Approach for Short-Term Traffic Flow Prediction Based on Empirical Mode Decomposition and Combination Model Fusion

Zhongda Tian

***Abstract*— Accurate prediction of the traffic state can help to address the issue of traffic congestion, providing guiding advices for people's travel and traffic regulation. In this paper, we propose a novel short-term traffic flow prediction approach based on empirical mode decomposition and combination model fusion. First, we explore the amplitude-frequency characteristics of short-term traffic flow series, and use empirical mode decomposition to decompose traffic flow to several components with different frequency. Second, based on the results of self-similarity analysis of each component, improved extreme learning machine, seasonal auto regressive integrated moving average and auto regressive moving average are selected to predict different components. Meanwhile, an improved fruit fly optimization algorithm is proposed to optimize the weight coefficient of the combination model. Third, the prediction results of each prediction model are multiplied by their respective weight coefficient to get the final prediction results. We evaluate our prediction approach by doing thorough experiment on a real traffic data set. Moreover, experimental results show that the proposed approach has superior performance than state-of-the-art prediction methods or models in short-term traffic flow prediction.**

***Index Terms*— Short-term traffic flow, prediction, empirical mode decomposition, combination model fusion, improved fruit fly optimization algorithm.**

I. Introduction

With the rapid development of the city, traffic accidents are caused by traffic congestion, low traffic efficiency and other factors, which have become a common problem faced by many cities [1]. Traffic congestion leads to social, economic and environmental problems, rationale for which public and private organizations have attempted at addressing them for more than 50 years [2]. Through traffic flow prediction, the timely adjustment of trunk roads has become one of the important ways to solve the problem of urban traffic congestion. Traffic flow prediction can effectively alleviate traffic congestion, reduce the accident rate and give travelers a comfortable and safe traffic environment [3], [4]. According to the different prediction steps, traffic flow prediction can be divided into long-term traffic flow prediction and short-term traffic flow prediction. Because traffic flow control needs real-time, short-term prediction is more important in practice. Meanwhile, the short-term traffic flow is random and affected by traffic light period and type, and the numbers of lanes connected, so it is more difficult to model and predict [5].

The short-term traffic flow prediction has always been a hot topic studied by many scholars. Many methods and models are proposed and applied to traffic flow prediction. At present, many scholars have pointed out that the combination prediction model has better prediction effect than the single prediction model for traffic flow [6], [7]. This paper will also focus on the research of the combination prediction approach of short-term traffic flow. However, the currently research results of combination prediction model have the following disadvantages. First, there is blindness in the selection of prediction models. Different prediction models have their own applicable objects. Second, the weight coefficient of the combination prediction model all uses the same value. Namely, the prediction values of different prediction models adopt a direct superposition method. However, different prediction models have different effects on the final prediction value. The different weight coefficient should be assigned to different models. This paper will carry out the corresponding research for the above disadvantages. The main contributions of this research are as follows.

1. Empirical mode decomposition (EMD) is introduced to decompose the original short-term traffic flow time series. The Hurst exponent is used to determine the suitable prediction model for each decomposed component.
2. Improved extreme learning machine (ELM), seasonal auto regressive integration moving average (SARIMA), and auto regressive moving average (ARMA) is chosen as the prediction model.
3. An improved fruit fly optimization algorithm (FOA) is introduced to optimize the weight coefficient of each prediction model.

The overflow of the proposed approach is as shown in Fig. 1.

The rest contents of this paper are as follows. Section II summarizes the literature review on the recent popular techniques for the prediction of traffic flow. Section III introduces the data preprocess. Section IV introduces traffic flow decomposition method. Section V introduces the proposed prediction approach. The case studies and results analysis are provided in Section VI. The discussions are presented in Section VII. The conclusions are provided in Section VIII.

Manuscript received March 3, 2019; revised September 8, 2019, February 5, 2020, and April 9, 2020; accepted April 10, 2020. Date of publication May 8, 2020; date of current version September 1, 2021. This work was supported in part by the Science Research Project of Liaoning Education Department under Grant LGD2016009 and in part by the Natural Science Foundation of Liaoning Province under Grant 20170540686. The Associate Editor for this article was R. Malekian.

The author is with the College of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China (e-mail: tianzhongda@126.com).

Digital Object Identifier 10.1109/TITS.2020.2987909

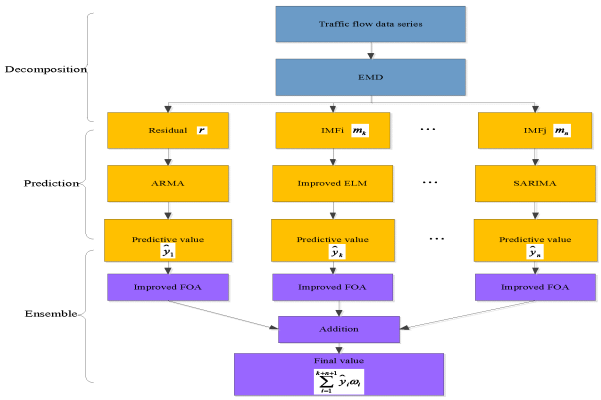

Fig. 1. The overflow of the proposed short-term traffic flow prediction approach.

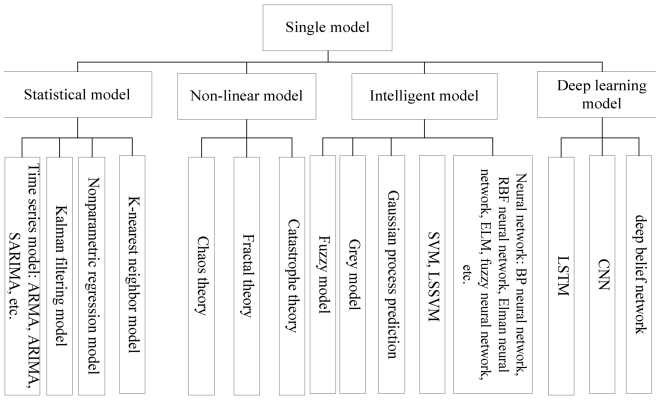

Fig. 2. The single prediction model for traffic flow.

II. Literature Review

At present, the traffic flow prediction models can be divided into single model and combination model. The single model can be roughly divided into the following four types as shown in Fig. 2.

Although these single prediction models or methods have achieved good prediction results, they also have their own disadvantages and limitations. These single traffic flow prediction models have their unique information characteristics and applicable conditions, which can only reflect the future situation from different aspects. Due to the complexity and uncertainty of the traffic system, the prediction results of a single prediction model cannot fully reflect the essential characteristics of the road traffic situation and traffic flow, which is one-sided. Therefore, some scholars have presented the combination models. Combination prediction models can be divided into three kinds: method-based combination, model-based combination and decomposition-based combination model.

1. The method-based combination model refers to incorporate a certain method into a single model to predict traffic flow and improve the prediction performance of the model. This optimization method cannot be used for traffic flow prediction alone, but can only improve the performance of the original model. These achievements include seasonal relevance vector regression with the hybrid chaotic simulated annealing algorithm [8], hybrid PSO-SVR [9], SVR and the Bayesian optimization [10], GA-PSO-LSSVM [11], fuzzy wavelet neural network and biogeography-based optimization [12], LSSVR with the Gaussian kernel function [13], and etc.
2. The model-based combination model refers to a new model composed of multiple models, such as neural network combined with the K-nearest neighbor [14], DNN and BTF [15], ARIMA and GARCH [16], layerwise structure and the Markov transition matrix [17], classification algorithm and neural network [18], etc.
3. Decomposition-based combination model. The traffic flow is decomposed to produce different components. These different components are predicted by the same or different models, and the predicted values of each component are superimposed. These decomposition methods include empirical mode decomposition (EMD) [19]–[21], principal component analysis [22], [23], dynamic tensor completion [24], spectral analysis [25], [26], wavelet methods [27], [28], variational mode decomposition [29], and etc. These methods can decompose the original traffic flow into many components with different characteristics. Then the appropriate prediction model is used to predict each component. Finally, the final prediction value is obtained by reconstruction.

Combination model has made good use of the advantages of various methods and models, so it has achieved good prediction results. However, due to the influence of weather conditions, holidays, emergencies and other random interference factors, the selection of prediction model and modeling method is still the current research hotspot.

III. Data Preprocess

In order to show the effectiveness of the proposed prediction approach, this paper takes the intersection as the research object. Segments and intersections are two major parts of the road network that carries traffic demands. The junctions of several segments form an intersection. The most typical intersection is the cross road. Compared with segments, intersections are more complex. The traffic flow at intersections is more random and is affected by traffic light period and type, and the number of lanes connected, so it is more difficult to model and predict.

In some previous research results, it is necessary to consider the current traffic flow in all directions of the intersections, as well as the traffic flow of several adjacent intersections. Many variables, such as the period and type of signal light, need to be considered, which increases the complexity of the prediction model. Meanwhile, there are a lot of coupling relationships between these variables. These variables are used as the input of the prediction model, which will reduce the prediction accuracy of the model. Moreover, some of these variables are difficult to obtain accurately in practice. This paper only considers the lane traffic flow to be predicted, and does not consider other variables.

A. Research Object

A typical traffic flow layout at intersections is shown in Fig. 3. This graph is based on Microsoft's Bing maps.

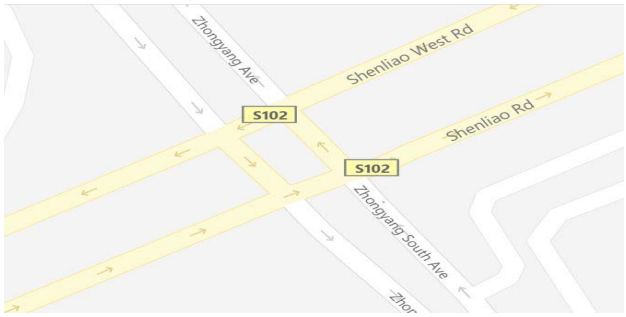

Fig. 3. The layout of intersections studied in this paper.

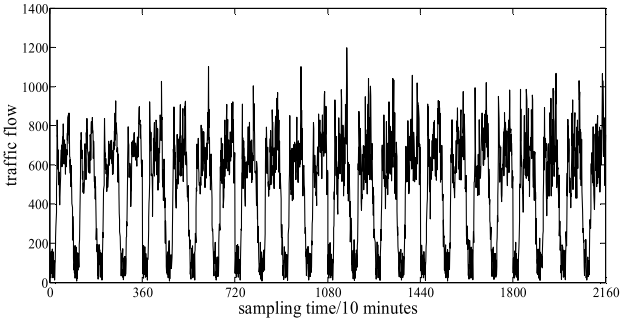

Fig. 4. Actual short-term traffic flow data time series.

The intersection is located at the junction of Shenliao West Road (five lanes dual carriageway, speed limit is 60 km/h, vehicle composition includes cars, trucks, buses, etc.) and Zhongyang Avenue (three lanes dual carriageway, speed limit is 60 km/h, vehicle composition includes cars, trucks, buses, etc.) in Shenyang, Liaoning Province, China.

In the Fig. 3, the traffic flow of Shenliao West Road from east to west is taken as the research object. A total of 15 days of traffic flow data from the above intersection is collected. The sampling period of data is 10 minutes. Therefore, the data sets have 2160 group data. In the actual applications, there are outlier data in the sampled samples due to external interference, sensor failure, communication error, etc. These outlier data will reduce the modeling accuracy, the outlier data need to be removed from the sample data. The Pauta criterion can be used to remove these outlier sample data [30]. The actual short-term traffic flow data processed by the Pauta criterion is as shown in Fig. 4.

B. Data Form of Modeling

Traffic flow can be regarded as a time series. It means that within a certain sampling interval, the future traffic flow can be predicted by the historical data. If the number of samples is large enough, the traffic flow in the future time can be predicted. Suppose that the currently sampling time is t, the traffic flow between the current and the next sampling time is counted by the sensor at the intersections. The traffic flow can be expressed as $T(t)$. The mapping relationship between future traffic flow and historical data is as follows.

$$[T(t+1), \cdots, T(t+p)] = f([T(t), T(t-1), \cdots, T(t-m+1)]) \tag{1}$$

As shown in Eq. (1), traffic flow of training set is collected as $T(t), t = 1, 2, \cdots, N$, N is the length of data. The input matrix $\boldsymbol{X}$ of training set is expressed as the next Eq. (2), as shown at the bottom of this page. $\boldsymbol{X}$ transforms the input traffic flow sequence with $1 \times N$ order into a matrix with $(N - p - m + 1) \times m$ order. The output of training set can be expressed as a matrix $\boldsymbol{Y}$ with $p \times (N - p - m + 1)$ order. $\boldsymbol{Y}$ can be expressed as the Eq. (3), as shown at the bottom of this page.

Therefore, the traffic flow prediction problem in this paper can be expressed as finding the mapping function of the following formula.

$$\boldsymbol{Y}^{\mathrm{T}}(k) = f(\boldsymbol{X}(k)) \tag{4}$$

where k is the row and column number. In the training process, a row of $\boldsymbol{X}$ and a column of $\boldsymbol{Y}$ constitute the input and output to complete a training process. Repeat the process above until the model training is completed.

IV. Traffic Flow Decomposition Method

A. The Characteristics of Short-Term Traffic Flow

In order illustrate the necessity of introducing EMD, this paper analyzes the amplitude-frequency characteristics

$$\boldsymbol{X} = \begin{bmatrix} T(1) & T(2) & \cdots & T(m-1) & T(m) \\ T(2) & T(3) & \cdots & T(m) & T(m+1) \\ T(3) & T(4) & \cdots & T(m+1) & T(m+2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ T(N-p-m) & T(N-p-m+1) & \cdots & T(N-p-2) & T(N-p-1) \\ T(N-p-m+1) & T(N-p-m+2) & \cdots & T(N-p-1) & T(N-p) \end{bmatrix} \tag{2}$$

$$\boldsymbol{Y} = \begin{bmatrix} T(m+1) & T(m+2) & \cdots & T(m+p-1) & T(m+p) \\ T(m+2) & T(m+3) & \cdots & T(m+p) & T(m+p+1) \\ T(m+3) & T(m+4) & \cdots & T(m+p+1) & T(m+p+2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ T(N-p) & T(N-p+1) & \cdots & T(N-2) & T(N-1) \\ T(N-p+1) & T(N-p+2) & \cdots & T(N-1) & T(N) \end{bmatrix}^{\mathrm{T}} \tag{3}$$

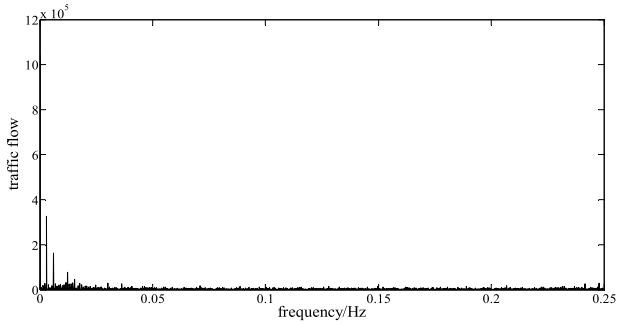

Fig. 5. Amplitude-frequency characteristic of traffic flow time series.

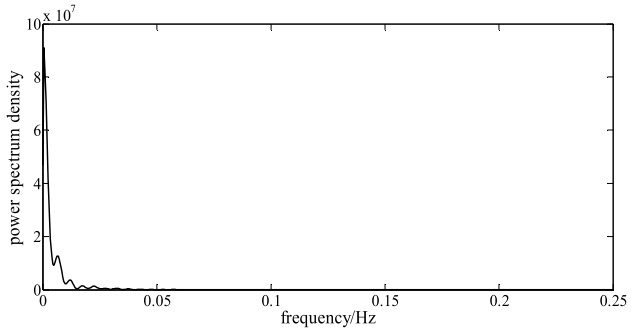

Fig. 6. The power spectrum density of short-term traffic flow.

of short-term traffic flow. Fast Fourier transform method is used to transform the time domain to the frequency domain, and the result of amplitude-frequency characteristic is shown in Fig. 5. Fig. 5 shows that short-term traffic flow is mainly concentrated in the low-frequency part; the high-frequency part of the short-term traffic flow is low. This means that the fluctuation information caused by external factors, or the periodic information of the traffic flow itself, is contained with different frequency variations.

In order to further analyze the characteristics of short-term traffic flow, the power spectrum density is calculated and is as shown in Fig. 6. It can be seen from Fig. 6 that although short-term traffic flow is random, but based on the above analysis results, it can be determined that short-term traffic flow can be decomposed into multiple components with different frequencies. At the same time, most of short-term traffic flow is slow change. The short-term traffic flow signal is nonlinear and non-stationary. Some decomposition methods can reduce the non-stationary of the signal and extract the information of different frequency signals. On this basis, decomposition-based combination prediction method can often improve the prediction effect. Therefore, short-term traffic flow time series is processed first, and the components reflecting the different information of the series are resolved. As a data analysis method, EMD is especially suitable for the decomposition and processing of nonlinear and non-stationary data. Therefore, EMD algorithm is used to decompose the short-term traffic flow signal into high frequency components and low frequency components. These components are predicted by the appropriate prediction method or algorithm. This reduces the influence of the instability of short-term traffic flow on the prediction accuracy.

B. EMD Algorithm

The essence of EMD is to smooth the time series, decompose the original signal into several intrinsic mode functions (IMFs), and determine the basic intrinsic characteristics of the effective signals in the data according to the experiences. It can more effectively reflect the distribution of energy on the space (or time) scale in the physics process. Through EMD decomposition, each IMF component of non-stationary complex signals is stable.

In the previous signal analysis, the actual signal is generally regarded as a linear stationary signal approximately. For example, Fourier spectrum analysis requires that the signal system to be analyzed must be linear and strictly periodic or stable, otherwise, the results obtained by Fourier spectrum analysis will lose physical meaning. EMD decomposes complex data into a limited, usually a small number of IMFs, and uses Hilbert transform to differential the phase to solve the instantaneous frequency, which makes these components have practical physical meaning.

Compared with discrete wavelet transform (DWT), variational mode decomposition (VMD), principal component analysis (PCA), and etc., EMD has some characteristics. It is difficult for DWT to choose proper wavelet basis function and decomposition level, so its adaptive ability is poor. The problem of VMD is that the decomposition level needs to be specified in advance, but it is difficult to determine the optimal decomposition level in practice. EMD is a decomposition method suitable for dealing with non-linear and non-stationary time series. PCA is a linear dimension reduction method which can retain the information content of original data to the maximum extent. EMD is more suitable for signal decomposition, while PCA is suitable for reducing the dimension of data.

Suppose the original time series to be decomposed is x(t). The final results of EMD can be expressed as.

$$x(t) = \sum_{i=1}^{n} c_i(t) + r(t) \tag{5}$$

where, $c_i(t)$ is the ith IMF component, $r(t)$ is residual component. Therefore, EMD can decompose the original signal x(t) into the sum of n different frequencies of IMF and a trend item.

C. EMD Decomposition of Short-Term Traffic Flow

Through EMD, 2160 group traffic flow data can be divided into 9 IMF components and 1 residual component. Fig. 7 shows the components after EMD decomposition. It can be seen from Fig. 7, except IMF 1 to IMF 3, the change of other components is relatively stable. Decomposition reduces the interference and coupling between different characteristic information. Therefore, EMD reduces the difficulty of modeling and improves the prediction accuracy. Fig. 8 shows the autocorrelation function of the original traffic flow and the components after EMD decomposition.

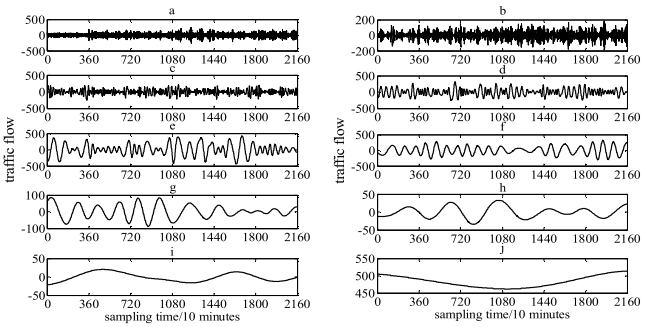

Fig. 7. The components after EMD decomposition (a: IMF1; b: IMF2; c: IMF3; d: IMF 4; e: IMF 5; f: IMF 6; g: IMF7; h: IMF8; i: IMF9; j: residual).

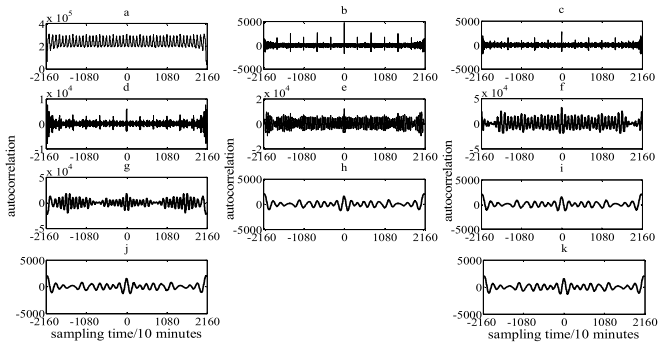

Fig. 8. The autocorrelation function results (a: original traffic flow; b: IMF1; c: IMF2; d: IMF3; e: IMF 4; f: IMF 5; g: IMF 6; h: IMF7; i: IMF8; j: IMF9; k: residual).

It can be seen from the Fig. 8 that the autocorrelation function of the original traffic flow time series has not been reduced to 0 in a long time. However, the autocorrelation function of each component after EMD quickly converges to 0 for the first time. It indicates that the components after EMD decomposition have the short correlation property. EMD can reduce the self-similarity of traffic flow time series. Compared with the long correlation model, the short correlation model has low complexity, so EMD decomposition can reduce the complexity of the model.

V. Prediction Approach

For traffic flow time series, several components are obtained after EMD processing. However, it is an important problem to choose the appropriate prediction model for these components. This section discusses the selection of prediction methods.

A. Self-Similar Analysis of Component After Decomposition

The self-similarity of time series can be measured by Hurst exponent - H. When H is 0.5, it shows that time series is random and the time series is uncorrelated. When $H \in [0, 0.5)$, it shows that the sequence is anti persistent. When $H \in (0.5, 1]$, it shows that the sequence is persistent and self-similarity. Therefore, the greater the Hurst exponent is, the higher the self-similarity of the time series. Table I gives the Hurst exponent of each component calculated by the rescaled range analysis method.

From Table I, it can be seen that IMF1, IMF 2, and IMF3 have smaller Hurst exponents, indicating that

TABLE I
The Hurst Exponent of Each Component

Components	Hurst exponent	Components	Hurst exponent
IMF1	0.5247	IMF6	0.7362
IMF2	0.5393	IMF7	0.7785
IMF3	0.5717	IMF8	0.8174
IMF4	0.6650	IMF9	0.8452
IMF5	0.6912	Residual	0.9015

TABLE II
The Z Calculation Results of the Components

Components	$\|Z\|$	Stationary (True/False)
IMF 4	6.4297	F
IMF 5	4.1682	F
IMF 6	3.8591	F
IMF 7	3.7099	F
IMF 8	2.6315	F
IMF 9	2.0108	F
Residual	1.6257	T

the corresponding components have strong random and nonlinear characteristics, so the prediction model needs to have good nonlinear regression ability. ELM has better regression prediction ability than ARMA, ARIMA, SVM and LSSVM. Although the prediction performance is weaker than that of LSTM or Bayes belief networks, its real-time performance is better than that of the deep learning model. Therefore, this paper proposes an improved ELM as the prediction model in terms of performance and real-time; IMFs 4 to 9 have the median Hurst exponents, which show that the corresponding components have periodic and non-stationary characteristics. SARIMA model can effectively analyze the correlation of periodic non-stationary data sequences. SARIMA model is chosen as the prediction model for IMFs 4 to 9; Residual component has a larger Hurst exponent, which shows that the residual component has stable change and linear characteristics. ARMA model has good prediction ability for linear sequences and is chosen as the prediction model for the residual component.

B. Determination of Non-Stationary and Non-White Noise Characteristics of Components

This paper further discusses the reasons for choosing SARIMA and ARMA. SARIMA model is generally suitable for non-stationary and seasonal time series. ARMA model is suitable for stationary time series. The run test method is adopted to determine the non-stationary characteristic of these IMF components and stationary characteristic of the residual component. The Z value of these components are calculated and shown in Table II. It can be seen from Table II, the Z values of IMF 4 to IMF9 components are larger than critical value 1.96 at the significance level 0.05. The Z value of the residual component is smaller than critical value 1.96 at the significance level 0.05. IMF 4 to IMF9 components are non-stationary time series and residual component is stationary time series.

LB statistic method is suitable for random characteristic test of white noise time series. When the P value of the LB statistic

TABLE III
THE P VALUE RESULTS OF THE COMPONENTS

Components	P value	White noise (True/False)
IMF 4	0.00235	F
IMF 5	0.00025	F
IMF 6	0.00211	F
IMF 7	0.00142	F
IMF 8	0.00105	F
IMF 9	0.00094	F
Residual	0.00042	F

is less than significance level α, it can be considered that the time series is not white noise sequence at the confidence level of $1-\alpha$. The calculation results of P value of the components are shown in Table III. The results in Table III show that the P value of each component is less than 0.05, so it can be concluded that each component does not have the characteristics of white noise and is not a random time series. From the calculation results in Table II and Table III, SARIMA is suitable for the prediction of IMF 4 to IMF9 components; ARMA is suitable for the prediction of the residual component.

C. Improved ELM

Based on standard ELM, this paper proposes an improved ELM model. Suppose that output weights $\boldsymbol{\beta}_k$ is calculated by sampling sample $(\boldsymbol{x}_1, \boldsymbol{t}_1), (\boldsymbol{x}_2, \boldsymbol{t}_2), \cdots (\boldsymbol{x}_k, \boldsymbol{t}_k)$. When new sample $(\boldsymbol{x}_{k+1}, \boldsymbol{t}_{k+1})$ is added into sample, then $\boldsymbol{\beta}_{k+1}$ can be expressed as

$$\begin{aligned}\boldsymbol{\beta}_{k+1} &= \left(\begin{bmatrix}\boldsymbol{H}_k \\ \boldsymbol{h}_{k+1}\end{bmatrix}\begin{bmatrix}\boldsymbol{H}_k \\ \boldsymbol{h}_{k+1}\end{bmatrix}\right)^{-1}\begin{bmatrix}\boldsymbol{H}_k \\ \boldsymbol{h}_{k+1}\end{bmatrix}\begin{bmatrix}\boldsymbol{T}_k \\ t_{k+1}\end{bmatrix} \\ &= (\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{H}_k + h_{k+1}^{\mathrm{T}}h_{k+1})^{-1}(\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{T}_k + \boldsymbol{h}_{k+1}^{\mathrm{T}}t_{k+1}) \quad (6)\end{aligned}$$

where
$\boldsymbol{h}_{k+1} = \left[\, f(\boldsymbol{a}_1\boldsymbol{x}_{k+1}+b_1)\ f(\boldsymbol{a}_2\boldsymbol{x}_{k+1}+b_2)\cdots f(\boldsymbol{a}_L\boldsymbol{x}_{k+1}+b_L)\,\right]$.

$\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{H}_k$ and $\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{T}_k$ are given weights, the above Eq. (6) can be rewritten as

$$\boldsymbol{\beta}_{k+1} = (\mu \mathrm{H}_k^{\mathrm{T}}\boldsymbol{H}_k + \boldsymbol{h}_{k+1})^{-1}(\mu \boldsymbol{H}_k^{T}\boldsymbol{T}_k + \boldsymbol{h}_{k+1}^{\mathrm{T}}t_{k+1}) \quad (7)$$

where μ, $0 < \mu < 1$ is weight coefficient. Let

$$\boldsymbol{P}_{k+1} = (\mu \boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{H}_k + \boldsymbol{h}_{k+1}^{\mathrm{T}}\boldsymbol{h}_{k+1})^{-1} \quad (8)$$

The inverse of Eq. (8) can be obtained

$$\boldsymbol{P}_{k+1}^{-1} = \mu \boldsymbol{P}_k^{-1} + \boldsymbol{h}_{k+1}^{\mathrm{T}}\boldsymbol{h}_{k+1} \quad (9)$$

Eq. (9) is substituted into Eq. (7), the next can be obtained.

$$\begin{aligned}\boldsymbol{\beta}_{k+1} &= \boldsymbol{P}_{k+1}(\mu \boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{T}_k + \boldsymbol{h}_{k+1}^{\mathrm{T}}t_{k+1}) \\ &= \boldsymbol{P}_{k+1}(\mu \boldsymbol{P}_k^{-1}\boldsymbol{\beta}_k + \boldsymbol{h}_{k+1}^{\mathrm{T}}t_{k+1}) \\ &= \boldsymbol{P}_{k+1}((\boldsymbol{P}_{k+1}^{-1} - \boldsymbol{h}_{k+1}^{\mathrm{T}}h_{k+1})\boldsymbol{\beta}_k + \boldsymbol{h}_{k+1}^{\mathrm{T}}t_{k+1}) \\ &= \boldsymbol{\beta}_k + \boldsymbol{P}_{k+1}\ \boldsymbol{h}_{k+1}^{\mathrm{T}}(t_{k+1} - \boldsymbol{h}_{k+1}\boldsymbol{\beta}_k) \quad (10)\end{aligned}$$

When a new sample x_{k+1} is obtained, it is necessary to judge the change trend of the error. When the error value is greater than a threshold value ε, $\boldsymbol{P}_k$ is updated, otherwise $\boldsymbol{P}_k$ remains unchanged. The update strategy is shown in the following formula.

$$\boldsymbol{P}_{k+1} = \begin{cases}(\mu \boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{H}_k + \boldsymbol{h}_{k+1}^{\mathrm{T}}\boldsymbol{h}_{k+1})^{-1} & (E_N > \varepsilon) \\ \boldsymbol{P}_k & (E_N \le \varepsilon)\end{cases} \quad (11)$$

where, $E_N = \sqrt{\sum_{i=1}^{N}(x_i - \widehat{x}_i)^2/N}$, $\widehat{x}_i$ is the predictive value of x_i.

D. Weight Coefficient Optimization

It can be seen from Fig. 1, the weight coefficient of each prediction model has a great influence on the final results. Therefore, this paper introduced an improved FOA to determine the weight coefficient of each prediction model [31]. Assuming the final predictive value of traffic flow is $\widehat{T}(t)$, which can be expressed as the follows.

$$\widehat{T}(t) = \sum_{i=1}^{k+n+1}\omega_i\widehat{y}_i(t) \quad (12)$$

where $\widehat{y}_i(t)$ is the predictive value of each prediction model, ω_i is the weight coefficient of each prediction model, k is the number of improved ELM models, n is the number of SARIMA models, the length of traffic flow time series is N. The fitness value of improved FOA is chosen as the root square mean error (RMSE) between actual and prediction value and is expressed the follows.

$$\begin{aligned}&\min(RMSE) = \min\left(\sqrt{\frac{1}{N}\sum_{i=1}^{k+n+1}(y_i(t) - \omega_i\widehat{y}_i(t))^2}\right) \\ &\text{s.t. } \omega_i \in [\omega_{\min}, \omega_{\max}] \quad (13)\end{aligned}$$

E. Performance Analysis of Prediction Method

For the optimization problem represented by Eq. (13), the common solution at present is the least square method or the gradient descent method. But the least square method needs to ensure that the matrix is reversible in the calculation process. The gradient descent method is slow. In this paper, improved FOA is used to obtain the optimal weight coefficient and make the RMSE the minimum. If the weight coefficient optimization strategy is not adopted, namely, all ω_i is 1 is one of the possible optimization results. For the optimization problem, the improved FOA can guarantee the global optimal solution within the range of weight coefficients. By way of conclusion, after improved FOA optimization, the final RMSE in this paper is less than the case of no optimization strategy.

VI. CASE STUDIES AND RESULTS ANALYSIS

In the case studies, the sample data in the Section III is used as training set. Through EMD, 2160 group traffic flow data can be divided into 9 IMF components and 1 residual component. The other collected 720 group data are used as test set to verify the predictive effect of the proposed prediction approach. In the improved ELM, the activation function is chosen as Sigmoid function (λ is 1). The data embedding

TABLE IV
THE DETAIL PARAMETERS OF SARIMA AND ARMA MODEL

Components	Parameters	AIC
IMF4	SARIMA (2,1,2)(2,1,1)	7.752
IMF5	SARIMA (2,1,1)(2,1,1)	4.367
IMF6	SARIMA (2,0,1)(1,1,1)	3.711
IMF7	SARIMA (2,0,1)(1,1,1)	4.676
IMF8	SARIMA (1,1,1)(0,1,1)	8.931
IMF9	SARIMA (1,0,1)(1,1,0)	3.259
Residual	ARMA (2,1)	1.118

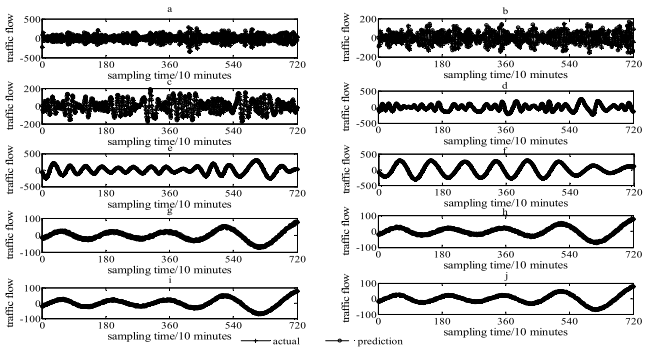

Fig. 9. The prediction value and actual value comparison of each component (a: IMF1; b: IMF2; c: IMF3; d: IMF 4; e: IMF 5; f: IMF 6; g: IMF7; h: IMF8; i: IMF9; j: Residual).

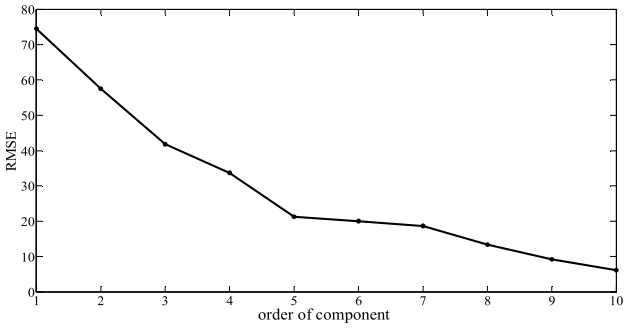

Fig. 10. RMSE of prediction and actual values of each component.

dimension m is determined as 30. The number of neurons in the hidden layer L is set to 200. The weight coefficient μ is set to 0.95. The threshold value ε is chosen as 2. The parameters of SARIMA and ARMA are determined and are shown in Table IV.

720 group traffic data is used as the test set. Fig. 9 shows a comparison between the prediction value and the actual value of each IMF component and residual by different prediction models. It can be seen from this figure that the prediction values are very close to the actual values of each component. Each prediction model has a good prediction effect.

Fig. 10 shows the RMSE of predictive and actual values of each component by these prediction models. From the Fig. 10, with the increase of the order of each component, it can be seen that RMSE decreases rapidly and the prediction accuracy becomes higher. The reason is that the randomness of the sequence decreases with the increase of the order of component, which improves the prediction accuracy.

After each prediction model of each decomposed component is established, the prediction values of these models

TABLE V
THE WEIGHT COEFFICIENT OF EACH PREDICTION MODEL

Weight coefficients	Value	Weight coefficients	Value
ω_1	0.2316	ω_6	0.8634
ω_2	0.3005	ω_7	0.9254
ω_3	0.4134	ω_8	1.2952
ω_4	0.5121	ω_9	1.8821
ω_5	0.8121	ω_{10}	2.8138

are multiplied by their weight coefficients to get the final prediction value. The improved FOA is adopted to obtain these weight coefficients. The change range of weight coefficient is $[-5, 5]$. The parameters of improved FOA are denoted as: the maximum number of iterations is 100, the size of the population is 20, p is 0.1, η is 0.2. After optimized, the weight coefficient of each component is shown in Table V.

In improved FOA, the strategy of population generation and position update has some random properties. At the same time, the minimum RMSE value between weight coefficients multiplied by the predicted value of each component and the actual value is taken as the optimization objective in the improved FOA. The constraint condition that the mean value of all the weighted coefficients is 1 is not included. Therefore, the mean value of the weight coefficients in Table V is not equal to 1. In this paper, the mean value of all the weight coefficients is 1.00489, which is close to 1.

The final prediction value can be obtained by these prediction models multiplied by the corresponding weight coefficient. In order to verify the prediction accuracy, the proposed approach is compared with GA-PSO-LSSVM [12], DNN-BTF [15], hybrid model in [18], wavelet transform and multiple models fusion [32], multi-kernel SVM [33], EMD and LSTM [34], deep learning algorithm in [35], EMD and Belief model [36], and proposed approach without improved FOA optimization. The detailed parameters of these prediction methods are as shown in the following Table VI.

Fig. 11 is the prediction error distribution histogram of the prediction methods mentioned in this paper.

From Fig. 11 can be known, because the prediction error is smaller and prediction error distribution is more uniform, the prediction effect and performance of the proposed prediction approach in this paper is better. The following 9 performance indicators are introduced to measure the prediction accuracy. The definitions of these performance indicators are shown in Table VII. In Table VII, N is the length of traffic flow sequence, T_k is the actual value of traffic flow at time k, $\widehat{T}_k$ is the predictive value of traffic flow at time k. $\xi^{(1-a)}$ is the number of confidence intervals in which the actual value falls under the confidence level $1-\alpha$, and $\overline{T}$ is the mean value of short-term traffic flow. Table VIII gives the comparison results of performance indicators of these prediction methods. From the comparison results in Table VIII, RMSE, MAE, MAPE, RRMSE, SSE, and TIC of the proposed prediction approach is smaller than the other prediction methods. At the same time, R^2 and IA value of the proposed prediction

TABLE VI
THE DETAILED PARAMETERS OF COMPARATIVE PREDICTION METHODS

Prediction methods	Parameters of method
[12]	$pop\_size = 100$; $max\_gen = 100$; $p_c = 0.8$; $p_m = 0.05$; $c_1 = c_2 = 2$; $M = 4 * pop\_size$; $k_{max} = 20$
[15]	The number of feature maps of each layer of the CNN is 40; Stacked GRUs with 3 layers were used to extract the temporal features; 50 states and 600 hidden neurons; The BPNN uses a single layer neural network with 1900 hidden neurons using RelU activation; The norm loss is 0.02 in LASSO
[18]	$\beta = 25$; $c = 11$; $m = 1.5$
[32]	Db3 wavelet transform; ARIMA(3,2,1); ARIMA(3,1,2); ARIMA(3,1,1); $\gamma = 25.362$; $\sigma^2 = 11.329$; $m = 27$
[33]	$C = 0.076$; $\gamma = 0.019$; $\varepsilon = 0.051$
[34]	9 IMFs and 1 residual component by EMD; Each component is predicted by different LSTM; Each LSTM has 6 LSTM layers and 1 fully connected layer; Hidden unit in the LSTM is 32; Adam algorithm is used to optimize the parameters; The learning rate is 0.001
[35]	The activation function is tanh; $n \in [1, 50]$; $N_l \in [1, 100]^n$; $\lambda \in [10^{-4}, 10^{-2}]$
[36]	9 IMFs and 1 residual component by EMD; DBN structure is 20-100-100-1; $\lambda_p = 0.6$; $\lambda_f = 2.5$; $m_p = 0.5$; $\eta_f = 0.4$

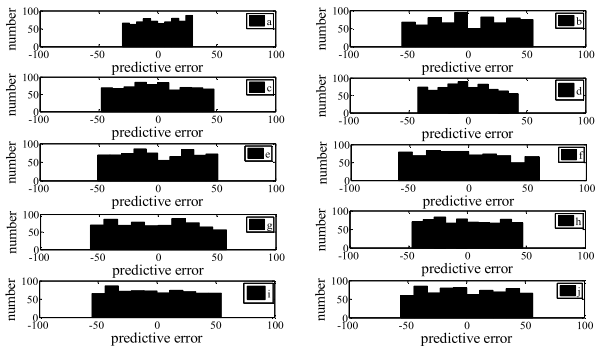

Fig. 11. The prediction error distribution histogram of the prediction methods (a: proposed approach; b: [33]; c: [34]; d: [35]; e: [36]; f: [12]; g: [15]; h: [18]; i: [32]; j: proposed approach without improved FOA).

approach is closer to 1 than the other prediction methods. The closer the value of R^2 and IA is to 1, the better the regression prediction performance of the model is. Therefore, the prediction accuracy of the proposed prediction approach for short-term traffic flow is better than the other prediction methods.

Fig. 12 is the reliability and confidence distribution of the prediction methods mentioned in this paper. It can be observed from this graph that the proposed prediction approach has higher reliability under the same confidence level. It can be known that the reliability of traffic flow prediction approach in this paper is better than other prediction methods.

To prove the prediction performance, the Diebold-Mariano (DM) test and Pearson's test are used to test prediction accuracy from the statistical perspective. Under the assumption that the two models have the same predictive performance, the loss functions of the two models have equal unconditional expectations.

$$\mathrm{E}(d_h) = E(L(\varepsilon_1) - L(\varepsilon_2)) = 0 \tag{14}$$

where L is the loss function. If $\mathrm{E}(d_h) > 0$, it means the predictive performance of the model 2 is better than the model 1. If $\mathrm{E}(d_h) = 0$, it means the predictive performance of the model 1 is the same as that of the model 2. Otherwise, the predictive performance of the model 1 is better than the model 2. The DM test values of the mentioned prediction methods are listed in Table IX. From the results in Table IX, the DM test value between the proposed prediction approach and the other prediction methods is greater than 0, the proposed prediction approach significantly outperforms the other prediction method at different significance levels. Thus, it can reasonably be concluded that the proposed prediction approach is superior to the other prediction methods.

Pearson's test is introduced to measure the association strength between the actual value and the predictive value. If Pearson's correlation coefficient is equal to 1, it indicates that the actual value and the prediction value have a linear relationship. On the other hand, if Pearson's correlation coefficient is equal to 0, there is no relationship between the actual value and the predictive value. In order to test the statistically significant of the prediction model, Wilcoxon Sign-Rank test and Ranksum test are introduced. Table X gives the results of Pearson's test, the Wilcoxon Sign-Rank and Ranksum test results between predictive values and actual values of each prediction method. It can be known from Table X, the results of Pearson's test of the proposed prediction approach are higher than those of the other prediction methods. In addition, P-value of the Wilcoxon Sign-Rank test of the proposed prediction approach is larger than the other prediction methods. It means that compared with other prediction methods, the median difference between the predictive value and the actual value of this paper is even less significant. Meanwhile, the P-value of the Wilcoxon Ranksum test of the proposed approach is also larger than the other methods. It means that compared with other prediction methods, the average probability that the predicted value is equal to the actual value of the proposed approach is greater.

In order to test the robustness of the proposed prediction approach, this paper adds some interference noise to the input data of the model. The random noise signal by 2%, 5% and 10% of the original traffic flow is added. RMSE, MAE, and MAPE are chosen as the evaluation indicators. Table XI gives the RMSE, MAE, and MAPE results of original traffic flow and traffic flow after adding noise predicted by the proposed prediction approach. As can be seen from Table XI, the difference between RMSE, MAE, and MAPE of traffic flow and traffic flow after adding noise is small, which shows that the proposed prediction approach has strong robustness to the noise signal.

VII. DISCUSSIONS

From the obtained experimental results, we can know that the proposed prediction approach has achieved better effect.

TABLE VII

THE DEFINITIONS OF THE PERFORMANCE INDICATORS

Performance indicators	Definition	Performance indicators	Definition	Performance indicators	Definition
RMSE	$\sqrt{\frac{1}{N}\sum_{k=1}^{N}(T_k-\hat{T}_k)^2}$	mean absolute error (MAE)	$\frac{1}{N}\sum_{k=1}^{N}\lvert(T_k-\hat{T}_k)\rvert$	mean absolute percentile error (MAPE)	$\frac{1}{N}\sum_{k=1}^{N}\lvert(T_k-\hat{T}_k)\rvert\times100\%/T_k$
relative root mean square error (RRMSE)	$\sqrt{\frac{1}{N}\sum_{k=1}^{N}(\frac{T_k-\hat{T}_k}{T_k})^2}$	square sum error (SSE)	$\sum_{k=1}^{N}(T_k-\hat{T}_k)^2$	R^2 (R Square)	$1-(\sum_{k=1}^{N}(T_k-\hat{T}_k)^2/\sum_{k=1}^{N}(T_k-\overline{T})^2)$
Theil inequality coefficient (TIC)	$\frac{\sqrt{\frac{1}{N}\sum_{k=1}^{N}(T_k-\hat{T}_k)^2}}{\sqrt{\frac{1}{N}\sum_{k=1}^{N}T_k^2}+\sqrt{\frac{1}{N}\sum_{k=1}^{N}\hat{T}_k^2}}$	The index of agreement (IA)	$1-\frac{\sum_{k=1}^{N}(T_k-\hat{T}_k)^2}{\sum_{k=1}^{N}(\lvert\hat{T}_k-\overline{T}\rvert+\lvert T_k+\overline{T}\rvert)^2}$	Reliability	$[\frac{\xi^{(1-a)}}{N}-(1-a)]\times100\%$

TABLE VIII

COMPARISON RESULTS OF PERFORMANCE INDICATORS

Methods	RMSE	MAE	MAPE	RRMSE	SSE(e+005)	R square	TIC	IA
Proposed approach	**17.2691**	**15.0283**	**8.0767**	**0.2154**	**2.1472**	**0.9963**	**0.0153**	**0.9997**
[12]	33.8642	29.0604	17.0745	0.4864	8.2568	0.9857	0.0300	0.9989
[15]	32.7188	28.4169	14.7213	0.3600	7.7077	0.9867	0.0290	0.9990
[18]	26.9977	23.4185	12.8380	0.3415	5.2479	0.9909	0.0239	0.9993
[32]	31.4516	27.1992	15.9108	0.4282	7.1223	0.9877	0.0279	0.9990
[33]	32.2652	27.9434	14.1332	0.3345	7.4950	0.9870	0.0287	0.9990
[34]	26.8944	23.0523	11.6449	0.2537	5.2078	0.9910	0.0239	0.9993
[35]	23.6991	20.2755	10.4486	0.2300	4.0439	0.9930	0.0210	0.9995
[36]	29.4695	25.5916	13.2307	0.2854	6.2528	0.9892	0.0262	0.9992
Without improved FOA	31.9509	27.8095	15.2607	0.4036	7.3502	0.9873	0.0283	0.9990

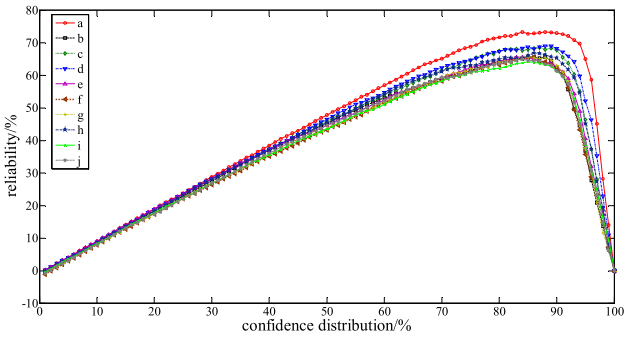

Fig. 12. The reliability and confidence distribution of the prediction methods (a: proposed approach; b: [33]; c: [34]; d: [35]; e: [36]; f: [12]; g: [15]; h: [18]; i: [32]; j: proposed approach without improved FOA).

TABLE IX

RESULTS OF THE DM TEST BETWEEN THE PROPOSED PREDICTION APPROACH AND OTHER PREDICTION METHODS

DM (model 1, model 2)	significance level		
	1%	5%	10%
DM ([12], proposed approach)	20.5028	8.6756	6.1027
DM ([15], proposed approach)	20.9472	8.8877	6.1471
DM ([18], proposed approach)	16.6051	8.5268	6.0868
DM ([32], proposed approach)	19.9739	8.6021	6.0991
DM ([33], proposed approach)	20.5284	8.7472	6.1641
DM ([34], proposed approach)	15.9045	8.1535	5.9619
DM ([35], proposed approach)	11.9467	7.7967	5.7452
DM ([36], proposed approach)	18.5760	8.4333	6.1035
DM (without improved FOA, proposed approach)	20.4856	8.5070	5.9963

TABLE X

THE PEARSON'S TEST, WILCOXON SIGN-RANK AND RANKSUM TEST RESULTS

Model	Pearson's test	Sign-Rank test	Ranksum test
Proposed approach	**0.9981**	**0.8873**	**0.9883**
[12]	0.9929	0.8496	0.8840
[15]	0.9934	0.8345	0.9029
[18]	0.9955	0.8400	0.9636
[32]	0.9938	0.8223	0.9301
[33]	0.9935	0.8250	0.9480
[34]	0.9955	0.8544	0.9777
[35]	0.9965	0.8458	0.9120
[36]	0.9946	0.8167	0.9383
Without improved FOA	0.9937	0.8359	0.9743

TABLE XI

THE EVALUATION INDICATORS BETWEEN ORIGINAL TRAFFIC FLOW AND TRAFFIC FLOW WITH NOISE

Dataset	RMSE	MAE	MAPE
original traffic flow	17.2691	15.0283	8.0767
original traffic flow with 2% noise	18.2538	15.3327	8.9903
original traffic flow with 5% noise	19.3368	16.2570	9.1100
original traffic flow with 10% noise	22.3640	19.6571	12.3064

We can safely draw the conclusion that the proposed prediction approach shows more powerful prediction ability than the compared models for short-term traffic flow. The main advantages of the proposed approach in relation to compared models are as follows. (a) The introduction of EMD. After EMD processing, the traffic flow time series is changed from long correlation sequence to short correlation, which highlights the change rule of each component, so the prediction difficulty and complexity are both reduced. (b) According to the different characteristics of each component, the appropriate

prediction model is determined. (c) For component with strong random and nonlinear characteristics, improved ELM with better performance is proposed as the prediction model. (d) For the final prediction results, improved FOA is used to optimize the weight instead of simple weighted combination.

VIII. Conclusion

A short-term traffic flow prediction approach based on EMD and combination model fusion is proposed. EMD can decompose the real-scale fluctuation or trend of the same scale in traffic flow time series step by step, and smooth the non-linear and non-stationary characteristic of the series. It produces a series of data sequences with the same characteristic scale. For each stationary sequence with approximate characteristics, self-similar analysis of each component after EMD processing is carried out. Different prediction models are determined according to the different characteristics of each component. These prediction models are respectively established, so as to reduce the influence of nonlinearity and non-stationary on the traffic flow prediction results. Improved ELM, SARIMA and ARMA are determined as the prediction models. Meanwhile, an improved FOA is proposed to optimize the weight coefficient of each prediction model. The final prediction value of traffic flow is determined by predictive value of these prediction models are multiplied by their weight coefficients. The case study results of actual traffic flow data show that the prediction approach proposed in this paper can better track the change rules of traffic flow, and effectively improve the prediction accuracy of short-term traffic flow.

In the future, this prediction approach will be applied to other scenarios such as work zone traffic flow prediction and passenger flow prediction. In addition, it is worth investigating which prediction model or decomposition method is more reliable in the combination model families (EMD-SVM, EMD-CNN, EEMD-ELM, VMD-ELM, LMD-ARMA, etc.). The proposed approach will be tested with more data (or less data) to discuss its sensitivity to the amount of data. Furthermore, we should study how to combine the research contents of this paper with control strategies such as model predictive control, and optimization of traffic congestion.

References

[1] X. Jin, M. Eom, and D. Chwa, "Adaptive sliding mode traffic flow control using a deadzoned parameter adaptation law for ramp metering and speed regulation," *J. Electr. Eng. & Technol.*, vol. 12, no. 5, pp. 2031–2042, 2017.

[2] I. Lana, J. Del Ser, M. Velez, and E. I. Vlahogianni, "Road traffic forecasting: Recent advances and new challenges," *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 2, pp. 93–109, 2018.

[3] E. Castillo, Z. Grande, A. Calviño, W. Y. Szeto, and H. K. Lo, "A State-of-the-Art review of the sensor location, flow observability, estimation, and prediction problems in traffic networks," *J. Sensors*, vol. 2015, pp. 1–26, Oct. 2015.

[4] Y. Xu, H. Chen, Q.-J. Kong, X. Zhai, and Y. Liu, "Urban traffic flow prediction: A spatio-temporal variable selection-based approach," *J. Adv. Transp.*, vol. 50, no. 4, pp. 489–506, Jun. 2016.

[5] A. Dhamaniya and S. Chandra, "Conceptual approach for estimating dynamic passenger car units on urban arterial roads by using simultaneous equations," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2553, no. 1, pp. 108–116, Jan. 2016.

[6] J. Tang, X. Chen, Z. Hu, F. Zong, C. Han, and L. Li, "Traffic flow prediction based on combination of support vector machine and data denoising schemes," *Phys. A, Stat. Mech. Appl.*, vol. 534, Nov. 2019, Art. no. 120642.

[7] S. Kamath, S. Singh, and M. S. Kumar, "Multiclass queueing network modeling and traffic flow analysis for SDN-enabled mobile core networks with network slicing," *IEEE Access*, vol. 8, pp. 417–430, 2020.

[8] Z. G. Shen, W. L. Wang, Q. Shen, and Z. C. Li, "Hybrid CSA optimization with seasonal RVR in traffic flow forecasting," *KSII Trans. Internet Inf. Syst.*, vol. 11, no. 10, pp. 4867–4887, Oct. 2017.

[9] W. Hu, L. Yan, K. Liu, and H. Wang, "A short-term traffic flow forecasting method based on the hybrid PSO-SVR," *Neural Process. Lett.*, vol. 43, no. 1, pp. 155–172, Feb. 2016.

[10] D. Wang, C. Wang, J. Xiao, Z. Xiao, W. Chen, and V. Havyarimana, "Bayesian optimization of support vector machine for regression prediction of short-term traffic flow," *Intell. Data Anal.*, vol. 23, no. 2, pp. 481–497, Apr. 2019.

[11] Q. Bing, D. Qu, X. Chen, F. Pan, and J. Wei, "Short-term traffic flow forecasting method based on LSSVM model optimized by GA-PSO hybrid algorithm," *Discrete Dyn. Nature Soc.*, vol. 2018, pp. 1–10, Nov. 2018.

[12] J.-F. Chen, S.-K. Lo, and Q. H. Do, "Forecasting short-term traffic flow by fuzzy wavelet neural network with parameters optimized by biogeography-based optimization algorithm," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–13, Oct. 2018.

[13] X. Chen, X. Cai, J. Liang, and Q. Liu, "Ensemble learning multiple LSSVR with improved harmony search algorithm for short-term traffic flow forecasting," *IEEE Access*, vol. 6, pp. 9347–9357, 2018.

[14] Z. Liu, J. Guo, J. Cao, Y. Wei, and W. Huang, "A hybrid short-term traffic flow forecasting method based on neural networks combined with K-Nearest neighbor," *PROMET Traffic Transp.*, vol. 30, no. 4, pp. 445–456, 2018.

[15] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transp. Res. C, Emerg. Technol.*, vol. 90, pp. 166–180, May 2018.

[16] H. Zhang, X. Wang, J. Cao, M. Tang, and Y. Guo, "A hybrid short-term traffic flow forecasting model based on time series multifractal characteristics," *Int. J. Speech Technol.*, vol. 48, no. 8, pp. 2429–2440, Aug. 2018.

[17] S. Zhang, Z. Kang, Z. Zhang, C. Lin, C. Wang, and J. Li, "A hybrid model for forecasting traffic flow: Using layerwise structure and Markov transition matrix," *IEEE Access*, vol. 7, pp. 26002–26012, 2019.

[18] J. Tang, L. Li, Z. Hu, and F. Liu, "Short-term traffic flow prediction considering spatio-temporal correlation: A hybrid model combing Type-2 fuzzy C-Means and artificial neural network," *IEEE Access*, vol. 7, pp. 101009–101018, 2019.

[19] Z. Diao *et al.*, "A hybrid model for short-term traffic volume prediction in massive transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 935–946, Mar. 2019.

[20] H. Wang, L. Liu, Z. Qian, H. Wei, and S. Dong, "Empirical mode Decomposition–Autoregressive integrated moving average: Hybrid short-term traffic speed prediction model," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2460, no. 1, pp. 66–76, Jan. 2014.

[21] L. Li, X. Qu, J. Zhang, H. Li, and B. Ran, "Travel time prediction for highway network based on the ensemble empirical mode decomposition and random vector functional link network," *Appl. Soft Comput.*, vol. 73, pp. 921–932, Dec. 2018.

[22] L. Li, X. Su, Y. Zhang, Y. Lin, and Z. Li, "Trend modeling for traffic time series analysis: An integrated study," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3430–3439, Dec. 2015.

[23] S. Feng, X. Wang, H. Sun, Y. Zhang, and L. Li, "A better understanding of long-range temporal dependence of traffic flow time series," *Phys. A, Stat. Mech. Appl.*, vol. 492, pp. 639–650, Feb. 2018.

[24] H. Tan, Y. Wu, B. Shen, P. J. Jin, and B. Ran, "Short-term traffic prediction based on dynamic tensor completion," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2123–2133, Aug. 2016.

[25] Y. Zhang, Y. Zhang, and A. Haghani, "A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 65–78, Jun. 2014.

[26] T. T. Tchrakian, B. Basu, and M. O'Mahony, "Real-time traffic flow forecasting using spectral analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 519–526, Jun. 2012.

[27] S. O. Mousavizadeh Kashi and M. Akbarzadeh, "A framework for short-term traffic flow forecasting using the combination of wavelet transformation and artificial neural networks," *J. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 60–71, Jan. 2019.
[28] H. Rabbouch, F. Saadaoui, and R. Mraihi, "A vision-based statistical methodology for automatically modeling continuous urban traffic flows," *Adv. Eng. Informat.*, vol. 38, pp. 392–403, Oct. 2018.
[29] E.-J. Kim, H.-C. Park, S.-Y. Kho, and D.-K. Kim, "A hybrid approach based on variational mode decomposition for analyzing and predicting urban travel speed," *J. Adv. Transp.*, vol. 2019, pp. 1–12, Dec. 2019.
[30] C. Shen, X. Bao, J. Tan, S. Liu, and Z. Liu, "Two noise-robust axial scanning multi-image phase retrieval algorithms based on pauta criterion and smoothness constraint," *Opt. Express*, vol. 25, no. 14, p. 16235, Jul. 2017.
[31] Z. Tian, "Short-term wind speed prediction based on LMD and improved FA optimized combined kernel function LSSVM," *Eng. Appl. Artif. Intell.*, vol. 91, May 2020, Art. no. 103573.
[32] T. Zhongda, L. Shujiang, W. Yanhong, and S. Yi, "A prediction method based on wavelet transform and multiple models fusion for chaotic time series," *Chaos, Solitons Fractals*, vol. 98, pp. 158–172, May 2017.
[33] X. Feng, X. Ling, H. Zheng, Z. Chen, and Y. Xu, "Adaptive multi-kernel SVM with Spatial–Temporal correlation for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2001–2013, Jun. 2019.
[34] Y. Tian, K. Zhang, J. Li, X. Lin, and B. Yang, "LSTM-based traffic flow prediction with missing data," *Neurocomputing*, vol. 318, pp. 297–305, Nov. 2018.
[35] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 1–17, Jun. 2017.
[36] L. Li, L. Qin, X. Qu, J. Zhang, Y. Wang, and B. Ran, "Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm," *Knowl.-Based Syst.*, vol. 172, pp. 1–14, May 2019.

Zhongda Tian received the Ph.D. degree in control theory and control engineering from Northeastern University, China, in 2013. He is currently an Associate Professor with the College of Information Science and Engineering, Shenyang University of Technology, China. His research interests include time series prediction and the predictive control of complex industrial systems.