IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 October 2023, accepted 23 October 2023, date of publication 26 October 2023, date of current version 1 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3327707

RESEARCH ARTICLE

A Novel Hybrid Model for PM2.5 Concentration Forecasting Based on Secondary Decomposition Ensemble and Weight Combination Optimization

YUAN HUANG, XIAOYU ZHANG, AND YANXIA LI
School of Information and Electrical Engineering, Hebei University of Engineering, Handan 056038, China

Corresponding author: Xiaoyu Zhang (zxyhebeu1998@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61772451, in part by the Hebei Natural Science Foundation Youth Project D2021402043, in part by the Science and Technology Project of Hebei Education Department under Grant QN2019168, and in part by the Project of Handan Science and Technology Bureau under Grant 21422093285.

ABSTRACT Accurate and efficient forecast of PM2.5 concentration is the primary prerequisite for promoting urban green development and improving residents' well-being. In this study, a hybrid model based on secondary decomposition ensemble and weight combination optimization is presented to materialize exact PM2.5 concentration prediction. First, the empirical wavelet transform (EWT) is adopted to disassemble the primeval PM2.5 concentration sequence to get high and low-frequency components. Considering the intricacy of high-frequency components and the difficulty of direct prediction. Therefore, it is further decomposed into a collection of modes with significant discrepancies by adaptive variational mode decomposition (AVMD). Second, the prediction network and meteorological data are determined respectively according to Hurst exponent. Then support vector regression (SVR) model and bidirectional long short-term memory (BILSTM) network are used to model each sequence separately. In addition, the weights of each forecast network were optimized by improved sparrow search algorithm (ISSA) to correct decomposition errors. Finally, all prediction results were weighted and integrated to receive the ultimate prediction values. The test results show that whether it is 1-step prediction, 3-step prediction or 5-step prediction, the proposed model has the best prediction effect in Beijing, Handan and additional Shanghai cases.

INDEX TERMS Adaptive variational mode decomposition (AVMD), hurst exponent, improved sparrow search algorithm (ISSA), PM2.5 concentration prediction, weighted combination model.

I. INTRODUCTION

With the accelerating process of urbanization and industrialization, contaminants are discharged unscrupulously, and the problem of air contamination is turning out to be increasingly serious [1], [2]. As the statistics of the World Health Organization, about 13 million people lose their precious lives due to environmental pollution every year, and about 1/2 of them die from respiratory diseases caused by air pollution [3]. The survey shows that the PM2.5 is the chief culprit of the tragedy. In order to control air pollution, many automatic monitoring systems of particulate matter have been established all over the world, which are used to monitor urban pollution in real time. However, some monitoring points are easily affected by external factors, resulting in biases in the obtained data, which in turn affect the judgment and decision-making of relevant departments [4]. For another, real-time monitoring stations are difficult to meet the human desire to predict future weather changes, as they can only provide real-time weather information [5]. In contrast, multi-step prediction can provide more long-term and comprehensive information, which can help us better plan outdoor travel and ensure the health of ourselves and our families. Therefore, the multi-step accurate prediction of PM2.5 concentration is of great purpose for preserving the ecological environment, improving the efficiency of

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos.

government decision-making, and ensuring residents' travel safety.

Currently, a great quantity of PM2.5 concentration forecast methods and models have been proposed. The early numerical methods used pollution sources and contaminant discharge data to imitate the generation, accumulation, proliferation and transformation of air pollution, which promoted the development of atmospheric science to a great extent [6]. However, the numerical method depends on the researchers' understanding of the polluter and the portrayal of tanglesome physical-chemical procedures [7]. Besides, the method has the characteristics of time-consuming and low accuracy. Therefore, numerical methods are not up to par in terms of accuracy and practicality. Subsequently, statistical learning methods gradually entered our vision due to their powerful linear approximation ability. However, the statistical learning methods can only capture the linear relationship, and have some limitations in the face of nonlinear time series data. Meanwhile, PM2.5 time series data generally reflect the characteristics of nonlinearity, volatility, and complexity. Therefore, the prediction performance of statistical learning methods is difficult to meet expectations.

In contrast, machine learning methods, like artificial neural networks [8], hidden markov models, random forests [9] and support vector regression (SVR), have strong nonlinear fitting ability and feature extraction performance. Lai et al. [10] proposed a SVR model based on causal characteristics analysis and particle swarm optimization for PM2.5 concentration prediction, and test results show that SVR is significantly superior to other machine learning methods. By introducing kernel function, SVR maps samples from low to high-dimensional space, which not only ensures computational efficiency, but also effectively avoids a series of consequences caused by dimensional disaster. As an offshoot of machine learning, deep learning has also been generally utilized in air quality forecasts. Zhai et al. [11] developed a long short-term memory (LSTM) model that combines meteorological-social factors, and achieved fine prediction results. However, LSTM only has one-way feature extraction capability, ignoring the influence of the future on the present. To this end, a PM2.5 concentration prediction model based on feature reduction and bidirectional long short-term memory (BILSTM) network is advanced by Zhang et al. [12]. A multivariable BILSTM model combining meteorological factors is presented by Zhang et al. [13], with a prediction accuracy of up to 95%. From the above literature, we can draw a conclusion that SVR and BILSTM networks are promising and outstanding prediction techniques.

In recent years, the hybrid prediction model based on ''decomposition and ensemble'' is generally verified to be better than the above single prediction model [14], [15], [16], [17]. A two-channel air quality index (AQI) prediction model based on variational mode decomposition (VMD) and wavelet transform is established by Wu et al. [18]. A urban AQI daily forecasting system based on VMD and LSTM has been developed by Wu et al. [19]. The results demonstrate that VMD can diminish the difficulty of forecasting to a great extent. VMD decomposes time series data according to the frequency domain, but the problem is how to certainly the optimum number of modes and penalty parameters [20]. To conquer this difficulty, many scholars have improved VMD in terms of evaluation indicators and search strategies. Cui et al. [21] first used curvature as an important indicator for quantitative curve analysis. A grid search method based on kurtosis and energy loss coefficient is proposed by Zhang et al. [22] to seek the optimum hyperparameters of the VMD algorithm, and the results are satisfactory. However, the characteristic analysis ability of a single evaluation indicator is limited, which may lead to the deviation of the results. Furthermore, simple search will inevitably fall into local optimization. Therefore, the reasonable selection of evaluation indicators and the effective formulation of search strategies are crucial for the VMD model.

When using the ''decomposition and ensemble'' method to forecast air quality, two key problems must be considered, one is the selection of forecast network and meteorological data, the other is the error correction of decomposition results [23]. On the one hand, there are significant differences in the instantaneous frequency and characteristics of different subsignals, and selecting an appropriate prediction network for these signals is a topic worth discussing. Additionally, the addition of meteorological data can enhance the forecast precision of the model, but it also introduces noise [24]. So, should we add meteorological data to all prediction networks? On the other hand, any form of decomposition can not achieve the real sense of decomposition, if not controlled, there will be a serious bias effects. For this purpose, an optimal combination forecasting model based on cuckoo search (CS) algorithm and grey wolf optimizer (GWO) is proposed by Zhu et al. [25]. An error correction model based on general regression neural network (GRNN) and SVR is presented by Zhu et al. [26]. Luo et al. [27] combines VMD with extreme learning machine (ELM) optimized by CS to correct the initial prediction sequence. Liu et al. [28] proposed a three channel PM2.5 concentration prediction model that combines deep neural networks and imperial competition algorithm (ICP). Sun et al. [29] introduces stacking-driven ensemble model (SDEM) to realize the double improvement of information utilization and feature extraction. Fan et al. [30] used LSTM and autoregressive integrated moving average (ARIMA) to predict the trend and residual components of PM2.5 sequences, and achieves remarkable results. Table 1 summarizes the above decomposition and ensemble methods. Although the above methods are sanguine in terms of prediction accuracy, they only focus on a one-sided perspective, which is obviously insufficient. Therefore, there is an urgent need for a robust, stable, and comprehensive hybrid prediction model to overcome the shortcomings of existing models and achieve synchronous

TABLE 1. Summary of related prediction work based on decomposition ensemble method.

Work	Method	Subject	Metric	Prediction network	Meteorological data	Error correction
[25]	• Complementary ensemble empirical mode decomposition (CEEMD) • SVR, CS, GWO	NO_2, SO_2	MAE, RMSE, R^2, Mean Absolute Percentage Error (MAPE)	✓	×	×
[26]	• CEEMD, GRNN, SVR • Swarm intelligence optimization algorithm	AQI	MAE,MAPE, Mean Square Error, Index of Agreement	×	×	✓
[27]	• VMD, CS, ELM	PM_{10}	MAE, RMSE, MAPE	×	×	✓
[28]	• CEEMD, LSTM, ICP • Deep belief network • Multilayer perceptron	$PM_{2.5}$	MAE, MAPE, RMSE, Pearson correlation coefficient	×	×	✓
[29]	• GRNN, ELM, SVR, SDEM • Fast ensemble empirical mode decomposition	$PM_{2.5}$	MAE, MAPE, RMSE, R^2	✓	×	×
[30]	• wavelet decomposition • ARIMA, LSTM	six major pollutants and AQI	MAE, RMSE, R^2	✓	×	×
Our work	• EWT AVMD, Hurst exponent • SVR, BILSTM, ISSA	$PM_{2.5}$	MAE, RMSE, SMAPE, R^2	✓	✓	✓

improvement of prediction accuracy and convergence speed [31].

In this context, this article presents a novel weighted combination forecast model based on empirical wavelet transform (EWT), adaptive variational mode decomposition (AVMD), Hurst exponent and improved sparrow search algorithm (ISSA), which aims to upgrade the accuracy of PM2.5 concentration prediction. In our model, SVR and BILSTM are selected as candidate prediction networks. Hence, the proposed model is called EWT-AVMD-Hurst-SVR-BILSTM-ISSA.

The predominant contributions of the research are provided as blew:

1) **Adaptive variational mode decomposition:** The proposed EOS comprehensive evaluation indicator and a novel grid search strategy realize the adaptive acquisition of the VMD parameters.
2) **Novel division mechanism:** The complexity of each subsignal is analyzed by Hurst exponent, and a signal divider is designed to simultaneously determine the prediction network and meteorological data. Moreover, it is worth mentioning that meteorological data is not considered in [32]
3) **Improved sparrow search algorithm:** In contrast with particle swarm optimization (PSO), fruit fly optimization algorithm (FOA) and sparrow search algorithm (SSA), the proposed ISSA has the best convergence precision and the fastest convergence rate. Moreover, the ISSA is utilized to optimize the weights of each prediction network, so as to correct the errors generated in the decomposition process and improve the forecast precision.

The remainder of this article is arranged like this: Section II shortly introduces the relevant methods and the general architecture of the proposed model. The relevant content of AVMD algorithm and Hurst exponent are depicted in Section III. Section IV supplies a detailed explanation of the ISSA algorithm. Section V is the experimental part. Finally, the conclusion is put forward in Section VI.

II. METHODOLOGY

A. EMPIRICAL WAVELET TRANSFORM (EWT)

EWT is a signal extraction technique advanced by J. Gilles [33] to tackle nonlinear and unsmooth data. This method combines empirical mode decomposition with wavelet transform and overcomes the modal aliasing problem of the empirical mode decomposition. It divides the primeval time series data $f(t)$ into several consecutive regions on the frequency domain, and constructs a band-pass filter bank on each region for filtration, and then reconstructs the signal to obtain a collection of mode components. The signal reconstruction expression is given by:

$$f(t) = W_f^{\varepsilon}(0,t) \times \delta_1(t) + \sum_{n=1}^{N} W_f^{\varepsilon}(n,t) \times \varphi_n(t), \qquad (1)$$

where $W_f^{\varepsilon}(n,t)$ and $W_f^{\varepsilon}(0,t)$ represent the detail coefficient and the approximation coefficient, respectively. $\varphi_n(t)$ and $\delta_1(t)$ signify the empirical wavelet and scaling function, respectively.

B. VARIATIONAL MODE DECOMPOSITION (VMD)

VMD is a disintegration technique advanced by Dragomiretskiy et al. [34], whose overall framework is essentially a procedure of constructing and settling variational problems. For each modal component u_k, the analytical signal is attained using the Hilbert transform, and the spectrum of the modal component is modulated to the corresponding baseband by multiplying the operator $e^{-jw_k t}$, and then the bandwidth is estimated utilizing H^1 Gaussian smoothness [34]. The construction results are as follows:

$$\begin{cases} \min \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\phi(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \\ s.t. \sum_{k=1}^{K} u_k = f(t), \end{cases} \tag{2}$$

where $\phi(t)$ stands for Dirac distribution, $\otimes$ denotes convolution operation.

By introducing the punishment factor and lagrange multiplier, the above restraint problem is converted into an unrestraint problem, which can be indicated as follows:

$$\begin{aligned} L(\{u_k\}, \{w_k\}, \lambda) = \alpha \sum_{k=1}^{K} \left\| \partial_t \left[\left(\phi(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-jw_k t} \right\|_2^2 \\ + \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 \\ + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle. \end{aligned} \tag{3}$$

VMD solves the above unrestrained problem by alternate direction method of multipliers and ultimately realizes the disintegration of the original signal.

C. FORECASTING NETWORKS

1) SUPPORT VECTOR REGRESSION (SVR)

SVR is a method of applying support vector machine to regression problems [4]. It has been generally utilized in air quality forecast [35], ultra-short-term wind speed prediction [32], rolling bearing fault diagnosis [36] and so on. The core of SVR is to utilize structural risk minimization theory to find the optimum fitting function: $f(x)$ so that as many training sets as possible fall near it [4]. The problem can be defined as:

$$\begin{aligned} \min \ & \frac{1}{2}\|w\|^2 + P \sum_{i=1}^{N} \left(\xi_i + \hat{\xi}_i \right) \\ s.t. \ & y_i - w^T \psi(x_i) - b \le \varepsilon + \xi_i \\ & w^T \psi(x_i) + b - y_i \le \varepsilon + \hat{\xi}_i \\ & \xi_i \ge 0 \\ & \hat{\xi}_i \ge 0, \end{aligned} \tag{4}$$

where P represents the penalty coefficient, ξ_i and $\hat{\xi}_i$ are the relaxation variables, ε stands for tolerance width of the fitting function $f(x)$, which means that if the training set is within the width, the error will not be calculated.

The introduction of kernel function can realize the transformation from low-dimensional to high-dimensional. In this paper, gaussian kernel function $K(x_i, x_j) = \exp(-||x_i - x_j||^2/2\gamma^2)$ is utilized, in which γ denotes the bandwidth.

2) BIDIRECTIONAL LONG SHORT-TERM MEMORY (BILSTM)

BILSTM network can be regarded as the stack of LSTM in different directions, which is made up of two independent LSTM. Each LSTM unit contains three gates internally: forget gate f_t, input gate i_t and output gate o_t. These three gates perform their duties to achieve the regulation and control of the network, and the specific formulas are as below:

$$f_t = \sigma(u_f z_{t-1} + v_f x_t + b_f) \tag{5}$$

$$i_t = \sigma(u_i z_{t-1} + v_i x_t + b_i) \tag{6}$$

$$o_t = \sigma(u_o z_{t-1} + v_o x_t + b_o) \tag{7}$$

$$c_t = f_t * c_{t-1} + i_t * \tanh(u_c z_{t-1} + v_c x_t + b_c) \tag{8}$$

$$z_t = o_t * \tanh(c_t), \tag{9}$$

where u_f, u_i, u_o, u_c are the output activation members, v_f, v_i, v_o, v_c are the input activation members, and b_f, b_i, b_o, b_c are the biases. c_t and z_t denote the unit state and output at time t, respectively. σ is the sigmoid activation function [37].

D. SPARROW SEARCH ALGORITHM (SSA)

Xue et al. [38] proposed a novel heuristic algorithm, SSA, based on careful observation of the foraging and anti-predation behaviors of sparrow populations. Sparrows are quite clever and have strong recall, with a strict exploration-follow-warning mechanism within the entire population, corresponding to three different individuals: producers, predators, and perceivers [39], [40].

Producer sparrows have high intelligence in the population and play a leading role, responsible for providing foraging information to other sparrows. The update formula is as follows:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} \cdot \exp\left(\frac{-i}{\tau \cdot T} \right) & if \ R_{af} < ST \\ X_{i,j}^{t} + R_n \cdot L & if \ R_{af} \ge ST, \end{cases} \tag{10}$$

where τ is a random number, T represents the maximum number of iterations, R_{af} and ST represent the alarm factor and the safety threshold, respectively. R_n is a random number with normal distribution, L denotes a $1 \times d$ dimensional vector.

Predators are particularly eager to replace producers. In order to achieve their own goals, while foraging with the producers, they regard the producers as competitors for food. This competitive behavior stupendously upgrades the convergence rate of SSA. The predator location is updated

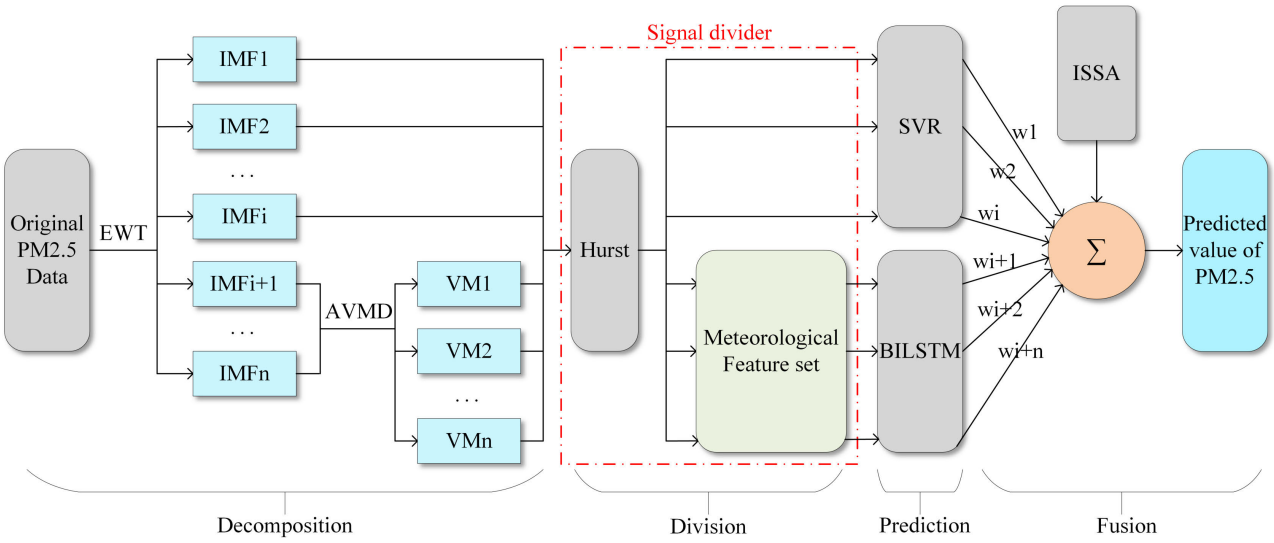


FIGURE 1. The overall framework of EWT-AVMD-Hurst-SVR-BILSTM-ISSA model.

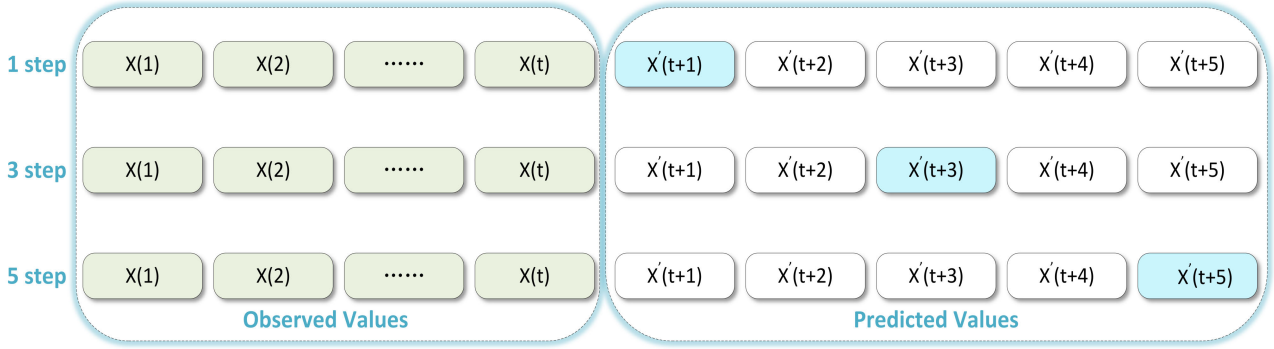


FIGURE 2. The data input and output structure of multi-step advance prediction.

below:

$$X_{i,j}^{t+1} = \begin{cases} R_n \cdot \exp\left(\dfrac{X_{worst}^t - X_{i,j}^t}{i^2}\right) & if\ i > n/2 \\ X_P^{t+1} + |X_{i,j}^t - X_P^{t+1}| \cdot A^+ \cdot L & otherwise, \end{cases} \tag{11}$$

where X_{worst} stands for worst population location, X_P denotes the location of the smartest producer. A is used to control the update direction of predator location, which is a $1 \times d$ dimensional matrix that can only take 1 or -1, and $A^+ = A^T(AA^T)^{-1}$.

When the sparrow population encounters danger, the perceiver are in charge of sounding the alarm signal to prompt the population out of danger. The perceiver location is updated below:

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^t + \beta \cdot |X_{i,j}^t - X_{best}^t| & if\ f_i > f_g \\ X_{i,j}^t + K \cdot \left(\dfrac{|X_{i,j}^t - X_{worst}^t|}{(f_i - f_w) + \varepsilon}\right) & if\ f_i = f_g, \end{cases} \tag{12}$$

where X_{best} represents the most intelligent sparrow position, K is a random number, the step size control factor is denoted by β.

E. THE PROPOSED HYBRID FORECASTING MODEL

In this section, the proposed EWT-AVMD-Hurst-SVR-BILSTM-ISSA hybrid model is presented at great length. The general architecture of the model is shown in Fig. 1, and the specific procedure can be broken down into the following four steps:

Step 1: The EWT is adopted to disassemble the primeval PM2.5 concentration sequence, and then the AVMD algorithm is utilized to further disassemble the high-frequency components, and eventually several intrinsic mode functions (IMF) are attained.

Step 2: The Hurst exponent of each IMF component is calculated, via the signal divider, the corresponding prediction network is determined and meteorological data are added.

Step 3: Each IMF component is inputted into the SVR model and BILSTM network for multi-step predic-

Algorithm 1 AVMD

Parameters:

- k: modal number
- α: penalty parameter
- Ep : envelope spectrum entropy
- f : high-frequency signal

Begin:

for $k = 2 : 20$ **do**

$\backslash * Rough\ search * \backslash$

$\alpha = 1000$

while $\alpha < 10000$ **do**

k modes are gained by VMD decomposition: $\{u^m\}_{m=1}^k = VMD\,(k, \alpha, f)$

Calculate the envelope spectrum entropy $\{Ep^m\}_{m=1}^k$ of all modes

$Ep_{k,\alpha} = \min\left(Ep^1, Ep^2, \ldots, Ep^k\right)$

$\alpha = \alpha + 1000$

Calculate the initial local optimal penalty parameter under the corresponding number of modes:

$\_, \alpha = \arg\min_{k,\alpha}\left(\left\{Ep_{k,\alpha}\right\}\right)$

$\backslash * Fine\ search * \backslash$

Determine the scope of fine search: $\alpha'_{\min} = \alpha - 500$, $\alpha'_{\max} = \alpha + 500$

Set $\alpha' = \alpha'_{\min}$

while $\alpha' < \alpha'_{\max}$ **do**

$\{u^m\}_{m=1}^k = VMD\left(k, \alpha', f\right)$

$Ep_{k,\alpha'} = \min\left(\{Ep^m\}_{m=1}^k\right)$

$\alpha' = \alpha' + 100$

The local optimal mode number and local optimal penalty parameters are obtained:

$k_{local}, \alpha_{local} = \arg\min_{k,\alpha'}\left(\left\{Ep_{k,\alpha'}\right\}\right)$

$\{u^m\}_{m=1}^{k_{local}} = VMD\,(k_{local}, \alpha_{local}, f)$

Calculate EOS indicator: $EOS_{k_{local},\alpha_{local}} = EOS\left(\{u_m\}_{m=1}^{k_{local}}, f\right)$

The global optimal hyperparameter of VMD algorithm are acquired by minimizing EOS indicators:

$k^*, \alpha^* = \arg\min_{k_{local},\alpha_{local}}\left(\left\{EOS_{k_{local},\alpha_{local}}\right\}\right)$

End

tion according to the structure of Fig. 2. Different IMF components have different characteristic information, so the hyperparameters of each prediction network are different. The hyperparameters of SVR model and BILSTM network are obtained by ISSA algorithm and trial-and-error method respectively.

Step 4: The ISSA algorithm is utilized to attain the optimal weight of each prediction network, then all the predicted results are weighted and integrated to get the ultimate predicted value of PM2.5 concentration.

III. DECOMPOSITION AND DIVISION OF DATA

A. ADAPTIVE VARIATIONAL MODE DECOMPOSITION

When decomposing signals employing VMD, the appropriate mode number k and penalty parameters α must be chosen ahead of time. However, there are many deficiencies in the extant methods. For instance, the manual calculation method requires researchers to have sufficient prior knowledge. Otherwise, unreasonable parameter settings will result in over-decomposition and owe-decomposition [32]. In addition, the manual calculation method has strong uncertainty and significant human disturbances. The swarm intelligent optimization method ignores the time cost, and the computational efficiency is relatively low. To this end, an AVMD algorithm based on EOS comprehensive evaluation indicator and rough-to-fine grid search strategy is proposed in this article to achieve the adaptive acquisition of VMD parameters on the premise of ensuring efficiency. On the one hand, we propose a new EOS indicator to measure the decomposition effect of VMD, which comprehensively considers the energy difference and modal aliasing problem of the signal, and can more accurately reflect the degree of signal decomposition. This helps to reduce the impact of external disturbances caused by inaccurate or inappropriate evaluation indicators on model performance. On the other hand, the rough-to-fine grid search strategy can dynamically adjust the search space range of α, and gradually refine from the whole to more specific parameter settings. To some

extent, the uncertainty and error caused by manual setting of hyperparameters are alleviated, thus improving the stability and robustness of the model. The specifics of the AVMD are elucidated in Algorithm 1.

As shown in Algorithm 1, firstly, the minimum envelope spectral entropy (Ep) under different α values is calculated to gain the approximate value of the local optimal penalty parameter, and the Ep can be denoted as follows:

$$Ep = -\sum_{i=1}^{N} P_i \log_2 P_i, \tag{13}$$

in which N denotes the signal length and P_i denotes the standardized form of the instantaneous envelope of the signal.

Then, the search range is narrowed, the final local optimal mode number k_{local} and local optimal penalty parameters α_{local} are obtained with the same steps, and the corresponding EOS indicator are calculated. This indicator combines energy loss coefficient (ELC), orthogonal index (IO) and spectrum degree of cross-correlation (SPC). Specifically, the ELC reflects the degree of difference among the primeval signal with the reconstructive signal. A smaller ELC results in less information loss and better decomposition. IO reveals the degree of similarity between modes. A smaller IO means strong independence between modes. However, it tends to produce bad decomposition results when k is fixed and α is large, at which point IO is small [41]. SPC is devoted to assess the degree of mode aliasing. A smaller SPC indicates a lower probability of modal aliasing, but the degree of information loss is large [42]. From the above analysis, the more moderate the IO and SPC, the better the decomposition effect. These indicators are expressed as follows:

1) ELC is either expressed as:

$$ELC = \left\| f(t) - \sum_{i=1}^{k} u_i \right\|_2^2 / \|f(t)\|_2^2, \tag{14}$$

2) IO is either expressed as:

$$IO = \frac{1}{k-1} \sum_{i=1}^{k-1} \left(\frac{u_i \cdot u_{i+1}}{|u_i||u_{i+1}|} \right), \tag{15}$$

3) SPC is either expressed as:

$$SPC = \frac{\sum_{i=1}^{k} \sum_{j=1}^{k} \left(F(u_i) \cdot F(u_j) \right)}{\sum \left(F(f(t)) \right)^2} (i \neq j), \tag{16}$$

4) EOS is either expressed as:

$$EOS = ELC + |IO - IO_{avg}| + |SPC - SPC_{avg}|, \tag{17}$$

where $f(t)$ represents the primeval signal and u_i represents the i-th mode component, $F(\cdot)$ stands for Fourier transform. IO_{avg} and SPC_{avg} denote the average value of the corresponding indicator, respectively.

The algorithm iterates continuously until the EOS indicator corresponding to all local optimal parameters are acquired. Finally, the optimal hyperparameters (k^*, α^*) of VMD algorithm are obtained by minimizing EOS.

In this article, EWT is used to disassemble the primeval data into six IMF components (IMF1, IMF2,..., IMF6), and then the high-frequency component IMF6 is further disassembled by AVMD. Table 2 and Fig. 3 show the procedure of VMD parameter optimization. We observed that the EOS indicator reaches the minimum when the mode number is 3, on the Beijing data set. Over-decomposition occurs when the mode number is greater than 8. Therefore, k and α are set to 3 and 600, respectively, and IMF6 is further disassembled into three independent IMFs. Eventually, the primeval data is disassembled into IMF1-IMF8. Similarly, on the Handan data set, the primeval data is disassembled into 11 IMF components.

B. DETERMINATION OF PREDICTION NETWORK AND METEOROLOGICAL DATA

It is crucial to select the corresponding prediction network and meteorological data for different components. To tackle this problem, the Hurst exponent (H) is introduced. The Hurst exponent reflects the long-term correlation of time series. When $H < 0.5$, the time series has anti-persistence and shows the characteristic of oscillation. When $H = 0.5$, the time series has no correlation. When $H > 0.5$, the time series has a long-term correlation. In general, A higher Hurst exponent indicates that time series are easier to predict [32]. Based on this characteristic, this article takes Hurst exponent as the reference of signal division and designs a signal divider. The concrete rules are as below:

1) Refer to the division criteria of Tian [32], when $H \geq 0.7$, it implies that the time series is monotonous and stationary. In this case, SVR is selected as the prediction network.
2) When $H < 0.7$, it implies that the time series has strong randomness, nonlinearity and unpredictability. In this case, BILSTM with excellent learning ability is selected as the prediction network, and meteorological data is added to upgrade the forecast precision.

Based on the results of the decomposition in Section III-A, Table 3 presents the Hurst exponents of each IMF component under different data sets, which are calculated by rescaled range analysis. We observed that IMF1, IMF2, and IMF3 follow rules (1), and IMF4, IMF5, IMF6, IMF7, and IMF8 follow rules (2) on Beijing data set. Similarly, on the Handan data set, each IMF component is divided into two categories: one is IMF1-5, the other is IMF6-11. These IMF components are input into the corresponding forecasting network to obtain their respective forecasting results.

IV. IMPROVED SPARROW SEARCH ALGORITHM

Although SSA algorithm has strong preponderances in convergence rate and optimization power, it is susceptible to

TABLE 2. All the local optimal parameters obtained based on AVMD.

Beijing	k_{local}	2	**3**	4	5	6	7	8	9	10
	α_{local}	900	**600**	10500	10500	10500	900	5400	2500	1900
	11	12	13	14	15	16	17	18	19	20
	10500	10500	10500	10500	10500	10500	10500	10500	10500	10500
Handan	k_{local}	2	3	4	5	**6**	7	8	9	10
	α_{local}	600	2500	6400	5500	**900**	500	500	500	500
	11	12	13	14	15	16	17	18	19	20
	8100	3000	4300	8200	1200	1600	6800	6700	8700	1300

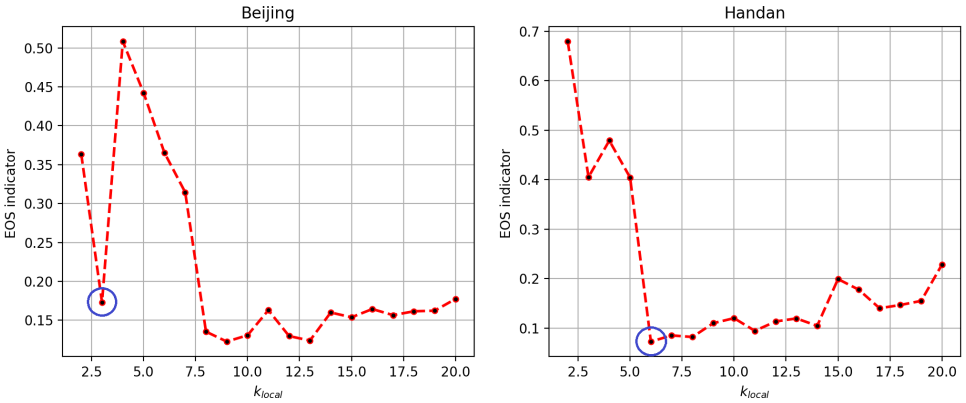


FIGURE 3. Corresponding EOS indicator values for all local optimum parameters.

TABLE 3. The Hurst exponents of each IMF component under different data sets.

City	IMF component										
	1	2	3	4	5	6	7	8	9	10	11
Beijing	0.9131	0.8300	0.7069	0.5910	0.3756	0.2417	0.1745	0.1396	-	-	-
Handan	0.9345	0.8879	0.8178	0.7452	0.7091	0.4889	0.4066	0.4199	0.2935	0.1586	0.2151

running into local optimum in late iterations. To surmount this problem, this article proposes an ISSA algorithm based on adaptive hyperparameter, Levy flight, mutation and lens opposition-based learning (LOBL). Besides, the validity of the ISSA was demonstrated adopting six benchmark functions with different dimensions.

A. ADAPTIVE HYPERPARAMETER

Since producer sparrows are the bellwether of the whole population, a larger foraging range is required during the iteration process. However, in the actual situation, the foraging range of producer sparrows will become increasingly smaller, which will inevitably lead to the imbalance among global with local search ability, and seriously affect the convergence precision. Therefore, this article defines an adaptive weight factor ω to amplify the foraging range of producers, and the position update of the producer can be rewritten as:

$$\omega = \omega_0 \cdot \exp\left(-\eta \frac{t}{T}\right) \tag{18}$$

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} \cdot \exp\left(\frac{-i}{\omega \cdot \tau \cdot T}\right) & if\ R_{af} < ST \\ X_{i,j}^{t} + R_n \cdot L & if\ R_{af} \geq ST, \end{cases} \tag{19}$$

where $\omega_0 = 1$ is the initial weight and $\eta = 20$ is the attenuation factor used to control the attenuation degree of ω, t denotes the present number of iterations.

B. LEVY FLIGHT STRATEGY

Levy flight is a stochastic walk method [43]. It simulates the flight behavior of various animals and insects in nature [44], with characteristics of randomness and intermittency.

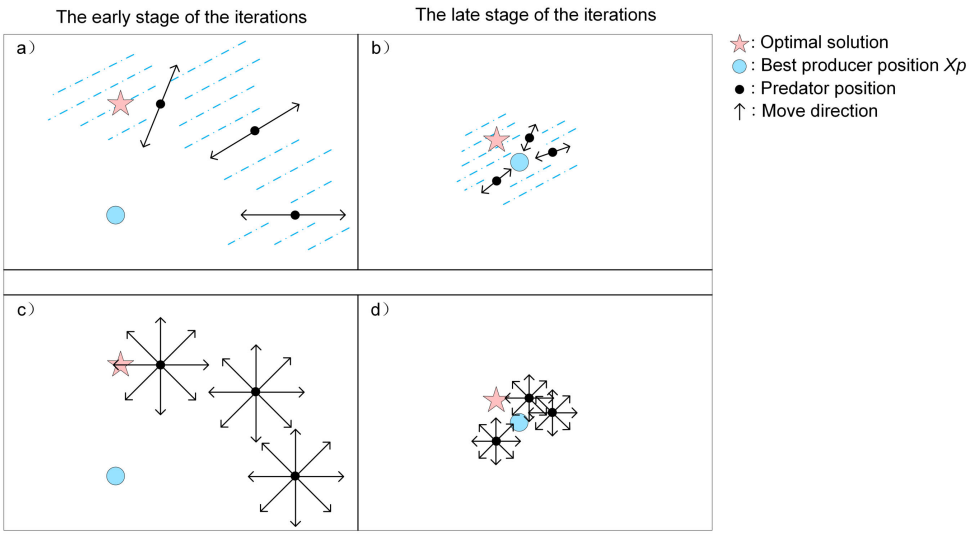


FIGURE 4. The displacement direction of predator sparrows before and after introducing levy flight.

The detailed formula is as below:

$$s \sim N\left(0, \sigma_s^2\right) \tag{20}$$

$$r \sim N\left(0, \sigma_r^2\right) \tag{21}$$

$$\sigma_r = 1 \tag{22}$$

$$\sigma_s = \left\{ \frac{\Gamma(1+\rho) \cdot \sin\left(\frac{\pi\rho}{2}\right)}{\rho \cdot \Gamma\left(\frac{1+\rho}{2}\right) \cdot 2^{\frac{(\rho-1)}{2}}} \right\}^{\frac{1}{\rho}} \tag{23}$$

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt \tag{24}$$

$$levy = \frac{\sigma_s \cdot s}{|r|^{\frac{1}{\rho}}}, \tag{25}$$

where $\rho = 1.5$ is a constant, Γ represents the gamma function, *levy* denotes the random step size.

In (14), restricted by the value of matrix A, predators can only update positions from two opposite directions, as shown in Fig. 4a) and 4b). We can see that the shadow part is a blind area, and the SSA algorithm cannot search for it sufficiently, which implies that the SSA algorithm is prone to miss the optimal solution and trap a temporary scam. In view of the strong randomness of levy flight, the introduction of levy flight can fully search the vicinity of predators and best producers, which merely improves the local search ability of the algorithm yet effectively avoids the problem of missing the optimal solution, as shown in Fig. 4c) and 4d). In this subsection, we utilize levy random step size to control the updating direction of predator position, and the new predator position can be indicated as:

$$X_{i,j}^{t+1} = \begin{cases} R_n \cdot \exp\left(\frac{X_{worst}^t - X_{i,j}^t}{i^2}\right) & if\ i > n/2 \\ X_P^{t+1} + |X_{i,j}^t - X_P^{t+1}| \cdot levy \cdot L & otherwise. \end{cases} \tag{26}$$

C. ALTERNATE SEARCH MECHANISM BASED ON LEVY MUTATION AND LOBL

Levy flight can alternate between the majority of short-range searches and occasional long-range searches. This intermittent behavior is profit for elevating the algorithm's local approximation power and enabling it to leap out of the local optima. To this end, we combine levy flight with mutation strategy and propose a novel levy mutation method, which is expressed as follows:

$$X_i^{t+1} = X_i^t * (1 + levy) \qquad rand \le P_m, \tag{27}$$

where X_i^t means the position of the ith sparrow in the t iteration, $rand \in (0, 1)$ is a random number, $P_m = 0.05$ indicates the mutation rate.

In addition, in the iterative process of SSA algorithm, the population diversity will decrease, which may lead to the problem of prematurity [45]. Fortunately, opposition-based learning methods can effectively overcome this problem. Among them, the LOBL method has great potential in improving population diversity due to its unique physical characteristics. It is a method to seek the opposite solution based on the principle of convex lens imaging [46], and the calculation identity is as below:

$$\hat{X}_{i,j}^t = \frac{a^j + b^j}{2} + \frac{a^j + b^j}{2k} - \frac{X_{i,j}^t}{k} \tag{28}$$

$$k = k_{\min} + 0.5\,(k_{\max} - k_{\min}) \cdot \left(1 - \cos\left(\frac{t\pi}{T}\right)\right), \tag{29}$$

where $X_{i,j}^t$ and $\hat{X}_{i,j}^t$ represents original solution and opposite solution the jth dimension of the ith sparrow in the t iteration, respectively. a^j and b^j stands for upper and lower bounds of the jth dimension, respectively. k is a nonlinear decreasing function, with its bound is [0.5, 1.5]. LOBL can produce a variety of different opposite solutions by adaptive adjusting the size of k, so that raise the diversity of the population [47].

TABLE 4. Benchmark function.

ID	Function	Search scope	Dimension	Minimum
F1	$f(x) = \sum_{i=1}^{n} x_i^2$	[-100,100]	30	0
F2	$f(x) = \sum_{i=1}^{n} ix_i^4 + rand(0,1)$	[-1.28,1.28]	30	0
F3	$f(x) = \sum_{i=1}^{n} -x_i sin(\lvert x_i \rvert^{\frac{1}{2}})$	[-500,500]	30	-12569.487
F4	$f(x) = \frac{\pi}{n}\{10\sin^2(\pi y_1) + \sum_{i=1}^{n-1}(y_i-1)^2[1+10\sin^2(\pi y_{i+1})] + (y_n-1)^2\} + \sum_{i=1}^{n} u(x_i,10,100,4)$, $u(x_i,m,n,p) = \begin{cases} n(x_i-m)^p x_i > m \\ 0 - m \le x_i \le m \\ n(-x_i-m)^p x_i < -m \end{cases}$	[-50,50]	30	0
F5	$f(x) = -\sum_{i=1}^{4} c_i \exp(-\sum_{j=1}^{3} a_{ij}(x_j - q_{ij})^2)$	[0,1]	3	-3.86
F6	$f(x) = -\sum_{i=1}^{10} [(X-a_i)(X-a_i)^T + c_i]^{-1}$	[0,10]	4	-10.5363

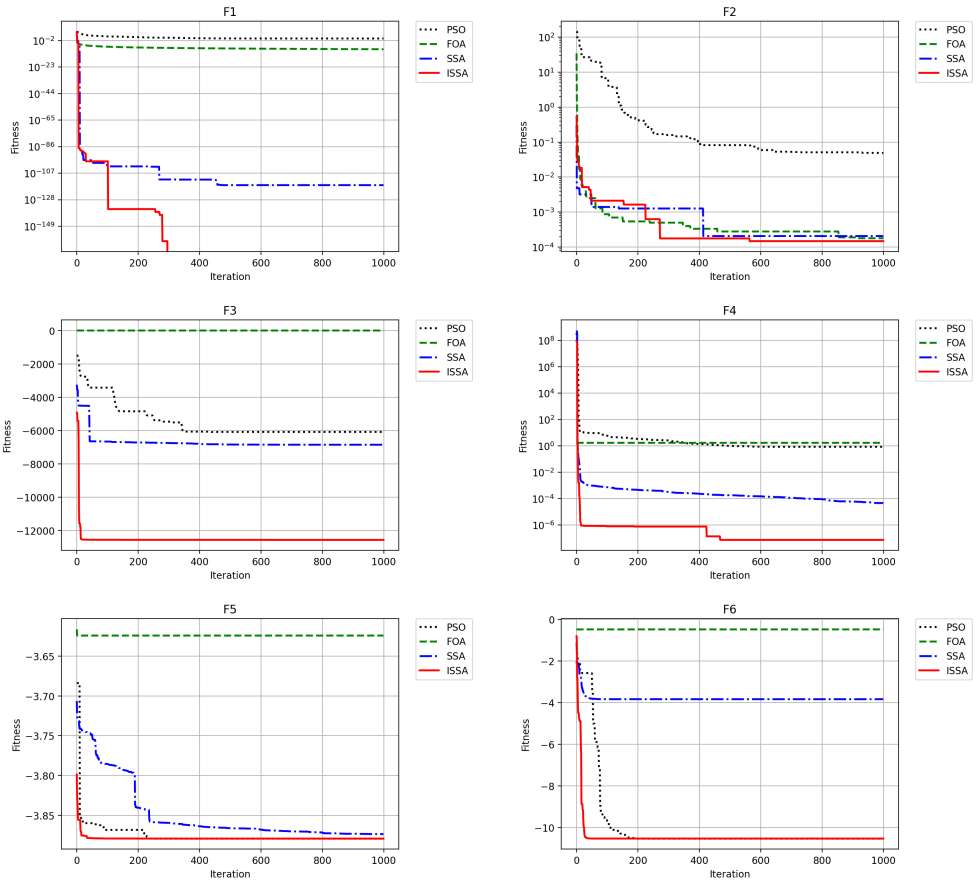


FIGURE 5. Fitness curves of PSO, FOA, SSA and ISSA on different benchmark functions.

Inspired by [48], we propose an alternate search mechanism based on levy mutation and LOBL. The core of this mechanism is to introduce a selection probability P_s, and selectively mutate or seek opposite solutions for the whole population after each iteration to further upgrade the performance of the SSA algorithm. Specifically, when $rand < P_s$, the levy mutation is executed. When $rand \ge P_s$, the LOBL is executed. Finally, the position of all populations is updated by greedy rule. P_s is a function that increases from 0 to 1 and can be denoted as follows:

$$P_s = 0.5 \cdot \left(1 - \cos\left(\frac{t\pi}{T}\right)\right). \tag{30}$$

D. COMPARISON AND VALIDATION

To authenticate the validity of the proposed ISSA, it was tested using six benchmark functions and the outcomes were contrasted with PSO, FOA, and SSA. Table 4 shows the details of the benchmark function. The comparison outcomes are shown in Fig. 5.

We can see that ISSA can accurately find the global optimal solution on the F1 - F6. On the F1 and F4 test functions, only ISSA successfully jumps out of the local optimum. On the F2 test function, ISSA, SSA, and FOA have similar convergence accuracy and are better than PSO. On the F3 test function, ISSA is significantly superior to SSA, PSO, and FOA. On the

TABLE 5. The optimal weight of each prediction network under different data sets.

City	Step	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}
Beijing	1	0.8426	0.8735	0.9941	0.9201	0.8984	1.0600	0.9926	1.0620	-	-	-
	3	0.9394	0.9914	1.1086	0.9288	1.0445	1.0482	1.0071	1.2725	-	-	-
	5	0.8005	0.9522	0.8826	1.0765	1.0084	1.0437	1.0539	1.0277	-	-	-
Handan	1	0.9542	0.9950	0.9037	1.0089	1.0308	0.9190	0.9016	0.9962	0.9893	1.2294	1.2026
	3	0.9474	0.9506	0.9477	0.9479	0.9433	0.9487	0.9510	0.9410	0.9513	0.9469	0.9475
	5	0.9449	0.9678	0.9323	0.9384	0.9851	0.9667	0.9379	0.9025	0.9281	0.9207	0.9748

F5 and F6 test functions, both ISSA and PSO have the highest convergence accuracy, but the time consumed by ISSA is much less than that of PSO. In conclusion, compared with PSO, FOA, and SSA, ISSA is optimal both convergence accuracy and convergence rate.

Based on the above tests and the results of Section III-B, we utilized the validated ISSA algorithm to optimize the weights of each prediction network and achieved satisfactory prediction outcomes. The optimal weights of each prediction network under different data sets are shown in Table 5.

V. EMPIRICAL STUDY

A. DATA DESCRIPTION

The PM2.5 concentration data used in this experiment come from China Air quality online Monitoring and Analysis Platform (https://www.aqistudy.cn/). The meteorological data come from the Weather Network (https://www.tianqi.com/). The sample data are PM2.5 concentration data and meteorological data from January 1, 2019 to December 31, 2021 in Beijing and Handan, which with an interval of one day. The meteorological data include highest temperature, lowest temperature, wind power, wind direction, air pressure, humidity and weather, a total of 7 meteorological related factors. Among them, the wind direction and weather belong to the nonnumerical type, which has been numerically processed in advance. In the process of data collection, there will inevitably be missing data. Here, the cubic spline interpolation approach is employed to preprocess the missing data. Table 6 shows the statistical information of PM2.5 concentration data after preprocessing. We beheld that there is a noteworthy discrepancy in PM2.5 concentration between these two regions, and both of which are heavily polluted cities. Despite the geographical proximity of Beijing and Handan, the PM2.5 concentration sequence of the two cities presents different characteristics due to their significant differences in regional scale, population, traffic, industrial structure and climate characteristics [27]. Therefore, the broad applicability and practicability of the presented model can be fully verified on these two data sets.

The dataset is separated into three portions in this paper, with a 6:2:2 rate: training set, verification set, and testing set. The first 657 data (2019/1/1-2020/10/18) is the training set, the next 219 data (2020/10/19-2021/5/25) is the verification set, and the last 220 data (2021/5/26-2021/12/31) is the test set, which is used for model training, avoid overfitting and performance evaluation, respectively.

TABLE 6. Statistical information of PM2.5 concentration data after preprocessing.

City	Maximum	Minimum	Mean	Standard Deviation	Skewness	Kurtosis
Beijing	217	3	37.68	33.36	2.01	5.11
Handan	329	5	56.14	43.47	2.23	6.29

B. PERFORMANCE EVALUATION CRITERIA

There are many different categories of indicators used to evaluate model performance, but there is no pervasive international evaluation criterion at present. Therefore, this article utilizes the four mainstream error indicator to authenticate the validity of the proposed model, including mean absolute error (MAE), root mean square error (RMSE), symmetric mean absolute percentage error (SMAPE), and R-square (R^2). The above indicators can be denoted as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_{pre}(i) - y_{obs}(i)| \tag{31}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_{pre}(i) - y_{obs}(i)\right)^2} \tag{32}$$

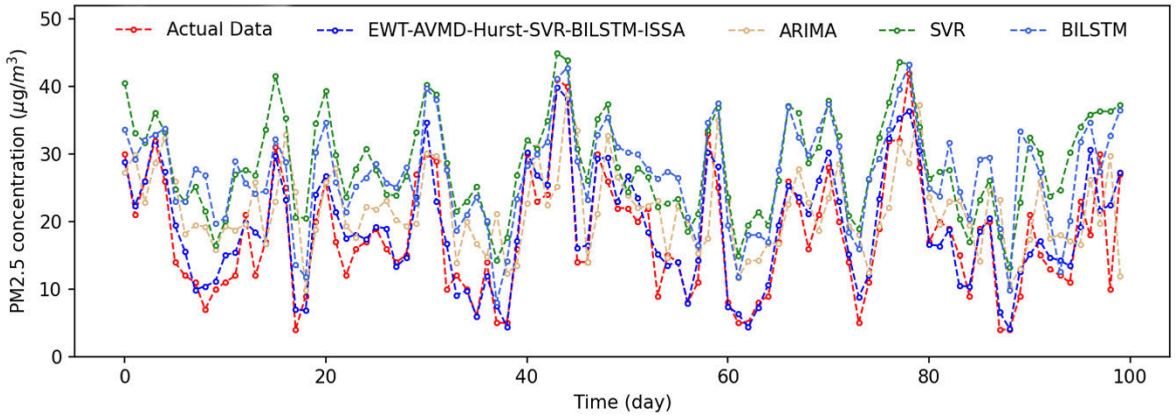
$$SMAPE = \frac{100\%}{N}\sum_{i=1}^{N}\frac{|y_{pre}(i) - y_{obs}(i)|}{\left(|y_{pre}(i)| + |y_{obs}(i)|\right)/2} \tag{33}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_{pre}(i) - y_{obs}(i)\right)^2}{\sum_{i=1}^{N}\left(y_{avg} - y_{obs}(i)\right)^2}, \tag{34}$$

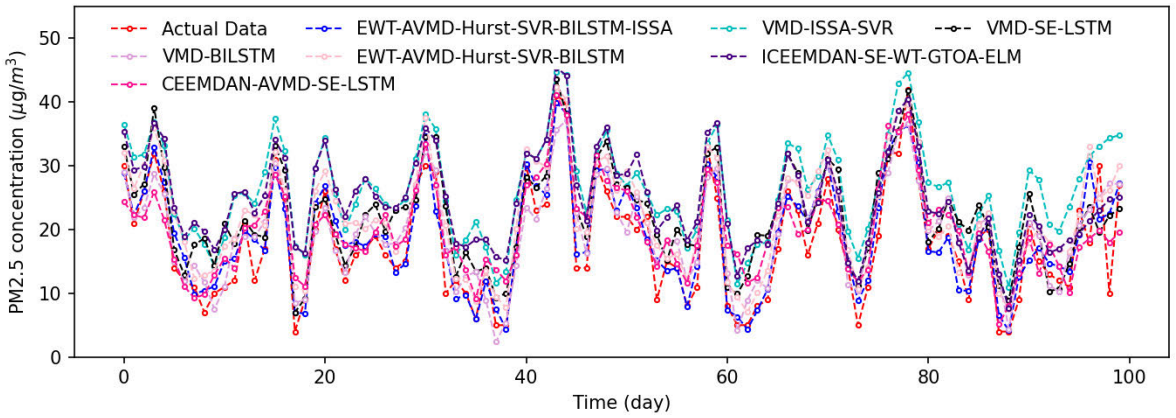
where y_{pre}, y_{obs}, and y_{avg} represent the predicted, observed, and average values of test sets, respectively. N stands for number of test sets.

C. MODEL COMPARISON

In order to authenticate the superiority of the proposed weighted combination forecast model, we compared the proposed model with nine contrast models, including ARIMA [49], SVR [10], BILSTM [13], VMD-ISSA-SVR, ICEEMDAN-SE-WT-GTOA-ELM [50], VMD-SE-LSTM



(a) Compared with single models



(b) Compared with hybrid models

FIGURE 6. 1-Step prediction outcomes of PM2.5 concentration in Beijing case.

TABLE 7. Performance statistics of 1-step prediction of different models in Beijing case.

City	Models	MAE	RMSE	SMAPE	R^2
Beijing	ARIMA	11.2172	16.2332	47.8744	0.4519
	SVR	10.4936	12.6797	49.3368	0.6656
	BILSTM	9.1371	10.9181	46.1128	0.7520
	VMD-ISSA-SVR	7.4484	9.7666	40.7380	0.8353
	ICEEMDAN-SE-WT-GTOA-ELM	7.4812	9.4462	42.3119	0.8297
	VMD-SE-LSTM	6.1860	8.0571	32.9826	0.8649
	CEEMDAN-AVMD-SE-LSTM	5.9070	7.6323	29.2288	0.8788
	VMD-BILSTM	5.0984	7.2428	29.1315	0.8908
	EWT-AVMD-Hurst-SVR-BILSTM	5.6728	6.7470	31.6812	0.8937
	EWT-AVMD-Hurst-SVR-BILSTM-ISSA	**4.0900**	**5.7810**	**23.4720**	**0.9304**

[19], CEEMDAN-AVMD-SE-LSTM [51], VMD-BILSTM [52], EWT-AVMD-Hurst-SVR-BILSTM. In addition, 1, 3, and 5-step predictions are carried out for all models to measure their multi-step prediction ability. The detailed comparison results of different cases will be presented in the following subsections.

1) PERFORMANCE COMPARISON IN BEIJING CASE

The outcomes of 1-step prediction of PM2.5 concentration for Beijing are shown in Fig. 6. The performance evaluation outcomes of different models are displayed in Table 7.

As shown in Fig. 6, compared with rest contrast models, the predicted value sequence of the proposed model is closer to the real value sequence and has the best prediction performance. The hybrid prediction model based on "decomposition and ensemble" method can accurately predict the overall trend of PM2.5 concentration sequences. The predicted value sequences of ARIMA, SVR, and BILSTM exhibit significant fluctuations, and their performance is far inferior to hybrid prediction models. The performance statistics in Table 7 further confirm the above results. Apparently, the proposed EWT-AVMD-Hurst-SVR-BILSTM-ISSA has

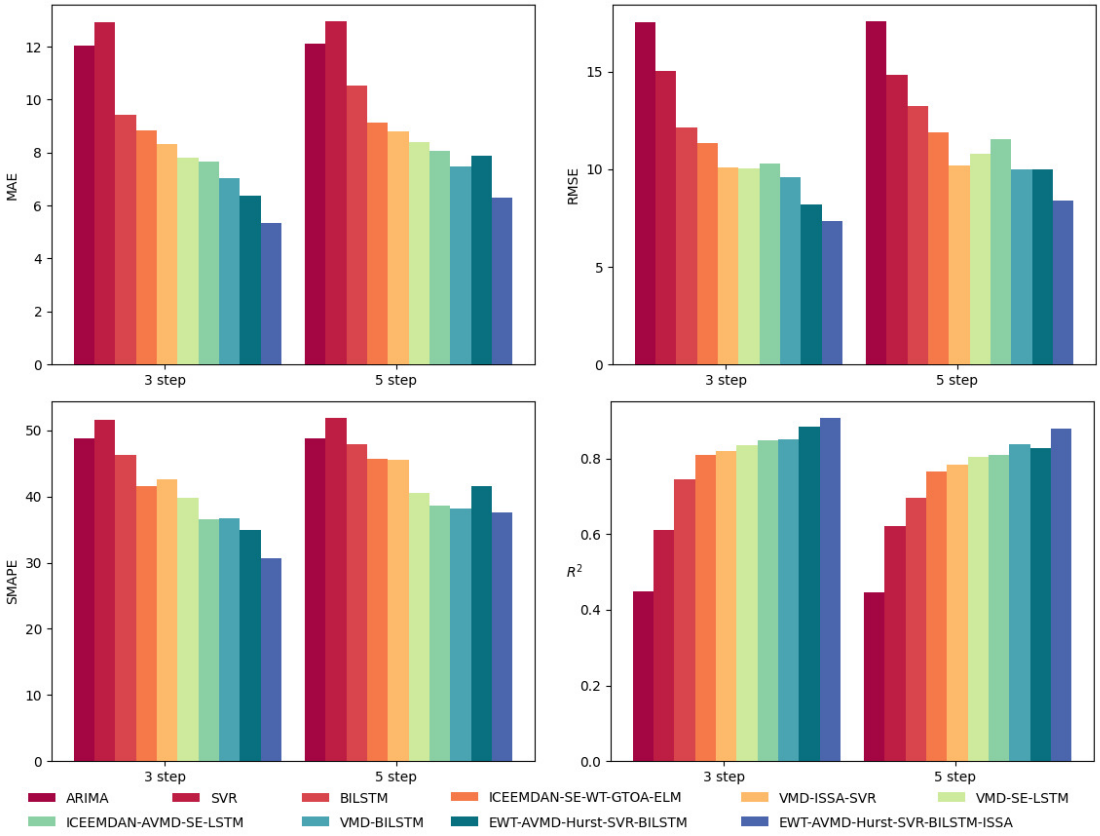


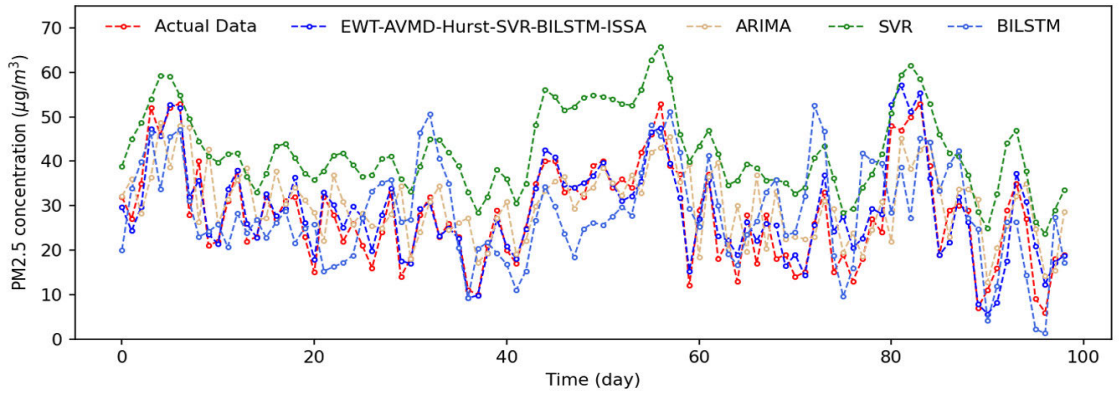
FIGURE 7. Performance statistical histogram of 3 and 5-step prediction in Beijing case.

TABLE 8. Performance statistics of 3 and 5-step prediction of different models in Beijing case.

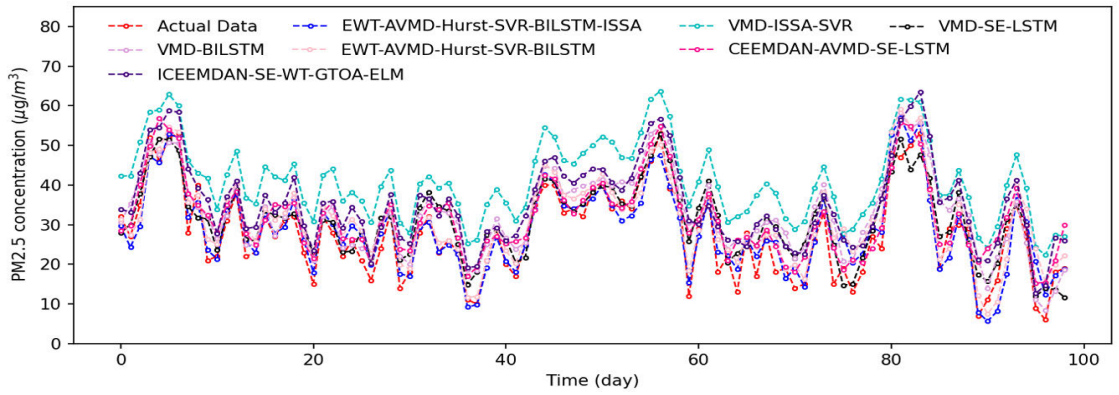
City	Models	3 Step				5 Step			
		MAE	RMSE	SMAPE	R^2	MAE	RMSE	SMAPE	R^2
Beijing	ARIMA	12.0478	17.5528	48.7272	0.4482	12.0967	17.5793	48.7952	0.4466
	SVR	12.9210	15.0314	51.5632	0.6100	12.9510	14.8178	51.8612	0.6210
	BILSTM	9.4350	12.1598	46.3053	0.7448	10.5132	13.2467	47.9586	0.6971
	VMD-ISSA-SVR	8.3332	10.1015	42.5426	0.8211	8.8101	10.1829	45.6101	0.7843
	ICEEMDAN-SE-WT-GTOA-ELM	8.8223	11.3255	41.6212	0.8088	9.1173	11.8997	45.7124	0.7668
	VMD-SE-LSTM	7.8086	10.0419	39.8733	0.8359	8.3885	10.7866	40.5004	0.8047
	CEEMDAN-AVMD-SE-LSTM	7.6565	10.3051	36.6130	0.8483	8.0608	11.5412	38.5626	0.8096
	VMD-BILSTM	7.0411	9.5804	36.7630	0.8516	7.4620	10.0008	38.1484	0.8373
	EWT-AVMD-Hurst-SVR-BILSTM	6.3630	8.1829	34.9148	0.8844	7.8894	10.0002	41.5963	0.8274
	EWT-AVMD-Hurst-SVR-BILSTM-ISSA	**5.3360**	**7.3399**	**30.6014**	**0.9070**	**6.2820**	**8.3916**	**37.6704**	**0.8784**

the lowest MAE, RMSE, SMAPE and the highest R^2, which are 4.0900, 5.7810, 23.4720, and 0.9304, respectively. For other hybrid prediction models, the prediction performance of VMD-ISSA-SVR and ICEEMDAN-SE-WT-GTOA-ELM is slightly worse, with MAE, RMSE, SMAPE, and R^2 are 7.4484, 9.7666, 40.7380, 0.8353 and 7.4812, 9.4462, 42.3119, 0.8297, respectively. The possible reason is that SVR and ELM cannot sufficiently extract the potential features of complex signals. Nevertheless, the prediction outcomes of the model are also satisfactory. In the single prediction model, BILSTM outperforms SVR and ARIMA, with MAE, RMSE, SMAPE, and R^2 of 9.1371, 10.9181, 46.1128, and 0.7520 respectively, but its performance is far inferior to that of VMD-ISSA-SVR and ICEEMDAN-SE-WT-GTOA-ELM. This demonstrates that the ''decomposition and ensemble'' method can efficaciously strengthen the expression power and generalization performance of the model, which has a enthusiastic impact on the forecast of PM2.5 concentration.

Taking into account that the multi-step prediction ability of model is more extensive in practical application [53]. To this end, Fig. 7 and Table 8 evaluate the multi-step



(a) Compared with single models



(b) Compared with hybrid models

FIGURE 8. 1-Step prediction outcomes of PM2.5 concentration in Handan case.

TABLE 9. Performance statistics of 1-step prediction of different models in Handan case.

City	Models	MAE	RMSE	SMAPE	R^2
Handan	ARIMA	13.0764	20.3681	33.1588	0.4781
	SVR	12.9657	16.0131	37.3750	0.6774
	BILSTM	12.0232	15.6249	36.9537	0.6928
	VMD-ISSA-SVR	11.7277	13.3442	32.2217	0.7960
	ICEEMDAN-SE-WT-GTOA-ELM	9.9322	13.2783	28.0788	0.8113
	VMD-SE-LSTM	7.1841	10.4146	21.5375	0.8635
	CEEMDAN-AVMD-SE-LSTM	6.7437	10.2102	23.3012	0.8655
	VMD-BILSTM	6.1190	10.1450	20.5994	0.8705
	EWT-AVMD-Hurst-SVR-BILSTM	5.1827	7.1545	15.9416	0.9356
	EWT-AVMD-Hurst-SVR-BILSTM-ISSA	**4.3635**	**6.3160**	**14.0291**	**0.9498**

prediction performance of the various models. Combining Fig. 7 and Table 8, we can clearly see that whether it is 3-step prediction or 5-step prediction, EWT-AVMD-Hurst-SVR-BILSTM-ISSA model procured the lowest MAE, RMSE, and SMAPE of 5.3360, 7.3399, 30.6014, 6.2820, 8.3916, 37.6704, and the highest R^2 of 0.9070, 0.8784. The performance of the single prediction model is seriously inadequate, which is consistent with the above conclusion. In short, the proposed EWT-AVMD-Hurst-SVR-BILSTM-ISSA model provides the best prediction accuracy and effect.

2) PERFORMANCE COMPARISON IN HANDAN CASE

This subsection takes the PM2.5 concentration in Handan as the research object to further verify the applicability and dependability of the proposed model. The fitting curve of the 1-step prediction are shown in Fig. 8, and Table 9 lists the performance statistics outcomes of different models. We observed that the MAE, RMSE, and SMAPE of EWT-AVMD-Hurst-SVR-BILSTM-ISSA are 4.3635, 6.3160, 14.0291, respectively, and R^2 is 0.9498, which is better than other contrast models. VMD-ISSA-

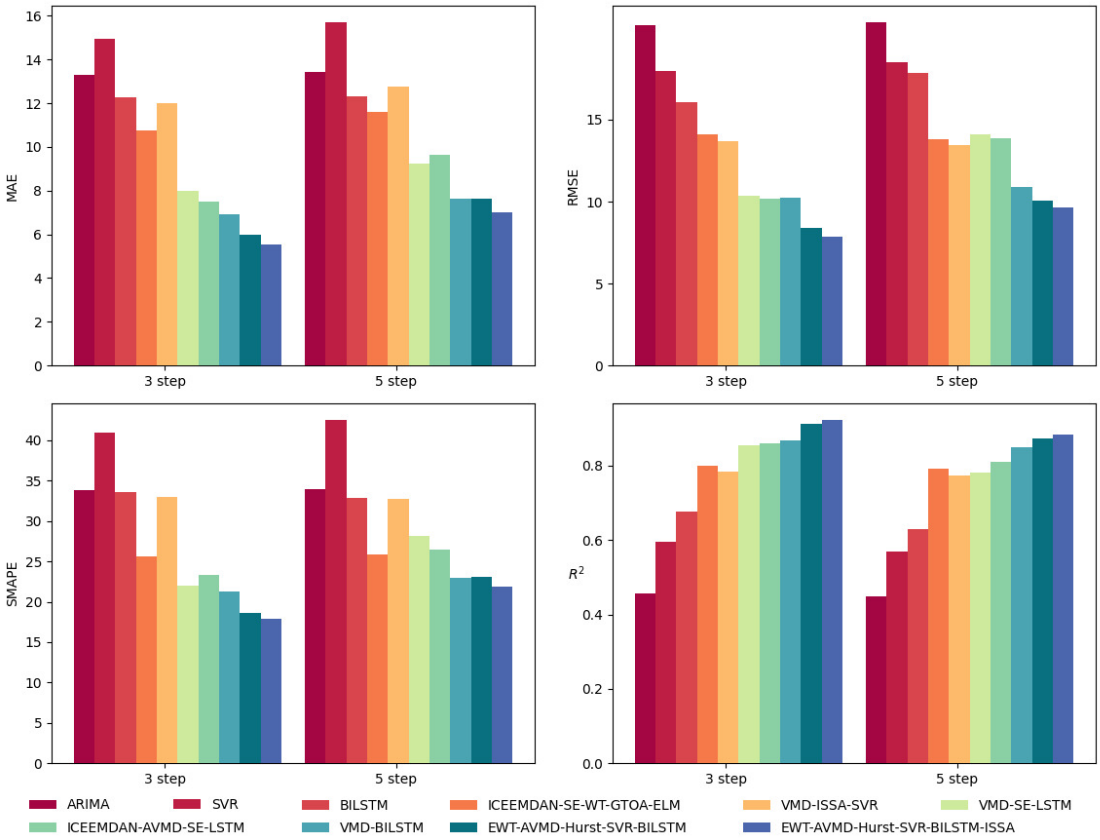


FIGURE 9. Performance statistical histogram of 3 and 5-step prediction in Handan case.

TABLE 10. Performance statistics of 3 and 5-step prediction of different models in Handan case.

City	Models	3 Step				5 Step			
		MAE	RMSE	SMAPE	R^2	MAE	RMSE	SMAPE	R^2
Handan	ARIMA	13.2927	20.7659	33.7950	0.4575	13.4217	20.9296	33.9416	0.4489
	SVR	14.9531	17.9462	40.9753	0.5948	15.6910	18.5135	42.4961	0.5688
	BILSTM	12.2784	16.0490	33.6219	0.6760	12.2929	17.8542	32.8345	0.6290
	VMD-ISSA-SVR	11.9911	13.6924	32.9914	0.7841	12.7636	13.4395	32.7794	0.7727
	ICEEMDAN-SE-WT-GTOA-ELM	10.7369	14.1101	25.6747	0.8001	11.6023	13.7951	25.8211	0.7917
	VMD-SE-LSTM	8.0000	10.3425	22.0548	0.8554	9.2268	14.0743	28.2017	0.7808
	CEEMDAN-AVMD-SE-LSTM	7.5094	10.1747	23.3347	0.8591	9.6538	13.8687	26.4504	0.8111
	VMD-BILSTM	6.9258	10.2409	21.2678	0.8680	7.6247	10.8979	22.9959	0.8506
	EWT-AVMD-Hurst-SVR-BILSTM	5.9838	8.4067	18.6400	0.9111	7.6235	10.0655	23.1035	0.8725
	EWT-AVMD-Hurst-SVR-BILSTM-ISSA	**5.5255**	**7.8595**	**17.9128**	**0.9222**	**6.9889**	**9.6328**	**21.9255**	**0.8832**

SVR and ICEEMDAN-SE-WT-GTOA-ELM exhibit a weak correlation between predicted and observed values, but vaguely superior to BILSTM. The error analysis outcomes of 3 and 5-step prediction are shown in Fig. 9 and Table 10. Obviously, the R^2 of the proposed model has been kept around 0.9, which has a strong fitting advantage. The results of the above analysis are intensively coherent with the Beijing case, which means that the proposed model can suit to different environments excellently and has significant practical application value.

D. DISCUSSION

The comparison results of the two city cases indicate that the proposed hybrid prediction model EWT-AVMD-Hurst-SVR-BILSTM-ISSA has outstanding performance in different environments. In VMD-ISSA-SVR, VMD-SE-LSTM, and VMD-BILSTM models, the parameters of VMD are universally affected by human intervention, resulting in the characteristics of the data cannot be adequately expressed. Our research ingeniously integrates the EOS indicator with a grid search strategy that progresses from rough-to-fine, and

TABLE 11. Performance statistics of 1, 3 and 5-step prediction of different models in Shanghai case.

Forecasting models	Step	MAE	RMSE	SMAPE	R^2
ARIMA	1	8.8015	12.0152	37.4283	0.4003
	3	8.8289	12.0007	37.6127	0.3926
	5	8.8226	12.0217	37.5056	0.3996
SVR	1	7.3684	8.9247	32.9371	0.6837
	3	7.8008	9.4138	34.3974	0.6574
	5	8.2943	9.6211	34.5778	0.6171
BILSTM	1	7.1523	9.3503	29.3092	0.7319
	3	7.4736	10.2946	31.3941	0.6984
	5	7.8670	10.6393	32.1060	0.6736
VMD-ISSA-SVR	1	4.8410	6.0567	22.3503	0.8532
	3	5.6187	6.2585	22.6669	0.8374
	5	6.0634	6.5182	23.8928	0.8223
ICEEMDAN-SE-WT-GTOA-ELM	1	4.9825	6.4120	23.9774	0.8275
	3	5.8821	6.4385	27.2283	0.8111
	5	6.4659	7.0121	29.8525	0.7765
VMD-SE-LSTM	1	4.6875	6.3872	18.0717	0.8842
	3	5.1536	6.9597	19.9698	0.8590
	5	5.9169	7.7722	24.0910	0.8245
CEEMDAN-AVMD-SE-LSTM	1	4.7864	6.5874	18.9494	0.8747
	3	5.5170	7.3383	21.8157	0.8437
	5	5.7802	7.5810	23.4555	0.8287
VMD-BILSTM	1	4.6103	6.3217	17.9663	0.8846
	3	5.1072	6.8271	20.4492	0.8647
	5	5.5973	7.3863	22.6210	0.8413
EWT-AVMD-Hurst-SVR-BILSTM	1	2.4672	3.1353	12.1558	0.9603
	3	3.4014	4.4701	17.3426	0.9193
	5	4.0323	5.2418	20.8100	0.8890
EWT-AVMD-Hurst-SVR-BILSTM-ISSA	1	2.2930	2.9687	11.1099	0.9644
	3	3.3358	4.3037	18.5832	0.9252
	5	4.0015	5.2313	20.6710	0.8894

realizes the accurate location of the optimum parameters. The superiority of the proposed model is indirectly confirmed by this as well. Furthermore, it is noteworthy that the proposed model shows a significant improvement in MAE, RMSE, SMAPE by 27.9%, 19.1%, 25.9% respectively compared to EWT-AVMD-Hurst-SVR-BILSTM, with an increase of R^2 by 3.9%. Similarly, in Handan case, MAE, RMSE, and SMAPE increased by 15.8%, 11.7%, 11.9% respectively, and R^2 increased by 1.5%. The above analysis results show that the optimized weight coefficient can preferably approximate the real value, and it is feasible to use ISSA algorithm for error correction.

We can see from Figs. 7 and 9 that the MAE, RMSE, and SMAPE of 5-step prediction for each model are generally taller than those of 3-step prediction, and R2 is on the contrary. For Beijing, the MAE predicted by EWT-AVMD-Hurst-SVR-BILSTM-ISSA model at 1, 3, and 5 steps are 4.0900, 5.3360, and 6.2820, respectively, showing an upward trend, while R^2 of 0.9304, 0.9070, and 0.8784 respectively, showing a decreasing trend. This phenomenon has also been further demonstrated in the case of Handan City. The cause of this phenomenon could be that the correlation between input variables and output variables weakens with the increase of prediction step size, resulting in a dramatic decline in prediction performance. Unfortunately, the proposed model has a weak advantage in multi-step prediction performance and needs to be improved, but it outperforms other contrast models on the whole.

To ensure that the proposed model can be effectively applied in different scenarios, we collected PM2.5 concentration and meteorological data in Shanghai, and as an additional case study. The experimental procedure is the same as above. Table 11 presents the performance evaluation results of different models. As can be seen from Table 11, the behavioral characteristics of each model in the Shanghai case are highly consistent with those in the Beijing and Handan cases, and we can draw conclusions similar to those in Section V. In conclusion, the proposed weighted combination forecast model is a promising model and can be taken as a energetic tool for PM2.5 concentration prediction.

VI. CONCLUSION

In this article, a hybrid prediction model of PM2.5 concentration based on secondary decomposition ensemble and weight combination optimization is presented to improve the shortcomings of the existing models. The validity of the model is authenticated by using real case data of Beijing, Handan and Shanghai. The experimental results reveal that compared with other contrast models, the proposed model primarily embodies the following three advantages: (1) the proposed AVMD algorithm alleviates the repercussions of external perturbance on the performance of the model to some extent, and effectively reveals the potential characteristics of PM2.5 concentration series; (2) the signal division mechanism based on Hurst exponent realizes the synchronous determination of prediction network and meteorological data, which promotes the adaptive capacity of the model to complex nonlinear dynamic series; (3) the learning paradigm of weight optimization balances the prediction performance of different networks and enhances the stability of the model. In the future, we hope to expand our approach to additional cities and investigate the underlying causes of air pollution.

REFERENCES

[1] N. Sirisumpun, K. Wongwailikhit, P. Painmanakul, and P. Vateekul, "Spatio-temporal $PM_{2.5}$ forecasting in Thailand using encoder–decoder networks," *IEEE Access*, vol. 11, pp. 69601–69613, Jul. 2023.

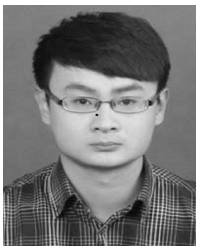
[2] Q. Guo, Z. He, and Z. Wang, "Long-term projection of future climate change over the twenty-first century in the Sahara region in Africa under four shared socio-economic pathways scenarios," *Environ. Sci. Pollut. Res.*, vol. 30, no. 9, pp. 22319–22329, Oct. 2022.

[3] WHO. *Billions of People Still Breathe Unhealthy Air: New WHO Data*. Accessed: May 5, 2023. [Online]. Available: https://www.who.int/news/item/04-04-2022-billions-of-people-still-breathe-unhealthy-air-new-who-data

[4] M. Niu, Y. Wang, S. Sun, and Y. Li, "A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term $PM_{2.5}$ concentration forecasting," *Atmos. Environ.*, vol. 134, pp. 168–180, Jun. 2016.

[5] Y. Li, Z. Liu, and H. Liu, "A novel ensemble reinforcement learning gated unit model for daily $PM_{2.5}$ forecasting," *Air Quality, Atmos. Health*, vol. 14, no. 3, pp. 443–453, Mar. 2021.
[6] A. Tella and A.-L. Balogun, "GIS-based air quality modelling: Spatial prediction of PM_{10} for Selangor state, Malaysia using machine learning algorithms," *Environ. Sci. Pollut. Res.*, vol. 29, no. 57, pp. 86109–86125, Sep. 2021.
[7] H. Liu, K. Jin, and Z. Duan, "Air $PM_{2.5}$ concentration multi-step forecasting using a new hybrid modeling method: Comparing cases for four cities in China," *Atmos. Pollut. Res.*, vol. 10, no. 5, pp. 1588–1600, Sep. 2019.
[8] Q. Guo, Z. He, and Z. Wang, "Simulating daily $PM_{2.5}$ concentrations using wavelet analysis and artificial neural network with remote sensing and surface observation data," *Chemosphere*, vol. 340, Nov. 2023, Art. no. 139886.
[9] K. Huang, Q. Xiao, X. Meng, G. Geng, Y. Wang, A. Lyapustin, D. Gu, and Y. Liu, "Predicting monthly high-resolution $PM_{2.5}$ concentrations with random forest model in the North China plain," *Environ. Pollut.*, vol. 242, pp. 675–683, Nov. 2018.
[10] X. Lai, H. Li, and Y. Pan, "A combined model based on feature selection and support vector machine for $PM_{2.5}$ prediction," *J. Intell. Fuzzy Syst.*, vol. 40, no. 5, pp. 10099–10113, Apr. 2021.
[11] W. Zhai and C. Cheng, "A long short-term memory approach to predicting air quality based on social media data," *Atmos. Environ.*, vol. 237, Sep. 2020, Art. no. 117411.
[12] B. Zhang, H. Zhang, G. Zhao, and J. Lian, "Constructing a $PM_{2.5}$ concentration prediction model by combining auto-encoder with bi-LSTM neural networks," *Environ. Model. Softw.*, vol. 124, Feb. 2020, Art. no. 104600.
[13] M. Zhang, D. Wu, and R. Xue, "Hourly prediction of $PM_{2.5}$ concentration in Beijing based on bi-LSTM neural network," *Multimedia Tools Appl.*, vol. 80, pp. 24455–24468, Apr. 2021.
[14] G. Huang, X. Li, B. Zhang, and J. Ren, "$PM_{2.5}$ concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition," *Sci. Total Environ.*, vol. 768, May 2021, Art. no. 144516.
[15] Y. Li, H. Wu, and H. Liu, "Multi-step wind speed forecasting using EWT decomposition, LSTM principal computing, RELM subordinate computing and IEWT reconstruction," *Energy Convers. Manage.*, vol. 167, pp. 203–219, Jul. 2018.
[16] H. Wang and H. Chen, "A novel particulate matter 2.5 concentration prediction model based on double-layer decomposition and feedback of model learning effect," *IEEE Access*, vol. 10, pp. 12164–12178, 2022.
[17] M. Niu, K. Gan, S. Sun, and F. Li, "Application of decomposition-ensemble learning paradigm with phase space reconstruction for day-ahead $PM_{2.5}$ concentration forecasting," *J. Environ. Manage.*, vol. 196, pp. 110–118, Jul. 2017.
[18] Q. Wu and H. Lin, "A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors," *Sci. Total Environ.*, vol. 683, pp. 808–821, Sep. 2019.
[19] Q. Wu and H. Lin, "Daily urban air quality index forecasting based on variational mode decomposition, sample entropy and LSTM neural network," *Sustain. Cities Soc.*, vol. 50, Oct. 2019, Art. no. 101657.
[20] L. Lv, Z. Wu, J. Zhang, L. Zhang, Z. Tan, and Z. Tian, "A VMD and LSTM based hybrid model of load forecasting for power grid security," *IEEE Trans. Ind. Informat.*, vol. 18, no. 9, pp. 6474–6482, Sep. 2022.
[21] J. Cui, R. Yu, D. Zhao, J. Yang, W. Ge, and X. Zhou, "Intelligent load pattern modeling and denoising using improved variational mode decomposition for various calendar periods," *Appl. Energy*, vol. 247, pp. 480–491, Aug. 2019.
[22] F. Zhang, W. Sun, H. Wang, and T. Xu, "Fault diagnosis of a wind turbine gearbox based on improved variational mode algorithm and information entropy," *Entropy*, vol. 23, no. 7, p. 794, Jun. 2021.
[23] Y. He and K. F. Tsang, "Universities power energy management: A novel hybrid model based on iCEEMDAN and Bayesian optimized LSTM," *Energy Rep.*, vol. 7, pp. 6473–6488, Nov. 2021.
[24] X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, C. You, and T. Chi, "Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation," *Environ. Pollut.*, vol. 231, pp. 997–1004, Dec. 2017.
[25] S. Zhu, X. Qiu, Y. Yin, M. Fang, X. Liu, X. Zhao, and Y. Shi, "Two-step-hybrid model based on data preprocessing and intelligent optimization algorithms (CS and GWO) for NO_2 and SO_2 forecasting," *Atmos. Pollut. Res.*, vol. 10, no. 4, pp. 1326–1335, Jul. 2019.
[26] S. Zhu, J. Sun, Y. Liu, M. Lu, and X. Liu, "The air quality index trend forecasting based on improved error correction model and data preprocessing for 17 port cities in China," *Chemosphere*, vol. 252, Aug. 2020, Art. no. 126474.
[27] H. Luo, D. Wang, C. Yue, Y. Liu, and H. Guo, "Research and application of a novel hybrid decomposition-ensemble learning paradigm with error correction for daily PM_{10} forecasting," *Atmos. Res.*, vol. 201, pp. 34–45, Mar. 2018.
[28] H. Liu and S. Dong, "A novel hybrid ensemble model for hourly $PM_{2.5}$ forecasting using multiple neural networks: A case study in China," *Air Quality, Atmos. Health*, vol. 13, no. 12, pp. 1411–1420, Aug. 2020.
[29] W. Sun and Z. Li, "Hourly $PM_{2.5}$ concentration forecasting based on mode decomposition-recombination technique and ensemble learning approach in severe haze episodes of China," *J. Cleaner Prod.*, vol. 263, Aug. 2020, Art. no. 121442.
[30] S. Fan, D. Hao, Y. Feng, K. Xia, and W. Yang, "A hybrid model for air quality prediction based on data decomposition," *Information*, vol. 12, no. 5, p. 210, May 2021.
[31] M. Zulfiqar, M. Kamran, M. B. Rasheed, T. Alquthami, and A. H. Milyani, "Hyperparameter optimization of support vector machine using adaptive differential evolution for electricity load forecasting," *Energy Rep.*, vol. 8, pp. 13333–13352, Nov. 2022.
[32] Z. Tian, "Modes decomposition forecasting approach for ultra-short-term wind speed," *Appl. Soft Comput.*, vol. 105, Jul. 2021, Art. no. 107303.
[33] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3999–4010, Aug. 2013.
[34] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
[35] Y. Wang, X. Wang, L. Wei, J. Su, and S. Zhu, "DPC-based combined model for $PM_{2.5}$ forecasting: Four cities in China," *Soft Comput.*, vol. 25, pp. 9199–9208, May 2021.
[36] M. Ye, X. Yan, and M. Jia, "Rolling bearing fault diagnosis based on VMD-MPE and PSO-SVM," *Entropy*, vol. 23, no. 6, p. 762, Jun. 2021.
[37] S. Wang, Z. Yu, G. Xu, and F. Zhao, "Research on tool remaining life prediction method based on CNN-LSTM-PSO," *IEEE Access*, vol. 11, pp. 80448–80464, 2023.
[38] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: Sparrow search algorithm," *Syst. Sci. Control Eng.*, vol. 8, no. 1, pp. 22–34, Jan. 2020.
[39] S. Zhou, H. Xie, C. Zhang, Y. Hua, W. Zhang, Q. Chen, G. Gu, and X. Sui, "Wavefront-shaping focusing based on a modified sparrow search algorithm," *Optik*, vol. 244, Oct. 2021, Art. no. 167516.
[40] Z. Zhang, R. He, and K. Yang, "A bioinspired path planning approach for mobile robots based on improved sparrow search algorithm," *Adv. Manuf.*, vol. 10, no. 1, pp. 114–130, Mar. 2022.
[41] Y. Zhao, C. Li, W. Fu, J. Liu, T. Yu, and H. Chen, "A modified variational mode decomposition method based on envelope nesting and multi-criteria evaluation," *J. Sound Vib.*, vol. 468, Mar. 2020, Art. no. 115099.
[42] J. Wang, Y. Zhang, F. Zhang, W. Li, S. Lv, M. Jiang, and L. Jia, "Accuracy-improved bearing fault diagnosis method based on AVMD theory and AWPSO-ELM model," *Measurement*, vol. 181, Aug. 2021, Art. no. 109666.
[43] D. Wu and C. Yuan, "Threshold image segmentation based on improved sparrow search algorithm," *Multimedia Tools Appl.*, vol. 81, no. 23, pp. 33513–33546, Apr. 2022.
[44] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *Proc. World Congr. Nature Biologically Inspired Comput. (NaBIC)*, Dec. 2009, pp. 210–214.
[45] C. Zhang and S. Ding, "A stochastic configuration network based on chaotic sparrow search algorithm," *Knowl.-Based Syst.*, vol. 220, May 2021, Art. no. 106924.
[46] R. A. Ibrahim, M. A. Elaziz, and S. Lu, "Chaotic opposition-based grey-wolf optimization algorithm based on differential evolution and disruption operator for global optimization," *Expert Syst. Appl.*, vol. 108, pp. 1–27, Oct. 2018.
[47] S. Yan, P. Yang, D. Zhu, W. Zheng, and F. Wu, "Improved sparrow search algorithm based on iterative local search," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–31, Dec. 2021.

[48] P. Wang, Y. Zhang, and H. Yang, "Research on economic optimization of microgrid cluster based on chaos sparrow search algorithm," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–18, Mar. 2021.
[49] U. Kumar and V. K. Jain, "ARIMA forecasting of ambient air pollutants (O_3, NO, NO_2 and CO)," *Stochastic Environ. Res. Risk Assessment*, vol. 24, no. 5, pp. 751–760, Jul. 2010.
[50] F. Jiang, Y. Qiao, X. Jiang, and T. Tian, "MultiStep ahead forecasting for hourly PM_{10} and $PM_{2.5}$ based on two-stage decomposition embedded sample entropy and group teacher optimization algorithm," *Atmosphere*, vol. 12, no. 1, p. 64, Jan. 2021.
[51] W. Sun and Z. Xu, "A hybrid daily $PM_{2.5}$ concentration prediction model based on secondary decomposition algorithm, mode recombination technique and deep learning," *Stochastic Environ. Res. Risk Assessment*, vol. 2022, pp. 1–20, Jan. 2022.
[52] Z. Zhang, Y. Zeng, and K. Yan, "A hybrid deep learning technology for $PM_{2.5}$ air quality forecasting," *Environ. Sci. Pollut. Res.*, vol. 28, no. 29, pp. 39409–39422, Mar. 2021.
[53] L. Wang and Y. He, "M2STAN: Multi-modal multi-task spatiotemporal attention network for multi-location ultra-short-term wind power multi-step predictions," *Appl. Energy*, vol. 324, Oct. 2022, Art. no. 119672.



YUAN HUANG was born in Hebei, China, in 1987. He received the bachelor's and master's degrees from the Hebei University of Engineering, in 2010 and 2013, respectively, and the Ph.D. degree from Yanshan University, in 2017. Since 2017, he has been a Teacher with the School of Information and Electrical Engineering, Hebei University of Engineering. He has published 11 articles. His research interests include data mining and machine learning.

XIAOYU ZHANG was born in Shanxi, China, in 1998. He received the bachelor's degree from the Business College, Shanxi University, in 2021. He is currently pursuing the master's degree with the School of Information and Electrical Engineering, Hebei University of Engineering. His research interests include data mining and machine learning.

YANXIA LI was born in Shandong, China, in 1995. She received the bachelor's degree from Shandong Technology and Business University, in 2017. She is currently pursuing the master's degree with the School of Information and Electrical Engineering, Hebei University of Engineering. Her research interests include data mining, deep learning, and NLP.