

A prediction model for hybrid air pollutants combining two-stage decomposition and two-scale feature selection

Yong Zhang

PowerChina Guizhou Electric Power Engineering Co., Ltd

Siti Wei

Wst99969@163.com

PowerChina Guizhou Electric Power Engineering Co., Ltd

Kaibo Shen

PowerChina Guizhou Electric Power Engineering Co., Ltd

Jindian Ma

Academy of Digital China (Fujian), Fuzhou University

Shuanghe Cao

PowerChina Guizhou Electric Power Engineering Co., Ltd

Dandan Liu

PowerChina Guizhou Electric Power Engineering Co., Ltd

Shuzou Xiao

Electric Power Research Institute of Guizhou Power Grid Co. Ltd

Research Article

Keywords:

Posted Date: October 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8983033/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

A prediction model for hybrid air pollutants combining two-stage decomposition and two-scale feature selection

Author :Yong Zhang¹, Siti Wei¹(Corresponding Author), Kaibo Shen¹, Jindian Ma², Shuanghe Cao¹, Dandan Liu¹, Shuzou Xiao³

1.PowerChina Guizhou Electric Power Engineering Co., Ltd, Guiyang 550008, China

2.Key Laboratory of Spatial Data Mining & Information Sharing of MOE, National & Local Joint Engineering Research Center of Satellite Geospatial Information Technology, Academy of Digital China (Fujian), Fuzhou University, Fuzhou 350108, China

3.Electric Power Research Institute of Guizhou Power Grid Co. Ltd, Guiyang 550002, China

Corresponding Author E-mail address : Wst99969@163.com Tel:+8615100326135

Abstract: The high ozone concentration will cause not only photochemical pollution but also cause significant harm to the human body. Therefore, ozone forecast accuracy is crucial for daily travel, health protection, and the avoidance of pollution-related natural calamities. This paper proposes a new air pollutant prediction model ICEEMDAN-VMD-FE-PACF-RF-TCN to predict ozone. First, the original sequence is decomposed through improved complete adaptive noise ensemble empirical mode(ICEEMDAN) to reduce the volatility and complexity of the data. Second, IMF1 is decomposed by variable mode decomposition(VMD), and the remaining subsequences are reconstructed with fuzzy entropy(FE). Third, the subsequences are selected by two-scale through partial autocorrelation function(PACF) and random forest(RF), and the optimal input dimension is determined, which reduces the input of irrelevant features. Finally, the prediction is made by the time convolution network(TCN). The experimental results show that the proposed model is superior and robust compared to the comparison model and other existing models when forecasting in two different scenarios, Beijing and Shanghai, with MAE, RMSE and SMAPE of 0.53509, 0.73346 and 0.8433%. This research can provide a valuable reference for the prediction of atmospheric pollutants.

Keyword: Two-stage decomposition, Two -scale feature selection, Air pollutant predict

1.Introduction

With the continuous development of society, air pollution has aroused widespread

concern(Wiwatanadate 2014; Xie et al. 2018). Ozone in the troposphere is a vital air pollutant and greenhouse gas, and excessive ozone concentration is the leading cause of photochemical pollution(Pires et al. 2011). Photochemical pollution will harm human health and lead to respiratory diseases(Kampa and Castanas 2008; Strosnider et al. 2019). Especially in densely populated areas, excessive air pollutants are one of the causes of respiratory diseases in humans, and this effect is more severe for children and the elderly. Other research has demonstrated that photochemical pollution not only negatively impacts the ecosystem but also reduces crop yields and, in turn, results in indirect economic losses(Wang XK et al. 2012; Cao et al. 2020). Therefore, regional prediction of ozone concentration can directly reflect changes in the spatial distribution of ozone, provide vital early warning information for patients with respiratory diseases, help individuals make prudent travel choices, and limit the harm caused by air pollutants. At the same time, it also provides a valuable reference for pollution control, urban planning and construction departments.

The existing models of atmospheric pollutant concentration prediction can be divided into three categories: traditional statistical methods, traditional models and intelligent computer methods. Standard models predict air pollutants by simulating mathematical formulations using various physical and chemical processes(Finardi et al. 2008; Corani and Scanagatta 2016). Among them, the Atmospheric Dispersion Modeling System (ADMS) model, the Community Multiscale Air Quality Modeling System (CMAQ) and the Weather Research and Prediction Model and Chemistry (WRF-Chem) model are very representative of the deterministic models(Garaga et al. 2018). Deterministic models have made significant progress in forecasting air pollutants. For example, (Yu S et al. 2014)used the CMAQ model to simulate PM and O₃ distribution across the Northern Hemisphere; (Hu et al. 2016) uses CMAQ to illustrate China's air pollution capabilities ; (Wang DD et al. 2018) analyzed the pollutants in Beijing winter in 2016 through the WRF-Chem model; Using the WRF-Chem model to analyze air pollution in the Amazon rainforest. On the basis of deterministic models, further increasing the use of statistical methods may lead to better results in terms of prediction accuracy and decision-making(Rybarczyk and Zalakeviciute 2018).Statistical forecasting methods build relevant models based on historical data(Goyal et al. 2006). Unlike deterministic models, statistical prediction methods are not based on comprehension of the mechanism of pollutant formation

but on mining the relationship between historical data; once discovered, these correlations can be applied to predictions of pollutants. For example; (Bai L et al. 2018) gave a detailed overview of pollutant statistical prediction methods. They describe the rationale for each statistical forecasting method, but the strengths, weaknesses and practical application of these models have not been proven.

With the continuous development of science and the deepening of research, traditional forecasting models and regression methods can no longer meet the current needs. Intelligent computer methods – including machine learning methods such as deep learning (DL), artificial neural networks (ANNs), support vector machines (SVMs), and extreme learning machines (ELMs), are data-driven at their core. In recent years, machine learning has been widely used in prediction. Numerous studies have shown that the accuracy of AI predictions outperforms traditional statistical methods (Taheri Shahraiyi and Sodoudi 2016). (Gardner and Dorling 1999) used an artificial neural network (ANN) combined with meteorological data such as temperature and wind direction to model NO₂ concentration, and the results were better than traditional regression models. The air pollutant prediction problem has the characteristics of typical time series. Models such as recurrent neural network (RNN) can arbitrarily process the memory of the input sequence for time series prediction. As a result, RNNs have developed many variant models, such as Long Short-Term Memory (LSTM) (Hochreiter et al. 2007), Gradient Boosted Decision Tree (GBDT), and Gate Recurrent Unit (GRU). A new prediction model is proposed, which combines the hybrid model of convolutional neural network (CNN) and LSTM, and has good accuracy in PM_{2.5} prediction (Huang and Kuo 2018). (Luna et al. 2014) demonstrated that nonlinear tools such as artificial neural networks (ANNs) and support vector machines (SVMs) are very effective for predicting ozone concentrations. With the deepening of research, the prediction accuracy of a single prediction model has been unable to meet people's needs. To improve its efficiency and practicability, various optimization algorithms are widely used to optimize prediction models, including Grey Wolf Algorithm (GWO), Genetic Algorithm (GA), Sparrow Algorithm (SSA), etc. The prediction accuracy of the hybrid optimization model is more convincing than that of a single model, which has been confirmed in many studies. (Cabaneros et al. 2020) proposed a hybrid artificial neural network and applied the model to the prediction of nitrogen dioxide on urban roads. (Chen et al.

2019)proposed optimization of support vector machines with mixed particle swarms and predicted air pollutants in Beijing.

However, air pollutant concentration data are highly complex and noisy. To reduce data noise, the key idea is to partition the complex original sequence into numerous modal components with similar sizes. Studies have proved that this method can effectively improve the prediction accuracy of the model(Sun W and Xu 2021). The applicability of the hybrid model on the original premise is substantially improved by this theoretical framework (Li et al. 2021).(Zhu et al. 2019) proposed a new model, which combines the optimized support vector machine with fully integrated empirical mode decomposition to form a new hybrid model. The results showed that the model has high prediction accuracy.(Ahani et al. 2020) built a multi-step prediction model for PM2.5, pre-processed the data by integrated empirical mode decomposition (EEMD), and then brought the decomposed sequence into the prediction model. Although the introduction of signal processing technology reduces the complexity of the original data to a large extent, it has been confirmed that the first modal function (IMF1) obtained by the decomposition still has high complexity(Sun W and Huang 2020). Therefore, it is a meaningful attempt to process IMF1 again. Also, after decomposition, we need to predict each subsequence. While the decomposition algorithm improves the prediction accuracy, it also dramatically increases the running time of the model. Meanwhile, as the feature space dimension grows in machine learning, so does the training time. The "curse of dimensionality" problem can arise as dimensions increase(Gheyas and Smith 2010). Most of the input dataset's features are redundant or irrelevant, which prolongs training time and lowers the trained classifier's classification accuracy(Xue et al. 2019).

To address these issues, an O3 pollution concentration prediction model based on two-stage decomposition, two-scale feature selection, and machine learning modules is proposed. Firstly, the daily O3 concentration in Beijing and Shanghai is taken as the prediction object; the original data are processed with outliers and missing values to eliminate their adverse effects. Secondly, the processed O3 data is decomposed into several subsequences by ICEEMDAN, and the first subsequence with higher complexity is introduced into VMD for further decomposition. Meanwhile, the fuzzy entropy of the remaining subsequences is calculated, and the subsequences with similar and low fuzzy entropy are reconstructed. Thirdly, PACF is

introduced to calculate possible input dimensions for all subsequences, and RF is used to eliminate irrelevant features to achieve dimensionality reduction. Finally, the sequence with the best input dimension is applied to TCN for prediction. The main contributions of this paper are as follows: (1) A new mixed model ICEEMDAN-VMD-FE-PACF-RF-TCN was proposed to predict O₃ concentration. Strengthen existing prediction models in data processing, prediction mechanism and feature selection to fill some research gaps. (2) Innovative application of ICEEMDAN to extract hidden linear and nonlinear features in O₃ sequences. Moreover, the high complexity of IMF1 is proved again by fuzzy entropy, and VMD is used for secondary processing of IMF1. Then, the applicability of the two-stage decomposition method in this field is discussed. (3) A two-scale feature selection is proposed to reduce the dimensionality of decomposed data, which solves the common problems of dimensionality and model overfitting in machine learning. The experimental results demonstrate the contribution of two-scale feature selection in air pollutant prediction. (4) TCN expansion convolution, which provides a flexible field of vision, may be tailored to specific tasks and characteristics. We offer an expansive view to ensure that long-term historical data is effectively covered, something that other models cannot. (5) Collect actual ozone data; set up four comparative experiments, and verify the superiority and stability of the hybrid optimization model in Beijing and Shanghai. Targeted analysis, feature capture and prediction of O₃ data can provide a more intuitive and helpful reference for ozone control, prevention and early warning of photochemical pollution caused by ozone while reducing the harm caused by people's exposure to O₃.

2. Methods

2.1 Improved complete adaptive noise ensemble empirical mode decomposition

The decomposition algorithm was proposed by (Colominas et al. 2014). It is improved on the basis of adaptive noise complete set empirical mode decomposition (CEEMDAN). The main purpose of ICEEMDAN is to solve the problem of residual noise and pseudo-modes in CEEMDAN. The specific steps are as follows:

1. X is defined as the signal to be decomposed, $F_K(\cdot)$ represents the K -order modal component generated by EMD decomposition, $L(\cdot)$ represents the local mean of the generated signal, and $W^{(l)}$ represents Gaussian white noise.

-
2. Add I groups of white noise $W^{(i)}$ to the original sequence to construct a sequence $x^{(i)} = x + \beta_0 F(W^{(i)})$. Get the first set of residuals $R_1 = (L(x^{(i)}))$.
 3. Calculating the first modal component $s_1 = x - R_1$.
 4. Continue to add white noise, and use local mean decomposition to calculate the second group of residual $R_1 + \beta_1 F(W^{(i)})$. Define the second modal component S2. $s_2 = R_1 - R_2 = R_1 - (L(R_1 + \beta_1 F(W^{(i)})))$
 5. Calculate the k-th residual $R_K = (L(R_{K-1} + \beta_{K-1} F(W^{(i)})))$ and modal component $s_K = R_{K-1} - R_K$.
 6. Until the end of calculation and decomposition, all modes and residual numbers are obtained.

2.2 Variable modal decomposition

Variational mode decomposition is a signal decomposition estimating technique used in signal processing. When obtaining the decomposed components, the method iteratively searches for the variational model's optimal solution in order to determine the frequency center and bandwidth of each component. This allows for the adaptive frequency domain division of the signal and the effective separation of each component.

Assuming that the signal is composed of K modal components $v_k(t)$ with limited bandwidth, the center frequency of each component IMF is ω_k , and the constraint condition is that the input signal and the modal components are equal. The specific steps are as follows:

1. Obtain $v_k(t)$ through Hilbert transform, calculate its unilateral general. Modulate the center band of $v_k(t)$ to the corresponding baseband by multiplying with $e^{-j\omega_k t}$:

$$\left[\left(\delta(t) + \frac{j}{\pi t} \right) * v_k(t) \right] e^{-j\omega_k t}$$

2. Calculate the squared norm L^2 of the demodulation gradient and estimate the band tube of each IMF as follows:

$$\begin{cases} \min_{\{v_k\}\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * v_k(t) \right] e^{-j\omega_k t} \right\|^2 \right\} \\ s.t. \sum_k v_k = s \end{cases}$$

Among them, $v_k = \{v_1, \dots, v_k\}$ 为 is the decomposed IMF component, and $\omega_k =$

$\{\omega_1, \dots, \omega_k\}$ represents the center frequency of each part. The Lagrange multiplier $\tau(t)$ and the second-order penalty factor α are introduced to transform the constrained variational problem into an unconstrained variational problem. The extended Lagrangian expression is as follows:

$$L(v_k, \omega_k, \tau) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * v_k(t) \right] e^{-j\omega_k t} \right\|^2 + \left\| s(t) - \sum_k v_k(t) \right\|^2 + \left\langle \tau(t), s(t) - \sum_k v_k(t) \right\rangle$$

3. The alternating direction multiplier method ADMM is used to continuously update each component and its center frequency, and finally the optimal solution to the problem is obtained. All components can be obtained according to the following formula according to the frequency domain space:

$$\hat{v}_k^{n+1}(\omega) = \frac{\hat{s}(\omega) - \sum_{i \neq k} \hat{v}_i(\omega) + \hat{\tau}(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2}$$

ω represents the frequency, $v_k^{n+1}(\omega)$, $\hat{s}(\omega)$, $\hat{\tau}(\omega)$ are the Fourier transforms corresponding to $v_k^n(t)$, $s(t)$, $\tau(t)$, respectively transform.

4. $v_k^{n+1}(\omega)$ is the residual amount of $\hat{s}(\omega) - \sum_{i \neq k} \hat{v}_i(\omega)$ after Wiener filtering, the algorithm re-estimates the center of gravity according to the power spectrum center of each component frequency, as follows:

$$\begin{cases} \omega_k^{n+1} = \frac{\int_{-\infty}^{\infty} \omega |v_k^{n+1}(\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |v_k^{n+1}(\omega)|^2 d\omega} \\ \hat{\tau}^{n+1}(\omega) = \hat{\tau}^n + \tau(\hat{s}(\omega) - \sum_k \hat{v}_k^{n+1}(\omega)) \\ \sum_k \|\hat{v}_k^{n+1} - \hat{v}_k^n\|_2^2 / \|\hat{v}_k^n\|_2^2 < \varepsilon \end{cases}$$

2.3 Fuzzy entropy (FE)

Entropy is used to measure the uncertainty of a system or a piece of information. Fuzzy entropy is used to measure the complexity of time series (Farhadinia 2013). The greater the fuzzy entropy, the greater the complexity of the series. The details are as follows:

1. For a given n-dimensional time series $[t(1), t(2), \dots, t(N)]$
2. Define phase space dimension S ($S \leq N - 2$) and similarity tolerance y , and reconstruct the space: $E(i) = [t(i), t(i+1), \dots, t(i+s-1)] - t_0(i), i =$

1, 2, \dots, N - s + 1

3. Introducing fuzzy membership function

$$M(d) = \begin{cases} 1, & d = 0 \\ \exp\left[-\ln(2)\left(\frac{d}{y}\right)^2\right], & d > 0 \end{cases}$$

Where y is the similar tolerance, and for $i = 1, 2, \dots, N - s + 1$, calculate

$$M_{ij}^k = \exp\left[-\ln(2) \cdot (p_{ij}^k/y)^2\right], j = i = 1, 2, \dots, N - s + 1, \text{ and } j \neq i$$

Among them,

$$p_{ij}^k = p[E(i), E(j)] = \max_{q=1,2,\dots,s} (|t(i+q-1) - x_0(i)| - |t(j+q-1) - t_0(j)|)$$

is the maximum absolute distance between window vectors $E(i)$ and $E(j)$.

4. For each i , get the average value: $C_i^y(s) = \frac{1}{N-s+1} \sum_{j=1, j \neq i}^{N-s+1} M_{ij}^s$

5. Definition $\phi^s(y) = \frac{1}{N-s+1} \sum_{i=1}^{N-s+1} C_i^s(y)$

6. Therefore, the fuzzy entropy of the original time series is

$$FE(s, y) = \lim_{N \rightarrow \infty} [\ln \phi^s(y) - \ln \phi^{s+1}(y)]$$

For limited data sets, the fuzzy entropy is estimated as

$$FE(s, y, N) = \ln \phi^s(y) - \ln \phi^{s+1}(y)$$

2.4 Feature selection

2.4.1 Partial autocorrelation coefficient

The partial correlation of time series with lag values is frequently calculated using the PACF method, which is a useful technique for describing the structural properties of stochastic processes. The k -order autoregressive model can be given as follows:

$$\delta_t = \phi k_1 \delta_{t-1} + \phi k_2 \delta_{t-2} + \dots + \phi k_k \delta_{t-k} + \lambda_t$$

2.4.2 Random forest

Random forest (RF) is a classical machine learning algorithm (Speiser et al. 2019). For each decision tree, select the corresponding out-of-bag (OOB) data to calculate the out-of-bag data error, denoted as errOOB1 . The so-called out-of-bag data means that each time a decision tree is built, data is obtained by repeated sampling. It is used to train the decision tree. At this time, about 1/3 of the data is not used and does not participate in establishing the decision tree. This part of the data can be used to evaluate the decision tree's

performance and calculate the model's prediction error rate, which is called the out-of-bag data error.

Randomly add noise interference to the feature X of all samples of the out-of-bag data OOB (the value of the sample at the feature X can be randomly changed), and calculate the error of the out-of-bag data again, which is recorded as $errOOB2$.

Suppose there are N trees in the forest, then the importance of feature $x = \sum(errOOB2 - errOOB1) / N$. The reason why this value can explain the importance of the feature is that if the random noise is added, the accuracy of the out-of-pocket data will decrease significantly (that is, $errOOB2$ will increase), which indicates that this feature has a significant influence on the prediction result of the sample. Further, indicates that the degree of importance is relatively high.

2.5 Temporal Convolutional Neural Network

A convolution network was proposed by (Bai SJ et al. 2019), which can be used for time series data processing. The temporal convolution network introduces causal convolution, expansion convolution and residual module based on the convolution network. TCN can be used as an alternative to the recurrent neural network RNN, and it can effectively deal with the time series prediction problem in time series modelling.

2.5.1 Based on Time Series Modeling

The input time series is designated as $\{x_0, x_1 \dots x_t\}$, while the model's anticipated output value is designated as $\{y_0, y_1 \dots y_t\}$. The input $\{x_0, x_1 \dots x_t\}$ is historical data, and its core idea is to infer new possible information from historical data. The predicted value of is obtained by function F , as follows:

$$F\{x_0, x_1 \dots x_t\} = \{y_0, y_1 \dots y_t\}$$

Among them, F function represents the expansion convolution proposed in this paper.

To forecast the hourly PM2.5 concentration, (Zhang et al. 2021) suggested a spatio-temporal causal convolution neural network. The results suggest that causal convolution performs well in predicting PM2.5 levels in the short run. In causal convolution, for each element in sequence $\{0, \dots, input_length - 1\}$, the n th element of the output sequence only

depends on the element in the input sequence with index $\{0, 1, \dots, n\}$. The element in the output sequence is only affected by the element in the input sequence before it. Although causality convolution can broaden the scope of historical data, history can only be examined on a linear time scale in the depth of the network. As a result, capturing long-term historical data is challenging. According to the study of (Oord et al. 2016; Yu F and Koltun 2016), we used expansion convolution to allow a broad receiving domain to capture the long-term historical information better. Expansion convolution is defined as follows:

$$(F_d^* X)(x_t) = \sum_k^K f_k x_{t - (K - k)d}$$

The visual field of the convolution layer can be enlarged in long-term prediction by altering the number k of convolution kernels or widening the expansion factor d . As a result, we pick a constant b that will allow us to compute the expansion d of a given layer based on the number of layers i underneath it. $d = b^i$. Generally speaking, for each additional layer, the width of the currently accepted field of vision increases by a $d * (k - 1)$. Therefore, the receptive field width T of TCN based on b exponential expansion, the kernel size k and the number of layers n are given as follows:

$$T = 1 + \sum_{i=0}^{n-1} (k - 1) \cdot b^i = 1 + (k - 1) \cdot \frac{b^n - 1}{b - 1}$$

i is for the number of levels underneath the new layer.

As a result, we require a constant B to broaden our field of vision and better capture long-term historical data. To guarantee that the long-term history can be successfully covered, we set the constant $b=2$, and d rises exponentially proportional to the network depth $d = 2^i$ in this study. Figure 1 depicts the TCN prediction model's particular mechanism.

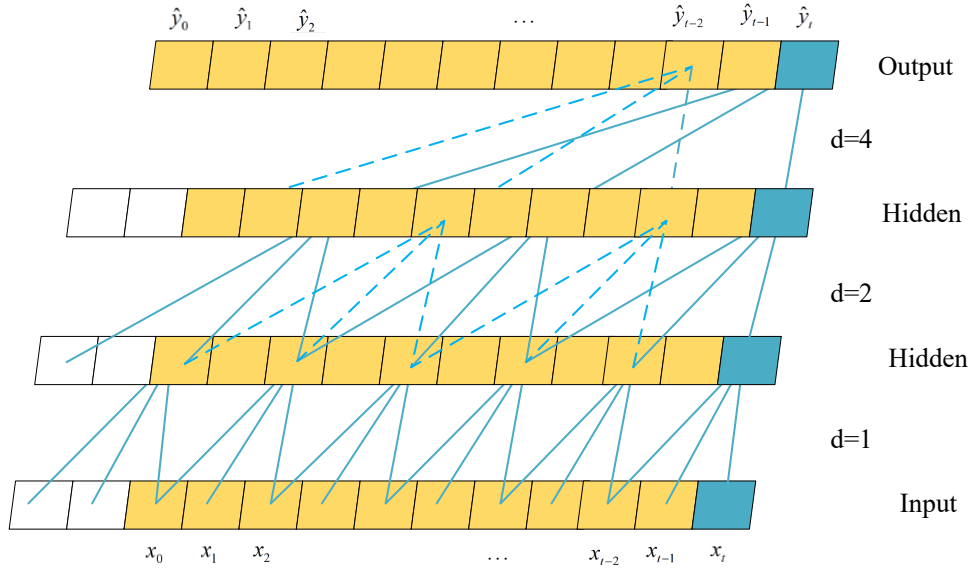


Fig. 1. Expansion convolution network with kernel size = 4 and d (dilations) = 1, 2, and 4.

2.5.2 residual connection

The network depth n , filter size k and cavity expansion factor d determine the receiving domain of TCN. Therefore, it is necessary to ensure that the coefficients of n , k and d allow TCN to obtain a large enough receiving area. Experimentally, it is an effective architecture to make the network deeper and narrower, that is, to superimpose many layers and select a more acceptable filter size. The residual connection has been proved to be a very effective training depth network. In the residual network, the whole network uses a jump connection, which reduces the training time and avoids the gradient vanishing of the depth model. Figure 2 shows the whole process of residual connection.

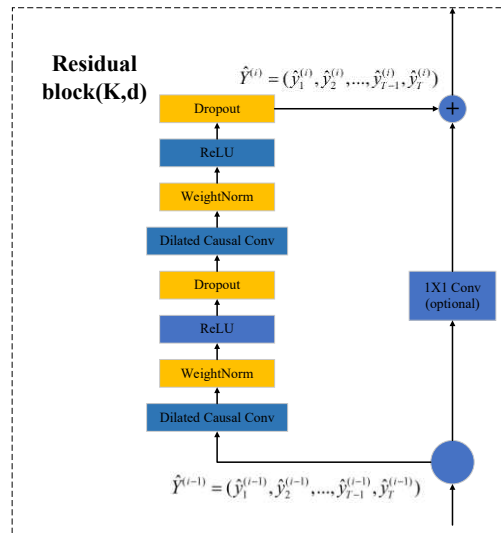


Fig. 2.The process of residual connection.

2.6 Model architecture

The flowchart of the proposed mixed ozone concentration prediction model ICEEMDAN-VMD-FE-PACF-RF-TCN is shown in Fig. 3.The specific processes are as follows:

- (1) First, to eliminate the adverse effects of outliers and missing values in O3 data, we will perform outlier and missing value processing on the collected O3 raw data.
- (2) Second, the O3 concentration time series is decomposed into several subsequences ($IMF1, IMF2, \dots, IMF_n$) by the ICEEMDAN decomposition algorithm, and n is the number of decomposed subsequences. The VMD decomposition algorithm is used to further decompose $IMF1$, and a new sequence is obtained. At the same time, a fuzzy entropy method is introduced to quantify the complexity of the remaining subsequences and reconstruct subsequences with similar complexity.
- (3) Third, perform a two-scale feature selection on the sequence obtained by decomposition. PACF is used to identify the possible input dimensions of the decomposition sequence and then apply random forest screening to further advance valid information from the input variables determined by PACF to determine the optimal input dimension of the prediction model.
- (4) Finally, the obtained subsequence of the best input dimension is substituted into the TCN model for prediction. The final prediction result is then obtained by adding the prediction results of each subsequence.

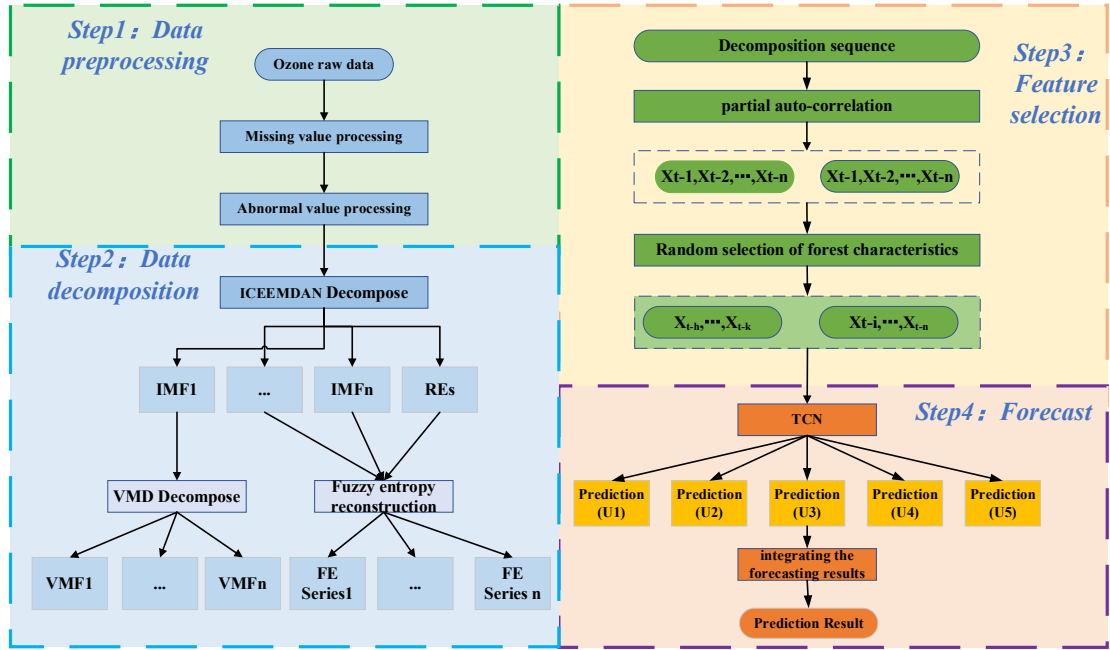


Fig. 3. Flow chart of the forecast system.

3. Data sources and evaluation indicators

3.1 Data sources

In this paper, the O3 data series of two representative cities in Beijing and Shanghai are selected as case studies, and the proposed hybrid optimization model has been verified. The data on O3 are from Tian tan Observation Station in Beijing and Zhang jiang Observation Station in Pudong, Shanghai. Data from: <https://www.aqistudy.cn/historydata/>. The data is daily O3 data from January 1, 2015, to November 21, 2021, with a total of 2495 observations. The original sequence of the data set is shown in Figure 4. The data is described in Table 1.

Table 1

Dataset description in this study.

Cities	Pollutants	Time	Count	Mean	Std.	Min	Max
Beijing	O3	2015/1/1-	2495	95.52	59.50	0	311
		2021/11/21					
Shanghai	O3	2015/1/1-	2495	100.96	43.32	0	295
		2021/11/21					

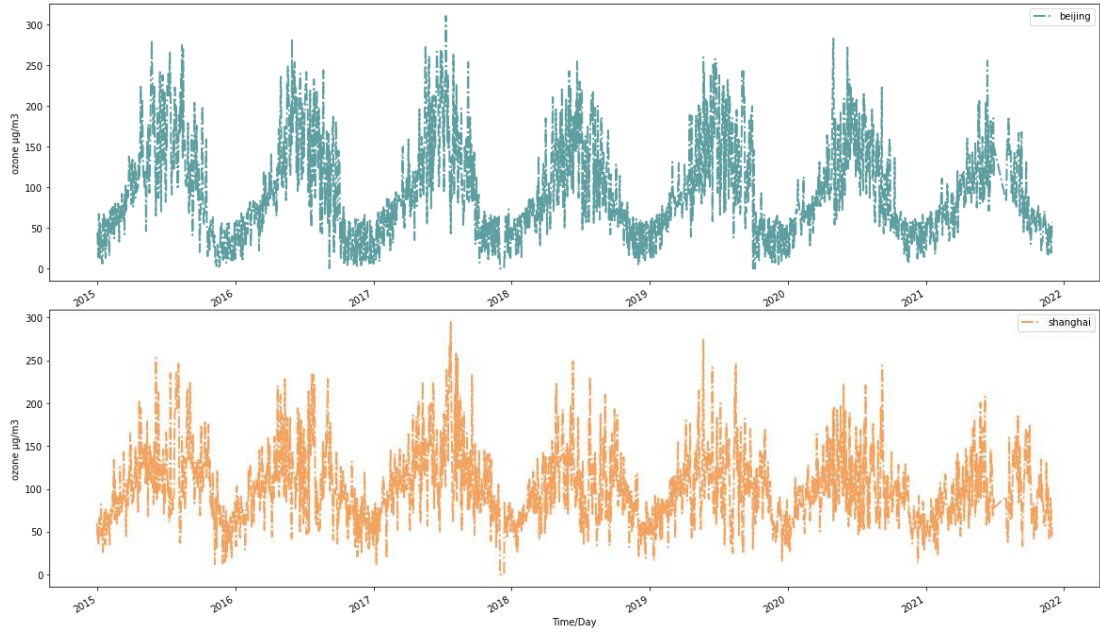


Fig. 4. Original ozone concentration sequences in Beijing and Shanghai.

3.2 Evaluation indicators

To evaluate the performance of the proposed model more intuitively and effectively, three typical indicators MAE, RMSE and SMAPE, are introduced for evaluation. The smaller the above three values, the better the model's prediction accuracy. The specific calculation formula is shown in Table 2.

Table 2

Description of comprehensive evaluation index

Evaluation criteria	Definition	Formula
MAE	Mean Absolute Error	$\frac{1}{n} \sum_{i=1}^n X_i - X'_i $
RMSE	Root Mean Square Error, RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - X'_i)^2}$
SMAPE	Symmetric Mean Absolute Percentage Error	$\frac{100\%}{n} \sum_{i=1}^n \frac{ X_i - X'_i }{(X_i + X'_i) / 2}$

X_i is the actual value of time series I period, X'_i is the predicted value.

4. Case analysis

4.1 Data preprocessing

There will inevitably be missing values and abnormal values in the historical data. As a result, we use KNN to deal with missing values and the Euclidean/Mahalanobis equidistant approach to finding the k samples closest to the samples with missing data and the weighted average of these k values to estimate the missing data of this sample. For the judgment of outliers, we adopt the 3σ criterion. According to the standard distribution definition, the probability beyond the average value of 3σ is $P(|x-\delta|>3\sigma)\leq 0.003$, which is a small probability event. Therefore, we conclude that the sample's distance exceeds the average value of 3σ does not exist. Therefore, when the average distance of the sample data is more significant than 3σ , the sample is considered an abnormal value.

4.2 Sequence decomposition

Due to the volatility and non-stationarity of O3 original sequences, to reduce the volatility and non-stationarity of data. In this paper, ICEEMDAN is introduced to decompose the original sequence, and the decomposition diagram is shown in Figure 5. After ICEEMDAN processing, the original sequence was decomposed into 11 subsequences. Compared with the original sequence, the subsequence has more obvious regular characteristics.

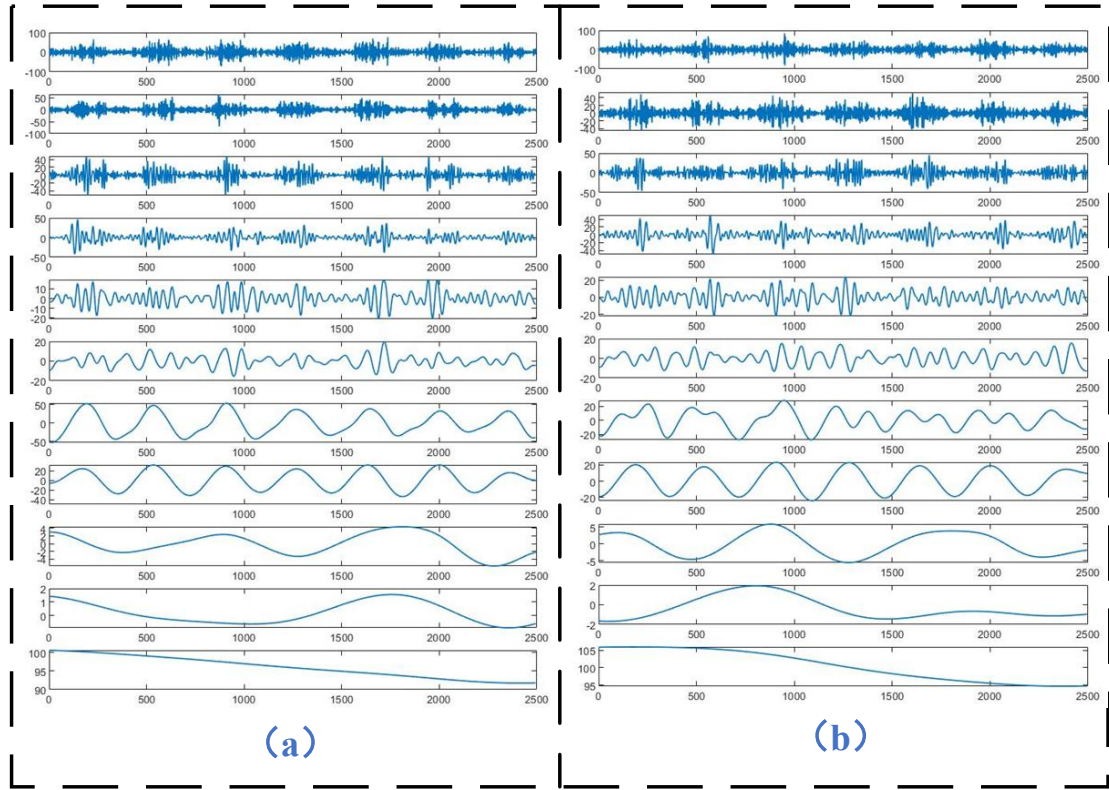


Fig. 5. Decomposition results in Beijing(a) and Shanghai(b) by ICEEMDAN

4.3 Secondary decomposition and reconstruction

After the two datasets' decomposition results are obtained, all subsequences' fuzzy entropy is calculated. As shown in Table 3, the fuzzy entropy of the decomposed IMF1 is the highest among all subsequences. In order to further verify that the result is not caused by the ICEEMDAN algorithm, standard decomposition algorithms EMD, EEMD and CEEMDAN are introduced to decompose the original data, and the results are shown in Table 3. Although the fuzzy entropy of IMF1 obtained by different decomposition algorithms is different, the difference is insignificant. This result further demonstrates the widespread nature of the high complexity of IMF1. Therefore, the quadratic decomposition of IMF1 is a reasonable attempt.

Then, the decomposed IMF1 is introduced into VMD for secondary decomposition. Figure 6 shows the results of VMD quadratic decomposition of IMF1. As can be seen from Figure 6, IMF1 is decomposed into five new subsequences. Compared with the original IMF1, the subsequences obtained by VMD decomposition have more obvious fluctuation laws, which reduces the complexity of IMF1.

Another step in this section is reconstructing the remaining subsequences except for IMF1.

According to the result of fuzzy entropy, subsequences with similar fuzzy entropy values are reconstructed into a new sequence. For the convenience of description, we denote the sequence obtained by VMD decomposition and the reconstructed sequence as FEn, and n is the number of sequences. FE1-FE5 are the new sequences of IMF1 decomposed by VMD, and FE6-FE10 are the new reconstruction sequences. The reconstruction sequence is shown in Table 4. Taking Beijing as an example, IMF2 becomes FE 6, IMF3 and IMF4 form FE 7, FE 8 consists of IMF5, IMF6 and IMF7, IMF8 and IMF9 form FE 9, and FE 10 consists of the original IMF10 and R.

Table 3

EMD, EEMD, CEEMDAN and ICEEMDAN algorithms get the fuzzy entropy of each component.

Dataset	Cites	Method	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
O3	Beijing	EMD	2.862293	2.002407	1.134893	0.659802	0.404474	0.105392
O3	Beijing	EEMD	2.765143	2.064401	1.095472	0.63318	0.381543	0.090587
O3	Beijing	CEEMDAN	2.786467	1.780028	2.194324	1.425125	0.794475	0.513701
O3	Beijing	ICEEMDAN	2.667224	2.165636	1.182462	0.665447	0.433135	0.16367
Dataset	Cites	Method	IMF7	IMF8	IMF9	IMF10	IMF11	IMF12
O3	Beijing	EMD	0.102117	0.00413	0.000831	4.89E-05	4.24E-07	
O3	Beijing	EEMD	0.093756	0.001139	5.54E-05	2.66E-06	3.03E-07	
O3	Beijing	CEEMDAN	0.261901	0.086163	0.068422	0.002042	0.001097	2.81E-05
O3	Beijing	ICEEMDAN	0.086685	0.052956	0.001073	0.000131	4.73E-06	
Dataset	Cites	Method	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
O3	Shanghai	EMD	2.800139	2.186273	1.099874	0.635976	0.335535	0.10312
O3	Shanghai	EEMD	2.831183	2.15336	1.133035	0.64491	0.396266	0.158094
O3	Shanghai	CEEMDAN	2.826112	1.709308	2.124296	1.326291	0.745122	0.505814
O3	Shanghai	ICEEMDAN	2.754652	2.20495	1.214403	0.683404	0.463341	0.200639
Dataset	Cites	Method	IMF7	IMF8	IMF9	IMF10	IMF11	IMF12
O3	Shanghai	EMD	0.033494	0.014209	0.000325	2.69E-07		

O3	Shanghai	EEMD	0.063863	0.014921	0.000569	1.74E-05	1.60E-06	9.06E-09
O3	Shanghai	CEEMDAN	0.236295	0.081954	0.045857	0.001776	0.000581	2.77E-05
O3	Shanghai	ICEEMDAN	0.069956	0.038489	0.001668	0.000165	2.60E-05	

389 **Table 4**

390 Fuzzy entropy reconstruction sequence

Cites	FE6	FE7	FE8	FE9	FE10
Beijing	IMF2	IMF3、 IMF4	IMF5、 IMF6、 IMF7	IMF8、 IMF9	IMF10、 IMF11
Shanghai	IMF2	IMF3、 IMF4	IMF5、 IMF6	IMF7、 IMF8	IMF9、 IMF10、 IMF11

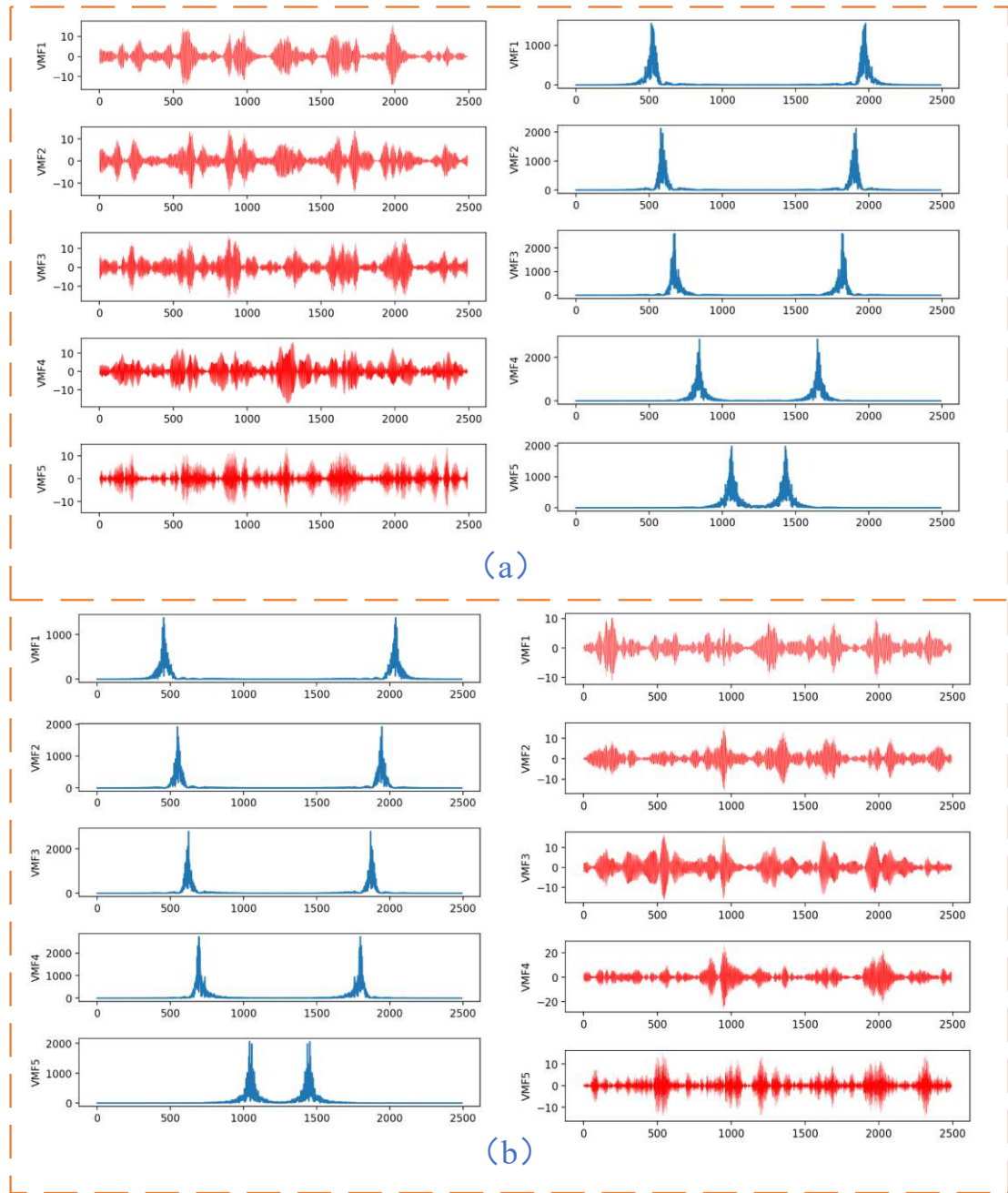


Fig. 6. IMF1 VMD decomposition result. (a is Beijing, b is Shanghai.)

4.4 Feature selection

To obtain the best input dimension, PACF and RF two-scale feature selection are introduced to select features as the input of the prediction model. Using the PACF algorithm, which calculates the potential input dimensions for each series with a 95% confidence interval from lag 0 to 20. The PACF results are shown in Table 5. Then, RF feature selection is introduced to all possible input features of the subsequence to determine the optimal input dimension to reduce the input of irrelevant features. Table 6 shows the RF ranking results of all

Shanghai	FE 1	Xt-1	Xt-2	Xt-3	Xt-4	Xt-6	Xt-7	Xt-8	Xt-11	Xt-15						
Shanghai	FE 2	Xt-1	Xt-2	Xt-3	Xt-4	Xt-5	Xt-6	Xt-7	Xt-8	Xt-12	Xt-13	Xt-15				
Shanghai	FE 3	Xt-1	Xt-2	Xt-3	Xt-4	Xt-5	Xt-6	Xt-7	Xt-8	Xt-9	Xt-12	Xt-13	Xt-14			
Shanghai	FE 4	Xt-1	Xt-2	Xt-3	Xt-4	Xt-5	Xt-6	Xt-7	Xt-8	Xt-9	Xt-10	Xt-11	Xt-12	Xt-14	Xt-15	
Shanghai	FE 5	Xt-1	Xt-2	Xt-3	Xt-5	Xt-6	Xt-7	Xt-9	Xt-10	Xt-13	Xt-14	Xt-15				
Shanghai	FE 6	Xt-1	Xt-2	Xt-3	Xt-4	Xt-5	Xt-6	Xt-8	Xt-10							
Shanghai	FE 7	Xt-1	Xt-2	Xt-3	Xt-4	Xt-5	Xt-6	Xt-8	Xt-9	Xt-12	Xt-13	Xt-14				
Shanghai	FE 8	Xt-1	Xt-2	Xt-3	Xt-4	Xt-5	Xt-6	Xt-7	Xt-12	Xt-13	Xt-14	Xt-15				
Shanghai	FE 9	Xt-1	Xt-2	Xt-3	Xt-4	Xt-5	Xt-6	Xt-7	Xt-8							
Shanghai	FE 10	Xt-1	Xt-2	Xt-3	Xt-4	Xt-5	Xt-6	Xt-7								

406 **Table 6**

407 Random forest ranking order for all features of each subseries.

Beijing

Rank number	FE 1	FE 2	FE 3	FE 4	FE 5	FE 6	FE 7	FE 8	FE 9	FE 10
1	Xt-2	Xt-2	Xt-3	Xt-5	Xt-1	Xt-3	Xt-1	Xt-1	Xt-1	Xt-1
2	Xt-5	Xt-4	Xt-6	Xt-1	Xt-2	Xt-1	Xt-6	Xt-2	Xt-2	Xt-2
3	Xt-7	Xt-3	Xt-12	Xt-4	Xt-3	Xt-4	Xt-4	Xt-3	Xt-3	Xt-3
4	Xt-3	Xt-1	Xt-1	Xt-3	Xt-5	Xt-2	Xt-5	Xt-5	Xt-9	Xt-15

5	Xt-1	Xt-5	Xt-15	Xt-11	Xt-6	Xt-6	Xt-3	Xt-4	Xt-8	Xt-11
6	Xt-4	Xt-7	Xt-2	Xt-14	Xt-7	Xt-5	Xt-2	Xt-6	Xt-7	Xt-12
7	Xt-12	Xt-9	Xt-4	Xt-6	Xt-8	Xt-9	Xt-10	Xt-7	Xt-5	Xt-8
8	Xt-6	Xt-6	Xt-14	Xt-2	Xt-9	Xt-11	Xt-14	Xt-9	Xt-6	Xt-9
9	Xt-8	Xt-8	Xt-10	Xt-9	Xt-15	Xt-8	Xt-9	Xt-11	Xt-4	Xt-14
10	Xt-9	Xt-10	Xt-11	Xt-13	Xt-12	Xt-10	Xt-8	Xt-15		Xt-7
11	Xt-13	Xt-15	Xt-5	Xt-12		Xt-14		Xt-12		Xt-4
12		Xt-13	Xt-7	Xt-7				Xt-14		Xt-5
13		Xt-11	Xt-8	Xt-8				Xt-13		Xt-6
14		Xt-14								
Shanghai										
Rank number	FE 1	FE 2	FE 3	FE 4	FE 5	FE 6	FE 7	FE 8	FE 9	FE 10
1	Xt-2	Xt-2	Xt-3	Xt-5	Xt-1	Xt-3	Xt-1	Xt-1	Xt-1	Xt-1
2	Xt-4	Xt-7	Xt-6	Xt-4	Xt-2	Xt-1	Xt-6	Xt-2	Xt-2	Xt-2
3	Xt-3	Xt-4	Xt-9	Xt-1	Xt-3	Xt-2	Xt-5	Xt-3	Xt-3	Xt-3
4	Xt-7	Xt-5	Xt-2	Xt-3	Xt-5	Xt-4	Xt-4	Xt-7	Xt-8	Xt-4
5	Xt-1	Xt-3	Xt-12	Xt-10	Xt-6	Xt-6	Xt-3	Xt-6	Xt-7	Xt-7
6	Xt-6	Xt-1	Xt-5	Xt-14	Xt-7	Xt-5	Xt-2	Xt-12	Xt-6	Xt-6
7	Xt-11	Xt-6	Xt-8	Xt-9	Xt-9	Xt-8	Xt-12	Xt-5	Xt-5	Xt-5
8	Xt-8	Xt-8	Xt-1	Xt-2	Xt-10	Xt-10	Xt-8	Xt-13	Xt-4	
9	Xt-15	Xt-13	Xt-14	Xt-15	Xt-15		Xt-13	Xt-4		
10		Xt-15	Xt-7	Xt-8	Xt-13		Xt-14	Xt-15		
11		Xt-12	Xt-13	Xt-12	Xt-14		Xt-9	Xt-14		
12			Xt-4	Xt-6						
13				Xt-7						
14				Xt-11						

The best input dimension for each subsequence

Cites		Best feature					
Beijing	FE 1	Xt-2	Xt-5	Xt-7	Xt-3	Xt-1	Xt-4
Beijing	FE 2	Xt-2	Xt-4	Xt-3	Xt-1	Xt-5	Xt-7
Beijing	FE 3	Xt-3	Xt-6	Xt-12	Xt-1	Xt-15	Xt-2
Beijing	FE 4	Xt-5	Xt-1	Xt-4	Xt-3	Xt-11	Xt-14
Beijing	FE 5	Xt-1	Xt-2	Xt-3	Xt-5	Xt-6	Xt-7
Beijing	FE 6	Xt-3	Xt-1	Xt-4	Xt-2	Xt-6	Xt-5
Beijing	FE 7	Xt-1	Xt-6	Xt-4	Xt-5	Xt-3	Xt-2
Beijing	FE 8	Xt-1	Xt-2	Xt-3	Xt-5	Xt-4	Xt-6
Beijing	FE 9	Xt-1	Xt-2	Xt-3	Xt-9	Xt-8	Xt-7
Beijing	FE 10	Xt-1	Xt-2	Xt-3	Xt-15	Xt-11	Xt-12
Cites		Best feature					
Shanghai	FE 1	Xt-2	Xt-4	Xt-3	Xt-7	Xt-1	Xt-6
Shanghai	FE 2	Xt-2	Xt-7	Xt-4	Xt-5	Xt-3	Xt-1
Shanghai	FE 3	Xt-3	Xt-6	Xt-9	Xt-2	Xt-12	Xt-5
Shanghai	FE 4	Xt-5	Xt-4	Xt-1	Xt-3	Xt-10	Xt-14
Shanghai	FE 5	Xt-1	Xt-2	Xt-3	Xt-5	Xt-6	Xt-7
Shanghai	FE 6	Xt-3	Xt-1	Xt-2	Xt-4	Xt-6	Xt-5
Shanghai	FE 7	Xt-1	Xt-6	Xt-5	Xt-4	Xt-3	Xt-2
Shanghai	FE 8	Xt-1	Xt-2	Xt-3	Xt-7	Xt-6	Xt-12
Shanghai	FE 9	Xt-1	Xt-2	Xt-3	Xt-8	Xt-7	Xt-6
Shanghai	FE 10	Xt-1	Xt-2	Xt-3	Xt-4	Xt-7	Xt-6

4.5. Experimental result analysis and discussion

For thorough analysis and discussion, ELM, LSSVM, and LSTM models are introduced to compare with TCN models proposed in this paper, and four groups of experiments are carried out. 80% of the data are selected as training sets and 20% as test sets to emphasize the ability of the proposed models to explain O3 sequences better. In the first experiment, the original data

is not decomposed but directly brought into the model for direct prediction. This is done to demonstrate the efficacy of various single models and choose the best single prediction model out of all utilized ones. In the second experiment, the ICEEMDAN algorithm was introduced to decompose the original sequence, then TCN and other contrast models were established, and predictions were made. In the third experiment, based on the second experiment, we added the VMD decomposition algorithm to decompose IMF1 twice to verify whether IMF1 with high decomposition complexity can reduce its complexity and improve prediction accuracy. In the fourth experiment, the two-scale feature selection of PACF and RF was added, and RF further screened the possible input dimensions obtained by PACF to obtain the optimal input dimension of the prediction model to achieve the purpose of dimensionality reduction. The Beijing O3 concentration prediction experiment results are shown in Table 8, and the prediction results are also shown in Figure 7. The Shanghai prediction results are shown in Figure 8, and Table 9 shows the experimental results. The following analysis results and discussion are the charts and supplementary materials from the above two scenarios.

- (1) To verify the excellent performance of the TCN chosen in this paper, several classical single predictors, such as ELM, LSSVM, and LSTM, are used as comparison models. Without any decomposition of the data, the MAE, RMSE, and SMAPE of TCN are 16.0464, 20.6254, and 16.59%, respectively, and the MAE, RMSE, and SMAPE of LSTM are 18.66, 23.4022, and 20.37%, respectively. The prediction results of TCN are significantly better than LSTM and other prediction models. The experimental results also demonstrate that the expansion convolution in TCN can capture the historical information of the data more effectively and flexibly while producing more accurate predictions. Due to the high complexity of O3 concentration data, the prediction accuracy of a single model is often challenging to meet expectations when predicting nonlinear data. This conclusion is consistent with (Cabaneros et al. 2020). Therefore, the compound model may be a better choice.
- (2) Experiment 2 demonstrates the contribution of the ICEEMDAN decomposition algorithm in O3 concentration prediction. ICEEMDAN was introduced to break down and forecast the initial sequence based on experiment 1. The data becomes

smooth and regular after adding the ICEEMDAN decomposition technique, and the prediction accuracy is significantly increased. Taking the Beijing scene as an example, compared with TCN, the MAE, RMSE, and SMAPE of ICEEMDAN-TCN are reduced by 12.8132, 16.3351, and 14.52%, respectively. The MAE, RMSE, and SMAPE of the other three contrasting models are reduced by 15.5186, 21.1828, and 15.09% on average. The second experimental finding demonstrates the ICEEMDAN decomposition algorithm's advantage in decomposing vital aspects of data. Consequently, all models' prediction performance was enhanced when the ICEEMDAN algorithm was used in various prediction studies (Ali and Prasad 2019). Contrarily, the O₃ series' lack of preprocessing will make it challenging for the model to capture the information included in the data itself; as a result, the prediction accuracy is subpar.

- (3) The VMD algorithm is applied to the third experiment to decompose IMF1 further and, at the same time, reconstruct the remaining sequence generated by the ICEEMDAN decomposition. The subsequences obtained by quadratic decomposition of IMF1 are less nonlinear and more accessible to predict than the subsequences obtained by primary decomposition. Taking the Beijing scene as an example: ICEEMDAN-VMD-FE-LSTM compared with ICEEMDAN-LSTM, the MAE, RMSE, and SMAPE of the former are 2.5405, 3.3716, and 4.34%, respectively, which are 2.011, 3.044 and 3.36% lower than the latter. Similarly, when comparing ICEEMDAN-VMD-FE-TCN with ICEEMDAN-TCN, the MAE, RMSE, and SMAPE of the former are 1.9942, 2.5361, and 3.34% lower than those of the latter, respectively. The evaluation indicators of other comparative models have been improved, respectively, and the same results can also be obtained from the Shanghai scene. The third experiment's findings demonstrate that processing IMF1 once more can successfully decrease its complexity, increasing prediction accuracy (Sun W and Huang 2020). At the same time, the reconstruction of subsequences other than IMF1 will not increase the complexity of subsequences, which has been confirmed by (Sun SL et al. 2021) and has been proved again on two datasets..

-
- (4) In the fourth experiment, PACF and RF are introduced for two-scale feature selection on the foundation of experiment three to confirm the superiority of the two-scale feature selection suggested in this study. Taking the Beijing scene as an example: Comparing ICEEMDAN-VMD-FE-PACF-RF-TCN with ICEEMDAN-VMD-FE-TCN, the MAE, RMSE, and SMAPE of the former are 0.5350, 0.7334 and 0.8433%, respectively, which are higher than those of the latter respectively. Reduced by 1.5283, 2.0335, and 2.4284%. Similarly, the MAE, RMSE, and SMAPE of Shanghai ICEEMDAN-VMD-FE-PACF-RF-TCN and ICEEMDAN-VMD-FE-TCN are 2.222, 2.990, and 2.3189% lower than those of the latter, respectively. The other comparison models' MAE, RMSE, and SMAPE decreased. A model that has too many input dimensions is more likely to overfit and be subject to duplication and correlation between variables, which makes prediction challenging(Hao et al. 2020). The experimental findings demonstrate once more that the two-scale feature selection decreases the input of unimportant characteristics, prevents overfitting of the model, enhances prediction accuracy, and, to a certain extent, shortens the model's running time.
- (5) The proposed hybrid ICEEMDAN-VMD-FE-PACF-RF-TCN still outperforms all competing models when predicting O3 sequences with significant fluctuations, with MAE of 0.5351, RMSE of 0.7346 and SMAPE of 0.8433%. The experimental findings demonstrate that the hybrid model exhibits excellent generalizability even in challenging O3 sequences.

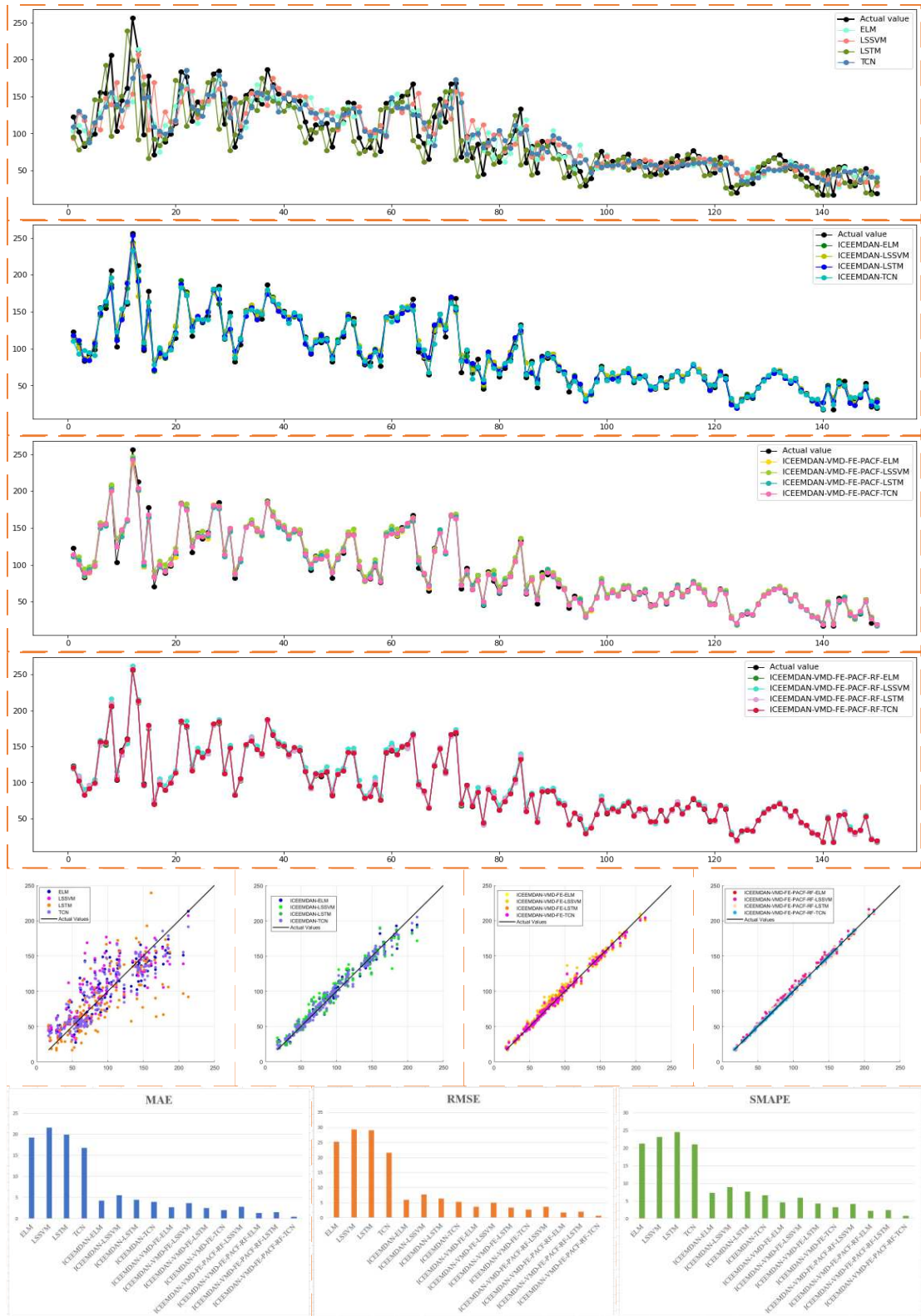


Fig. 7. Forecasting results of different models for Beijing ozone sequence.

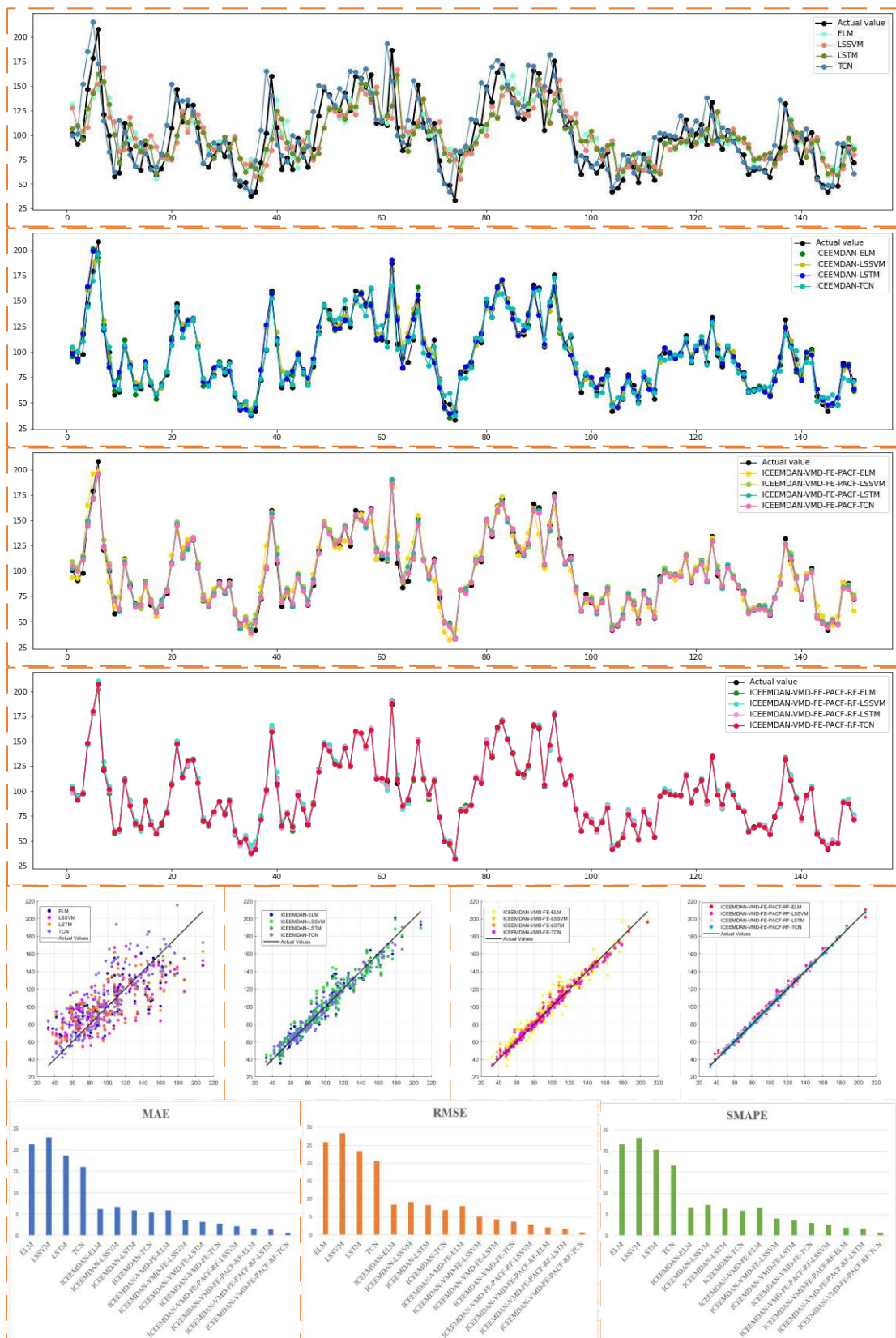


Fig. 8. Forecasting results of different models for Shanghai ozone sequence.

Table 7

Evaluation index of O3 prediction value of various models in Beijing.

Models	MAE	RMSE	SMAPE (%)
ELM	19.27385269	25.2759429	21.38779754%
LSSVM	21.69898756	29.34698985	23.20655983%
LSTM	20.01001333	29.14884597	24.66371402%
TCN	16.87089365	21.63823144	21.14072064%
ICEEMDAN-ELM	4.287483552	5.996457600	7.288447443%
ICEEMDAN-LSSVM	5.587484843	7.8110207626	9.001965015%
ICEEMDAN-LSTM	4.551944908	6.415844726	7.705829264%
ICEEMDAN-TCN	4.057714219	5.30313003	6.613925607%
ICEEMDAN-VMD-FE-ELM	2.709856582	3.669460547	4.615295266%
ICEEMDAN-VMD-FE-LSSVM	3.748559731	5.075067759	5.89566487%
ICEEMDAN-VMD-FE-LSTM	2.540533891	3.371673681	4.343873635%
ICEEMDAN-VMD-FE-TCN	2.063414735	2.766973756	3.271753153%
ICEEMDAN-VMD-FE-PACF-RF- LSSVM	2.848833385	3.700191002	4.21229321%
ICEEMDAN-VMD-FE-PACF-RF- ELM	1.363206332	1.813749442	2.250081933%
ICEEMDAN-VMD-FE-PACF-RF- LSTM	1.530342762	1.988192571	2.481575524%
ICEEMDAN-VMD-FE-PACF-RF- TCN	0.53509618	0.733460954	0.843300731%

Table 8

Evaluation index of O3 prediction value of various models in Shanghai.

Models	MAE	RMSE	SMAPE (%)
ELM	21.30910969	25.93005057	21.56900232%
LSSVM	22.97882691	28.40680017	23.11842781%
LSTM	18.66168934	23.40228129	20.37848879%
TCN	16.04641867	20.62547072	16.5874702%
ICEEMDAN-ELM	6.152631172	8.504478663	6.712899514%
ICEEMDAN-LSSVM	6.749401389	9.25089	7.211323046%
ICEEMDAN-LSTM	5.888175837	8.290664038	6.46883928%
ICEEMDAN-TCN	5.329633847	6.912293411	5.921024258%
ICEEMDAN-VMD-FE-ELM	5.89970184	8.144554222	6.683544991%
ICEEMDAN-VMD-FE-LSSVM	3.636405674	5.089916775	4.088991919%
ICEEMDAN-VMD-FE-LSTM	3.219932833	4.400264376	3.574824594%
ICEEMDAN-VMD-FE-TCN	2.826696727	3.789384835	3.020896143%
ICEEMDAN-VMD-FE-PACF-RF- LSSVM	2.212406577	2.988710644	2.616568544%
ICEEMDAN-VMD-FE-PACF-RF- ELM	1.613867481	2.059598506	1.879047265%
ICEEMDAN-VMD-FE-PACF-RF- LSTM	1.423188505	1.766533882	1.696489526%
ICEEMDAN-VMD-FE-PACF-RF- TCN	0.603800185	0.799242072	0.701991375%

511

512

513 5. Conclusion and future work

514 A new mixed prediction model is proposed in this paper. The model combines
515 ICEEMDAN、VMD、FE、PACF、RF and TCN。Two-stage decomposition, two-scale feature
516 selection, and TCN are first applied in the field of ozone emission. Through four simulation
517 experiments in two independent scenes in Beijing and Shanghai and a comparison with existing

models, the following conclusions can be drawn from the experimental results:

First, in terms of O₃ prediction, the ICEEMDAN-VMD-FE-PACF-RF-TCN prediction model proposed in this paper has excellent prediction performance. In four comparative experiments, this model's MAE, RMSE, and SMAPE are the best among all models, with 0.6038, 0.7992, and 0.70%, respectively, and are validated on two data in Beijing and Shanghai. The experimental results prove that the model can accurately capture the regular features when faced with O₃ sequences with intense fluctuations.

Secondly, it is necessary to decompose the original data when forecasting nonlinear data. The empirical analysis shows a significant error in the direct prediction of O₃ concentration. In this study, ICEEMDAN was used to decompose O₃, and MAE, RMSE, and SMAPE after decomposition decreased by 15.5186, 21.1828, and 15.09% on average. The experimental findings demonstrate that the O₃ data properties may be successfully extracted using the ICEEMDAN decomposition technique.

Third, quadratic decomposition of IMF1 is an effective exploration. In this study, fuzzy entropy is introduced to quantify IMF1, and combined with the prediction results; it is proved that the complexity of IMF1 is higher than that of other subsequences. After decomposing IMF1, all models' MAE, RMSE, and SMAPE are reduced by 2.1342, 2.8835, and 2.24% on average. In addition, adding fuzzy entropy eliminates operational redundancy brought on by decomposition and boosts prediction accuracy.

Fourth, the two-scale feature selection can effectively improve prediction accuracy. The two-scale feature selection can not only accurately and effectively identify the sequence features, but also select the best features to reduce the model's input dimension and avoid the model's overfitting. After two-scale feature selection, the model evaluation indicators MAE, RMSE, and SMAPE are reduced by 2.4323, 3.4525, and 2.62% on average. The first use of the two-scale feature selection method for daily ozone concentration also closes a research gap.

Fifth, TCN demonstrated its superiority in four comparative experiments. Such good prediction results are attributed to the large enough field of view of TCN, which enables TCN to capture data features better. The experimental results further demonstrate that TCN can more deeply interpret the feature information of historical data and effectively utilize these features to obtain better predictions.

The experimental results show that the hybrid optimization model proposed in this paper is an effective and accurate air pollution prediction model. Even though the model described in this research performs well in empirical analysis, it can yet be improved. This study did not consider other meteorological factors such as temperature and wind speed. Future work can take meteorological factors into account. In addition, we can explore the application of this model in other fields.

References

Ahani IK, Salari M, Shadman A (2020) An ensemble multi-step-ahead forecasting system for fine particulate matter in urban areas. *Journal of Cleaner Production* 263. <https://doi.org/10.1016/j.jclepro.2020.120983>

Ali M, Prasad R (2019) Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition. *Renewable & Sustainable Energy Reviews* 104:281-295. <https://doi.org/10.1016/j.rser.2019.01.014>

Bai L, Wang JZ, Ma XJ, Lu HY (2018) Air Pollution Forecasts: An Overview. *International Journal of Environmental Research and Public Health* 15(4). <https://doi.org/10.3390/ijerph15040780>

Bai SJ, Kolter JZ, Koltun V (2019) Deep Equilibrium Models. 33rd Conference on Neural Information Processing Systems (NeurIPS). 32.

Cabaneros SM, Calautit JK, Hughes B (2020) Spatial estimation of outdoor NO₂ levels in Central London using deep neural networks and a wavelet decomposition technique. *Ecological Modelling* 424. <https://doi.org/10.1016/j.ecolmodel.2020.109017>

Cao YF, Qiao X, Hopke PK, Ying Q, Zhang YY, Zeng YY, Yuan YP, Tang Y (2020) Ozone pollution in the west China rain zone and its adjacent regions, Southwestern China: Concentrations, ecological risk, and Sources. *Chemosphere* 256. <https://doi.org/10.1016/j.chemosphere.2020.127008>

Chen SX, Wang JQ, Zhang HY (2019) A hybrid PSO-SVM model based on clustering algorithm for short-term atmospheric pollutant concentration forecasting. *Technological Forecasting and Social Change* 146:41-54. <https://doi.org/10.1016/j.techfore.2019.05.015>

Colominas MA, Schlotthauer G, Torres ME (2014) Improved complete ensemble EMD:

578 A suitable tool for biomedical signal processing. *Biomedical Signal Processing and Control*
579 14:19-29. <https://doi.org/10.1016/j.bspc.2014.06.009>

580 Corani G, Scanagatta M (2016) Air pollution prediction via multi-label classification.
581 *Environmental Modelling & Software* 80:259-264.
582 <https://doi.org/10.1016/j.envsoft.2016.02.030>

583 Farhadinia B (2013) Information measures for hesitant fuzzy sets and interval-valued
584 hesitant fuzzy sets. *Information Sciences* 240:129-144.
585 <https://doi.org/10.1016/j.ins.2013.03.034>

586 Finardi S, De Maria R, D'Allura A, Cascone C, Calori G, Lollobrigida F (2008) A
587 deterministic air quality forecasting system for Torino urban area, Italy. *Environmental*
588 *Modelling & Software* 23(3):344-355. <https://doi.org/10.1016/j.envsoft.2007.04.001>

589 Garaga R, Sahu SK, Kota SH (2018) A Review of Air Quality Modeling Studies in India:
590 Local and Regional Scale. *Current Pollution Reports* 4(2):59-73.
591 <https://doi.org/10.1007/s40726-018-0081-0>

592 Gardner MW, Dorling SR (1999) Neural network modelling and prediction of hourly NO_x
593 and NO₂ concentrations in urban air in London. *Atmospheric Environment* 33(5):709-719.
594 [https://doi.org/https://doi.org/10.1016/S1352-2310\(98\)00230-1](https://doi.org/https://doi.org/10.1016/S1352-2310(98)00230-1)

595 Gheyas IA, Smith LS (2010) Feature subset selection in large dimensionality domains.
596 *Pattern Recognition* 43(1):5-13. <https://doi.org/10.1016/j.patcog.2009.06.009>

597 Goyal P, Chan AT, Jaiswal N (2006) Statistical models for the prediction of respirable
598 suspended particulate matter in urban cities. *Atmospheric Environment* 40(11):2068-2077.
599 <https://doi.org/10.1016/j.atmosenv.2005.11.041>

600 Hao Y, Tian CS, Wu CY (2020) Modelling of carbon price in two real carbon trading
601 markets. *Journal of Cleaner Production* 244. <https://doi.org/10.1016/j.jclepro.2019.118556>

602 Hochreiter S, Heusel M, Obermayer K (2007) Fast model-based protein homology
603 detection without alignment. *Bioinformatics* 23(14):1728-1736.
604 <https://doi.org/10.1093/bioinformatics/btm247>

605 Hu JL, Chen JJ, Ying Q, Zhang HL (2016) One-year simulation of ozone and particulate
606 matter in China using WRF/CMAQ modeling system. *Atmospheric Chemistry and Physics*
607 16(16):10333-10350. <https://doi.org/10.5194/acp-16-10333-2016>

608 Huang CJ, Kuo PH (2018) A Deep CNN-LSTM Model for Particulate Matter (PM_{2.5})
609 Forecasting in Smart Cities. *Sensors* 18(7). <https://doi.org/10.3390/s18072220>

610 Kampa M, Castanas E (2008) Human health effects of air pollution. *Environmental*
611 *Pollution* 151(2):362-367. <https://doi.org/10.1016/j.envpol.2007.06.012>

612 Li HT, Jin F, Sun SL, Li YW (2021) A new secondary decomposition ensemble learning
613 approach for carbon price forecasting. *Knowledge-Based Systems* 214.
614 <https://doi.org/10.1016/j.knosys.2020.106686>

615 Luna AS, Paredes MLL, de Oliveira GCG, Correa SM (2014) Prediction of ozone
616 concentration in tropospheric levels using artificial neural networks and support vector machine
617 at Rio de Janeiro, Brazil. *Atmospheric Environment* 98:98-104.
618 <https://doi.org/10.1016/j.atmosenv.2014.08.060>

619 Oord A, Dieleman S, Zen H, Simonyan K, Kavukcuoglu K (2016) WaveNet: A Generative
620 Model for Raw Audio.

621 Pires JCM, Alvim-Ferraz MCM, Pereira MC, Martins FG (2011) Prediction of
622 tropospheric ozone concentrations: Application of a methodology based on the Darwin's Theory
623 of Evolution. *EXPERT SYSTEMS WITH APPLICATIONS* 38(3):1903-1908.
624 <https://doi.org/10.1016/j.eswa.2010.07.122>

625 Rybarczyk Y, Zalakeviciute R (2018) Machine Learning Approaches for Outdoor Air
626 Quality Modelling: A Systematic Review. *Applied Sciences-Basel* 8(12).
627 <https://doi.org/10.3390/app8122570>

628 Speiser JL, Miller ME, Tooze J, Ip E (2019) A comparison of random forest variable
629 selection methods for classification prediction modeling. *Expert Systems with Applications*
630 134:93-101. <https://doi.org/10.1016/j.eswa.2019.05.028>

631 Strosnider HM, Chang HH, Darrow LA, Liu Y, Vaidyanathan A, Strickland MJ (2019)
632 Age-Specific Associations of Ozone and Fine Particulate Matter with Respiratory Emergency
633 Department Visits in the United States. *American Journal of Respiratory and Critical Care*
634 *Medicine* 199(7):882-890. <https://doi.org/10.1164/rccm.201806-1147OC>

635 Sun SL, Jin F, Li HT, Li YW (2021) A new hybrid optimization ensemble learning
636 approach for carbon price forecasting. *Applied Mathematical Modelling* 97:182-205.
637 <https://doi.org/10.1016/j.apm.2021.03.020>

638 Sun W, Huang CC (2020) A hybrid air pollutant concentration prediction model combining
639 secondary decomposition and sequence reconstruction. *Environmental Pollution* 266.
640 <https://doi.org/10.1016/j.envpol.2020.115216>

641 Sun W, Xu C (2021) Carbon price prediction based on modified wavelet least square
642 support vector machine. *Science of the Total Environment* 754.
643 <https://doi.org/10.1016/j.scitotenv.2020.142052>

644 Taheri Shahraini H, Sodoudi S (2016) Statistical Modeling Approaches for PM10
645 Prediction in Urban Areas; A Review of 21st-Century Studies. *Atmosphere* 7(2).
646 <https://doi.org/10.3390/atmos7020015>

647 Wang DD, Jiang BL, Li FZ, Lin WS (2018) Investigation of the Air Pollution Event in
648 Beijing-Tianjin-Hebei Region in December 2016 Using WRF-Chem. *Advances in Meteorology*
649 2018. <https://doi.org/10.1155/2018/1634578>

650 Wang XK, Zhang QQ, Zheng FX, Zheng QW, Yao FF, Chen Z, Zhang WW, Hou PQ, Feng
651 ZZ, Song WZ et al. (2012) Effects of elevated O₃ concentration on winter wheat and rice yields
652 in the Yangtze River Delta, China. *Environmental Pollution* 171:118-125.
653 <https://doi.org/10.1016/j.envpol.2012.07.028>

654 Wiwatanadate P (2014) Acute Air Pollution-Related Symptoms Among Residents in
655 Chiang Mai, Thailand. *Journal of Environmental Health* 76(6):76-84.

656 Xie R, Zhao G, Zhu BZ, Chevallier J (2018) Examining the Factors Affecting Air Pollution
657 Emission Growth in China. *Environmental Modeling & Assessment* 23(4):389-400.
658 <https://doi.org/10.1007/s10666-018-9593-7>

659 Xue Y, Xue B, Zhang MJ (2019) Self-Adaptive Particle Swarm Optimization for Large-
660 Scale Feature Selection in Classification. *Acm Transactions on Knowledge Discovery from*
661 *Data* 13(5). <https://doi.org/10.1145/3340848>

662 Yu F, Koltun V (2016) Multi-Scale Context Aggregation by Dilated Convolutions. *ICLR*.

663 Yu S, Mathur R, Pleim J, Wong D, Gilliam R, Alapaty K, Zhao C, Liu X (2014) Aerosol
664 indirect effect on the grid-scale clouds in the two-way coupled WRF-CMAQ: model description,
665 development, evaluation and regional analysis. *Atmospheric Chemistry and Physics*
666 14(20):11247-11285. <https://doi.org/10.5194/acp-14-11247-2014>

667 Zhang L, Na JM, Zhu J, Shi ZK, Zou CX, Yang L (2021) Spatiotemporal causal

668 convolutional network for forecasting hourly PM2.5 concentrations in Beijing, China.
669 Computers & Geosciences 155. <https://doi.org/10.1016/j.cageo.2021.104869>
670 Zhu SL, Qiu XL, Yin YR, Fang M, Liu XR, Zhao XJ, Shi YJ (2019) Two-step-hybrid
671 model based on data preprocessing and intelligent optimization algorithms (CS and GWO) for
672 NO2 and SO2 forecasting. Atmospheric Pollution Research 10(4):1326-1335.
673 <https://doi.org/10.1016/j.apr.2019.03.004>
674