

Research Paper

Air quality predictions with a semi-supervised bidirectional LSTM neural network

Luo Zhang ^{a,b}, Peng Liu ^{a,*}, Lei Zhao ^c, Guizhou Wang ^{a,**}, Wangfeng Zhang ^d, Jianbo Liu ^a^a Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China^b School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China^c School of Information Science and Technology, Beijing Forestry University, Beijing, 100083, China^d Key Laboratory of Space Utilization, Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

ABSTRACT

Keywords:

Air quality
PM_{2.5} predictions
Bidirectional long short-term memory
Empirical mode decomposition
Semi-supervised

Efficient and accurate air quality predictions can contribute to public health protection and policy decision making. Fine particulate matter (PM_{2.5}) is an important index for measuring and controlling the degree of air pollution. Recent studies have obtained satisfactory PM_{2.5} predictions by designing complex models or adding numerous auxiliary data sets to models, and few studies have effectively extracted the spatiotemporal features of PM_{2.5} time-series data. In this study, a semi-supervised model was proposed for predicting PM_{2.5} concentrations. The approach includes empirical mode decomposition (EMD) and bidirectional long short-term memory (BiLSTM) neural networks. This model only requires PM_{2.5} time-series data as inputs, which are regarded as signal data. EMD is applied as an unsupervised feature learning method to decompose the data and extract the frequency and amplitude features. This approach improved short-term trend predictions, especially for sudden changes. BiLSTM was used in the supervised learning stage. Beijing hourly and daily PM_{2.5} datasets collected from the China National Environmental Monitoring Centre were used to validate the prediction performance of the proposed model. The results demonstrated that this model was more accurate than the other standard LSTM-based model, with four better indicator values at the hourly (RMSE: 6.86 $\mu\text{g}\cdot\text{m}^{-3}$, MAE: 4.92 $\mu\text{g}\cdot\text{m}^{-3}$, MAPE: 10.66%, R²: 0.989) and daily (RMSE: 22.58 $\mu\text{g}\cdot\text{m}^{-3}$, MAE: 16.67 $\mu\text{g}\cdot\text{m}^{-3}$, MAPE: 60.87%, R²: 0.742) scales. Furthermore, this study proposed a new method of multiscale PM_{2.5} predictions by reconstructing hourly PM_{2.5} datasets to form multi-hour datasets. This method could reduce error accumulation in PM_{2.5} multi-step predictions using LSTM-based models and captured at least 70% of the explained variance in this study, demonstrating the feasibility of the model.

1. Introduction

PM_{2.5} (particulate matter with a diameter less than 2.5 μm) is an important index for measuring and controlling the degree of air pollution, and it has attracted considerable attention in recent years. Researchers have found that exposure to pollutants such as PM_{2.5} increases the risk for cardiovascular and respiratory diseases (Brook et al., 2010; Brunekreef and Holgate, 2002; Dominici et al., 2006), which can result in mortality and decrease the related economic benefits (Giannadaki et al., 2013). In recent decades, China is striving for sustainable development but still experiences severe haze pollution as the result of the

intensive emission of air pollutants (Cheng et al., 2013). According to the 'China Ecological Environment Bulletin (2017)' among the 338 monitored cities in China, 239 cities (70.7%) have exceeded the air quality standards, and the average number of days below the air quality standard reached 22.0%. Therefore, efficient and accurate PM_{2.5} concentration predictions can aid in developing air quality management strategies and protecting public health.

Previous studies that predicted air pollutant concentrations mainly fell into two categories: deterministic methods and statistical methods. The deterministic methods, such as CMAQ (Community Multiscale Air Quality) (Byun and Schere, 2006) and WRF-Chem (Weather Research

Peer review under responsibility of Turkish National Committee for Air Pollution Research and Control.

* Corresponding author.

** Corresponding author.

E-mail address: liupeng@radi.ac.cn (P. Liu).

and Forecasting model coupled with Chemistry) (Shrivastava et al., 2011), using theoretical meteorological, physical and chemical models to simulate the translation, diffusion, or elimination process of contaminants. These interpretable, theoretically based models are widely used in the environmental and atmospheric community (Wang et al., 2014; Zhang et al., 2019). However, due to unreliable emission data, complicated underlying surface conditions, and an imperfect theoretical foundation, these methods often result in limited accuracy (Jin et al., 2019; Kumar et al., 2014).

Benefiting from the rapid development of remote sensing technology, continuous improvements in space and ground remote sensing detection systems have provided abundant data support for landslide detection (Chen et al., 2017a), glacial lake outlines (Chen et al., 2017b), active fire detection (Lin et al., 2018, 2019) as well as air pollutant (Fan et al., 2020; Wei et al., 2020). Many statistical methods such as the multiple linear regression (MLR) model (Tai et al., 2010), the autoregressive moving average (ARMA) method (Zhang et al., 2017), and generalized linear models (GLMs) (Tong et al., 2018) have been developed for air quality predictions. The parameters of these models have clear statistical significance; moreover, these models are simpler and more clearly than deterministic models. Other linear statistical models that consider combined land use (Son et al., 2018), meteorological conditions (Shi et al., 2018), multi-satellite data (Lin et al., 2015) and multi-angle implementation of atmospheric correction (Li et al., 2018) have been developed and displayed promising performance. Deterministic methods could benefit from these studies to obtain improved simulation results. However, these methods are difficult to apply in nonlinear models. Notably, the relationships between PM_{2.5} and other variables are not strictly linear. Therefore, the accuracy of statistical methods can also be limited.

Currently, some advanced artificial intelligence methods are used to establish nonlinear relationships and extract complex features in many geoscience and remote sensing applications such as remote image classification (Yan et al., 2019) (Huang et al., 2019) (Liu et al., 2017b) (Liu et al., 2017a), unmixing for hyperspectral Imagery (Feng et al., 2018) (Feng et al., 2017) (Feng et al., 2019), spatial data mining (Huang et al., 2017) (Liu et al., 2018), and spatial infrastructure (Yan et al., 2018) (Fan et al., 2018) etc. In the field of environmental science, support vector machine (SVM) (Sun and Sun, 2016), random forest (Chen et al., 2018; Hu et al., 2017; Huang et al., 2018), and artificial neural network (ANN) (Elangasinghe et al., 2014) methods are widely used in air pollution prediction and have achieved better performance than conventional statistical models. Some researchers have further considered combined methods to retain their respective advantages. For example, based on ARMA and an improved BP neural network (Zhu and Lu, 2016), air mass trajectory analysis and wavelet transformation were combined to obtain an improved ANN (Feng et al., 2015). Extensive research has shown that the long short-term memory neural network model has achieved significant improvements in temporal prediction compared to traditional methods. LSTM creatively mediates the balance between memorizing and forgetting by adding some multiple threshold gates. As one of the state-of-the-art method has demonstrated excellent performance in both short-term and long-term PM_{2.5} contamination prediction (Li et al., 2017; Zhao et al., 2019). Moreover, by adding auxiliary data (e.g., meteorological, satellite, geographical, and simulation data) and incorporating convolutional neural networks (CNNs) to efficiently extract spatial features, LSTM-based methods have yielded excellent results in PM_{2.5} spatial prediction (Qi et al., 2019; Wen et al., 2019). However, most previous studies did not extract the spatiotemporal features of PM_{2.5} time-series data and did not specify the effect of auxiliary data when using LSTM-based models. Based on an analysis of existing studies, we found that adding numerous auxiliary data sets, such as meteorological data, had more effect on frequency and amplitude changes. If features cannot be effectively extracted, auxiliary data do no improve predictions and actually increase interference. In addition, the prediction results of LSTM-based methods are continuations of

historical trends and do not accurately reflect short-term changes based on multi-step predictions. For high-frequency and irregular changes in time-series data, the prediction ability of LSTM is even more limited, and this limitation can be clearly shown in the analysis of the hourly and daily PM_{2.5} prediction performance given in section 3.

To overcome these restrictions, this study proposes a semi-supervised model named EMD-BiLSTM for predicting PM_{2.5} concentrations. This approach includes empirical mode decomposition (EMD) and bidirectional long short-term memory (BiLSTM) neural networks as the main components. The model only requires PM_{2.5} time-series data sets as inputs, which are regarded as signal data. In the unsupervised stage, EMD is applied for unsupervised feature learning and used to decompose the data into several components representing different frequency and amplitude features. In the supervised stage, the BiLSTM neural networks outperform unidirectional networks in efficiently extracting temporal correlation features, and the proposed method is much faster and more accurate than the other standard LSTM-based method. Although the proposed model has unique inputs and a simple model structure, it yields high accuracy and stability in air quality prediction and has a satisfactory effect on short-term trend predictions, especially sudden changes. Furthermore, this study proposes a new method of multiscale PM_{2.5} predictions by reconstructing hourly PM_{2.5} data sets to multi-hour data sets. This approach utilizes the advantages of LSTM and can reduce error accumulation in PM_{2.5} multi-step predictions.

The remainder of this paper is organized as follows. Section 2 illustrates the study area and available data and then describes the basic method and the proposed EMD-BiLSTM model in detail. Section 3 discusses the predicted performance of different methods and datasets using mean root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and coefficient of determination (R^2). A temporal comparison and detailed analysis are also given. Finally, Section 4 concludes the paper with the main contributions of this study.

2. Data and methods

2.1. Study area and available data

Beijing, a city located on the North China Plain (115° 25' to 117° 30' E; 39° 26' to 41° 03' N, Fig. 1A), is the capital of China. Fig. 1A displays all the environmental monitoring stations in Beijing. This study area is an urban area of Beijing that has a concentrated population and severe traffic. We select 13 monitoring stations in the major urban area shown in Fig. 1B. These sites are located in a region with local commercial and residential areas and traffic sources that is representative of the urban environment. In addition, we pick two monitoring stations shown in 1 S¹ (116.832 E, 40.370 N) and S2 (116.407 E, 40.003 N), as control experiment points for proving the robustness of the proposed model.

Open air quality observation data from the ground monitoring stations in China have been published by the China National Environmental Monitoring Centre (<http://www.cnemc.cn/>). We collected PM_{2.5} data from all 13 urban environmental monitoring stations and the two control experiment monitoring stations. We used these 13 observation stations average PM_{2.5} concentrations represent the Beijing urban area air quality. Overall, we collected hourly and daily PM_{2.5} datasets from January 1, 2014, to December 31, 2018, from the Beijing urban area. Among these datasets, we selected an hourly scaled dataset for the time period from February 1, 2017, to March 31, 2018, and this data set include approximately 9500 records from each station. Then, we removed invalid observations and calculated the average values of hourly PM_{2.5} in urban Beijing; the corresponding data set included 9073 records. Beijing urban daily PM_{2.5} datasets from January 1, 2014, to December 31, 2018, were obtained directly from the website, resulting in approximately 1815 valid observations. There is well known that daily PM_{2.5} concentrations are usually the average of the 24 hourly PM_{2.5} observations each day. Therefore, the reconstructed multi-hour

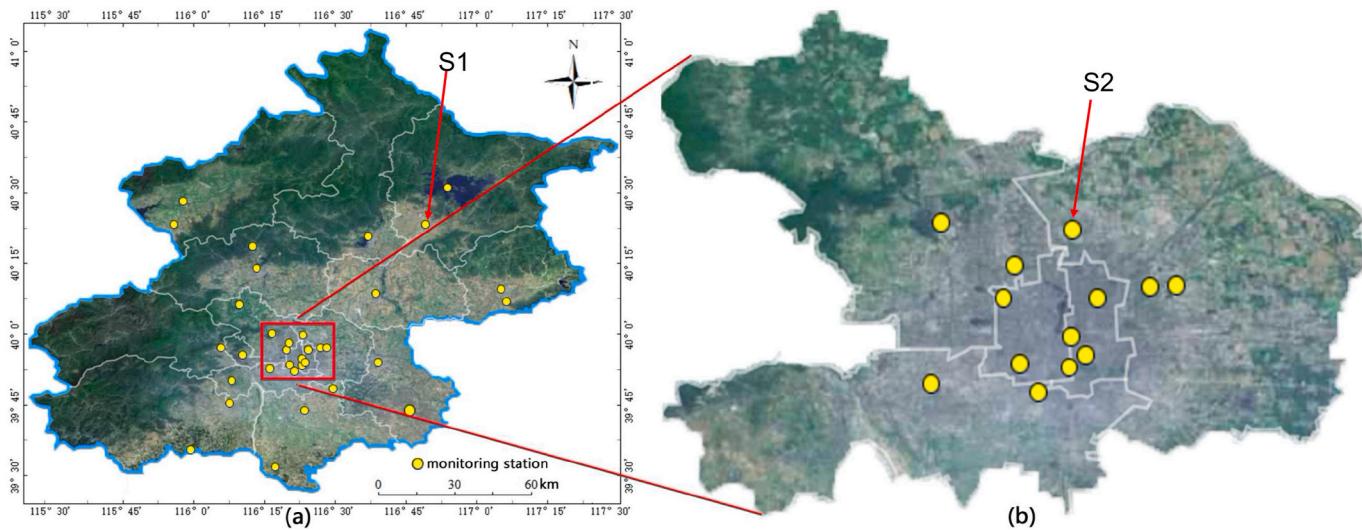


Fig. 1. (A) The locations of Beijing and all the environmental monitoring stations. (B) 13 monitoring stations in the Beijing urban area. Control experiment monitoring stations S1 and S2.

data sets were based on the Beijing urban hourly PM_{2.5} datasets from January 1, 2014, to December 31, 2018, to form a series of multi-hour PM_{2.5} data sets, including averages and maximum and minimum concentrations.

2.2. Multiscale predictors

Numerous previous studies have attempted to make multi-step predictions of air quality using LSTM-based methods. Actually, since the prediction result of LSTM-based methods is only a continuation of the historical trend, these methods cannot accurately predict short-term trend changes based on multi-step predictions. In other words, for an LSTM model using hourly data for multi-hour prediction, all the prediction results are similar because the real trend of observations may not be appropriately considered. For example 2–6 h predictions may be similar to the 1-h prediction.

To avoid this restriction, we achieved multiscale predictions by restructuring hourly data to obtain multi-hour data sets. The Beijing urban PM_{2.5} data sets used in this study mainly include three categories of data: hourly, daily and restructured multi-hour data. The format of the multiscale predictors can be generalized as follows.

As shown in Fig. 2, the x_t means PM_{2.5} concentrations at time t, the unit of t is hour. The top line of Fig. 2 illustrates the format of the hourly PM_{2.5} predictions. The time-series data ($x_{t-n} \dots x_t$) are used to predict x_{t+1} . For the multiscale predictions. Here, the X_t means PM_{2.5} concentrations during a period T. For a given period T, hourly data can be used to generate a series of T-hour PM_{2.5} data sets. X_t is generated from the hourly time-series data ($x_{t-T+1} \dots x_t$) for a time segment. Thus, the average, maximum and minimum PM_{2.5} values at this period are determined. In this way, we achieve multiscale predictions using the

resampled time-series of data ($X_{t-n} \dots X_t$) to predict X_{t+1} . Usually, when the $T = 24$, Fig. 2 could represent the format of the daily PM_{2.5} predictions.

In this study, we reconstructed hourly PM_{2.5} data sets to T-hour ($T = 3, 6, 12$) PM_{2.5} data sets to perform experiments. In this approach, the new datasets represent the temporal changes in PM_{2.5} at different time scales. Moreover, the multiscale predictions obtained in this way could reduce the error accumulation in PM_{2.5} multi-step data when using the LSTM-based model.

2.3. Temporal modeling with BiLSTM

LSTMs have displayed excellent ability in general-purpose time series modeling and prediction (Hochreiter and Schmidhuber, 1997) based on the original recurrent neural network (RNN), which provides a balance between memorizing and forgetting by adding multiple threshold gates. LSTMs are regarded as state-of-the-art methods for addressing time series prediction problems. As shown in Fig. 3, each LSTM block consists of a memory cell and three gates: the input gate i_t , forget gate f_t and output gate o_t . The key to LSTMs is the memory cell state C_t , as shown by the horizontal line running through the top of Fig. 3. Additionally, X_t and h_t are the input and output of the LSTM.

LSTM methods can be effectively applied in time-series prediction problems because of the use of the memory cell state C_t , which essentially acts as an accumulator of state information. When a new input is received, the forget gate f_t decides what information are going to throw away from the memory cell state. Next, the input gate i_t decides which values will be updated, and a tanh layer creates a vector of new candidate values \tilde{C}_t , which can be added to the memory cell state. Then, these

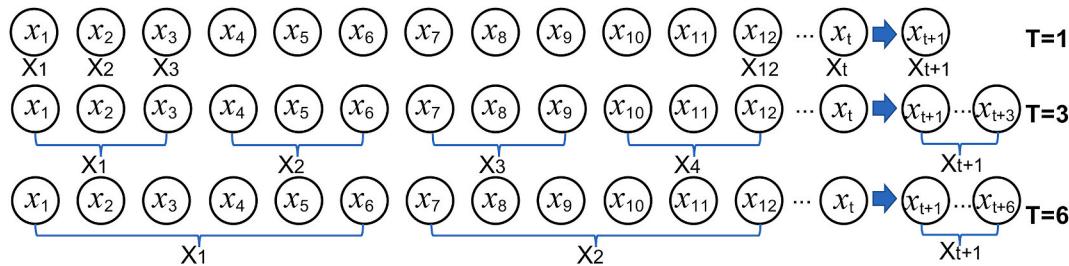


Fig. 2. Illustration of the multiscale predictors. The T means period. The X_t means PM_{2.5} concentrations at time t. The X_{-t} means PM_{2.5} concentrations during a period T.

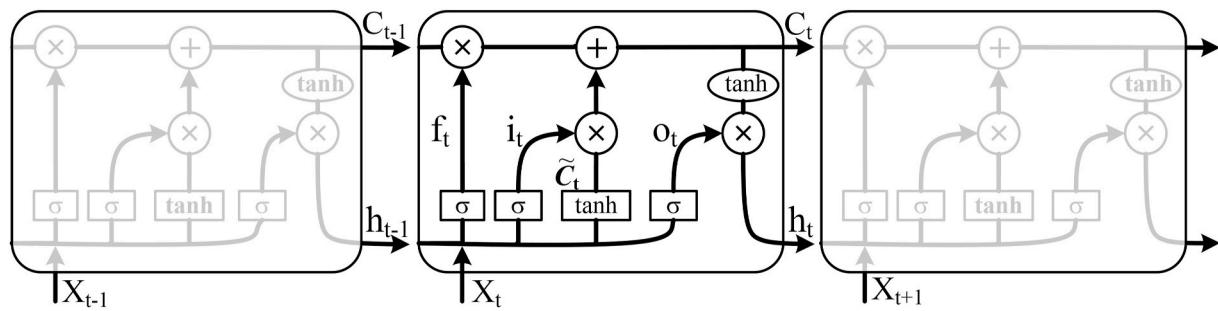


Fig. 3. The LSTM network structure.

results are combined to create a state update. Finally, the output gate o_t and C_t generate a final state h_t . The common key equations in this method are shown below, where \circ denotes the Hadamard product; W and b denote the corresponding weights and bias vectors, respectively; $\sigma(\cdot)$ denotes the sigmoid function; and $\tanh(\cdot)$ denotes the tanh function (Shi et al., 2015).

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci} \circ C_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf} \circ C_{t-1} + b_f) \quad (2)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc}x_t + W_{hc} \circ h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co} \circ C_{t-1} + b_o) \quad (4)$$

$$h_t = o_t \circ \tanh(C_t) \quad (5)$$

Multiple LSTMs can be stacked and temporally concatenated to form complex structures, such as BiLSTM networks. Graves et al. proposed a BiLSTM network and found that it outperformed unidirectional models (Graves and Schmidhuber, 2005). Additionally, some researchers have obtained satisfactory results in context-sensitive keyword detection (Wöullmer et al., 2010) and PM_{2.5} prediction (Tong et al., 2019).

Although bidirectional networks outperform unidirectional networks, LSTM methods are much faster and more accurate than both standard RNN and time window multilayer perceptron (MLP) methods (Graves and Schmidhuber, 2005). For high-frequency and irregular changes in time-series data, the prediction ability of BiLSTM is still limited. For example, for a BiLSTM model, no matter how the parameters are adjusted, the prediction accuracy of hourly PM_{2.5} data is better than that of daily PM_{2.5} data; moreover, the time delay phenomenon caused by the low accuracy of daily data is more obvious. Notably, the fluctuations in daily PM_{2.5} time-series are more dramatic than those in hourly data, and daily PM_{2.5} time-series convey more high-frequency information. These frequency and irregular change characteristics can lead to low accuracy.

With LSTM-based methods, previous studies obtained satisfactory PM_{2.5} prediction results through designing complex models combined with multiple methods or numerous auxiliary data sets (Li et al., 2017; Qi et al., 2019). We consider a complex model and numerous auxiliary data sets, such as meteorological data sets, to be more effective in considering frequency and amplitude changes. Therefore, if we can separate the high-frequency and irregular change components from the original data before making predictions, the accuracy of the results will be improved.

2.4. Unsupervised frequency extraction with EMD

To improve the BiLSTM prediction ability for high-frequency time-series of data with irregular changes, the essential step is separating the high-frequency and irregular change components from the original data. We apply a simple method that uses EMD as an unsupervised learning

method for frequency feature extraction. The original PM_{2.5} time-series data are decomposed to obtain some components with different frequency features. Thus, competitive performance can be achieved in PM_{2.5} predictions.

The adaptive signal time-frequency processing method called EMD, which was first proposed by Huang in 1998 (Huang et al., 1998), can directly decompose nonlinear and nonstationary time series into components. The fluctuations or trends of different scales in the original signal are sequentially decomposed. These series of components with different feature scales are called Intrinsic Mode Functions (IMFs), and the remaining components are called residues. The basic concept of EMD is to decompose a wave of an irregular frequency into two parts: multiple single-frequency waves and a residual wave. If we assume that $X(t)$ is the original data sequence, the relationship described above can be expressed as follows.

$$X(t) = \sum_{i=1}^n IMF_i + \text{Residue} \quad (6)$$

The EMD algorithm is not defined by a specific theoretical formula but an algorithmic process. The general algorithmic process is as follows. (1) Find all the maxima and minima points of the original sequence data $X(t)$; then, fit them to the upper and lower envelopes of the original sequence using the cubic spline function. (2) The mean of the upper envelope and the lower envelope is recorded as m_1 . Next, m_1 is subtracted from $X(t)$, i.e., the low frequency is subtracted, to obtain the new data sequence h_1 . Usually, h_1 is not a stationary sequence; thus, the above process must be repeated until a stationary sequence is obtained. (3) Repeat the above process to obtain the first intrinsic mode function component IMF_1 , which represents the component of the highest frequency of the signal data sequence $X(t)$. (4) Subtract IMF_1 from $X(t)$ to obtain a new data sequence $X_1(t)$ without the high-frequency component; then, decomposition is performed for $X_1(t)$ to obtain IMF_2 . This process is repeated until the final data sequence cannot be decomposed, and this sequence is recorded as *Residue*.

EMD is widely applied in diverse complex system analyses due to its superiority in forecasting time series, such as oil prices (Wu et al., 2019) and wind speeds (Qin et al., 2019). EMD uses a strategy called divide and conquer (Bai et al., 2019). Bai (Bai et al., 2019) first applied this concept for hourly PM_{2.5} concentration forecasting but ignored the combined advantages of EMD and LSTM. Moreover, there are two serious problems with the EMD-based modeling framework for time series prediction. First, the number of decomposition results will change when new data are added; as a result, the model trained previously will be invalid. In addition, the restraining end effect of EMD influences the prediction results (Xiong et al., 2014). To remove these restrictions of EMD, we added two processing steps before and after the EMD process.

First, when beginning the EMD process, we use standard LSTM to obtain several forecasting values. Then, we extend the original time series data by adding the predicted results. These components will reduce the restraining end effect of EMD and maintain the original frequency and amplitude features at the endpoint. In addition, we added

a filtering process at the end of EMD processing. This process includes the removal of IMFs that have invalid values from the entire prediction process to avoid a negative effect on the predicted results, retraining the trained model in which IMFs values display no changes after new data are added, and retraining the model in which IMF values lead to clearly changes after new data are added. Finally, we will obtain a fixed number of decomposition results that better fit the trained model.

2.5. Proposed model architecture

A schematic diagram of the EMD-BiLSTM model is displayed in Fig. 4. This model is a semi-supervised model based on EMD and BiLSTM. EMD can be considered an unsupervised feature learning method, and it was used to decompose the data and extract the frequency and amplitude features. BiLSTM was used in the supervised learning stage, and it is an effective architecture that can capture temporally correlated data. The entire process consists of three main stages.

The first stage are mainly about data preprocessing. Collected data are preprocessed to obtain serval different time scale data sets, including hourly, multil-hours (including averages and maximum and minimum concentrations of period T) and daily. Base on those data to achieve multiscale predictions.

In the second stage, an unsupervised frequency feature extraction process is implemented based on EMD. At the beginning of this unsupervised learning process, we use standard BiLSTM to obtain several prediction data sets for the given inputs and then extend the original time-series data by adding the predicted results. This process reduces the end effect in EMD. Then, the frequency and amplitude features are extracted from the extended data to generate several IMFs and residue components. Finally, Through the IMF filter discussed above, changes to the results based on adding new data are evaluated, and the components are filtered accordingly. This process adapts existing models and components to reduce the number of calculations and training time.

The final stage is using the assigned model parameters which the models are trained. The BiLSTM model parameters are adjusted based on the overall characteristics of the IMF components, and the model is retrained if the components changed after new data are added. Finally, each predicted IMF is summed to obtain the final predicted PM_{2.5} result. At each time scale, including hourly, daily and multi-hour scales, a prediction result is obtained. Aggregating each PM_{2.5} prediction of each time-series data to form multiscale predictions.

3. Results and discussion

3.1. Experimental settings

The datasets in this study mainly include hourly, daily and restructured multi-hour datasets. All the datasets are divided into three parts. For example, hourly data from February 1, 2017, to March 31, 2018, the top 60% (February 1, 2017, to October 15, 2017, approximately) are selected as the training dataset, and the remaining data are average shared, 20% (October 16, 2017, to January 15, 2018, approximately) for verification dataset and 20% (January 16, 2018, to March 31, 2018, approximately) for testing dataset. This study conducts experiments involving three methods, the ordinary LSTM model, BiLSTM model, and EMD-BiLSTM model. Compared with other methods, many previous studies achieved significant results in time series prediction using LSTM-based deep learning methods. Therefore, this study chose the ordinary LSTM model as the baseline model for a performance comparison with the proposed EMD-BiLSTM model.

To improve the EMD-BiLSTM model accuracy, the model parameters should be adjusted to fit the corresponding IMF data. Overall, the BiLSTM model has three layers: the first layer is a bidirectional LSTM with 128 nodes; the second layer is a dense layer with 64 nodes; and the last layer is a dense layer with 10 nodes. The dropout was set to 0.5, 0.3, and 0.1 for layers 1, 2, and 3, respectively. According to previous studies, a small time lag cannot guarantee enough long-term memory inputs for the model, and a large time lags permit an increased number of unrelated inputs. Many studies have used a time lag of 8 for PM_{2.5} predictions. However, in this study, we found that high-frequency IMFs set time lag to an extremum period would achieved great performance. In other words, the high-frequency IMFs should set a small time lag that ranged from 3 or 4. The normal-frequency IMFs set a time lag of 8. The other methods (BiLSTM and LSTM), which were used as controls in the experiment, used as variable lag based on frequency changes; e.g., daily data sets had a time lag of 4, and hourly data sets had a time lag of 8.

Finally, the RMSE, MAE, MAPE, and R² values were selected to evaluate the effectiveness of the proposed method. We assume that O_i and P_i are the observed (ground truth) and predicted values, respectively. \bar{O}_i is the average value of n observed samples. These indicators can be formulated as follows.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - Q_i)^2} \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - Q_i| \quad (8)$$

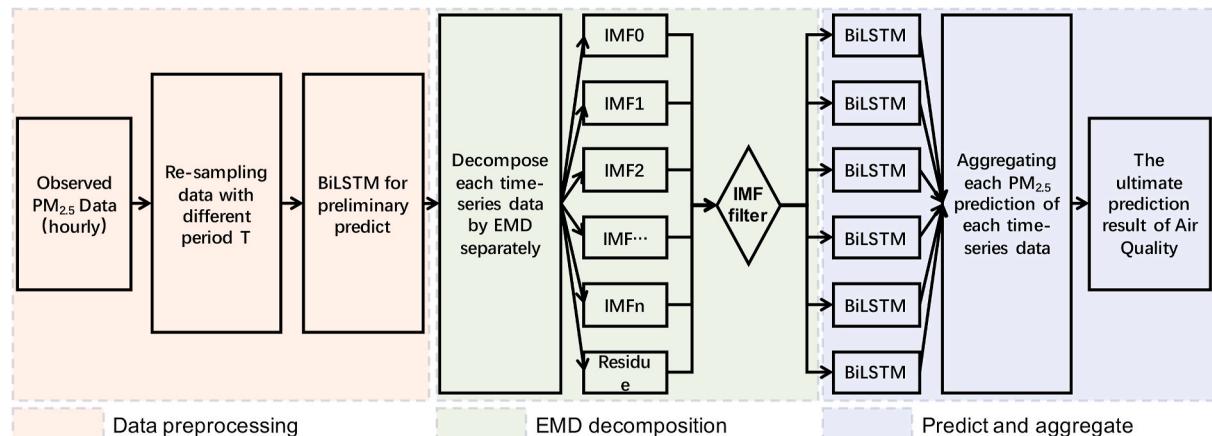


Fig. 4. The flowchart of the EMD-BiLSTM model proposed in this study. Re-sampled data include hourly, Mutil-hour (including averages and maximum and minimum concentrations of period T) and daily. They are predicted separately, then aggregated for the ultimate prediction result of air quality.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|P_i - Q_i|}{Q_i} \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (P_i - \bar{Q}_i)^2}{\sum_{i=1}^n (O_i - \bar{Q}_i)^2} \quad (10)$$

3.2. Prediction performance

Clearly comparison of these 13 observation stations hourly and daily average PM_{2.5} concentrations prediction performance are shown in Fig. 5 and Fig. 6. The quantitative performance of our approach and those used as baselines is summarized in Table 1. The best results are denoted in bold, indicating that the predictions of EMD-BiLSTM are more consistent with the observed data. Meanwhile, the two control experiment results displayed on Table 2, Fig. 7 and Fig. 8 further proved the robustness of the proposed model. The details of EMD and the component prediction performance are shown in Fig. 9 and Table 4. These results demonstrate the advantage of unsupervised frequency extraction. More comparison details of the these 13 observation stations hourly average PM_{2.5} concentrations prediction performance are shown in Fig. 5 (A1,A2,A3,B1,B2), and they reveal that the proposed model better predicts sudden changes than do traditional models. In addition, the quantitative results shown in Tables 1 and 4 indicate that the proposed model achieves high accuracy in multiscale prediction and always retains components accounting for at least 70% of the explained

variance, resulting in a greater than two-fold improvement in daily PM_{2.5} predictions compared to traditional methods.

3.3. Comparison of experiments

A comprehensive analysis of the experiments was performed in the following order. We used the hourly and daily Beijing urban datasets to compare the traditional LSTM model, BiLSTM model and EMD-BiLSTM model. The quantitative results shown in Table 1 agree with those of Tong (Tong et al., 2019), suggesting that BiLSTM indeed improved the accuracy of PM_{2.5} prediction, but the effect is not very obvious. Moreover, as shown in Figs. 5 and 6, a clearly comparison of the hourly and daily predictions demonstrates the competitive predictive ability of the proposed EMD-BiLSTM model, which performs better in short-term trend prediction, especially for sudden changes, as well as the two control experiment results displayed on Table 2, Figs. 7 and 8, those results further proved the robustness of the proposed model. Notably, daily data include more high-frequency details and irregular changes than hourly data, and the results indicate that the prediction ability of the traditional LSTM model is limited when predicting high-frequency and irregular changes in time-series data.

For the proposed EMD-BiLSTM model, the prediction performance was significantly improved compared to that of traditional models, achieving the best indicator values at the hourly (RMSE: 6.86 µg·m⁻³, MAE: 4.92 µg·m⁻³, MAPE: 10.66%, and R²: 0.989) and daily (RMSE: 22.58 µg·m⁻³, MAE: 16.67 µg·m⁻³, MAPE: 60.87%, and R²: 0.742) scales. Fig. 6 compares the daily prediction results of the EMD-BiLSTM and

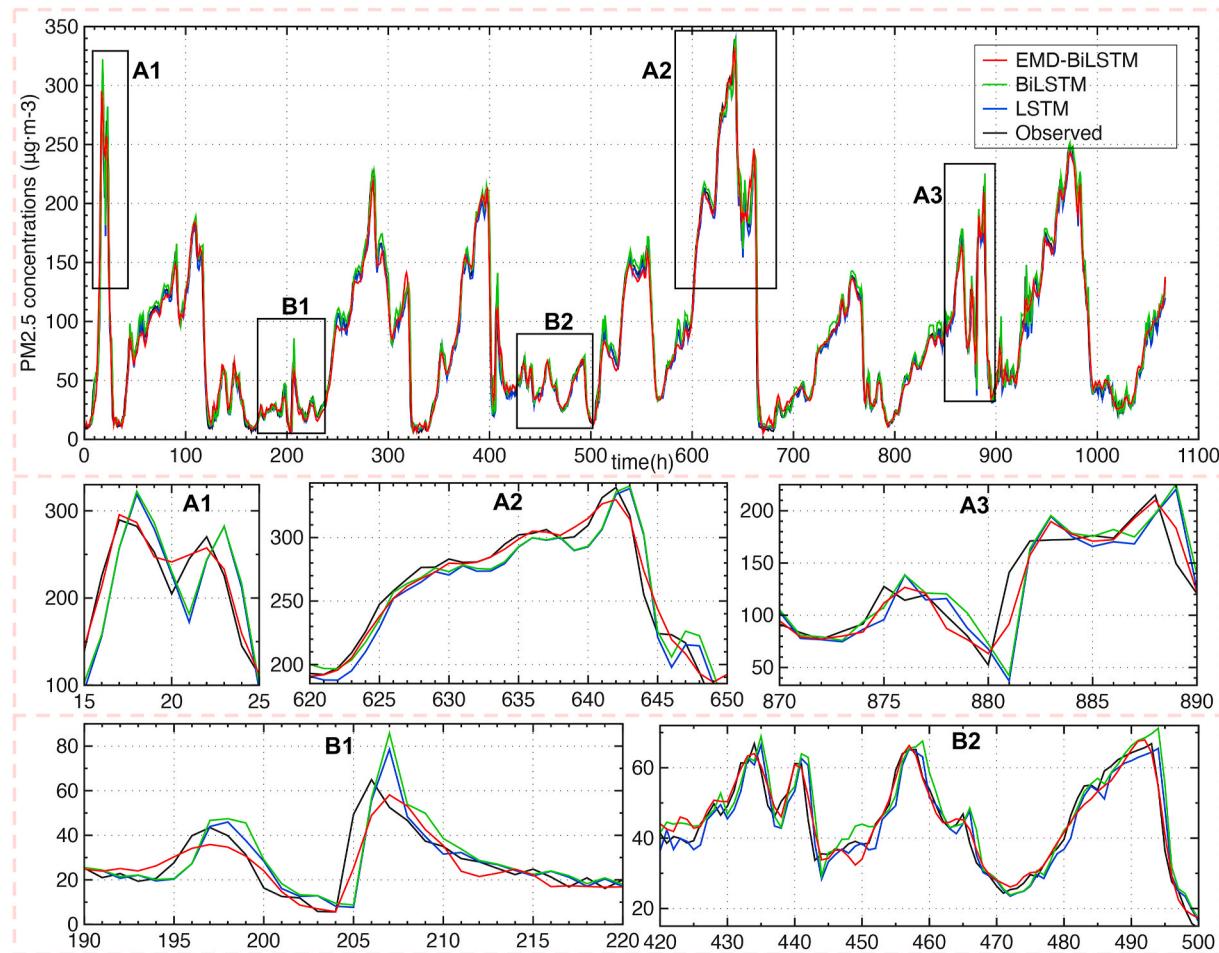


Fig. 5. Comparison of the different model prediction results with these 13 observation stations average hourly PM_{2.5} concentrations (date range: February 11 to March 31, 2018). The three areas of sudden change (A1, A2 and A3) and the two areas of smooth change (B1 and B2) as comparison of details.

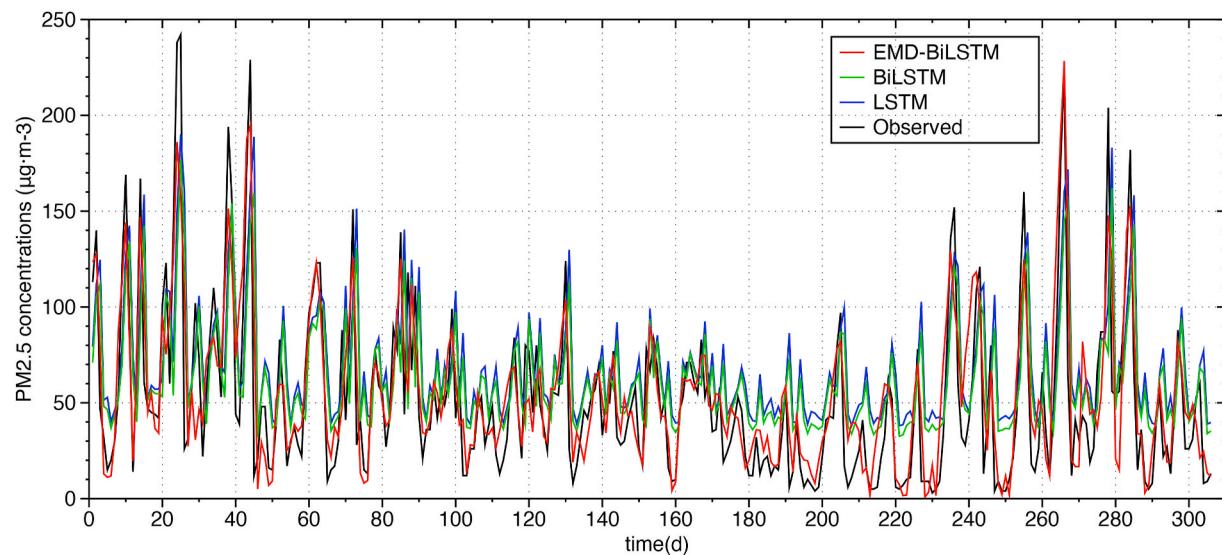


Fig. 6. Comparison of the EMD-BiLSTM and LSTM model prediction results with these 13 observation station average daily PM_{2.5} concentrations (date range: February 18 to December 31, 2018).

Table 1

Comparison of the performance of different methods with these 13 observation stations average PM_{2.5} concentrations using four indicators: the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2).

	Hourly				Daily			
	RMSE ($\mu\text{g}\cdot\text{m}^{-3}$)	MAE ($\mu\text{g}\cdot\text{m}^{-3}$)	MAPE (%)	R^2	RMSE ($\mu\text{g}\cdot\text{m}^{-3}$)	MAE ($\mu\text{g}\cdot\text{m}^{-3}$)	MAPE (%)	R^2
	EMD-BiLSTM	6.86	4.92	10.66	0.989	22.58	16.67	60.87
BiLSTM	11.23	6.94	11.35	0.973	35.89	29.36	129.84	0.375
LSTM	12.52	7.89	10.74	0.971	37.24	30.76	137.17	0.343

Table 2

Comparison of the performance of different methods with the S1 and S2 observation stations hourly PM_{2.5} concentrations using four indicators: the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2).

	Monitoring Stations S1				Monitoring Stations S2			
	RMSE ($\mu\text{g}\cdot\text{m}^{-3}$)	MAE ($\mu\text{g}\cdot\text{m}^{-3}$)	MAPE (%)	R^2	RMSE ($\mu\text{g}\cdot\text{m}^{-3}$)	MAE ($\mu\text{g}\cdot\text{m}^{-3}$)	MAPE (%)	R^2
	EMD-BiLSTM	9.72	5.41	17.09	0.979	8.09	4.91	15.38
BiLSTM	17.42	8.97	32.39	0.932	14.87	7.83	20.46	0.948
LSTM	17.77	9.13	35.65	0.928	14.92	8.03	23.34	0.937

LSTM models with the 13 observation stations hourly and daily average PM_{2.5} concentrations. Although the LSTM model prediction results in trend are generally consistent with the observed data, the obtained R^2 value is only 0.343, indicating that the model only captured 34.3% of the explained variance. Because daily PM_{2.5} concentrations have more frequent fluctuations and higher amplitudes than hourly data, traditional LSTM models struggle to obtain satisfactory results based on daily data. However, as shown in Fig. 6 and Table 1, the EMD-BiLSTM model yielded a considerable improvement in the R^2 value to 0.742. Fig. 6 illustrates this significant achievement and shows that the EMD-BiLSTM model prediction results are consistent with the observed data based on the frequency and amplitude. Specifically, the accuracy is high for high-frequency and irregular changes in the data.

To further demonstrate the advantage of the EMD-BiLSTM model, the three areas of sudden change (A1, A2 and A3) and the two areas of smooth change (B1 and B2) were shown in Fig. 5. Overall, benefiting from the prediction ability of the standard LSTM method, both EMD-BiLSTM and standard LSTM yield accurate overall trend prediction, despite some small accuracy deviations near trend variation points. A

detailed comparison of Fig. 5 suggests that EMD-BiLSTM significantly improve PM_{2.5} predictions compared to those of the standard LSTM model. Fig. 5(A1, A2 and A3) compares the prediction results of the EMD-BiLSTM model and other models with the observed PM_{2.5} concentrations based on the shape change. The standard LSTM and BiLSTM prediction results are similar, they are more consistent with the observed trend shape, but it displays an obvious time delay phenomenon. The time delay indicates that the standard LSTM and BiLSTM method cannot make correct predictions at turning points; as a result, the prediction is usually underestimated for upward trends or overestimated for downward trends. In other words, for PM_{2.5} concentration predictions in the next hour, the original trend is preserved, but the predicted values lag the actual change. In contrast, EMD-BiLSTM usually yields accurate predictions at turning points, and it can accurately reflect trend changes. However, as shown in Fig. 5(B1 and B2), this competitive advantage is not obvious in smooth change areas. Overall, unsupervised frequency extraction with EMD has a significant effect on the predictions, and BiLSTM outperforms the standard LSTM model based on the temporal prediction results.

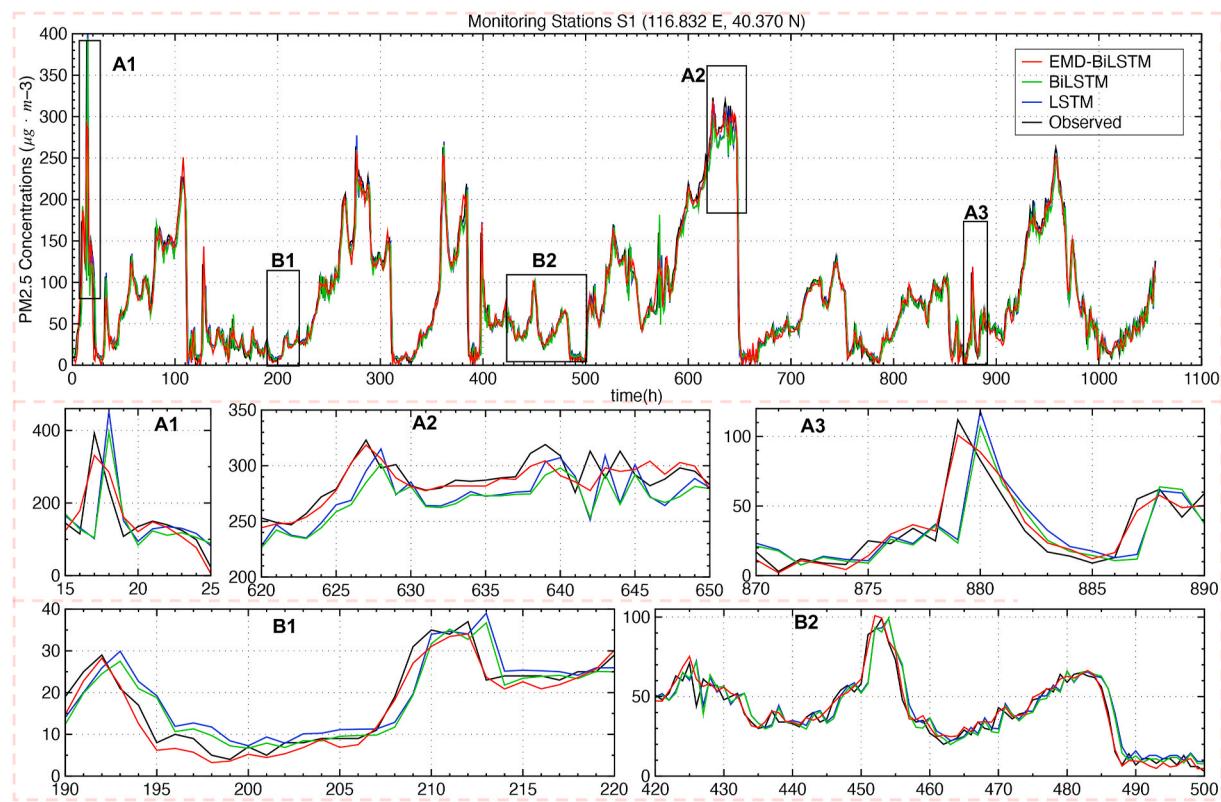


Fig. 7. Comparison of the different model prediction results with the S1 observation station hourly PM_{2.5} concentrations (date range: February 11 to March 31, 2018). The three areas of sudden change (A1, A2 and A3) and the two areas of smooth change (B1 and B2) as comparison of details.

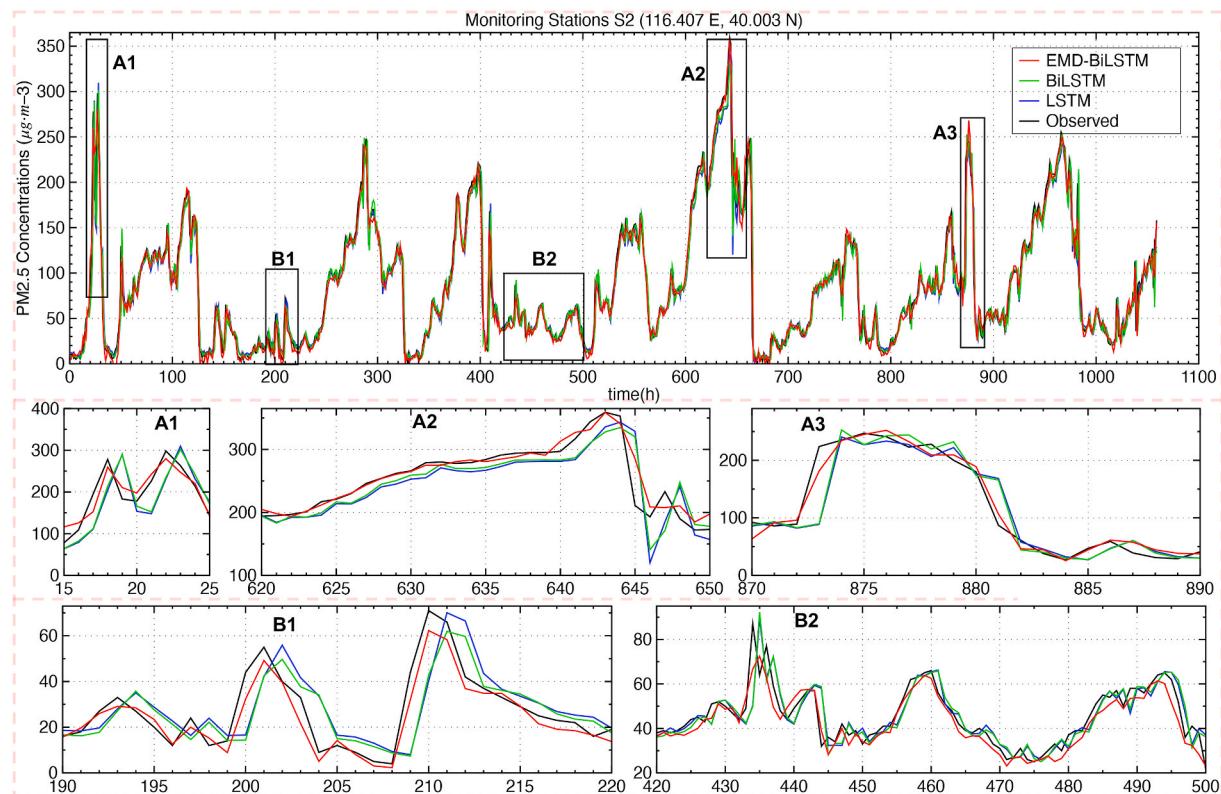


Fig. 8. Comparison of the different model prediction results with the S2 observation station hourly PM_{2.5} concentrations (date range: February 11 to March 31, 2018). The three areas of sudden change (A1, A2 and A3) and the two areas of smooth change (B1 and B2) as comparison of details.

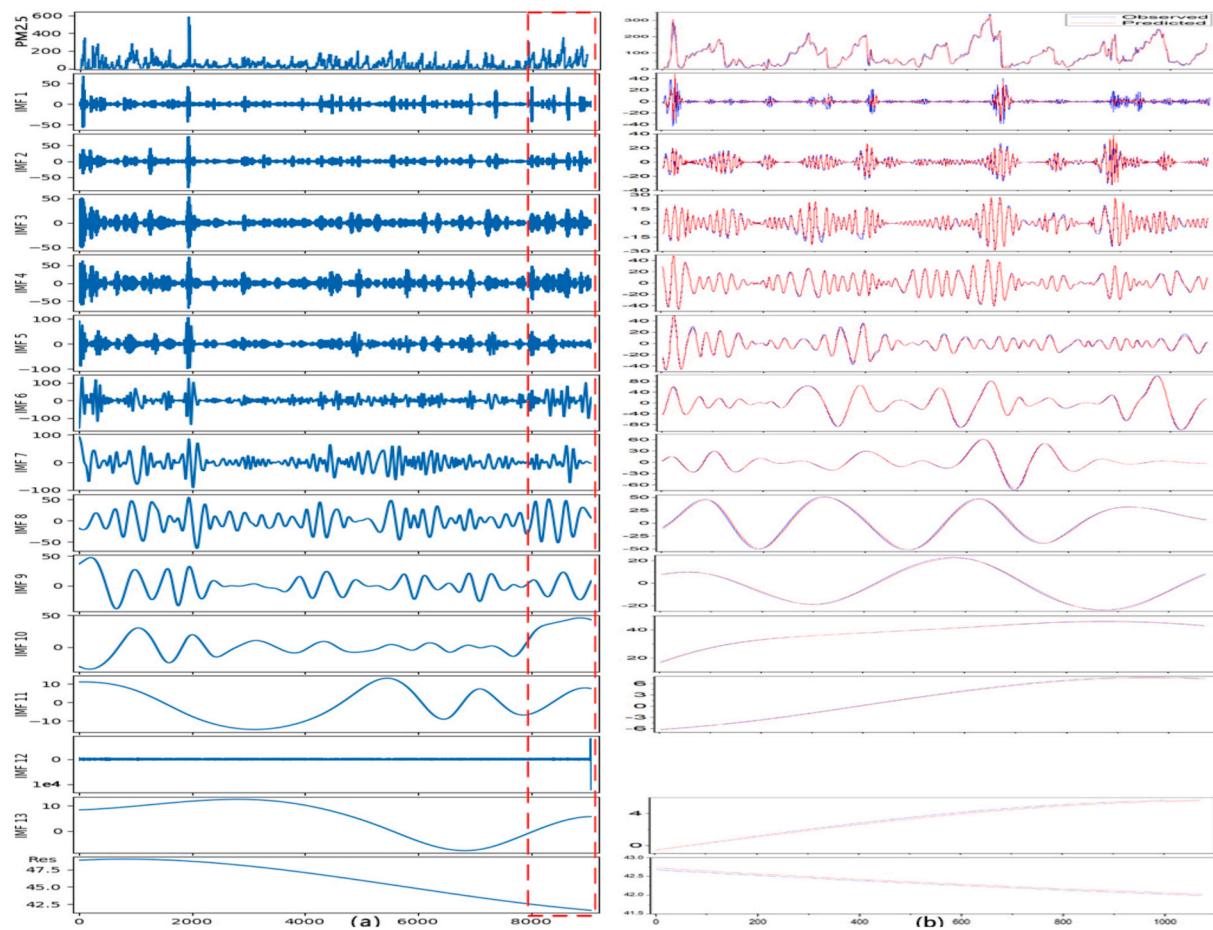


Fig. 9. (A) All hourly Beijing urban PM_{2.5} time-series data and the components based on EMD. The dotted box indicates the scope of the test dataset. (B) Comparison of the prediction performance of PM_{2.5} and each component in the test dataset, dotted box in (A). The IMFs represents different frequency features.

In order to prove the robustness of the proposed model, we pick two monitoring stations data for independent experiments. The results are shown in Table 2, Figs. 7 and 8. Their achievements are similar to average prediction performances in Fig. 5. The proposed EMD-BiLSTM model significantly improve PM_{2.5} predictions compared to those of the standard LSTM model. Combined Tables 1 and 2 for Comparing, the four indicators were changed along with the different datasets. But this approach always keep a good performance. Furthermore, Table 4 displayed more datasets predict results with different datasets, which has different frequency and amplitude features. Considering those different datasets have different frequency and amplitude features, the results proved that standard LSTM model is more difficult to predict data with high-frequency and irregular changes. However, we apply a simple method that uses EMD as an unsupervised learning method for frequency feature extraction. The original PM_{2.5} time-series data are decomposed to obtain some components with different frequency features, most of those components are suitable to be predicted by LSTM. Thus, competitive performance can be achieved in PM_{2.5} predictions.

We attribute the above excellent prediction performance to the advantage of EMD, which can effectively extract frequency features with unsupervised learning. This unsupervised learning process considers hidden temporally variable factors that influence the results, and limited auxiliary data are required, thus simplifying the model structure and reducing the number of calculation while achieving high accuracy and stability in PM_{2.5} predictions. There is no doubt that auxiliary data (e.g., meteorological, satellite, geographical, and simulation data) actually perform an important role in PM_{2.5} prediction, especially for long-time prediction (e.g., 24 h). However, when using LSTM-based models, if

features cannot be effectively extracted, auxiliary data do no improve predictions and actually increase interference. For example, the experiment in Wen et al. (2019) research shown that did not specify the effect of auxiliary data when using LSTM-based models had little effect for prediction, thus some models combined convolutional neural network (CNN) to extract auxiliary data features, such as Conv-LSTM (Wen et al., 2019), GCN-LSTM (Qi et al., 2019), those used for air pollution prediction achieved better performance. This approach extract frequency features with unsupervised learning to achieve similar performance. A detailed analysis and an experimental comparison of this advantage are performed with these 13 observation stations average hourly PM_{2.5} concentrations predictions as follows.

The EMD and component prediction performance values are shown in Fig. 9 and Table 3. As shown in Fig. 9A, after unsupervised frequency extraction by EMD, the original PM_{2.5} time-series data were decomposed into several components that represented different frequencies. IMF₁ had the highest frequency, and the others gradually decreased in frequency. IMF₁₂ was removed because it had no effect on the predictions. Fig. 7B clearly compares the PM_{2.5} and component prediction performance for the test dataset. According to the expressed frequency and amplitude features, only when PM_{2.5} displayed dramatic and sharp changes did IMF₁ exhibit large fluctuations. Similarly, the other IMFs retained different levels of PM_{2.5} information and fluctuated accordingly. The quantitative results are shown in Table 3. Due to the differentiation of the IMF values, which yielded few high values and many values near zero, the coefficient of determination R^2 is highly credible. The results indicate that the components generally capture approximately 88.4%–99.9% of the explained variance. Therefore, after

Table 3

Comparison of the prediction performance of these 13 observation stations average hourly PM_{2.5} concentrations using four indicators: the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2).

	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅	IMF ₆	IMF ₇
RMSE ($\mu\text{g}\cdot\text{m}^{-3}$)	5.63	2.49	1.25	1.06	1.39	1.75	1.14
MAE ($\mu\text{g}\cdot\text{m}^{-3}$)	2.85	1.57	0.90	0.83	1.06	1.34	0.69
MAPE (%)	202.98	171.95	65.91	34.79	28.95	18.83	21.97
R^2	0.319	0.884	0.982	0.996	0.992	0.999	0.998
	IMF ₈	IMF ₉	IMF ₁₀	IMF ₁₁	IMF ₁₂	IMF ₁₃	Residue
RMSE ($\mu\text{g}\cdot\text{m}^{-3}$)	2.41	0.74	0.21	0.18	NaN	0.85	0.41
MAE ($\mu\text{g}\cdot\text{m}^{-3}$)	2.06	0.66	0.15	0.17	NaN	0.83	0.40
MAPE (%)	40.24	21.42	0.47	11.41	NaN	92.38	0.95
R^2	0.994	0.998	0.999	0.999	NaN	0.999	0.999

Table 4

Comparison of the performance of the EMD-BiLSTM model and LSTM model prediction results with restructured multi-hour average datasets of 13 observation stations using four indicators: the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2).

		EMD-BiLSTM				LSTM			
		RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2
		($\mu\text{g}\cdot\text{m}^{-3}$)	($\mu\text{g}\cdot\text{m}^{-3}$)	(%)		($\mu\text{g}\cdot\text{m}^{-3}$)	($\mu\text{g}\cdot\text{m}^{-3}$)	(%)	
3h	Min	16.71	10.57	23.58	0.974	35.66	19.35	50.58	0.835
	Max	20.41	12.18	27.69	0.947	44.21	24.75	54.64	0.808
	Average	17.48	10.93	25.94	0.965	34.25	19.71	45.28	0.862
6h	Min	23.93	15.23	62.47	0.911	47.67	26.14	99.07	0.642
	Max	33.06	19.48	42.15	0.905	58.68	36.12	73.64	0.703
	Average	25.32	16.53	43.22	0.921	48.03	29.28	69.64	0.724
12h	Min	27.73	16.03	83.47	0.702	44.23	26.46	177.52	0.247
	Max	35.24	21.79	66.35	0.855	66.08	43.41	94.59	0.463
	Average	27.01	16.89	51.19	0.837	47.71	29.44	88.66	0.429

unsupervised frequency extraction with EMD, for IMF₁, the prediction results are relatively poor (R^2 : 0.319), but other IMFs achieved good prediction performance, thus improving the overall result. Finally, combining all IMF predictions results in relatively accurate PM_{2.5} predictions.

Table 4 presents a quantitative comparison of the performance of the EMD-BiLSTM model and standard LSTM model with reconstructed multi-hour average datasets of 13 observation stations. Daily PM_{2.5} concentrations are usually the average of the 24 hourly PM_{2.5} observations each day, and the above analyses reveal that the EMD-BiLSTM model proposed in this study significantly improves the performance of daily PM_{2.5} predictions. Moreover, error accumulation occurs in multi-step prediction when using a standard LSTM-based model. Therefore, this study proposed a new method for short-term PM_{2.5} predictions by reconstructing hourly PM_{2.5} datasets to multi-hour average datasets, including average, maximum and minimum concentrations. The results demonstrated the excellent prediction performance of EMD-BiLSTM. The RMSE, MAE and R_2 values were within acceptable ranges, even as the time interval decreased, and the values were better than those of the standard LSTM model, especially for the R_2 value. Some unusual MAPE values were observed, and they were likely due to the uneven numerical distribution of the PM_{2.5} data sets, which included many low values that resulted in a large absolute percentage error. These four indicators suggest that the EMD-BiLSTM model outperforms the standard LSTM model.

Overall, comparing the hourly, daily and multi-hour predictions, it is not difficult to conclude that we can obtain a series of multi-hour PM_{2.5} predictions from hourly to daily, including average, maximum and minimum concentrations, with the proposed approach. Moreover, the coefficient of determination of the predictions remains above 0.70, indicating that at least 70% of the explained variance is captured by the obtained multiscale predictions.

4. Conclusions

This study proposed a semi-supervised model named EMD-BiLSTM to predict PM_{2.5} concentrations. This approach includes EMD and a BiLSTM NN and only requires PM_{2.5} time-series data sets as inputs, thus the model structure is simpler than other approaches such as SVM or random forest, but still provides accurate PM_{2.5} predictions. The main contributions of this study are as follows.

- The EMD-BiLSTM model considers hidden temporally variable factors that influence the results, and limited auxiliary data are required, thus simplifying the model structure and reducing the number of calculation while achieving high accuracy and stability in PM_{2.5} predictions.
- PM_{2.5} time-series data were regarded as signal data through EMD to extract frequency features through unsupervised feature learning. This method can considerably improve the performance of short-term trend prediction, especially for sudden changes. Regarding to the high-frequency and irregular changes in time-series data, the prediction ability of the supervised LSTM model was improved by introducing unsupervised feature learning.
- Hourly PM_{2.5} data were reconstructed to form a series of multi-hour PM_{2.5} datasets, including averages and maximum and minimum concentrations. This approach can reduce error accumulation in the multi-step predictions of LSTM-based models. Thus, a credible series of multi-hour PM_{2.5} predictions can be obtained. EMD-BiLSTM yields high accuracy and stability for PM_{2.5} predictions at different temporal scales.

Funding

This research was funded by the National Natural Science Foundation of China, grant numbers 41971397, 61731022, 41701468, U1711266.

Author contributions

Conceptualization, Luo Zhang; Methodology, Luo Zhang and Peng Liu; Formal Analysis, Luo Zhang, Guizhou Wang and Peng Liu; Writing-Original Draft Preparation, Luo Zhang; Writing-Review & Editing, Peng Liu and Jianbo Liu; Experiment, Luo Zhang, Peng Liu, Guizhou Wang, Wanfeng Zhang; Project Administration, Peng Liu.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the China Environmental Monitoring Centre (<http://www.cnemc.cn>) for providing the data used in this study.

References

- Bai, Y., Zeng, B., Li, C., Zhang, J., 2019. An ensemble long short-term memory neural network for hourly pm_{2.5} concentration forecasting. Chemosphere 222, 286–294. <https://doi.org/10.1016/j.chemosphere.2019.01.121>.
- Brook, R.D., Rajagopalan, S., Pope, C.A., Brook, J.R., Bhatnagar, A., Diez-Roux, A.V., et al., 2010. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the american heart association. Circulation 121, 2331–2378. <https://doi.org/10.1161/CIR.0b013e3181dbece1>.
- Bruneekreef, B., Holgate, S.T., 2002. Air pollution and health. Lancet 360, 1233–1242. [https://doi.org/10.1016/S0140-6736\(02\)11274-8](https://doi.org/10.1016/S0140-6736(02)11274-8).
- Byun, D., Schere, K.L., 2006. Review of the governing equations, computational algorithms, and other components of the models-3 community multiscale air quality (cmaq) modeling system. Appl. Mech. Rev. 59, 51. <https://doi.org/10.1115/1.2128636>.
- Chen, F., Yu, B., Li, B., 2017a. A practical trial of landslide detection from single-temporal landsat 8 images using contour-based proposals and random forest: a case study of national Nepal. Landslides 15, 453–464. <https://doi.org/10.1007/s10346-017-0884-x>.
- Chen, F., Zhang, M.M., Tian, B.S., Z, L., 2017b. Extraction of glacial lake outlines in tibet plateau using landsat 8 imagery and google earth engine. IEEE J. Selected Top. Appl. Earth Observ. Rem. Sens. 10, 4002–4009. <https://doi.org/10.1109/JSTARS.2017.2705718>.
- Chen, G., Li, S., Knibbs, L.D., Hamm, N.A.S., Cao, W., Li, T., et al., 2018. A machine learning method to estimate pm_{2.5} concentrations across China with remote sensing, meteorological and land use information. Sci. Total Environ. 636, 52–60. <https://doi.org/10.1016/j.scitotenv.2018.04.251>.
- Cheng, Z., Wang, S., Jiang, J., Fu, Q., Chen, C., Xu, B., et al., 2013. Long-term trend of haze pollution and impact of particulate matter in the yangtze river delta, China. Environ. Pollut. 182, 101–110. <https://doi.org/10.1016/j.envpol.2013.06.043>.
- Dominici, F., Peng, R.D., Bell, M.L., Pham, L., McDermott, A., Zeger, S.L., et al., 2006. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. JAMA 295, 1127. <https://doi.org/10.1001/jama.295.10.1127>.
- Elangasinghe, M.A., Singhal, N., Dirks, K.N., Salmon, J.A., 2014. Development of an annāt-based air pollution forecasting system with explicit knowledge through sensitivity analysis. Atmosph. Pollut. Res. 5, 696–708. <https://doi.org/10.5094/APR.2014.079>.
- Fan, J., Yan, J., Ma, Y., Wang, L., 2018. Big data integration in remote sensing across a distributed metadata-based spatial infrastructure. Rem. Sens. 10, 7. <https://doi.org/10.3390/rs10010007>.
- Fan, W., Qin, K., Cui, Y., Li, D., Bilal, M., 2020. Estimation of hourly ground-level pm_{2.5} concentration based on himawari-8 apparent reflectance. IEEE Trans. Geosci. Rem. Sens. 99, 1–10. <https://doi.org/10.1109/TGRS.2020.2990791>.
- Feng, R., Wang, L., Zhong, Y., 2018. Least angle regression-based constrained sparse unmixing of hyperspectral remote sensing imagery. Rem. Sens. 10, 1546. <https://doi.org/10.3390/rs10101546>.
- Feng, R., Wang, L., Zhong, Y., 2019. Joint local block grouping with noise-adjusted principal component analysis for hyperspectral remote-sensing imagery sparse unmixing. Rem. Sens. 11, 1223. <https://doi.org/10.3390/rs11101223>.
- Feng, R., Zhong, Y., Wang, L., Lin, W., 2017. Rolling guidance based scale-aware spatial sparse unmixing for hyperspectral remote sensing imagery. Rem. Sens. 9, 1218. <https://doi.org/10.3390/rs9121218>.
- Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., Wang, J., 2015. Artificial neural networks forecasting of pm_{2.5} pollution using air mass trajectory based geographic model and wavelet transformation. Atmos. Environ. 107, 118–128. <https://doi.org/10.1016/j.atmosenv.2015.02.030>.
- Giannadaki, D., Giannakis, E., Pozzer, A., Lelieveld, J., 2013. Estimating health and economic benefits of reductions in air pollution from agriculture. Sci. Total Environ. 622, 1304–1316. <https://doi.org/10.1016/j.scitotenv.2017.12.064>.
- Graves, A., Schmidhuber, J., 2005. Framewise phoneme classification with bidirectional lstm and other neural network architectures. Neural Network. 18, 602–610. <https://doi.org/10.1016/j.neunet.2005.06.042>.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Hu, X.F., Belle, J.H., Meng, X., Wildani, A., Waller, L.A., Strickland, M.J., et al., 2017. Estimating pm_{2.5} concentrations in the conterminous United States using the random forest approach. Environ. Sci. Technol. 51, 6936–6944. <https://doi.org/10.1021/acs.est.7b01210>.
- Huang, F., Lu, J., Tao, J., Li, L., Tan, X., Liu, P., 2019. Research on optimization methods of ELM classification algorithm for hyperspectral remote sensing images. IEEE Access 7, 108070–108089. <https://doi.org/10.1109/ACCESS.2019.2932909>.
- Huang, F., Zhu, Q., Zhou, J., Tao, J., Zhou, X., Jin, D., Tan, X., Wang, L., 2017. Research on the parallelization of the DBSCAN clustering algorithm for spatial data mining based on the spark platform. Rem. Sens. 9, 1301. <https://doi.org/10.3390/rs9121301>.
- Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., et al., 2018. Predicting monthly high-resolution pm_{2.5} concentrations with random forest model in the north China plain. Environ. Pollut. 242, 675–683. <https://doi.org/10.1016/j.envpol.2018.07.016>.
- Huang, N.E., Shen, Z., Long, S.R., Wu, M.L.C., Shih, H.H., Zheng, Q.N., et al., 1998. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. Roy. Soc. Lond. A. 454, 903–995. <https://doi.org/10.1098/rspa.1998.0193>.
- Jin, X., Fiore, A.M., Curci, G., Lyapustin, A., Civerolo, K., Ku, M., et al., 2019. Assessing uncertainties of a geophysical approach to estimate surface fine particulate matter distributions from satellite-observed aerosol optical depth. Atmos. Chem. Phys. Discuss. 19 <https://doi.org/10.5194/acp-2018-962>, 295–313.
- Kumar, R., Barth, M.C., Pfister, G.G., Naja, M., Brasseur, G.P., 2014. Wrf-chem simulations of a typical pre-monsoon dust storm in northern India: influences on aerosol optical properties and radiation budget. Atmos. Chem. Phys. 14, 2431–2446. <https://doi.org/10.5194/acp-14-2431-2014>.
- Li, L., Zhang, J., Meng, X., Fang, Y., Ge, Y., Wang, J., et al., 2018. Estimation of pm_{2.5} concentrations at a high spatiotemporal resolution using constrained mixed-effect bagging models with maiac aerosol optical depth. Rem. Sens. Environ. 217, 573–586. <https://doi.org/10.1016/j.rse.2018.09.001>.
- Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., et al., 2017. Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation. Environ. Pollut. 231, 997–1004. <https://doi.org/10.1016/j.envpol.2017.08.114>.
- Lin, C., Li, Y., Yuan, Z., Lau, A.K.H., Li, C., Fung, J.C.H., 2015. Using satellite remote sensing data to estimate the high-resolution distribution of ground-level pm_{2.5}. Rem. Sens. Environ. 156, 117–128. <https://doi.org/10.1016/j.rse.2014.09.015>.
- Lin, Z.Y., F, C., Li, B., Yu, B., Jia, H.C., Zhang, M.M., et al., 2019. A contextual and multitemporal active-fire detection algorithm based on fengyun-2g s-vissr data. IEEE Trans. Geosci. Rem. Sens. 57, 8840–8852. <https://doi.org/10.1109/TGRS.2019.2923248>.
- Lin, Z.Y., F, C., Niu, Z., Li, B., Yu, B., Jia, H.C., et al., 2018. An active fire detection algorithm based on multi-temporal fengyun-3c virr data. Rem. Sens. Environ. 211, 376–387. <https://doi.org/10.1016/j.rse.2018.04.027>.
- Liu, P., Choo, K.K.R., Wang, L., Huang, F., 2017a. Svm or deep learning? a comparative study on remote sensing image classification. Soft Comput 21, 7053–7065.
- Liu, P., Di, L., Du, Q., Wang, L., 2018. Remote sensing big data: theory, methods and applications. Rem. Sens. 10, 711.
- Liu, P., Zhang, H., Eom, K.B., 2017b. Active deep learning for classification of hyperspectral images. IEEE J. Selected Top. Appl. Earth Observ. Rem. Sens. 10, 712–724.
- Qi, Y., Li, Q., Karimian, H., Liu, D., 2019. A hybrid model for spatiotemporal forecasting of pm_{2.5} based on graph convolutional neural network and long short-term memory. Sci. Total Environ. 664, 1–10. <https://doi.org/10.1016/j.scitotenv.2019.01.333>.
- Qin, Q., Lai, X., Zou, J., 2019. Direct multistep wind speed forecasting using lstm neural network combining cemd and fuzzy entropy. Appl. Sci. Basel 9, 126. <https://doi.org/10.3390/app9010126>.
- Shi, C., Yuan, R., Wu, B., Meng, Y., Zhang, H., Zhang, H., et al., 2018. Meteorological conditions conducive to pm_{2.5} pollution in winter 2016/2017 in the western yangtze river delta, China. Sci. Total Environ. 642, 1221–1232. <https://doi.org/10.1016/j.scitotenv.2018.06.137>.
- Shi, X.J., Chen, Z.R., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.C., 2015. Convolutional lstm network: a machine learning approach for precipitation nowcasting. Advances in Neural Information Processing Systems, NEURAL INFORMATION PROCESSING SYSTEMS, p. 28. <https://doi.org/10.1007/978-3-319-21233-3-6>.
- Shrivastava, M., Fast, J., Easter, R., Gustafson, W.I., Hodzic, A., 2011. Modeling organic aerosols in a megacity: comparison of simple and complex representations of the volatility basis set approach. Atmos. Chem. Phys. 11, 6639–6662. <https://doi.org/10.5194/acpd-10-30205-2010>.
- Son, Y., Osornio-Vargas, A.R., O'Neill, M.S., Hystad, P., Texcalac-Sangrador, J.L., Ohman-Strickland, P., et al., 2018. Land use regression models to assess air pollution exposure in Mexico city using finer spatial and temporal input parameters. Sci. Total Environ. 639, 40–48. <https://doi.org/10.1016/j.scitotenv.2018.05.144>.
- Sun, W., Sun, J., 2016. Daily pm_{2.5} concentration prediction based on principal component analysis and lssvm optimized by cuckoo search algorithm. J. Environ. Manag. 188, 144–152. <https://doi.org/10.1016/j.jenvman.2016.12.011>.
- Tai, A.P.K., Mickley, L.J., Jacob, D.J., 2010. Correlations between fine particulate matter (pm_{2.5}) and meteorological variables in the United States: implications for the

- sensitivity of pm2.5 to climate change. *Atmos. Environ.* 44, 3976–3984. <https://doi.org/10.1016/j.atmosenv.2010.06.060>.
- Tong, C.H.M., Yim, S.H.L., Rothenberg, D., Wang, C., Lin, C.Y., Chen, Y.D., et al., 2018. Assessing the impacts of seasonal and vertical atmospheric conditions on air quality over the pearl river delta region. *Atmos. Environ.* 180, 69–78. <https://doi.org/10.1016/j.atmosenv.2018.02.039>.
- Tong, W., Li, L., Zhou, X., Hamilton, A., Zhang, K., 2019. Deep learning pm2.5 concentrations with bidirectional lstm rnn. *Air Qual. Atmosph. Health* 12, 411–423. <https://doi.org/10.1007/s11869-018-0647-4>.
- Wang, L.T., Wei, Z., Yang, J., Zhang, Y., Zhang, F.F., Su, J., et al., 2014. The 2013 severe haze over southern hebei, China: model evaluation, source apportionment, and policy implications. *Atmos. Chem. Phys.* 14, 3151–3173. <https://doi.org/10.5194/acp-14-3151-2014>.
- Wei, J., Li, Z.Q., Cribb, M., Huang, W., Xue, W.H., Sun, L., et al., 2020. Improved 1 km resolution pm2.5 estimates across China using enhanced space-time extremely randomized trees. *Atmos. Chem. Phys.* 20, 3273–3289. <https://doi.org/10.5194/acp-20-3273-2020>.
- Wen, C., Liu, S., Yao, X., Peng, L., Li, X., Hu, Y., et al., 2019. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *Sci. Total Environ.* 654, 1091–1099. <https://doi.org/10.1016/j.scitotenv.2018.11.086>.
- WÄUllmer, M., Eyben, F., Graves, A., Schuller, B., Rigoll, G., 2010. Bidirectional lstm networks for context-sensitive keyword detection in a cognitive virtual agent framework. *Cog. Computat.* 2, 180–190. <https://doi.org/10.1007/s12559-010-9041-8>.
- Wu, Y.X., Wu, Q.B., Zhu, J.Q., 2019. Improved eemd-based crude oil price forecasting using lstm networks. *Phys. Stat. Mech. Appl.* 516, 114–124. <https://doi.org/10.1016/j.physa.2018.09.120>.
- Xiong, T., Bao, Y., Hu, Z., 2014. Does restraining end effect matter in emd-based modeling framework for time series prediction? some experimental evidences. *Neurocomputing* 123, 174–184. <https://doi.org/10.1016/j.neucom.2013.07.004>.
- Yan, J., Ma, Y., Wang, L., Choo, K.R., Jie, W., 2018. A cloud-based remote sensing data production system. *Future Generat. Comput. Syst.* 86, 1154–1166. <https://doi.org/10.1016/j.future.2017.02.044>.
- Yan, J., Wang, L., Song, W., Chen, Y., Chen, X., Deng, Z., 2019. A time-series classification approach based on change detection for rapid land cover mapping. *ISPRS J. Photogrammetry Remote Sens.* 158, 249–262. <https://doi.org/10.1016/j.isprsjprs.2019.10.003>.
- Zhang, H., Zhang, S., Wang, P., Qin, Y., Wang, H., 2017. Forecasting of particulate matter time series using wavelet analysis and wavelet-arma/arima model in taiyuan, China. *J. Air Waste Manag. Assoc.* 67, 776–788. <https://doi.org/10.1080/10962247.2017.1292968>.
- Zhang, Q., Xue, D., Liu, X., Gong, X., Gao, H., 2019. Process analysis of pm2.5 pollution events in a coastal city of China using cmaq. *J. Environ. Sci.* 31, 225–238. <https://doi.org/10.1016/j.jes.2018.09.007>.
- Zhao, J., Deng, F., Cai, Y., Chen, J., 2019. Long short-term memory - fully connected (lstm-fc) neural network for pm2.5 concentration prediction. *Chemosphere* 220, 486–492. <https://doi.org/10.1016/j.chemosphere.2018.12.128>.
- Zhu, H., Lu, X., 2016. The prediction of pm2.5 value based on arma and improved bp neural network model. In: International Conference on Intelligent Networking and Collaborative Systems. IEEE, pp. 515–517. <https://doi.org/10.1109/INCoS.2016.81>.