Data Science and Management 6 (2023) 46–54

Contents lists available at ScienceDirect

Data Science and Management

journal homepage: www.keaipublishing.com/en/journals/data-science-and-management

Research article

Forecasting hourly $PM_{2.5}$ concentrations based on decomposition-ensemble-reconstruction framework incorporating deep learning algorithms

Peilei Cai, Chengyuan Zhang *, Jian Chai

School of Economics and Management, Xidian University, Xi'an, 710126, China

A R T I C L E I N F O

Keywords:
$PM_{2.5}$ concentration prediction
Decomposition-ensemble-reconstruction framework
Variational mode decomposition method
Deep learning

A B S T R A C T

Accurate predictions of hourly $PM_{2.5}$ concentrations are crucial for preventing the harmful effects of air pollution. In this study, a new decomposition-ensemble framework incorporating the variational mode decomposition method (VMD), econometric forecasting method (autoregressive integrated moving average model, ARIMA), and deep learning techniques (convolutional neural networks (CNN) and temporal convolutional network (TCN)) was developed to model the data characteristics of hourly $PM_{2.5}$ concentrations. Taking the $PM_{2.5}$ concentration of Lanzhou, Gansu Province, China as the sample, the empirical results demonstrated that the developed decomposition-ensemble framework is significantly superior to the benchmarks with the econometric model, machine learning models, basic deep learning models, and traditional decomposition-ensemble models, within one-, two-, or three-step-ahead. This study verified the effectiveness of the new prediction framework to capture the data patterns of $PM_{2.5}$ concentration and can be employed as a meaningful $PM_{2.5}$ concentrations prediction tool.

1. Introduction

The $PM_{2.5}$ (particulates with a dynamic diameter less than 2.5 μm) is the most prominent air pollutant, attracting extensive attention from the government and public (Huang et al., 2021; Liu et al., 2020; Wang et al., 2022a). Since $PM_{2.5}$ particles are small, they enter the lungs and blood easily, causing harm to many organs in the human body (Wood, 2022; Wu et al., 2020). Therefore, the public is eager to build an air pollution warning system and formulate targeted healthy travel plans to effectively deal with the harmful effects of $PM_{2.5}$ (Wang et al., 2022b). Thus, through the monitoring and data collection of $PM_{2.5}$ concentration observations, it is particularly necessary to employ advanced prediction technology to capture, identify, model, and predict air pollution data patterns (Liu et al., 2021). Accordingly, in the academic community, whether the prediction model can predict $PM_{2.5}$ concentrations more accurately has been a hot issue which leads to the continuous updating of the prediction framework in this field to obtain more accurate and faster prediction results (Kow et al., 2022).

In fact, $PM_{2.5}$ concentration prediction frameworks are roughly divided into two groups based on the modeling processes: physical models and data-driven models (Bai et al., 2019; Liu et al., 2020; Wang et al., 2022a). In terms of physical models, these models mainly analyze and mine the relationship between the chemical and physical composition of the atmosphere and the concentration of $PM_{2.5}$ (Djalalova et al., 2015). As for data-driven models, popular prediction models using statistical rules are extensively employed to model the data patterns of $PM_{2.5}$ observations (Wang et al., 2022a). Interestingly, the latter classification can be approximately divided into three categories based on model complexity and principles: traditional statistical models, artificial intelligence (AI) models, and hybrid models (Liu et al., 2021; Yang et al., 2022). However, any single AI-based model has limitations when considering robustness (Yin et al., 2021).

The hybrid prediction strategy, consisting of combination models (Wang et al., 2021b) and decomposition-ensemble models (Ausati and Amanollahi, 2016), has been progressively utilized and has become a trend in an effort to overcome the disadvantages of the aforementioned two groups (Zhu and Xie, 2023). Therefore, as a promising approach in the field of decomposition, variational mode decomposition (VMD) is adopted to decompose the nonlinear and non-stationary $PM_{2.5}$ concentration data in the principle of "divide-and-conquer" in this study. More importantly, to improve the prediction efficiency (reduce model complexity and computational cost), it is a meaningful strategy to

Peer review under responsibility of Xi'an Jiaotong University.
* Corresponding author.
E-mail address: chyzhang@xidian.edu.cn (C. Zhang).

https://doi.org/10.1016/j.dsm.2023.02.002
Received 15 October 2022; Received in revised form 26 February 2023; Accepted 27 February 2023
Available online 1 March 2023

reconstruct the decomposed sub-components by using entropy computational complexity, because the decomposition technique may produce more than three subsequences, and some of them have similar volatility and variance (Liu and Chen, 2020; Niu et al., 2017).

In addition, the choice of prediction technology is crucial to the modeling and prediction of different components because the fluctuations of different component series are different (Sun and Huang, 2020; Yu et al., 2015; Yu and Ma, 2021). For example, some components show high-frequency characteristics owing to the influence of noise or irregular factors, whereas some components show trend terms based on the basic trend of the sequence (Yu et al., 2015). Econometric models are more suitable for stationary series, that is, trend terms, whereas AI-based forecasting techniques are more suitable for high-frequency subseries. Therefore, it is very important to select appropriate prediction techniques according to different trends to improve prediction accuracy (Yu and Ma, 2021).

In this context, the normal decomposition-ensemble prediction framework decomposes the raw observations into a series of sub-components, resulting in the prediction calculation of all the sub-components requiring a more time-consuming process and loss efficiency (Yu et al., 2015; Yu and Ma, 2021). Thus, this study proposes a modified and improved decomposition-ensemble prediction framework with a reconstruction process involving a remarkable decomposition technique, econometric model, and deep learning models to predict hourly $PM_{2.5}$ concentrations.

In particular, VMD is employed to decompose the original data, which can adaptively realize the frequency domain division of time-series data and effective separation of each component (Zhang et al., 2021). As for the prediction techniques, autoregressive integrated moving average model (ARIMA), convolutional neural networks (CNN), and temporal convolutional network (TCN) are chosen to model and predict the low-, mid-, and high-sub-components, respectively. Three main steps are included in the proposed framework. First, the VMD is used to decompose the $PM_{2.5}$ data into sub-components (intrinsic mode function components and residual), and then all the sub-components are reconstructed based on permutation entropy. Second, the reconstructed components of $PM_{2.5}$ concentration time series are modeled and predicted, and the final prediction result is calculated via a simple addition approach. Third, the empirical results are computed and demonstrated.

The main contributions of the proposed prediction framework are as follows.

(1) A new decomposition-ensemble-reconstruction prediction framework considered the decomposition method and reconstruction strategy, in which the sub-components are calculated and investigated using entropy.
(2) Different effective prediction tools (ARIMA, CNN, and TCN) are used to capture the different timescale patterns of the reconstructed sub-time series data, which included the low-, mid-, and high-frequency data.
(3) Using Lanzhou, Gansu Province, China as the target data, the proposed prediction framework achieves better performance than other benchmarks, including statistics-based models, AI-based models, and the basic decomposition-based model, in terms of three evaluation criteria.

The remainder of this paper is organized as follows. In Section 2, the proposed prediction framework is described. Section 3 presents the empirical design of the study. Section 4 demonstrates and summarizes the empirical results. Section 5 concludes the study and offers major directions for future research.

2. Literature review

In this section, the related literature on forecasting techniques and combination forecasting strategies is summarized.

2.1. Forecasting techniques

Traditional statistical models include regression models, time-series models, and econometric models, such as linear regression (Yuan and Che, 2022), exponential smoothing (Jiang et al., 2021a), and ARIMA models (Zhang et al., 2018). This method has the advantages of simple principles, convenient calculations, and good effects on stationary series (Yang et al., 2022). However, it is difficult to deal with complex systems with strong randomness and volatility, such as $PM_{2.5}$ concentration (Yang et al., 2022). For AI-based models, the popular support vector machine (SVM) model (Zhou et al., 2019), machine learning models (Dong et al., 2022) and deep learning models (Menares et al., 2021) have been extensively used to forecast $PM_{2.5}$ data (Yin et al., 2021). Correspondingly, AI-based models have high prediction accuracy because of their ability to describe nonlinear and non-stationary data efficiently, which has become an effective and popular nonlinear prediction technology in the field of $PM_{2.5}$ (Dong et al., 2022).

In terms of deep learning models, the deep learning models of the AI-based groups are popularly employed to forecast the complex system, such as $PM_{2.5}$ data (Ma et al., 2020; Sun et al., 2021; Tan et al., 2022). Examples include variant long short-term memory (LSTM) (Li et al., 2017), gated recurrent units (GRU) (Yeo et al., 2021), and CNN (Wang et al., 2021a). Deep learning neural networks are good at capturing and modeling the data patterns of high-frequency and long-term data using a large number of layers and a wide width. In fact, deep learning techniques are highly dependent on data; the greater the amount of data, the better the prediction performance. However, deep-learning techniques such as recurrent neural networks (RNN) and their variants (LSTM and GRU) may suffer from gradient explosion, information loss, and other problems (Deng et al., 2019; Wang et al., 2021a). In contrast, the temporal convolutional neural network (TCN), as a general architecture for convolution sequence prediction and a promising tool, can effectively model all historical observations and relevant variable data based on the causality of convolution in the architecture (Chen et al., 2020).

2.2. Combination forecasting strategy

For combination models (Wang et al., 2021b) and decomposition-ensemble models, on one hand, the former models can average the results of various prediction techniques to effectively reduce the defects of different algorithms. In addition, intelligent optimization algorithms can be used to determine the combined model weights (Wang et al., 2022a). For example, Liu et al. (2019) used the particle swarm optimization (PSO) and AdaBoost algorithms to optimize the back-propagation neural network (BPNN) in the field of air $PM_{2.5}$ concentration multi-step forecasting. Samal et al. (2021) combined the rapid feature extraction capability of a CNN and the sequential temporal modeling feature of a RNN in parallel to improve forecasting accuracy. Wang et al. (2021a) used the convolutional network (ConvNet) and dense-based bidirectional GRU (Bi-GRU) to predict $PM_{2.5}$, which combined ConvNet, Dense, and Bi-GRU. Wu et al. (2020) proposed a hybrid computing framework consisting of three modules: hybrid data pre-treatment, multi-objective feature selection, and ensemble prediction. Hybrid data pre-treatment can smooth the original series and generate more predictable sublayers.

On the other hand, the latter models decompose nonlinear and non-stationary complex system data into subsequences of different time scales, and then select appropriate prediction techniques to predict each subsequence and obtain the final predicted value (Zhang et al., 2021). For example, Chen et al. (2016) utilized the wavelet technique to decompose raw $PM_{2.5}$ observations, and basic prediction techniques, including econometric, artificial intelligence, and machine learning models, were selected to model and forecast the sub-components. Dong et al. (2022) proposed a two-stage decomposition technology consisting of complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and VMD to forecast daily $PM_{2.5}$. Du et al. (2020) used ICEEMDAN to filter high-frequency noise and extract the dominant characteristics of the time-series frequency. Jiang et al. (2021a) used complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and a deep temporal convolutional network (DeepTCN) to predict $PM_{2.5}$ concentration. Similarly, Niu et al. (2017) employed ensemble empirical mode decomposition and least square support vector machine (EEMD-LSSVM) to forecast daily $PM_{2.5}$ concentration.

3. Methodology

In this section, a new decomposition-ensemble framework is proposed to predict hourly $PM_{2.5}$ concentration, by employing the VMD, ARIMA, CNN, and TCN. Section 3.1 shows an overview of the proposed framework, and Sections 3.2, 3.3 and 3.4 describe the employed prediction models, VMD, TCN, and benchmarks, respectively.

3.1. Framework

Fig. 1 shows the general prediction framework. Accordingly, three sub-steps are involved in the framework: data decomposition and reconstruction, $PM_{2.5}$ concentration prediction, and empirical results evaluation, as follows:

3.1.1. Data decomposition and reconstruction

In this step, the time series of the research target data are decomposed and reconstructed. In particular, the $PM_{2.5}$ data $y(t)$ is decomposed into $n - 1$ intrinsic mode function (IMF) components $C_{j,t}$ $(j = 1, 2, \ldots, n-1)$ and one residual (r_t) using the VMD with self-adaptive decomposition. Then, the complexity of all the decomposed components is calculated by permutation entropy, and they are reconstructed into three "new" components, that is, low-, mid-, and high-frequency components.

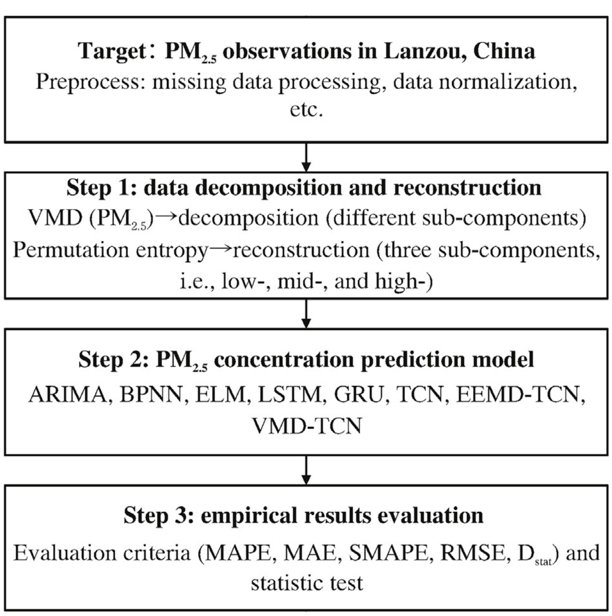

Fig. 1. General framework of hourly $PM_{2.5}$ concentration prediction in Lanzhou, China. Variational mode decomposition (VMD); autoregressive integrated moving average model (ARIMA); backpropagation neural network (BPNN); extreme learning machine (ELM); long short-term memory (LSTM); gated recurrent units (GRU); temporal convolutional network (TCN); ensemble empirical mode decomposition (EEMD); temporal convolutional network (TCN); mean absolute percent error (MAPE); mean absolute error (MAE); symmetric mean absolute percentage error (SMAPE); root mean square error (RMSE); and directional statistic (D_{stat}).

3.1.2. $PM_{2.5}$ concentration prediction model

In this section, three popular and excellent forecasting techniques are utilized to model and predict the reconstructed three "new" components by effectively capturing the different frequency time-series data pattern, namely, ARIMA, CNN and TCN. Specifically, ARIMA is introduced to model the low-frequency subsequence, owing to its advantages in dealing with low-frequency time-series patterns, CNN is adopted to model mid-frequency component features, and TCN is used for modeling high-frequency components.

3.1.3. Empirical results evaluation

In this step, the empirical results evaluation and statistical test are used to reveal the superiority of the proposed framework. Accordingly, first, the popular evaluation criteria, that is, mean absolute percent error (MAPE), mean absolute error (MAE), symmetric mean absolute percentage error (SMAPE), root mean square error (RMSE), and direction accuracy (D_{stat}), were employed to compare the prediction accuracy among all the forecasting models. Second, a typical statistical test, that is, the Diebold-Mariano (DM) test, was used to reveal the effectiveness of the proposed prediction framework.

3.2. VMD

As a promising decomposition technique, VMD decomposes non-stationary time-series into subseries with different frequencies (or different timescales) by constructing and solving the variational problem to deal with the center frequency and bandwidth (Yuan and Che, 2022). Correspondingly, the center frequency and finite bandwidth of each component are matched adaptively to solve the constrained variational model. The specific calculation process is as follows:

(1) The IMF (the components, $u_k(t)$) is taken as the amplitude modulated-frequency modulated signal (AM-FM).
(2) Calculate the center frequency of the components, and tune the spectrum of each mode signal.
(3) The constrained variational problem can be transformed into dealing with the optimization problem as follows:

$$min_{\{u_k,\omega_k\}} = \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}$$
$$s.t. \sum_k u_k = y_t \tag{2}$$

where * denotes convolution operator and Lagrange multipliers λ are used to transform the constraint variation problem into the following unconstrained problem:

$$L(\{u_k\},\{\omega_k\},\lambda) =$$
$$\alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| y_t - \sum_k u_k(t) \right\|_2^2 + \langle \lambda(t), y(t) - \sum_k u_k(t) \rangle \tag{3}$$

where α is the quadratic penalty parameter, $\lambda(t)$ denotes the Lagrange multiplier, u_k and ω_k represent k-th mode and the k-th center frequency, respectively, and * denotes convolution operator. A specific solution process was described by Dragomiretskiy and Zosso (2013).

3.3. TCN model

As an advanced variant of the standard CNN, TCN, which combines the time-series modeling capabilities of the RNN with the parallel computing capabilities of the CNN, is a promising deep learning technique for capturing the data patterns of nonlinear data (Mi et al., 2022).

Therefore, the TCN was introduced into this prediction framework based on its advantages, as shown in Fig. 2.

The expansion convolution operation is defined as (Hu et al., 2022; Meka et al., 2021):

$$F(t) = \sum_{i=0}^{k-1} f(i) \times x_{t-d\times i} \tag{4}$$

where $f(i)$ denotes the basic information of the filter, k is the size of the kernel, and d is the dilation factor which is generally increased exponentially with the depth of the network, i is the level of the network, $t-d\times i$ is the direction of the past.

Residual block is another component of TCN which is defined as:

$$o = ReLU(conv_{1\times1}(x) + F(x)) \tag{5}$$

where o is the output of the layer, and F represents the residual mapping to be learned. Moreover, rectified linear unit (ReLU) activation was applied, followed by a weight normalization layer.

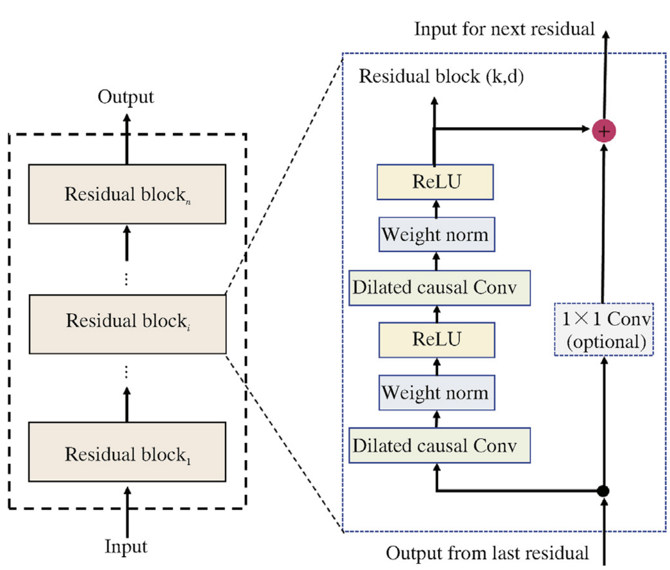

Fig. 2. Architecture of temporal convolutional network (TCN). ReLU: Rectified linear unit; Conv: convolution.

3.4. Benchmarks

To verify the effectiveness of the prediction method, basic and popular decomposition and forecasting benchmark models, including EMD-based, VMD-based, econometric forecasting, artificial intelligence forecasting, machine learning, deep learning, and combined models, are considered in the framework.

3.4.1. ARIMA

As a meaningful and basic econometric forecasting tool, the ARIMA (p, d, q) method includes the autoregressive and moving average information, the equation is as follows (Shahriar et al., 2021):

$$y_t = \varepsilon + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \mu_t + \psi_1 \mu_{t-1} + \cdots + \psi_q \mu_{t-q} \tag{6}$$

where p, d and q denote the autoregressive, degree of differencing, and moving average terms, respectively. Meanwhile, φ represents the parameter of the autoregressive part, ψ is the parameter of the moving average part, ε is a constant term, and μ is white noise.

3.4.2. BPNN

BPNN is a typical and popular artificial intelligence method in the field of forecasting, and it has been extensively used for forecasting time-series data (Liu et al., 2019). As a neural network, it requires two computing steps, that is, the positive propagation of information and the reverse propagation of error, to recursively tune the parameters and reduce the calculation error to acceptable levels. Normally, three layers of BPNN were constructed, which includes an input, hidden, and output layer. The calculation equation is as follows:

$$h_j = f_1\left(\sum_{i=1}^{n} w_{j,i} x_i + \theta_j\right), (\theta_j \geq 0, w_{j,i} \leq 1) \tag{7}$$

$$y = f_0\left(\sum_{j=1}^{m} w_{0,j} h_j + \lambda_0\right), (\lambda_0 \geq 0, w_{0,j} \leq 1) \tag{8}$$

where Eq. (7) indicates that the input information is mapped to the hidden layer through f_1, and x_i represents the i-th input layer node. Eq. (8) indicates that the hidden information is mapped to the output layer, h_j and y are j-th hidden layer node and the output, respectively. Correspondingly, θ_j and λ_0 are biases. n and m are the node numbers in the input and hidden layers, respectively. Then, w denotes the weight and f is the activation function. Furthermore, gradient descent is selected as the tuning method.

3.4.3. LSTM

As an improved recurrent neural network, LSTM, consisting of an forget gate (f_t), input gate (i_t), output gate (o_t) and cell (C_t), can deal with long-distance information. The architecture of a set of LSTM units is as follows (Li et al., 2017):

Forget gate:

$$f_t = \sigma\left[W_f \cdot (h_{t-1}, x_t) + b_f\right] \tag{9}$$

Input gate:

$$i_t = \sigma[W_i \cdot (h_{t-1}, x_t) + b_i] \tag{10}$$

$$\tilde{C}_t = \tanh[W_C \cdot (h_{t-1}, x_t) + b_C] \tag{11}$$

$$C_t = C_{t-1} \odot f_t + i_t \odot \tilde{C}_t \tag{12}$$

Output gate:

$$o_t = \sigma[W_o \cdot (h_{t-1}, x_t) + b_o] \tag{13}$$

$$h_t = o_t \odot \tanh(C_t) \tag{14}$$

Forecasted value:

$$\hat{y}_t = \sigma\left(W_{fc} h_t + b_{fc}\right) \tag{15}$$

where W and b represent the weight matrices and vectors of biases among the LSTM architecture, h is the hidden state, $\odot$ is point-wise multiplication, $\cdot$ denotes matrix multiplication, and $\sigma(\cdot)$ denotes the sigmoid activation function.

3.4.4. GRU

Similar to LSTM, GRU is also a widely employed deep learning tool, which is a simpler variant based on LSTM (Tao et al., 2019). In particular, two gates (update gate z_t and reset gate r_t) are constructed of the GRU for solving problems, such as the long-term memory and gradients in back-propagation when compared with the three gates of LSTM. The details are as follows:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{16}$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{17}$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t \odot h_{t-1}, x_t]) \tag{18}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{19}$$

$$\hat{y}_t = \sigma(W_{fc} h_t + b_{fc}) \tag{20}$$

For all the parameters of GRU, please refer to the above LSTM.

3.4.5. EEMD-TCN

To verify the effectiveness of the proposed decomposition method, the basic multi-scale decomposition method was utilized for forecasting the hourly $PM_{2.5}$ concentration. Correspondingly, the prediction target data are decomposed into different components, and the reputable deep learning algorithm, that is, TCN, is used to predict all the IMF components and residues. Then, the final predicted value is calculated via addition.

3.4.6. VMD-TCN

To reveal the advantage of the reconstruction of the decomposition method (i.e., the proposed ARIMA-CNN-TCN), the VMD-TCN was used to predict the $PM_{2.5}$ time series. Accordingly, first, VMD is employed to decompose the original data, and then TCN is used to predict all components. Finally, the predicted value of each component was added to obtain the final prediction result.

4. Empirical design

For illustration and verification, the proposed decomposition-ensemble framework was selected to forecast the hourly $PM_{2.5}$ concentration in Lanzhou, Gansu Province, China. In this section, the data source, evaluation criteria, and related forecasting parameters are described.

4.1. Data descriptions

To validate the effectiveness of the proposed forecasting framework, hourly observations of $PM_{2.5}$ concentration were collected in Lanzhou (36°03′N, 103°40′E), Gansu Province, China. Lanzhou is located in the west of China and in the upper reaches of the Yellow River. Lanzhou's air quality has deteriorated due to the high volume of industrial and automobile exhaust emissions, resulting in a pervasive haze, as a result of the influence of its constrained geographical environment. In terms of hourly $PM_{2.5}$ concentration, the dataset covers the period from January 1, 2020, to April 30, 2022, and is available on the website (https://quotsoft.net/air/). Accordingly, in this study, a popular and typical $PM_{2.5}$ concentration forecasting approach was utilized, which selects the historical time-series value ($X_i, X_{i+1}, \ldots, X_{i+L}$) as prediction inputs to forecast the predicted value $X_{(i+L+h)}$. L and h denote the lagged information and prediction horizon, respectively.

4.2. Evaluation criteria

In this study, the basic prediction accuracy criteria, including MAPE, RMSE, MAE, SMAPE, and D_{stat}, were selected to test the model's capability to forecast the level and direction of movement (Bai et al., 2019; Du et al., 2020; Jiang et al., 2021b; Kleine Deters et al., 2017; Ventura et al., 2019).

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} |1 - \hat{y}_t / y_t| \tag{21}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2} \tag{22}$$

$$MAE = \frac{1}{N} \sum_{t=1}^{N} |y_t - \hat{y}_t| \tag{23}$$

$$SMAPE = \frac{100\%}{N} \sum_{t=1}^{N} \frac{|y_t - \hat{y}_t|}{(|\hat{y}_t| + |y_t|)/2} \tag{24}$$

$$D_{stat} = \frac{1}{T} \sum_{t=1}^{T} a_t \times 100\% \tag{25}$$

where N denotes the amount of original data, y_t and $\hat{y}_t$ represent the actual and predicted values at time t, and D_{stat} $a_t = 1$ if $(\hat{y}_{t+1} - y_t)(y_{t+1} - y_t) \geq 0$, or $a_t = 0$ otherwise.

To prove the effectiveness of the proposed prediction framework at the statistical level, the DM test is employed to reveal the statistical significance of all forecasting models in the framework, and the loss function chooses the MAPE (Du et al., 2022). Accordingly, the null hypothesis compares the similar prediction accuracies of these forecasting models. The DM statistic can be calculated as follows:

$$DM = \frac{\overline{D}}{(\hat{V}_{\overline{D}}/N)^{1/2}} \tag{26}$$

where $\overline{D} = \frac{1}{N}\sum_{t=1}^{N}(D_t)$, $D_t = |1 - \hat{y}_{A,t}/y_t| - |1 - \hat{y}_{B,t}/y_t|$, $\hat{V}_{\overline{D}} = \gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k(\gamma_k = cov(D_t, D_{t-1}))$, h denotes the horizon, and $\hat{y}_{A,t}$ and $\hat{y}_{B,t}$ are the prediction values for y_t generated by proposed approaches A and benchmark model B, respectively, at time t.

4.3. Model specification

The dataset was split into 80% for the training set and 20% for the test set to ensure the reliability of the proposed method. In ARIMA, according to the Akaike information criterion (AIC), select ARIMA (1,1,2) for prediction. In the BPNN and ELM, the network was built with 5% hidden nodes in the training set. In the LSTM and GRU, the hidden size was set to 200, the learning rate was 0.005, the training epochs were 250, and the optimizer was Adam. In our experiment, the number of TCN blocks was set to 1, the kernel size was set to 3, RMSE has the advantages of scale-independence and interpretability, and RMSE was adopted as the loss function. Adam was used as the optimizer for our models. The number of training epochs was 1,000, the batch size was 64, the learning rate was 0.01, and the optimizer was Adam. In the CNN-TCN, the parameters were the same as those of the TCN, the kernel size was set to 3, the training epochs were 500, and the learning rate was 0.05. With the exception of ARIMA, all models were run an average of 100 times.

5. Empirical results

A comprehensive comparison of the proposed decomposition-ensemble framework, involving popular forecasting techniques and a basic decomposition-ensemble framework, was proposed. Thus, to illustrate and verify, the hourly $PM_{2.5}$ concentration observations' decomposed components are calculated based on VMD, a comprehensive comparison is presented in Section 5.1 and Section 5.2. The statistical test is conducted in Section 5.3, and Section 5.4 summarized the major conclusions of the empirical results.

5.1. Decomposition results

As a multi-scale decomposition technique, VMD can effectively deal with the mode-aliasing phenomenon and is robust to sampling and noise.

In the first step of the methodology, the different frequencies of the decomposed components and residues are shown in Fig. 3. In particular, all IMF components and residues were sorted from highest to lowest frequency. Meanwhile, four IMFs and one residue were extracted from the original $PM_{2.5}$ concentration time series using the VMD. The reconstruction results of the decomposed components (IMF and residue) are also shown in Fig. 4. In particular, three reconstructed decomposed subsequences based on complexity are obtained using permutation entropy, that is, high- (IMF 1 with the entropy as 0.9714, IMF 2 0.9611, IMF 3 0.9771), mid- (IMF 4 0.8509), and low-frequency (residue 0.5127).

5.2. Prediction results

Focusing on the prediction performance, a comprehensive comparison among the basic econometric forecasting methods, artificial intelligence methods, machine learning methods, deep learning techniques, decomposition-ensemble-based methods, and the proposed forecasting framework was conducted. For different level and direction prediction evaluation criteria, the prediction results with one-, two-, and three-step-ahead forecast values are presented in Table 1. The DM test, which can verify the effectiveness of the proposed prediction framework statistically, is presented in Table 2. As shown in Table 1, the bold font indicates that the proposed prediction framework achieves the best prediction performance in terms of all evaluation criteria at every horizon. Additionally, the statistical test (the DM test) also demonstrated that the proposed framework achieved a better predicted value than the other forecasting models, at a confidence level of 99%.

In terms of the MAPE in Table 1, the decomposition-ensemble-based prediction models, that is, EEMD-TCN, VMD-TCN, and VMD-TCN-ARIAM-CNN-TCN, are significantly superior to the single forecasting techniques for all horizons. More importantly, the proposed decomposition-ensemble framework achieved better prediction performance than other similar decomposition-ensemble models. In particular, the MAPE of the VMD-ARIMA-CNN-TCN prediction framework made a more accurate forecast value than the lowest value of the corresponding MAPE among all the other eight forecasting models, which improved the average performance across the three horizons by 53.53%. For the corresponding MAPE values of all the benchmarks, the proposed framework is approximately 70.45%, 74.17%, 74.95%, 75.41%, 76.85%, 74.53%, 67.59%, and 52.47% lower than the eight benchmarks when using ARIMA, BPNN, ELM, LSTM, GRU, TCN, EEMD-TCN, and VMD-TCN, respectively. Thus, it is proved that the proposed framework can achieve relatively stable and more accurate hourly $PM_{2.5}$ concentration prediction ability within one to three horizons. Moreover, the VMD-TCN

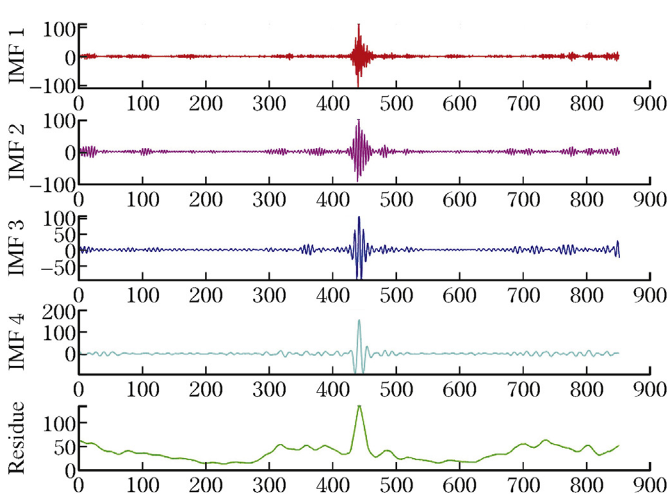

Fig. 3. Decomposed components of $PM_{2.5}$. IMF: intrinsic mode function.

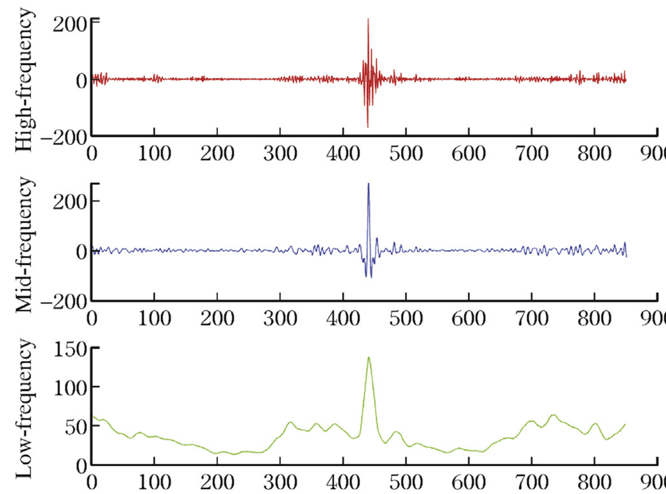

Fig. 4. Reconstructed decomposed components of $PM_{2.5}$.

achieved a better prediction performance than the EEMD-TCN, and the MAPE of the VMD-TCN method was approximately 31.81% lower than that of the EEMD-TCN, which demonstrates that the VMD can perform better in decomposing the original target data.

As for the RMSE in Table 1, based on the lowest number of bold outcomes reported by the proposed model, it is indicated that the employed decomposition technique, basic econometric model, and advanced deep learning technique at all horizons (VMD-ARIMA-CNN-TCN) can obtain the best performance when compared with the other models. Accordingly, the RMSE values of the benchmarks were approximately 76.74%, 77.22%, 78.02%, 76.26%, 77.68%, 76.78%, 71.38%, and 54.39% higher than the corresponding RMSE values of the proposed framework. Similarly, the first basic decomposition-ensemble prediction framework (EEMD-TCN) is almost 23.04%, 25.62%, 30.21%, 20.53%, 28.20%, and 23.24% lower than the benchmarks using ARIMA, BPNN, ELM, LSTM, GRU, and TCN, respectively. The second basic decomposition-ensemble prediction framework (VMD-TCN) was almost 49.00%, 50.05%, 51.81%, 47.94%, 51.05%, 49.08%, and 37.25% lower than those of the benchmarks using ARIMA, BPNN, ELM, LSTM, GRU, TCN, and EEMD-TCN, respectively. The above results repeatedly revealed the effectiveness of the decomposition-ensemble framework for modeling $PM_{2.5}$. Interestingly, the decomposition-ensemble-reconstruction strategy was superior to the simple decomposition-ensemble framework when comparing the last three prediction frameworks.

From the perspective of directional evaluation criteria, the VMD-ARIMA-CNN-TCN outperforms ARIMA, BPNN, ELM, LSTM, GRU, TCN, EEMD-TCN, and VMD-TCN by comparing the D_{stat} at all horizons, as shown in Table 1. Specifically, the values of the D_{stat} of the proposed framework are approximately 53.85%, 57.52%, 55.40%, 52.82%, 63.70%, 55.93%, 22.99%, and 13.02% higher than those of the benchmarks, respectively. All three decomposition-ensemble frameworks achieved better accuracy than the single forecasting techniques (i.e., ARIMA, BPNN, ELM, LSTM, GRU, and TCN). In addition, the VMD-TCN is superior to the EEMD-TCN's performance, with the former directional criteria being almost 8.82% higher than the latter framework.

Through all levels (MAPE, RMSE, MAE, and SMAPE) and directional (D_{stat}) evaluation criteria at each horizon, the following can be clearly revealed when forecasting $PM_{2.5}$ concentration. First, the decomposition-ensemble prediction framework can better forecast nonlinear and nonstationary $PM_{2.5}$ concentration observations than a single prediction model. Second, and most importantly, the decomposed prediction framework can be reconstructed well according to the complexity of each decomposed subsequence, and the formed prediction framework can effectively model the subsequence and significantly improve the

Table 1
The evaluation criteria of different prediction models.

Horizon	Criteria	Different prediction models								
		ARIMA	BPNN	ELM	LSTM	GRU	TCN	EEMD-TCN	VMD-TCN	VMD-ARIMA-CNN-TCN
One	MAPE	0.2179	0.2266	0.2647	0.2485	0.2688	0.2527	0.1782	0.1220	**0.0488**
	RMSE	16.4309	16.2735	19.8645	16.4511	16.8062	17.1787	11.7466	6.9856	**2.8009**
	D_{stat}	0.6058	0.6000	0.5529	0.6058	0.5529	0.5705	0.8058	0.8588	**0.9294**
	MAE	10.8224	10.2348	11.6018	10.0276	10.4410	10.0829	8.1891	7.1654	**5.9575**
	SMAPE	0.0557	0.0557	0.0587	0.0567	0.0576	0.0528	0.0446	0.0438	**0.0358**
Two	MAPE	0.2827	0.2873	0.3382	0.3475	0.3627	0.3293	0.2424	0.1804	**0.0737**
	RMSE	19.6476	18.4467	20.8315	18.8819	20.8984	20.0886	15.7872	10.5202	**3.9137**
	D_{stat}	0.5470	0.5647	0.5882	0.5529	0.5176	0.5764	0.7352	0.7882	**0.9058**
	MAE	15.0393	14.9998	15.7115	14.0694	14.1042	13.9994	12.9431	9.6380	**7.7306**
	SMAPE	0.0730	0.0730	0.0753	0.0734	0.0763	0.0754	0.0652	0.0525	**0.0453**
Three	MAPE	0.2954	0.3967	0.3361	0.3605	0.3846	0.3415	0.3052	0.1925	**0.1127**
	RMSE	20.5144	23.0602	19.1936	20.1039	21.2603	19.4156	18.4588	11.3526	**6.4458**
	D_{stat}	0.6058	0.5529	0.6000	0.6117	0.5823	0.5882	0.6588	0.7470	**0.8705**
	MAE	17.5855	17.1019	18.2773	16.3353	17.8967	16.0721	13.9928	9.6636	**8.3135**
	SMAPE	0.0743	0.0723	0.0772	0.0743	0.0781	0.0773	0.7422	0.0592	**0.0573**

Note: autoregressive integrated moving average model (ARIMA); backpropagation neural network (BPNN); extreme learning machine (ELM); long short-term memory (LSTM); gated recurrent units (GRU); temporal convolutional network (TCN); ensemble empirical mode decomposition-temporal convolutional network (EEMD); temporal convolutional network (TCN); variational mode (VMD); convolutional neural network (CNN); mean absolute percent error (MAPE); root mean square error (RMSE); directional statistic (D_{stat}); mean absolute error (MAE); symmetric mean absolute percentage error (SMAPE). Bold font indicates that the proposed prediction framework achieves the best prediction performance in terms of all evaluation criteria at every horizon.

Table 2
The Diebold-Mariano (DM) test (*p*-value).

Models	Horizon	ARIMA	BPNN	ELM	LSTM	GRU	TCN	EEMD-TCN	VMD-TCN
EEMD-TCN	1	2.39 (0.01)	2.97 (0.00)	2.95 (0.00)	3.17 (0.00)	3.12 (0.00)	2.95 (0.00)	-	-
	2	2.83 (0.00)	3.63 (0.00)	3.90 (0.00)	3.75 (0.00)	3.72 (0.00)	3.91 (0.00)	-	-
	3	2.49 (0.01)	2.77 (0.00)	2.65 (0.00)	2.14 (0.00)	3.31 (0.00)	2.67 (0.00)	-	-
VMD-TCN	1	3.50 (0.00)	3.92 (0.00)	3.85 (0.00)	4.04 (0.00)	3.86 (0.00)	3.85 (0.00)	2.75 (0.00)	-
	2	4.72 (0.00)	5.35 (0.00)	5.10 (0.00)	5.26 (0.00)	5.03 (0.00)	5.10 (0.00)	2.62 (0.00)	-
	3	4.81 (0.00)	5.90 (0.00)	4.95 (0.00)	4.70 (0.00)	5.29 (0.00)	4.98 (0.00)	2.59 (0.01)	-
VMD-ARIMA-CNN-TCN	1	3.71 (0.00)	4.14 (0.00)	4.07 (0.00)	4.25 (0.00)	4.06 (0.00)	4.07 (0.00)	3.35 (0.00)	4.03 (0.00)
	2	5.13 (0.00)	5.76 (0.00)	5.52 (0.00)	5.62 (0.00)	5.37 (0.00)	5.53 (0.00)	3.84 (0.00)	5.19 (0.00)
	3	5.31 (0.00)	6.48 (0.00)	5.68 (0.00)	5.42 (0.00)	5.64 (0.00)	5.72 (0.00)	4.42 (0.00)	5.17 (0.00)

Note: autoregressive integrated moving average model (ARIMA); backpropagation neural network (BPNN); extreme learning machine (ELM); long short-term memory (LSTM); gated recurrent units (GRU); temporal convolutional network (TCN); ensemble empirical mode decomposition (EEMD); variational mode decomposition (VMD); convolutional neural network (CNN).

prediction accuracy. The consistency of these results proves the effectiveness of the proposed decomposition ensemble reconstruction prediction framework (VMD-ARIMA-CNN-TCN) in modeling data patterns and improving the prediction accuracy, particularly for $PM_{2.5}$ concentration.

5.3. Statistic test

To further verify the validity of the proposed framework in terms of statistical significance, the popular statistical test (the DM test) was used to test the out-of-sample prediction results with statistical significance.

Specifically, the DM test demonstrated the superiority of the proposed VMD-ARIMA-CNN-TCN when compared with all the other benchmarks, at a confidence level of 99% (Table 2). Accordingly, three meaningful outcomes were identified. First, considering the prediction values of the proposed framework as the testing target, the *p*-values are all smaller than 1% (under a confidence level of 99%), indicating that the decomposition-ensemble-reconstruction prediction framework is statistically better than the benchmarks. Second, the EEMD-TCN and VMD-TCN can achieve better forecasting results than the six single models (ARIMA, BPNN, ELM, LSTM, GRU, and TCN) at a significance level of 1%, revealing the superiority of the basic decomposition-ensemble prediction framework. Third, the VMD-TCN can obtain a better prediction performance than the EEMD-TCN under a confidence level of 99%, which shows that the VMD can produce a good result for decomposing the target time series data.

5.4. Summary

Based on the prediction results and statistical test results of all forecasting methods, three major conclusions can be drawn.

(1) The proposed decomposition-ensemble-reconstruction forecasting framework, that is, VMD-ARIMA-CNN-TCN, is profoundly better than eight benchmarks (ARIMA, BPNN, ELM, LSTM, GRU, TCN, EEMD-TCN, and VMD-TCN) from the perspective of level and directional evaluation criteria.
(2) By decomposing hourly $PM_{2.5}$ observations, the predictive power of the decomposition-ensemble framework can be demonstrated repeatedly and statistically. It mainly models and forecasts the sub-components for single results, and then calculates the final prediction result via the rule of addition.
(3) Focusing on the decomposition technique, the proposed framework and VMD-TCN framework are statistically superior to the basic EEMD-TCN framework in terms of all three evaluation criteria, which reveals that the VMD-based prediction framework can comply well with the “divide-and-conquer” principle.

6. Conclusions

Due to the intrinsic complexity of $PM_{2.5}$ data, the main purpose of this study is to explore how the proposed decomposition-ensemble-reconstruction prediction framework model predicts nonlinear and non-stationary hourly $PM_{2.5}$ concentration data from a theoretical perspective. Based on this context, a comprehensive comparison of econometric, artificial intelligence, machine learning, deep learning, and basic decomposition-ensemble prediction models was conducted. Correspondingly, the empirical results indicate that the utilized VMD-ARIMA-CNN-TCN can repeatedly and significantly improve the prediction performance compared to the benchmarks (ARIMA, BPNN, ELM, LSTM, GRU, TCN, EEMD-TCN, and VMD-TCN). The comparison results of all statistical tests also reveal the effectiveness of the proposed framework.

Therefore, based on the complex data patterns of $PM_{2.5}$ historical observations, such as high volatility and irregularity, the above prediction results based on the decomposition-related framework prove that it is a promising strategy and tool to improve prediction accuracy in the field of $PM_{2.5}$ forecasts. In particular, the VMD method is a powerful decomposition processing technique for reducing the instability of the original data, and permutation entropy is adopted to regroup the components based on the different frequencies. Moreover, ARIMA was employed to model and predict the low-frequency components, CNN was used to deal with the mid-frequency components, and TCN was utilized to predict the high-frequency components. For illustration and verification purposes, Lanzhou $PM_{2.5}$ data are considered in this decomposition-ensemble-reconstruction prediction framework.

The developed forecasting framework could be potentially exploited for monitoring and predicting air quality conditions, such as $PM_{2.5}$ concentration, and detecting main air pollution sources based on the data pattern. This model is simple in structure, easy to implement, and practical, and can provide a theoretical basis and technical support for government departments to formulate reasonable measures in reducing $PM_{2.5}$.

Based on the above prediction results, the proposed framework (VMD-ARIMA-CNN-TCN) can improve the prediction with better robustness. However, different $PM_{2.5}$ related factors could be considered in this framework, such as the weather and wind speed variables. Furthermore, for the prediction horizon, the long-term hourly prediction of $PM_{2.5}$ concentration could also be explored in the future. As for robustness, the different prediction techniques, prediction strategies, and parameters in this framework require further consideration. In addition to $PM_{2.5}$ concentration, this prediction framework could be applied to complex systems, such as crude oil price, electricity consumption, and non-ferrous metal price.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This research work was partly supported by the National Natural Science Foundation of China (Grant Nos.: 71874133 and 72201201), the Research Program of Shaanxi Soft Science, China (Grant No.: 2022KRM015), the Youth Innovation Team of Shaanxi Universities (2020-68), and Shaanxi Province Qin Chuangyuan “scientist + engineer” team building project (Grant No.: 2022KXJ-007).

References

Ausati, S., Amanollahi, J., 2016. Assessing the accuracy of ANFIS, EEMD-GRNN, PCR, and MLR models in predicting $PM_{2.5}$. Atmos. Environ. 142 (Oct.), 465–474.

Bai, Y., Li, Y., Zeng, B., et al., 2019. Hourly $PM_{2.5}$ concentration forecast using stacked autoencoder model with emphasis on seasonality. J. Clean. Prod. 224 (3), 739–750.

Chen, T., He, J., Lu, X., et al., 2016. Spatial and temporal variations of $PM_{2.5}$ and its relation to meteorological factors in the urban area of Nanjing, China. Int. J. Environ. Res. Publ. Health 13 (9), 921.

Chen, Y., Kang, Y., Chen, Y., et al., 2020. Probabilistic forecasting with temporal convolutional neural network. Neurocomputing 399 (Jul), 491–501.

Deng, S., Zhang, N., Zhang, W., et al., 2019. Knowledge-driven stock trend prediction and explanation via temporal convolutional network. In : Companion Proceedings of the 2019 World Wide Web Conference. Association for Computing Machinery, New york, pp. 678–685.

Djalalova, I., Delle Monache, L., Wilczak, J., 2015. $PM_{2.5}$ analog forecast and Kalman filter post-processing for the community multiscale air quality (CMAQ) model. Atmos. Environ. 108 (May), 76–87.

Dragomiretskiy, K., Zosso, D., 2013. Variational mode decomposition. IEEE Trans. Signal Process. 62 (3), 531–544.

Dong, L., Hua, P., Gui, D., et al., 2022. Extraction of multi-scale features enhances the deep learning-based daily $PM_{2.5}$ forecasting in cities. Chemosphere 308 (Pt 2), 136252.

Du, P., Wang, J., Yang, W., et al., 2022. A novel hybrid fine particulate matter ($PM_{2.5}$) forecasting and its further application system: case studies in China. J. Forecast. 41 (1), 64–85.

Du, P., Wang, J., Hao, Y., et al., 2020. A novel hybrid model based on multi-objective Harris hawks optimization algorithm for daily $PM_{2.5}$ and PM_{10} forecasting. Appl. Soft Comput. 96 (Nov.), 106620.

Hu, J., Luo, Q., Tang, J., et al., 2022. Conformalized temporal convolutional quantile regression networks for wind power interval forecasting. Energy 248 (C), 123497.

Huang, G., Li, X., Zhang, B., et al., 2021. $PM_{2.5}$ concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition. Sci. Total Environ. 768 (May), 144516.

Jiang, F., Zhang, C., Sun, S., et al., 2021a. Forecasting hourly $PM_{2.5}$ based on deep temporal convolutional neural network and decomposition method. Appl. Soft Comput. 113 (Pt B), 107988.

Jiang, X., Luo, Y., Zhang, B., 2021b. Prediction of $PM_{2.5}$ Concentration based on the LSTM-TSLightGBM variable weight combination model. Atmosphere-basel. 12 (9), 1211.

Kleine Deters, J., Zalakeviciute, R., Gonzalez, M., et al., 2017. Modeling $PM_{2.5}$ urban pollution using machine learning and selected meteorological parameters. J. Electr. Comput. Eng. 2017 (Jan.), 5106045.

Kow, P., Chang, L., Lin, C., et al., 2022. Deep neural networks for spatiotemporal $PM_{2.5}$ forecasts based on atmospheric chemical transport model output and monitoring data. Environ. Pollut. 306 (Aug.), 119348.

Li, X., Peng, L., Yao, X., et al., 2017. Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation. Environ. Pollut. 231 (Pt 1), 997–1004.

Liu, H., Jin, K., Duan, Z., 2019. Air $PM_{2.5}$ concentration multi-step forecasting using a new hybrid modeling method: comparing cases for four cities in China. Atmos. Pollut. Res. 10 (5), 1588–1600.

Liu, H., Chen, C., 2020. Prediction of outdoor $PM_{2.5}$ concentrations based on a three-stage hybrid neural network model. Atmos. Pollut. Res. 11 (3), 469–481.

Liu, H., Duan, Z., Chen, C., 2020. A hybrid multi-resolution multi-objective ensemble model and its application for forecasting of daily $PM_{2.5}$ concentrations. Inf. Sci. 516 (C), 266–292.

Liu, X., Qin, M., He, Y., et al., 2021. A new multi-data-driven spatiotemporal $PM_{2.5}$ forecasting model based on an ensemble graph reinforcement learning convolutional network. Atmos. Pollut. Res. 12 (10), 101197.

Ma, J., Ding, Y., Cheng, J., et al., 2020. A Lag-FLSTM deep learning network based on Bayesian optimization for multi-sequential-variant $PM_{2.5}$ prediction. Sustain. Cities Soc. 60 (Sep.), 102237.

Meka, R., Alaeddini, A., Bhaganagar, K., 2021. A robust deep learning framework for short-term wind power forecast of a full-scale wind farm using atmospheric variables. Energy 221 (C), 119759.

Menares, C., Perez, P., Parraguez, S., et al., 2021. Forecasting $PM_{2.5}$ levels in Santiago de Chile using deep learning neural networks. Urban Clim. 38 (Jul.), 100906.

Mi, X., Yu, C., Liu, X., et al., 2022. A dynamic ensemble deep deterministic policy gradient recursive network for spatiotemporal traffic speed forecasting in an urban road network. Digit. Signal Prog. 129 (C), 103643.

Niu, M., Gan, K., Sun, S., et al., 2017. Application of decomposition-ensemble learning paradigm with phase space reconstruction for day-ahead $PM_{2.5}$ concentration forecasting. J. Environ. Manag. 196 (Jul.), 110–118.

Samal, K.K.R., Babu, K.S., Das, S.K., 2021. Multi-directional temporal convolutional artificial neural network for $PM_{2.5}$ forecasting with missing values: a deep learning approach. Urban Clim. 36 (Mar.), 100800.

Shahriar, S.A., Kayes, I., Hasan, K., et al., 2021. Potential of ARIMA-ANN, ARIMA-SVM, DT and CatBoost for atmospheric $PM_{2.5}$ forecasting in Bangladesh. Atmosphere-basel 12 (1), 100.

Sun, K., Tang, L., Qian, J., et al., 2021. A deep learning-based $PM_{2.5}$ concentration estimator. Displays 69 (4), 102072.

Sun, W., Huang, C., 2020. A hybrid air pollutant concentration prediction model combining secondary decomposition and sequence reconstruction. Environ. Pollut. 266 (Pt 3), 115216.

Tan, J., Liu, H., Li, Y., et al., 2022. A new ensemble spatio-temporal $PM_{2.5}$ prediction method based on graph attention recursive networks and reinforcement learning. Chaos, Solit. Fractals 162 (Sep.), 112405.

Tao, Q., Liu, F., Li, Y., et al., 2019. Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU. IEEE Access 7 (Jan.), 76690–76698.

Ventura, L.M.B., de Oliveira Pinto, F., Soares, L.M., et al., 2019. Forecast of daily $PM_{2.5}$ concentrations applying artificial neural networks and Holt–Winters models. Air Qual. Atmos. Health 12 (3), 317–325.

Wang, B., Kong, W., Zhao, P., 2021a. An air quality forecasting model based on improved convnet and RNN. Soft Comput. 25 (14), 9209–9218.

Wang, J., Wang, R., Li, Z., 2022a. A combined forecasting system based on multi-objective optimization and feature extraction strategy for hourly $PM_{2.5}$ concentration. Appl. Soft Comput. 114 (Jan.), 108034.

Wang, P., Feng, H., Bi, X., et al., 2021b. Phase objectives analysis for $PM_{2.5}$ reduction using dynamics forecasting approach under different scenarios of PGDP decline. Ecol. Indicat. 129 (Oct.), 108003.
Wang, Z., Li, H., Chen, H., et al., 2022b. Linear and nonlinear framework for interval-valued $PM_{2.5}$ concentration forecasting based on multi-factor interval division strategy and bivariate empirical mode decomposition. Expert Syst. Appl. 205 (Nov.), 117707.
Wood, D.A., 2022. Trend decomposition aids forecasts of air particulate matter ($PM_{2.5}$) assisted by machine and deep learning without recourse to exogenous data. Atmos. Pollut. Res. 13 (3), 101352.
Wu, H., Liu, H., Duan, Z., 2020. $PM_{2.5}$ concentrations forecasting using a new multi-objective feature selection and ensemble framework. Atmos. Pollut. Res. 11 (7), 1187–1198.
Yang, H., Wang, C., Li, G., 2022. A new combination model using decomposition ensemble framework and error correction technique for forecasting hourly $PM_{2.5}$ concentration. J. Environ. Manag. 318 (Sep.), 115498.
Yeo, I., Choi, Y., Lops, Y., et al., 2021. Efficient $PM_{2.5}$ forecasting using geographical correlation based on integrated deep learning algorithms. Neural Comput. Appl. 33 (22), 15073–15089.
Yin, S., Liu, H., Duan, Z., 2021. Hourly $PM_{2.5}$ concentration multi-step forecasting method based on extreme learning machine, boosting algorithm and error correction model. Digit. Signal Prog. 118 (Nov.), 103221.
Yu, L., Wang, Z., Tang, L., 2015. A decomposition–ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. Appl. Energy 156 (Oct.), 251–267.
Yu, L., Ma, M., 2021. A memory-trait-driven decomposition-reconstruction-ensemble learning paradigm for oil price forecasting. Appl. Soft Comput. 111 (C), 107699.
Yuan, F., Che, J., 2022. An ensemble multi-step M-RMLSSVR model based on VMD and two-group strategy for day-ahead short-term load forecasting. Knowl. Base Syst. 252 (C), 109440.
Zhang, L., Lin, J., Qiu, R., et al., 2018. Trend analysis and forecast of $PM_{2.5}$ in Fuzhou, China using the ARIMA model. Ecol. Indicat. 95 (Pt 1), 702–710.
Zhang, Z., Zeng, Y., Yan, K., 2021. A hybrid deep learning technology for $PM_{2.5}$ air quality forecasting. Environ. Sci. Pollut. Res. 28 (29), 39409–39422.
Zhou, Y., Chang, F., Chang, L., et al., 2019. Multi-output support vector machine for regional multi-step-ahead $PM_{2.5}$ forecasting. Sci. Total Environ. 651 (Pt 1), 230–240.
Zhu, M., Xie, J., 2023. Investigation of nearby monitoring station for hourly $PM_{2.5}$ forecasting using parallel multi-input 1D-CNN-biLSTM. Expert Syst. Appl. 211 (Jan.), 118707.