* 11(2),121-135. [https://doi.org/10.1016/S0968-090X\(03\)00004-4](https://doi.org/10.1016/S0968-090X(03)00004-4)
- Teresa P, Renata Ž, 2023, Estimation and prediction of the OD matrix in uncongested urban road network based on traffic flows using deep learning, *Engineering Applications of Artificial Intelligence*, Volume 117, ISSN 0952-1976, <https://doi.org/10.1016/j.engappai.2022.105550>.
- Tsai, T.H., Lee, C.K., Wei, C.H., 2005. Design of dynamic neural networks to forecast short-term railway passenger demand. *Journal of the Eastern Asia Society for Transportation Studies* 6, 1651-1666. doi: 10.11175/easts.6.1651
- Vaishnav, V., & Vajpai, J. 2018. Seasonal time series forecasting by group method of data handling. In 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) (pp. 1-5). IEEE, doi: 10.1109/SCEECS.2018.8546886.
- Vlahogianni, E.L., Karlaftis, M.G., 2004. Optimization of hybrid neural network predictor for spatial traffic flow data treatment: Genetic phase-space reconstruction. In: Proceedings of the International Conference on Applications of Advanced Technologies in Transportation Engineering, pp. 18-22. doi:10.1061/40730(144)4. [https://doi.org/10.1061/40730\(144\)4](https://doi.org/10.1061/40730(144)4)
- Vlahogianni, E.L., Karlaftis, M.G., Golias, J.C., 2014. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies* 43, 3-19. doi:10.1016/j.trc.2014.01.005. <https://doi.org/10.1016/j.trc.2014.01.005>
- Wang, Y., Papageorgiou, M., 2005. Real-time freeway traffic state estimation based on extended Kalman filter: A general approach. *Transportation Research Part B: Methodological* 39(2), 141-167. doi:10.1016/j.trb.2004.03.003. <https://doi.org/10.1016/j.trb.2004.03.003>
- Whittaker, J., Garside, S., Lindveld, K., 1997. Tracking and predicting a network traffic process. *International Journal of Forecasting* 13(1), 51-61. doi:10.1016/S0169-2070(96)00700-5. [https://doi.org/10.1016/S0169-2070\(96\)00700-5](https://doi.org/10.1016/S0169-2070(96)00700-5)
- Williams, B., 2001. Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modeling. *Transportation Research Record: Journal of the Transportation Research Board* 1776, 194-200. doi:10.3141/1776-25. <https://doi.org/10.3141/1776-25>
- Williams, B.M., Hoel, L.A., 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of Transportation Engineering* 129(6), 664-672. doi:10.1061/(ASCE)0733-947X(2003)129:6(664).
- Yan, W. 2012. Toward automatic time-series forecasting using neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 23(7), doi: 10.1109/TNNLS.2012.2198074.
- Zhang, Q., Basseville, M., Benveniste, A., 1998. Fault detection and isolation in nonlinear dynamic systems: A combined input-output and local approach. *Automatica* 34(11), 1359-1373. doi:10.1016/S0005-1098(98)00085-5. [https://doi.org/10.1016/S0005-1098\(98\)00085-5](https://doi.org/10.1016/S0005-1098(98)00085-5)

- Zheng Q., Yan P., Hamidreza Z., Niya C., 2019, A review and discussion of decomposition-based hybrid models for wind energy forecasting applications, *Applied Energy*, 235, 2019, P939-953, 0306-2619, <https://doi.org/10.1016/j.apenergy.2018.10.080>.
- Zhang, X., Onieva, E., Perallos, A., Osaba, E., Lee, V.C., 2014. Hierarchical fuzzy rule-based system optimized with genetic algorithms for short term traffic congestion prediction. *Transportation Research Part C: Emerging Technologies* 43, 127-142. doi:10.1016/j.trc.2014.02.013. <https://doi.org/10.1016/j.trc.2014.02.013>
- Zhang, Y., Zhang, Y., Haghani, A., 2014. A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model. *Transportation Research Part C: Emerging Technologies* 43, 65-78. doi:10.1016/j.trc.2013.11.011. <https://doi.org/10.1016/j.trc.2013.11.011>