Chemosphere 308 (2022) 136252

Contents lists available at ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

Extraction of multi-scale features enhances the deep learning-based daily $PM_{2.5}$ forecasting in cities

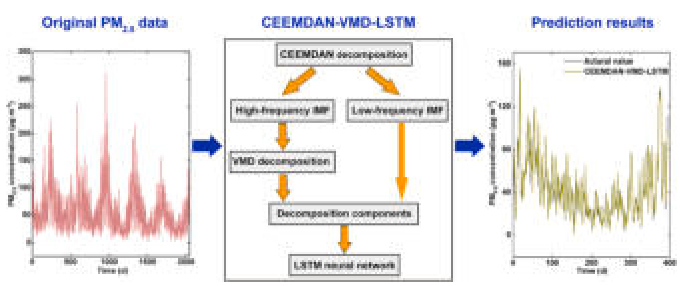
Liang Dong [a], Pei Hua [b,c], Dongwei Gui [e], Jin Zhang [d,e,*]

[a] South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou, 510535, China
[b] SCNU Environmental Research Institute, Guangdong Provincial Key Laboratory of Chemical Pollution and Environmental Safety & MOE Key Laboratory of Theoretical Chemistry of Environment, South China Normal University, 510006, Guangzhou, China
[c] School of Environment, South China Normal University, University Town, 510006, Guangzhou, China
[d] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Yangtze Institute for Conservation and Development, Hohai University, Nanjing, 210098, China
[e] State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, 830011, China

HIGHLIGHTS

- Multi-scale features of original data set were extracted by dual decomposition.
- Decomposition could effectively extract the sequence feature information.
- Secondary decomposition could further improve the model performance.
- Long short-term memory neural networks outperformed other shallow networks.
- Deep learning based on dual decomposition made effective predictions of $PM_{2.5}$

GRAPHICAL ABSTRACT



ARTICLE INFO

Handling Editor: Volker Matthias

Keywords:
Multi-scale features extractions
Deep learning
Hybrid modelling
$PM_{2.5}$
Two-stage decomposition

ABSTRACT

Characterising the daily PM2.5 concentration is crucial for air quality control. To govern the status of the atmospheric environment, a novel hybrid model for PM2.5 forecasting was proposed by introducing a two-stage decomposition technology of complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and variational mode decomposition (VMD); subsequently, a deep learning approach of long short-term memory (LSTM) was proposed. Five cities with unique meteorological and economic characteristics were selected to assess the predictive ability of the proposed model. The results revealed that PM2.5 pollution was generally more severe in inland cities (66.98 ± 0.76 μg m^{-3}) than in coastal cities (40.46 ± 0.40 μg m-3). The modelling comparison showed that in each city, the secondary decomposition algorithm improved the accuracy and prediction stability of the prediction models. When compared with other prediction models, LSTM effectively extracted featured information and achieved relatively accurate time-series prediction. The hybrid model of CEEMDAN-VMD-LSTM achieved a better prediction in the five cities ($R2$ = 0.9803 ± 0.01) compared with the benchmark models ($R2$ = 0.7537 ± 0.03). The results indicate that the proposed approach can identify the inherent correlations and patterns among complex datasets, particularly in time-series analysis.

* Corresponding author. State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Yangtze Institute for Conservation and Development, Hohai University, Nanjing, 210098, China.
E-mail address: jin.zhang@hhu.edu.cn (J. Zhang).

https://doi.org/10.1016/j.chemosphere.2022.136252
Received 30 March 2022; Received in revised form 14 July 2022; Accepted 26 August 2022
Available online 30 August 2022

1. Introduction

Air pollution caused by $PM_{2.5}$ (particles with aerodynamic diameter <2.5 μm) has been considered a critical problem threatening human health. These particles may trigger a variety of diseases, including asthma, nasal diseases, and cardiovascular diseases (Chen et al., 2020; Gong et al., 2019; Loftus et al., 2015; Sugiyama et al., 2020). According to recent studies, ambient $PM_{2.5}$, the fifth-ranked risk factor for global death, caused the premature death of 4.2 million people worldwide in 2015, thereby accounting for 7.6% of the total global deaths (Cohen et al., 2017; Zhang et al., 2020b). Under the influence of an increased exposure rate, population growth, and aging, it was estimated that 1.1 million people in China died prematurely due to $PM_{2.5}$-related diseases in 2015 (Cohen et al., 2017). In this regard, timely and accurate prediction of $PM_{2.5}$ concentration helps decision-makers implement effective early warning activities, which can help the public effectively arrange travel times and type of transportation, thereby reducing the impact of $PM_{2.5}$ on their daily lives (Huang et al., 2021).

In terms of modelling approaches, numerous deterministic and statistical methods exist for air pollution prediction. Deterministic methods based on theoretical meteorological emissions and chemical models can effectively simulate the distribution and dispersion of pollutants (Chen and Li, 2021; Hong et al., 2020). However, the model-building process of deterministic methods requires an appropriate application area, comprehensive emission data, sufficient information regarding the physicochemical processes of pollutants, and numerical calculations (Samal et al., 2021). Thus, deterministic methods for $PM_{2.5}$ predictive modelling are complex and time-consuming, and the generality and accuracy of the obtained results may be lower than those of other methods (Lightstone et al., 2017; Masood and Ahmad, 2021). In contrast, statistical model-based data mining methods that aim to develop dependencies between input variables and air pollutant concentration, avoid complex modelling processes; thus, they can obtain accurate prediction results (Goudarzi et al., 2021).

With the development of artificial intelligence, artificial neural networks (ANNs) have become an effective and popular nonlinear statistical technology for the prediction of $PM_{2.5}$ concentration and they have been shown to have a high prediction accuracy (Teng et al., 2022). Some examples are back propagation (BP) neural networks (Liu et al., 2019), radial basis function (RBF) neural networks (Lu et al., 2004), general regression neural networks (GRNN) (Wang et al., 2019), and recurrent neural networks (RNNs) (Biancofiore et al., 2017). As an improved version of RNNs, long short-term memory (LSTM) introduces memory cells to explore the inherent abstract features and constant structure of time series (Boulila et al., 2021). Thus, LSTM has a strong ability to extract features from historical data and yield excellent results. With the development of deep learning technology in recent years, LSTM has been extensively used in the prediction of air pollutants, and has been shown to have clear advantages in dealing with time series (Aggarwal and Toshniwal, 2021; Menares et al., 2021).

In addition, accurate and effective prediction of $PM_{2.5}$ remains a substantial challenge owing to the limited understanding of the dynamic processes in the atmospheric environment. Previous studies have shown that an appropriate signal processing method is an effective way to extract hidden dynamic information from pollutant sequences and improve the prediction accuracy of the models. Qiao et al. (2019) in their study established a novel $PM_{2.5}$ prediction method using wavelet transform (WT), stacked autoencoder (SAE), and LSTM. The case study showed that the proposed model had a higher prediction accuracy than that of the SAE-LSTM model at five different study sites. The basic idea behind the decomposition-based hybrid model is to decompose time-series data into several components characterised by more linear and smoother trends, thus making model prediction easier (Liu et al., 2020a; Sharma et al., 2020).

However, owing to their irregularity and non-stationarity (Niu et al., 2016), original $PM_{2.5}$ data cannot be determined completely using a single decomposition technique (Wang et al., 2017), and the high-frequency characteristics extracted by the single decomposition may affect the overall prediction performance of the model. Thus, further secondary decomposition is necessary, and this technique has achieved good results in a series of fields, such as wind speed forecasting (Sun et al., 2021), temperature forecasting (Xu and Ren, 2019), and streamflow forecasting (Wang et al., 2021b). However, few studies have applied two-stage decomposition technology to air pollutant concentration prediction, especially in cities with unique meteorological and economic characteristics.

Consequently, with the aim of targeting a better prediction performance, a $PM_{2.5}$ concentration prediction model based on two-stage decomposition technology and a deep learning method was developed in this study. The main aims of this study were: (1) to explore the ability of CEEMDAN-VMD two-stage decomposition to identify and separate tendencies, harmonic components, and irregular components of raw data, as well as determine the potential of LSTM to process the dynamic characteristics of $PM_{2.5}$, (2) integrate CEEMDAN-VMD and LSTM to construct a hybrid model for $PM_{2.5}$, and (3) corroborate the performance of the CEEMDAN-VMD-LSTM models with different performance measures.

2. Materials and methods

2.1. Study area and data

Severe air pollution events in China are highly concentrated in four separate regions: the North China Plain, Yangtze River Delta, Pearl River Delta, and Sichuan Basin. Owing to the unique meteorological and geographical conditions in Northwest China (near the desert with four distinct seasons), $PM_{2.5}$ has become the main air pollutant in the area. Therefore, five representative cities were selected as study areas: Urumqi (Northwest China), Guangzhou (Pearl River Delta), Chengdu (Sichuan Basin), Shanghai (Yangtze River Delta), and Shijiazhuang (North China Plain) (Fig. S1). Detailed descriptions of the economic conditions, geographic environment, and $PM_{2.5}$ pollution of the five cities are presented in **Text S1** in the Supporting Information. The geographical location, climate conditions, industrial structure, and economic scale of these five cities differ, resulting in different pollutant concentrations and air quality. A comprehensive study of these five cities will help rationally verify the practicability of the model and obtain more comprehensive and reliable experimental conclusions.

The daily $PM_{2.5}$ data series for the given cities from December 2, 2013 to October 31, 2019 were collected from five air quality monitoring stations set up by the China National Environmental Monitoring Centre (CNEMC) (www.cnemc.cn, accessed December 2021). $PM_{2.5}$ data for each city were derived from measurements from the samplers based on the CNEMC reference method. The time interval for collecting all $PM_{2.5}$ data was one day. For the given cities, the amount of missing data for various reasons, such as instrument damage, was less than 5% of the total recording period. Therefore, missing data did not affect the prediction performance of the model. In each city, the first 80% of $_{\text{the}}$ data were used for model training, and the last 20% were used to verify the performance of the development model, similar to previous research (Wang et al., 2021a).

2.2. CEEMDAN decomposition technique

As an evolution of empirical mode decomposition (EMD) (Huang et al., 1998), the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) proposed by Torres et al. (2011) is an adaptive data analysis method. Original data can be decomposed into a series of intrinsic mode functions (IMFs). By adding adaptive white noise in each of the decomposition stages, CEEMDAN solves the mode-mixing problem in EMD, providing better separation of modes and accurate reconstruction of the original signal with lower computing costs. In this study, the standard deviations of Gaussian noise and the iterations were set to 0.2 and 200, respectively. The detailed implementation steps are presented in **Text S2** of the Supporting Information.

2.3. VMD decomposition technique

Variational mode decomposition (VMD) is an adaptive, quasi-orthogonal, and completely nonrecursive signal decomposition method proposed by Dragomiretskiy and Zosso (2013). In essence, VMD adaptively transfers the signal decomposition process into the variational model and then realises the effective separation of components with different frequencies by iteratively searching for the optimal solution of the variational model. The details are provided in **Text S3** of the Supporting Information.

2.4. Tree-based feature selection

Random forest (RF) is an integrated machine learning method that can be used as a feature selection tool for high-dimensional data to analyse complex features and the importance of measurement variables (Liu et al., 2020b; Yang et al., 2020). Based on predictive performance degradation when the values of descriptive variables in the nodes of the tree are randomly arranged, RF can provide information about the importance of each input variable when calculating the values of the output variables (Wang et al., 2021a). Based on the variable importance measure, RF can be used to discard insignificant input variables (Fang et al., 2021; Rodriguez-Galiano et al., 2018). The detailed working mechanism of the RF is presented in **Text S4** in the Supporting Information.

2.5. LSTM deep learning technique

LSTM was introduced as a deep neural network to excavate the dynamic characteristics of time series and make effective predictions (Hochreiter and Schmidhuber, 1997). The LSTM unit mainly includes a memory cell and three types of gates (input, output, and forget gates) that control the state change in the memory unit (Fig. S2). With the help of the gate units and memory cells, the LSTM unit can control long-time information and establish long-time delays between input and feedback (Ma et al., 2019). The working mechanism of LSTM is shown in **Text S5** in the Supporting Information.

2.6. CEEMDAN-VMD-LSTM hybrid model

To achieve effective $PM_{2.5}$ prediction, we have proposed a novel hybrid model that integrates CEEMDAN-VMD secondary decomposition and LSTM neural networks. The detailed procedures of the CEEMDAN-VMD-LSTM hybrid model employed in this study are as follows:

Step 1. CEEMDAN was employed for the original $PM_{2.5}$ series into n IMFs with different frequencies and a residual component R.

Step 2. VMD was utilised to further decompose high-frequency IMF1 into m components (VMs) with clearer inherent characteristics.

Step 3. To improve the training speed and convergence of the model and avoid the adverse effects caused by singular sample data, the input data were normalised to a specific range [0, 1] before the modelling process began using Eq. (1).

$$x_{normalized} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Step 4. For each subseries, RF was applied to select the most important lags as a suitable input for the LSTM.

Step 5. An LSTM neural network model was established to predict each mode obtained by decomposition, including n IMF components, a residual component R, and m VM components.

Step 6. The predictions of all the modes were denormalized and summarised to obtain the final forecasting results for $PM_{2.5}$.

2.7. Performance evaluation

Appropriate model evaluation criteria are the premise for a reasonable evaluation and prediction. The mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) reflect the error and accuracy of the prediction model and they are commonly used evaluation indicators (Ding et al., 2020; Li et al., 2020). The standard deviation of the error (SDE) and Diebold Mariano (DM) tests were used to evaluate the prediction stability of the developed model. Detailed evaluation criteria are provided in **Text S6** of the Supporting Information.

To evaluate the effectiveness of the proposed model and the feasibility of the data processing method, we applied 14 p.m.$_{2.5}$ prediction models as comparison models, which were mainly composed of six classic models (the naive model, autoregressive integrated moving average (ARIMA), support vector regression (SVR), BP, extreme learning machine (ELM), and LSTM) and four typical decomposition algorithms (EMD, ensemble empirical mode decomposition (EEMD), VMD, CEEMDAN, and CEEMDAN-VMD). In this study, we continuously optimised the model by adjusting the parameters until we obtained the best-fit model using the training set; finally, we determined the optimal hyperparameters for each model. The parameter settings for the different algorithms are listed in Table S1. All simulation experiments were carried out on the MATLAB R2019b platform on a personal computer with an Intel(R) Core(TM) i7-9750H CPU @ 2.60 GHz 2.59 GHz and 8.00 GB RAM.

3. Results

3.1. Characteristics of $PM_{2.5}$ in different cities

During the study period, the average annual values of $PM_{2.5}$ in Urumqi, Guangzhou, Chengdu, Shanghai, and Shijiazhuang were 62.04 ± 1.32, 36.98 ± 0.47, 55.19 ± 0.83, 43.69 ± 0.63, and 84.05 ± 1.59 μg m^{-3}, respectively (Table S2). These results clearly showed that $PM_{2.5}$ pollution was more severe in inland cities (i.e., Urumqi, Chengdu, and Shijiazhuang, 66.98 ± 0.76 μg m^{-3}) than in coastal cities (i.e., Guangzhou and Shanghai, 40.46 ± 0.40 μg m-3). The daily $PM_{2.5}$ data series from the five study cities are illustrated in Fig. 1. We noticed that the $PM_{2.5}$ values in these five cities showed periodic fluctuations to varying degrees (Fig. 1). The average concentration in the five cities in winter (92.31 ± 1.39 μg m^{-3}) was significantly higher than that in spring (49.17 ± 0.64 μg m^{-3}), summer (34.14 ± 0.42 μg m^{-3}), and autumn (49.76 ± 0.82 μg m^{-3}). Analysis of variance ($p < 0.010$) showed that season had a significant effect on $PM_{2.5}$ concentration in the atmospheric environment (Table S3).

The seasonal variation and other multi-scale characteristics of $PM_{2.5}$ in the five typical regions were further analysed by continuous wavelet analysis. Fig. 2 shows the power spectrum of a wavelet transform of $PM_{2.5}$ data in the five study cities. It was clear that the $PM_{2.5}$ sequences in these study cities had significant wavelet power at a 1-year time scale in most cases, which revealed a strong annual signal. However, $PM_{2.5}$ had some regional differences in relatively minor periods. In addition to the above annual oscillations, $PM_{2.5}$ in Urumqi, Chengdu, and Shijiazhuang exhibited a semi-annual oscillation from day 400 to day 1900. For these longer time scales, the $PM_{2.5}$ concentration was mainly affected by seasonal emission periods and long-term socioeconomic factors (Zhong et al., 2018). On the contrary, it could be seen that $PM_{2.5}$ in Shanghai and Shijiazhuang had a series of quasi-unit weekly oscillations of 4–8 days, which fall within the confidence interval based on the

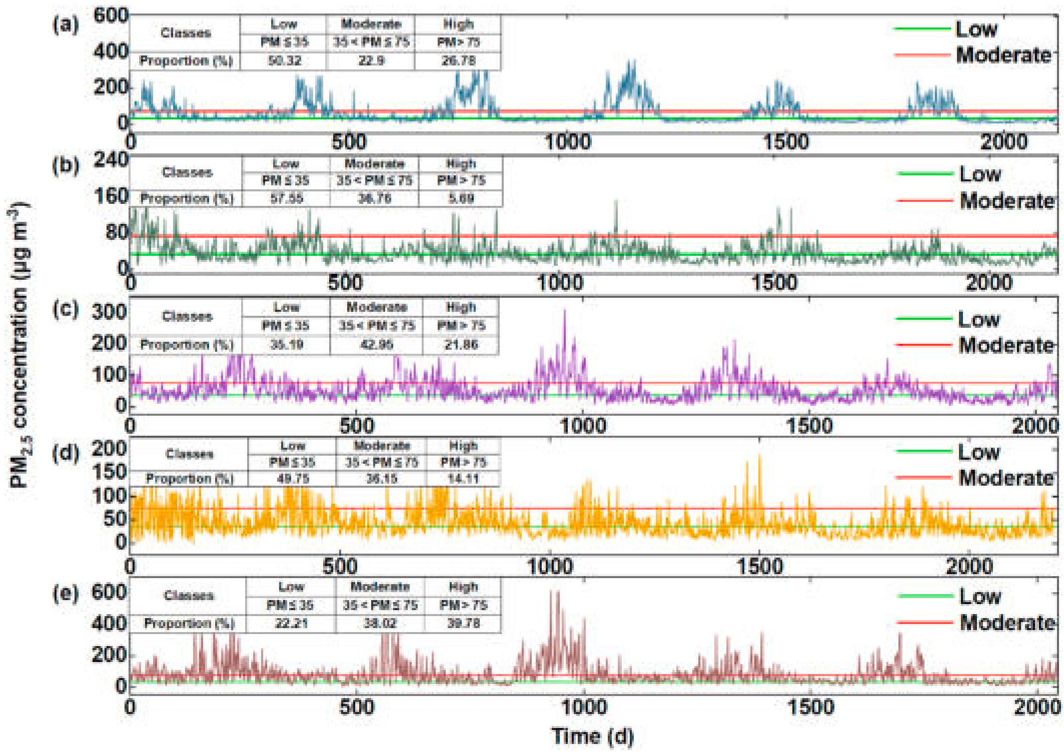


Fig. 1. $PM_{2.5}$ concentration of (a) Urumqi, (b) Guangzhou, (c) Chengdu, (d) Shanghai, and (e) Shijiazhuang (PM represents daily $PM_{2.5}$ concentration in the unit of μg m^3; Proportion represents the ratio of $PM_{2.5}$ days to total days at each level).

wave power spectrum and global wavelet spectrum. These shorter-period characteristics are closely related to low-frequency oscillations in the atmosphere and human activities. Therefore, it is challenging to characterise the variability of $PM_{2.5}$ in the five cities due to their diverse natural and anthropogenic impacts. To overcome the difficulty of coupling with complex information and to improve the feasibility of forecasting, pre-processing of the original data was essential.

3.2. Data decomposition and input selection

As there were many cities and methods studied in this study, the data processing process of daily $PM_{2.5}$ data in Urumqi was introduced as an example (Fig. 3). Pre-processing results for $PM_{2.5}$ data from other cities are shown in Fig. S3 to S6.

The original $PM_{2.5}$ sequence was first decomposed into a series of sub-signals after using the CEEMDAN decomposition technology. After the decomposition, the $PM_{2.5}$ data were changed into components with more obvious characteristics listed in the order from the highest to the lowest frequencies (Fig. 3a). The high-frequency IMF reflected random noise and irregular influencing factors in the $PM_{2.5}$. The lower frequency IMF and residual showed the periodic factors or long-term trends of the original series, respectively.

VMD was applied to further decompose high-frequency IMF, in which the normal distance (ND) was the criterion of the pre-set number of decompositions K. It can be seen that ND remains relatively stable when K increases to a certain extent (Fig. S7), which indicates that the appropriate decomposition number determined by ND can effectively extract the hidden features in the original data and therefore avoid the cumulative estimation error caused by too many intrinsic patterns. Fig. 3b shows the decomposed sequences of the IMF1 data using the VMD method.

RF was adopted to evaluate the importance of the lags of the original $PM_{2.5}$ sequence and its subsequence. Fig. 3 shows the RF ranking order for each time series for Urumqi. Owing to redundancy and correlation, more input features may affect the prediction accuracy and increase operating costs. Thus, the most relevant historical data were selected as the input variables for sequences that showed significant importance at a certain time. For the other sequences, five important historical data of the time series based on RF sorting were selected as the input variables of the prediction models.

3.3. Comparison with the classic individual models

In this study, the prediction performances of some classic individual models of the naive model, ARIMA, SVR, BP, ELM, and LSTM, were compared with the proposed hybrid model on the same test set. The prediction performance of $PM_{2.5}$ for the different models for the five cities, is shown in Fig. 4 and Table S4.

Although a good prediction ability of the individual models was shown in Urumqi (total $R^2 = 0.8079$) and Chengdu (total $R^2 = 0.6697$), the relatively low prediction accuracy in Guangzhou (total $R^2 = 0.4764$), Shijiazhuang (total $R^2 = 0.5774$), and Shanghai (total $R^2 = 0.2421$) indicated that individual models had poor recognition and analysis ability for information and learning patterns and could not capture the dynamic characteristics of $PM_{2.5}$ concentration series in different environments.

Taking Urumqi as an example, the proposed CEEMDAN-VMD-LSTM model had the smallest error (MAE = 2.606 and RMSE = 3.697) and highest accuracy ($R^2 = 0.9942$). When compared with the individual models, the MAE and RMSE of CEEMDAN-VMD-LSTM decreased by 80% and 83%, respectively, and R^2 increased by 23%. The developed model exhibited prominent advantages over traditional individual models. Moreover, in all of the urban environments, the prediction results of CEEMDAN-VMD-LSTM were satisfactory. Individual models can no longer meet the current requirements of high prediction accuracy and generalisation for air pollutants.

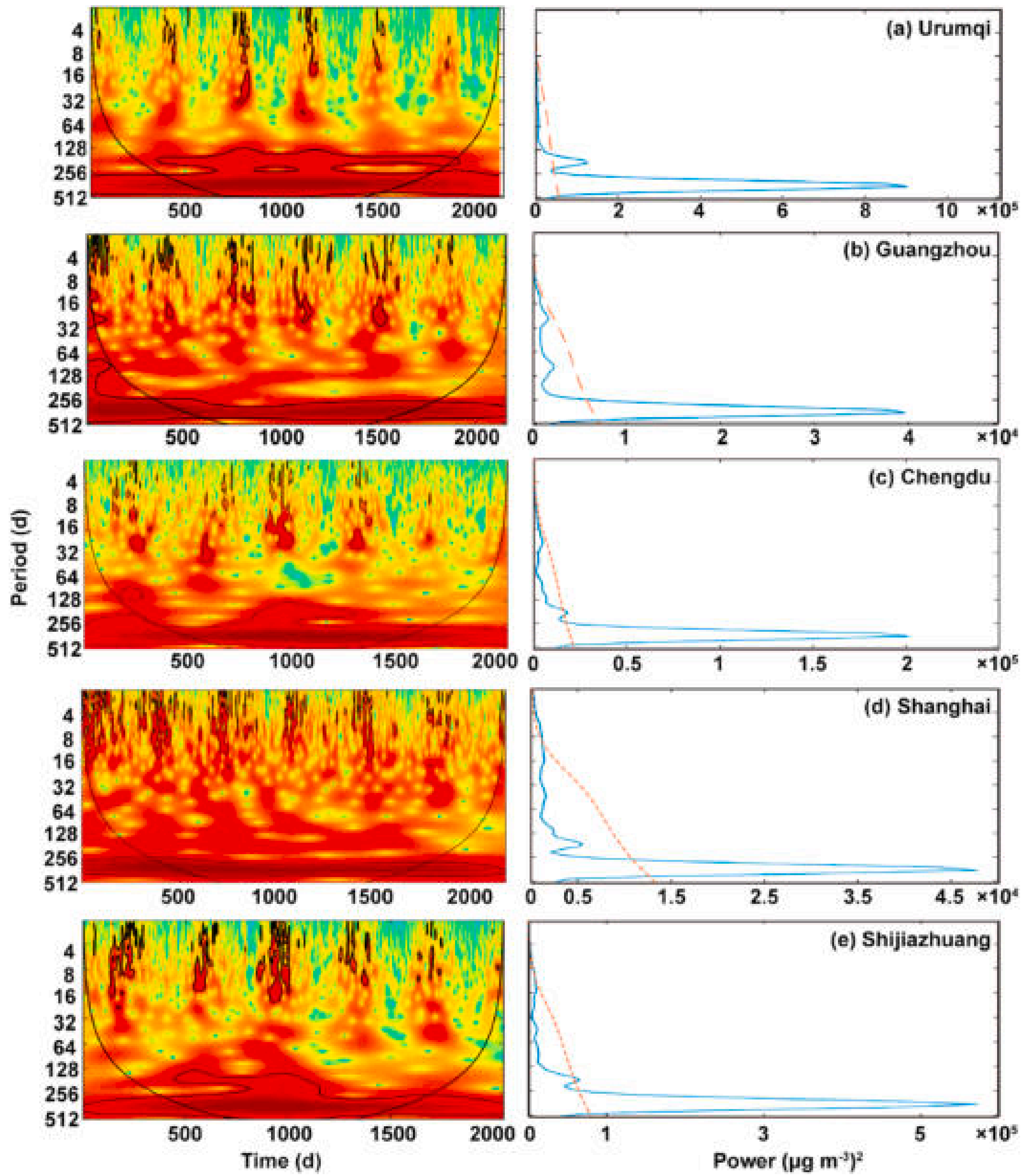


Fig. 2. Wavelet power spectra (left column) and the global wavelet power spectrum (right column) of daily $PM_{2.5}$ concentrations (μg m^{-3}) for (a) Urumqi, (b) Guangzhou, (c) Chengdu, (d) Shanghai, and (e) Shijiazhuang.

3.4. Comparison between the hybrid models with different decomposition techniques

In this section, we compare the performance differences between the proposed secondary decomposition and the commonly used single decomposition methods EMD, EEMD, CEEMDAN, and VMD. A comparison is presented in Fig. 5 and Table S5.

When compared with models without signal decomposition, all of the models combined with decomposition technology performed better with relatively small prediction errors. For example, in the Urumqi dataset, CEEMDAN-LSTM surpassed LSTM in three indicators, MAE, RMSE, and R^2, improving by 58%, 59%, and 18%, respectively. Furthermore, CEEMDAN and VMD showed better dissecting results because of the mode mixing in EMD and the inability of EEMD to eliminate the added white noise. However, no single decomposition technique always performed best in the dynamic feature extraction for $PM_{2.5}$ in all of the cities. In general, individual models might be detrimental to extracting useful information from pollutants, and inappropriate decomposition techniques might lead to elusive or irrelevant components, which might increase the prediction challenges.

Among all of the decomposition-based prediction models, the model combined with secondary decomposition performed significantly better in $PM_{2.5}$ prediction for all of the cities, with the total MAE, RMSE, and R^2 values of 3.063, 5.178, and 0.9818, respectively. When compared with the single decomposition-based model, the average improvement rates of R^2 of the secondary decomposition-based model in Urumqi, Guangzhou, Shanghai, Chengdu, and Shijiazhuang were 7%, 25%, 27%, 8%, and 11%, respectively. This indicates that the composite structure of the secondary decomposition and deep learning was quite effective, despite the differences in the improvement results due to the different dynamics in $PM_{2.5}$ in different environments.

3.5. Comparison between the hybrid models with different predictors

This study tested the effectiveness of a secondary decomposition combined with a classical data-driven predictor. ARIMA, SVR, BP, and

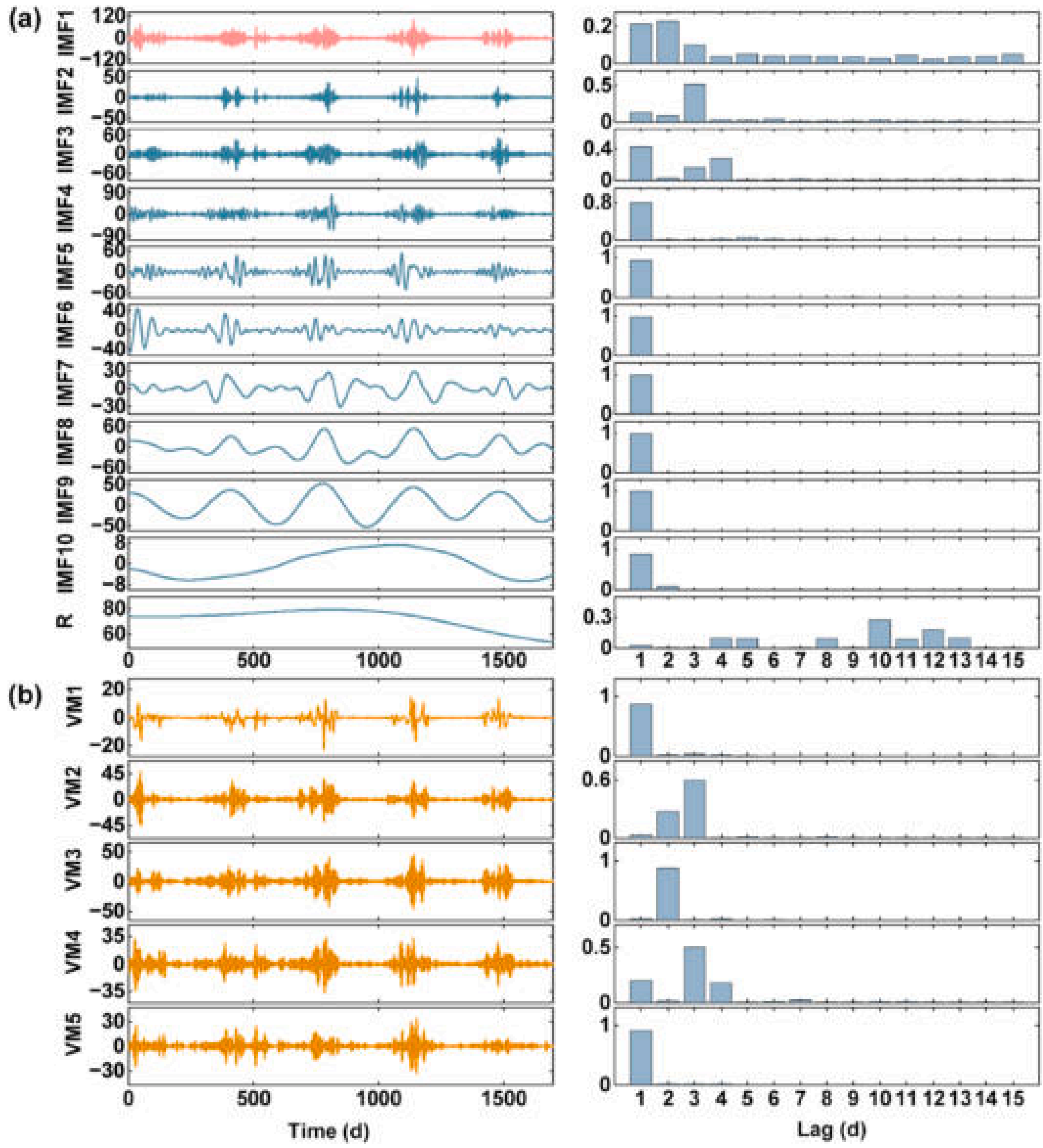


Fig. 3. (a) CEEMDAN decomposition results, (b) VMD decomposition results, and the importance of each lag in corresponding decomposition sequences for Urumqi.

ELM were introduced as predictors instead of LSTM in the integrated CEEMDAN-VMD-LSTM model, and their prediction performances are shown in Fig. 6 and Table S6.

Based on secondary decomposition, the performance of the three types of predictors conformed to the following order: deep learning technology (LSTM) > machine learning models (SVR, BP, and ELM) > empirical model (ARIMA). ARIMA predicted $PM_{2.5}$ with a high R^2 (total $R^2 = 0.9357$), indicating that some linear dynamic components of $PM_{2.5}$ obtained by secondary decomposition. The total R^2 of the machine learning models reached 0.9642, which was significantly higher than that of the ARIMA model. This result was reasonable because there were nonlinear components in the dynamics of $PM_{2.5}$, which were difficult for ARIMA to handle. Machine learning models have strong nonlinear and linear mapping capabilities. When compared with the traditional machine learning models, the R^2 of LSTM increased by an average of 3%, whereas the MAE and RMSE decreased by 29% and 25%, respectively. It could be seen that the prediction performance of LSTM was significantly higher than that of traditional machine learning due to its unique structure and the ability to process time series.

3.6. $PM_{2.5}$ prediction performance in the different cities

The CEEMDAN-VMD-LSTM model achieved the best performance when compared to the other 14 prediction models (Figs. 4–6). We drew predictions for the five urban areas to fully demonstrate the validation results (Fig. 7). It can be seen that the prediction performance of CEEMDAN-VMD-LSTM varies from region to region. R^2 in Urumqi was the highest (0.9942), whereas R^2 in Shanghai was the lowest of the five studied areas (0.9659), but it still performed very well. The average R^2 of the proposed model in the five regions was 0.9803 ± 0.01, while the average R^2 of the benchmark models was 0.7537 ± 0.03, indicating that our model was suitable for $PM_{2.5}$ prediction in China. In addition, the p-value of all DM tests based on the mean square error function was less than 0.01, which could indicate that the difference probability was 99%, indicating that CEEMDAN-VMD-LSTM was superior to the other methods (Table S7).

The forecasting accuracy of the developed $PM_{2.5}$ prediction method was demonstrated; however, for pollution data with a high fluctuation, the prediction model also requires low uncertainty to maintain the high prediction accuracy (Yu et al., 2022). In this study, SDE was used to assess the forecasting stability of the developed method. Among the five

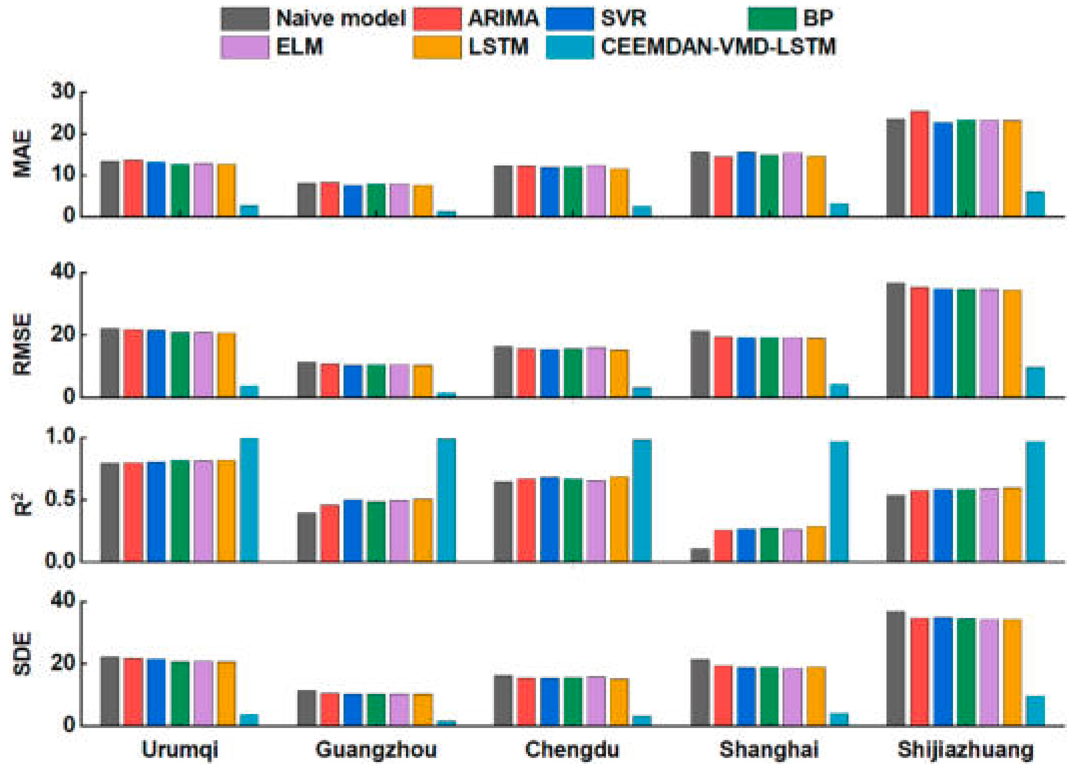


Fig. 4. Comparison of the forecasting performance of the proposed model with classic individual models.

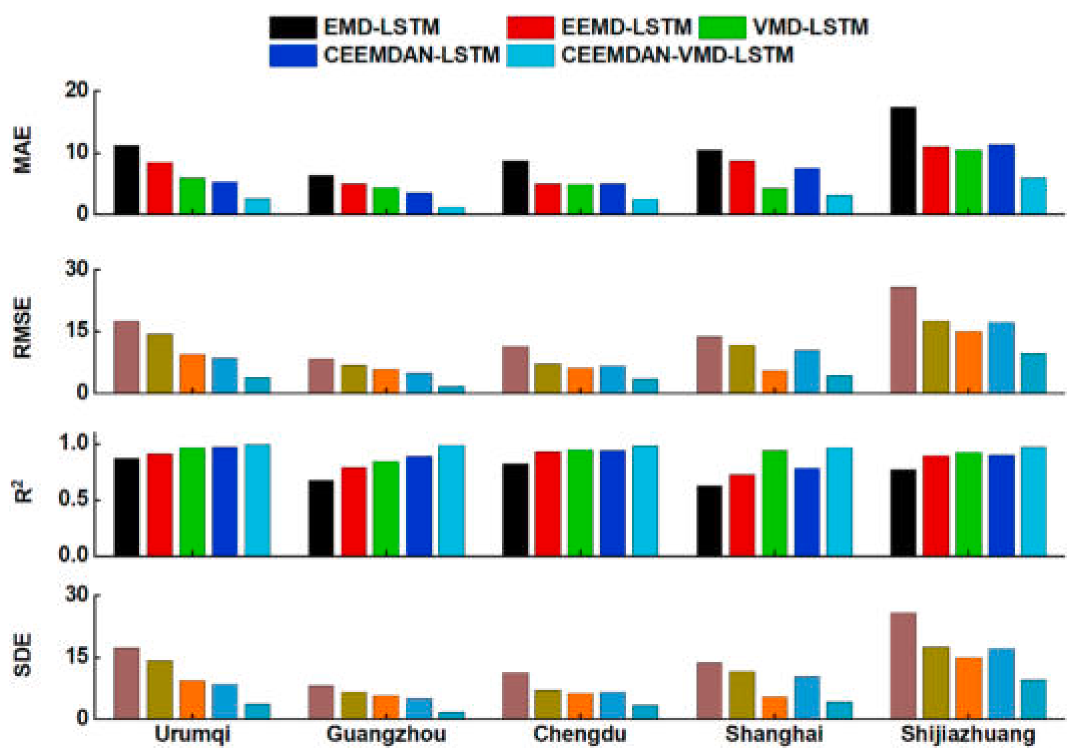


Fig. 5. Comparison of the forecasting performance of the proposed model with hybrid models with different decomposition techniques.

cities of Urumqi, Guangzhou, Chengdu, Shanghai, and Shijiazhuang, the SDE predicted by CEEMDAN-VMD-LSTM was the lowest in all models (3.695, 1.567, 3.325, 4.144, and 9.622, respectively). This indicates that the proposed model has a lower dispersion of the prediction error and higher stability (Zhang et al., 2020a).

The prediction results of different cities showed that CEEMDAN-VMD-LSTM could adapt well to the $PM_{2.5}$ characteristics of the different regions, resulting in superior prediction accuracy and prediction stability compared to those obtained using the benchmark methods.

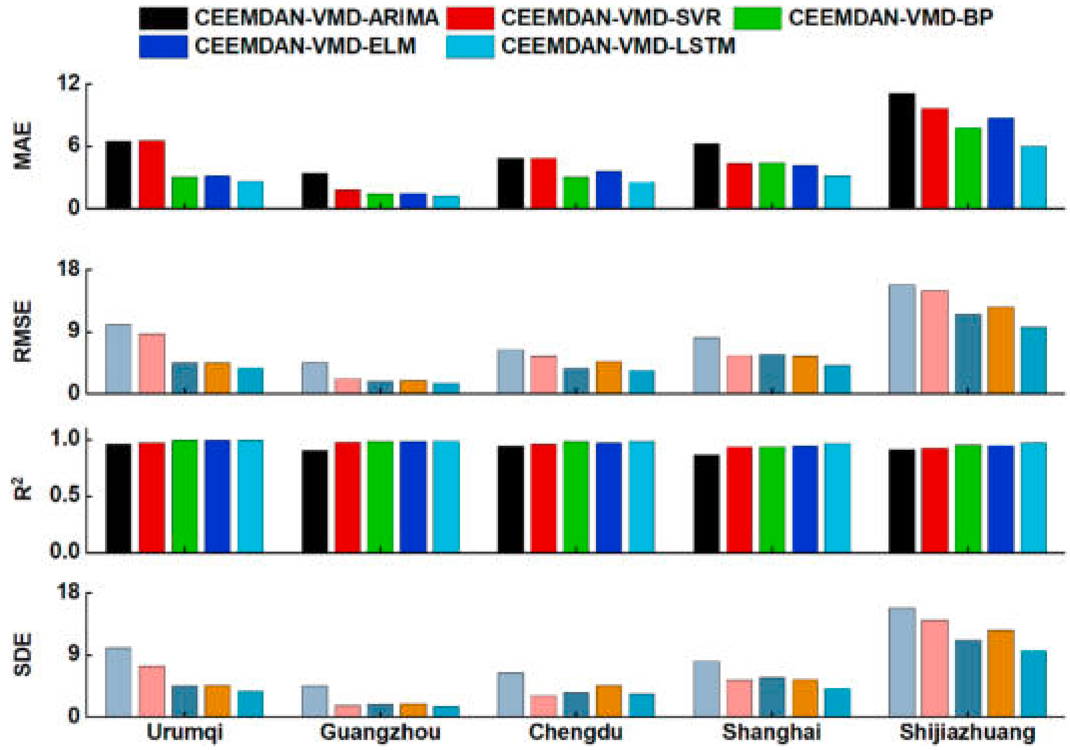


Fig. 6. Comparison of the forecasting performance of the proposed model with hybrid models with different predictor.

4. Discussion

4.1. Secondary decomposition-based deep learning guarantees the prediction accuracy

Based on the successful combination of secondary decomposition and deep learning algorithms, the CEEMDAN-VMD-LSTM performed best among all $PM_{2.5}$ prediction models used in this study.

Natural and social environmental factors such as pollutant emissions, local policies, and climate contribute to the complex dynamics of $PM_{2.5}$. For the highly complex and nonstationary $PM_{2.5}$, its dynamic characteristics are difficult to capture and learn by the model without preprocessing, resulting in a lower prediction accuracy. A single decomposition transformed the original sequence into several subsequences with distinct dynamic characteristics, which avoided the interference between the dynamic information at different time scales, made prediction easier, and improved the prediction accuracy of the model (Figs. 4 and 5). However, the high-frequency component obtained by single decomposition contained the main random component of the data, which was highly irregular and limited the improvement in the model accuracy. The secondary decomposition can handle the nonstationarity and non-linearity of the high-frequency components and fully extract the deep feature relationships of the time series in the time-frequency domain to improve the prediction performance of the model (Fig. 5).

LSTM exhibited a higher predictive power for the predictors used to develop decomposition-prediction models (Fig. 6). The subcomponents of $PM_{2.5}$ obtained by secondary decomposition had more distinct temporal characteristics. LSTM can preserve and retrieve input values and gradients as needed and automatically extract useful intrinsic features from historical time series. Therefore, LSTM was able to extract and store valuable historical information from the $PM_{2.5}$ time-frequency decomposition sequence to achieve accurate predictions. Furthermore, the deep-learning model adaptively adjusts and readjusts the features represented in the computing layer based on the previous representation. This unique hierarchy allows LSTM to automatically extract advanced and complex features related to datasets without manual intervention, which is a challenge for other models.

4.2. Differences in $PM_{2.5}$ between the cities affect the prediction results

We found that the model had different applicability in each city. In addition to differences in the algorithm stability of the predictors, the secondary decomposition algorithm plays a key role. Owing to the different fluctuation characteristics of $PM_{2.5}$ data in each city, the decomposition results were also different. From the results of the continuous wavelet analysis, it can be seen that $PM_{2.5}$ in Urumqi, Guangzhou, and Chengdu had stable annual changes. However, $PM_{2.5}$ in Shijiazhuang and Shanghai had significant high-frequency components (noise), such as quasi-unit weekly oscillations, in addition to annual scale cycles. By comparing the characteristics of cities in the secondary decomposition results, it can be seen that the decomposition sequence curves of Urumqi, Guangzhou, and Chengdu cities had a smaller amplitude and a higher smoothness, which indicated that the data decomposition results were better. Obvious noise characteristics remained in the high-frequency decomposition sequences in Shanghai and Shijiazhuang, resulting in large change amplitudes and poor curve smoothness. This difference might be due to the fact that Shijiazhuang was located in an industrial-intensive area and Shanghai had a dense population and frequent industrial activities, meaning that the impact of human activities on $PM_{2.5}$ in these two cities was reflected in the high-frequency time series component, which hindered secondary decomposition and resulted in a decrease in the prediction accuracy of the model. The above analysis shows that the intrinsic dynamic characteristics and secondary decomposition results can explain the prediction results of the model. When the internal dynamic characteristics of the $PM_{2.5}$ sequence fluctuated regularly, the second decomposition could completely extract the time information of pollutants, and the prediction performance of the model was often better. However, if the $PM_{2.5}$ sequence changed irregularly, the noise characteristics of the subsequences obtained by the secondary decomposition were more obvious, and the prediction results would be affected, resulting in a lower

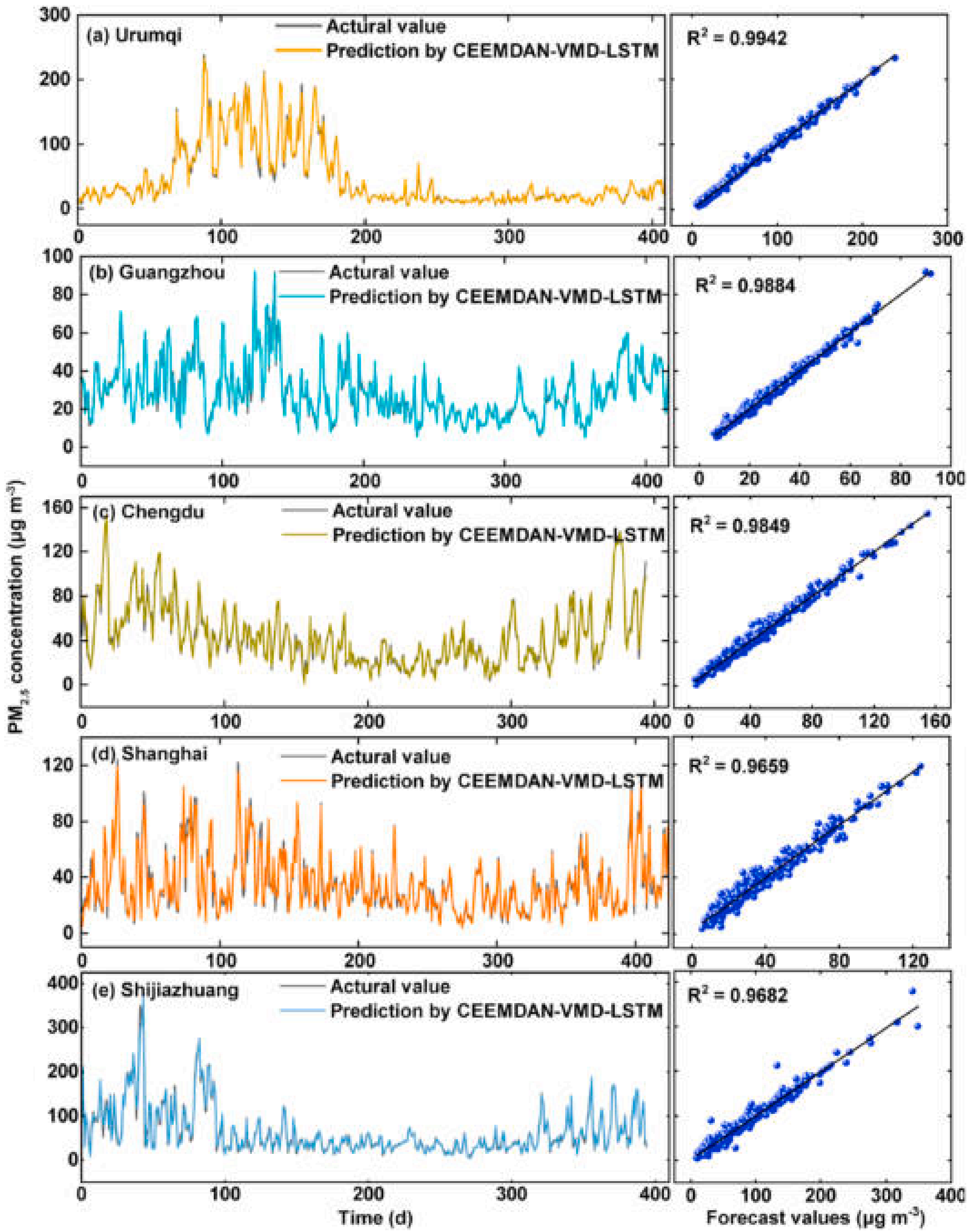


Fig. 7. Time series plots of predicted and actual values of CEEMDAN-VMD-LSTM model in (a) Urumqi, (b) Guangzhou, (c) Chengdu, (d) Shanghai, and (e) Shijiazhuang.

accuracy of the model prediction.

However, when compared with the single model (R^2 = 0.2420–0.8070), the single decomposition model (R^2 = 0.7750–0.9290), and secondary decomposition-prediction models based on other predictors (R^2 = 0.9207–0.9769), the CEEMDAN-VMD-LSTM model still had consistent estimation performance (R^2 = 0.9659–0.9942) regardless of which city was applied. One explanation is that prediction based on secondary decomposition could effectively reduce the difficulty of pollutant time-series prediction, while LSTM could effectively utilise the extracted time characteristics. Thus, the proposed model achieved better prediction results at all of the sites, thereby reducing the prediction error differences between cities.

4.3. Limitations and future work

Although this study has many important advantages, it has some limitations as well. The neural network is the black box in a sense, which means that the structure of the neural network cannot explain the interaction between pollutant concentration and determinants, nor can it explain the physical and chemical reactions that occur during the formation of $PM_{2.5}$ (Gu et al., 2022). Future research can consider integrating environmental factors and attention mechanisms to expand the developed prediction system to provide a cost-effective and transparent representation for estimating the contributions of specific sources and improving the interpretability of neural networks (Gu et al., 2022; Pyo et al., 2021). In addition, when compared with simple prediction models, the CEEMDAN-VMD-LSTM model needs to process more subsequences and has higher computation time and hardware costs. Future studies will recombine similar subpatterns to improve accurate information extraction and computing efficiency by ensuring the accuracy and stability of predictions.

5. Conclusions

When considering the complex dynamics of pollutants in the environment, this study has presented a hybrid $PM_{2.5}$ prediction approach. Secondary decomposition combining complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and variational mode decomposition (VMD) was introduced to process complex $PM_{2.5}$. A deep learning approach of long short-term memory (LSTM) processed the subsequences from the secondary decomposition and then obtained the $PM_{2.5}$ prediction results. The results showed that secondary decomposition technology could effectively extract the inherent dynamic characteristics of $PM_{2.5}$, thereby improving the accuracy and predictive stability of the learning model. The LSTM can effectively capture these dynamic characteristics and therefore make accurate predictions. By combining the advantages of secondary decomposition and deep learning, the proposed model could accurately predict $PM_{2.5}$ concentration under different environmental conditions.

Credit author statement

Liang Dong: Writing - Original draft preparation, Methodology, Software. Pei Hua: Writing - Review & Editing, Resources. Dongwei Gui: Writing - Review & Editing. Jin Zhang: Conceptualization, Writing - Reviewing and Editing.

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.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank the China Air Quality Online Monitoring and Analysis Platform for providing the dataset. This work was jointly supported by the National Natural Science Foundation of China (Grant No.: 42077156 and 52121006), Guangdong Basic and Applied Basic Research Foundation (Grant No.: 2020A1515011130), and Special Fund Project for Science and Technology Innovation Strategy of Guangdong Province (Grant No.: 2019B121205004). This manuscript has not been subject to peer review by the above agencies and does not, therefore, reflect the views of the above agencies nor should any official endorsement be inferred.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.chemosphere.2022.136252.

References

Aggarwal, A., Toshniwal, D., 2021. A hybrid deep learning framework for urban air quality forecasting. J. Clean. Prod. 329, 129660.
Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., Di Tommaso, S., Colangeli, C., Rosatelli, G., Di Carlo, P., 2017. Recursive neural network model for analysis and forecast of PM_{10} and $PM_{2.5}$. Atmos. Pollut. Res. 8, 652–659.
Boulila, W., Ghandorh, H., Khan, M.A., Ahmed, F., Ahmad, J., 2021. A novel CNN-LSTM-based approach to predict urban expansion. Ecol. Inf. 64, 101325.
Chen, H., Zhang, Z., van Donkelaar, A., Bai, L., Martin, R.V., Lavigne, E., Kwong, J.C., Burnett, R.T., 2020. Understanding the joint impacts of fine particulate matter concentration and composition on the incidence and mortality of cardiovascular disease: a component-adjusted approach. Environ. Sci. Technol. 54, 4388–4399.
Chen, Y.-C., Li, D.-C., 2021. Selection of key features for $PM_{2.5}$ prediction using a wavelet model and RBF-LSTM. Appl. Intell. 51, 2534–2555.
Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. Lancet 389, 1907–1918.
Ding, Y., Li, Z., Zhang, C., Ma, J., 2020. Prediction of ambient $PM_{2.5}$ concentrations using a correlation filtered spatial-temporal long short-term memory model. Appl. Sci. 10, 14.
Dragomiretskiy, K., Zosso, D., 2013. Variational mode decomposition. IEEE Trans. Signal Process. 62, 531–544.
Fang, H., Jamali, B., Deletic, A., Zhang, K., 2021. Machine learning approaches for predicting the performance of stormwater biofilters in heavy metal removal and risk mitigation. Water Res. 200, 117273.
Gong, T., Sun, Z., Zhang, X., Zhang, Y., Wang, S., Han, L., Zhao, D., Ding, D., Zheng, C., 2019. Associations of black carbon and $PM_{2.5}$ with daily cardiovascular mortality in Beijing, China. Atmos. Environ. 214, 116876.
Goudarzi, G., Hopke, P.K., Yazdani, M., 2021. Forecasting $PM_{2.5}$ concentration using artificial neural network and its health effects in Ahvaz, Iran. Chemosphere 283, 131285.
Gu, Y., Li, B., Meng, Q., 2022. Hybrid interpretable predictive machine learning model for air pollution prediction. Neurocomputing 468, 123–136.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.
Hong, J., Mao, F., Min, Q., Pan, Z., Wang, W., Zhang, T., Gong, W., 2020. Improved $PM_{2.5}$ predictions of WRF-Chem via the integration of Himawari-8 satellite data and ground observations. Environ. Pollut. 263, 114451.
Huang, G., Li, X., Zhang, B., Ren, J., 2021. $PM_{2.5}$ concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition. Sci. Total Environ. 768, 144516.
Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. Math. Phys. Eng. Sci. 454, 903–995.
Li, T., Hua, M., Wu, X., 2020. A hybrid CNN-LSTM model for forecasting particulate matter ($PM_{2.5}$). IEEE Access 8, 26933–26940.
Lightstone, S.D., Moshary, F., Gross, B., 2017. Comparing CMAQ forecasts with a neural network forecast model for $PM_{2.5}$ in New York. Atmosphere 8, 161.
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, 1588–1600.
Liu, H., Yin, S., Chen, C., Duan, Z., 2020a. Data multi-scale decomposition strategies for air pollution forecasting: a comprehensive review. J. Clean. Prod. 277, 124023.
Liu, N., Liu, X., Jayaratne, R., Morawska, L., 2020b. A study on extending the use of air quality monitor data via deep learning techniques. J. Clean. Prod. 274, 122956.
Loftus, C., Yost, M., Sampson, P., Arias, G., Torres, E., Vasquez, V.B., Bhatti, P., Karr, C., 2015. Regional $PM_{2.5}$ and asthma morbidity in an agricultural community: a panel study. Environ. Res. 136, 505–512.
Lu, W.-Z., Wang, W.-J., Wang, X.-K., Yan, S.-H., Lam, J.C., 2004. Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong. Environ. Res. 96, 79–87.
Ma, J., Cheng, J.C., Lin, C., Tan, Y., Zhang, J., 2019. Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. Atmos. Environ. 214, 116885.
Masood, A., Ahmad, K., 2021. A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: fundamentals, application and performance. J. Clean. Prod. 322, 129072.
Menares, C., Perez, P., Parraguez, S., Fleming, Z.L., 2021. Forecasting $PM_{2.5}$ levels in Santiago de Chile using deep learning neural networks. Urban Clim. 38, 100906.
Niu, M., Wang, Y., Sun, S., Li, Y., 2016. A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term $PM_{2.5}$ concentration forecasting. Atmos. Environ. 134, 168–180.
Pyo, J., Cho, K.H., Kim, K., Baek, S.-S., Nam, G., Park, S., 2021. Cyanobacteria cell prediction using interpretable deep learning model with observed, numerical, and sensing data assemblage. Water Res. 203, 117483.
Qiao, W., Tian, W., Tian, Y., Yang, Q., Wang, Y., Zhang, J., 2019. The forecasting of $PM_{2.5}$ using a hybrid model based on wavelet transform and an improved deep learning algorithm. IEEE Access 7, 142814–142825.
Rodriguez-Galiano, V.F., Luque-Espinar, J.A., Chica-Olmo, M., Mendes, M.P., 2018. Feature selection approaches for predictive modelling of groundwater nitrate pollution: an evaluation of filters, embedded and wrapper methods. Sci. Total Environ. 624, 661–672.
Samal, K.K.R., Babu, K.S., Das, S.K., 2021. Temporal convolutional denoising autoencoder network for air pollution prediction with missing values. Urban Clim. 38, 100872.
Sharma, E., Deo, R.C., Prasad, R., Parisi, A.V., 2020. A hybrid air quality early-warning framework: an hourly forecasting model with online sequential extreme learning machines and empirical mode decomposition algorithms. Sci. Total Environ. 709, 135934.

Sugiyama, T., Ueda, K., Seposo, X.T., Nakashima, A., Kinoshita, M., Matsumoto, H., Ikemori, F., Honda, A., Takano, H., Michikawa, T., 2020. Health effects of $PM_{2.5}$ sources on children's allergic and respiratory symptoms in Fukuoka, Japan. Sci. Total Environ. 709, 136023.

Sun, W., Tan, B., Wang, Q., 2021. Multi-step wind speed forecasting based on secondary decomposition algorithm and optimized back propagation neural network. Appl. Soft Comput. 113, 107894.

Teng, M., Li, S., Xing, J., Song, G., Yang, J., Dong, J., Zeng, X., Qin, Y., 2022. 24-Hour prediction of $PM_{2.5}$ concentrations by combining empirical mode decomposition and bidirectional long short-term memory neural network. Sci. Total Environ. 821, 153276.

Torres, M.E., Colominas, M.A., Schlotthauer, G., Flandrin, P., 2011. A complete ensemble empirical mode decomposition with adaptive noise. In: IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP), Prague, pp. 4144–4147, 2011.

Wang, D., Wei, S., Luo, H., Yue, C., Grunder, O., 2017. A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine. Sci. Total Environ. 580, 719–733.

Wang, J., Wang, R., Li, Z., 2021a. A combined forecasting system based on multi-objective optimization and feature extraction strategy for hourly $PM_{2.5}$ concentration. Appl. Soft Comput. 114, 108034.

Wang, R., Zhang, M., Jiang, Y., Yang, Y., 2019. Prediction model of insulator contamination degree based on adaptive mutation particle swarm optimisation and general regression neural network. J. Eng. 1423–1428, 2019.

Wang, W.-c., Du, Y.-j., Chau, K.-w., Xu, D.-m., Liu, C.-j., Ma, Q., 2021b. An ensemble hybrid forecasting model for annual runoff based on sample entropy, secondary decomposition, and long short-term memory neural network. Water Resour. Manag. 35, 4695–4726.

Xu, X., Ren, W., 2019. A hybrid model based on a two-layer decomposition approach and an optimized neural network for chaotic time series prediction. Symmetry 11, 610.

Yang, S., Chen, D., Li, S., Wang, W., 2020. Carbon price forecasting based on modified ensemble empirical mode decomposition and long short-term memory optimized by improved whale optimization algorithm. Sci. Total Environ. 716, 137117.

Yu, J.-W., Kim, J.-S., Li, X., Jong, Y.-C., Kim, K.-H., Ryang, G.-I., 2022. Water quality forecasting based on data decomposition, fuzzy clustering and deep learning neural network. Environ. Pollut. 303, 119136.

Zhang, J., Tan, Z., Wei, Y., 2020a. An adaptive hybrid model for day-ahead photovoltaic output power prediction. J. Clean. Prod. 244, 118858.

Zhang, L., Wilson, J.P., MacDonald, B., Zhang, W., Yu, T., 2020b. The changing $PM_{2.5}$ dynamics of global megacities based on long-term remotely sensed observations. Environ. Int. 142, 105862.

Zhong, Q., Ma, J., Shen, G., Shen, H., Zhu, X., Yun, X., Meng, W., Cheng, H., Liu, J., Li, B., 2018. Distinguishing emission-associated ambient air $PM_{2.5}$ concentrations and meteorological factor-induced fluctuations. Environ. Sci. Technol. 52, 10416–10425.