

PM2.5 Concentration Forecasting Based on Data Preprocessing Strategy and LSTM Neural Network

Tao Liang, Gaofeng Xie, Dabin Mi, Wen Jiang, and Guilin Xu

Abstract—In order to better grasp the change rule of PM2.5 concentration, this paper presents a prediction model of PM2.5 concentration based on Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)-Permutation Entropy (PE)-Long Short-Term Memory (LSTM). The PM2.5 concentration time series is decomposed into several sub-sequences with obvious complexity differences by CEEMDAN-PE. Then, the LSTM prediction model is built by adding meteorological parameters to each different sub-sequence. The final results are got by adding the prediction results. The data of four monitoring stations in Tangshan City, Hebei Province is used to implement simulation experiment. Experiment results confirm that the proposed prediction model compared with other combined and single forecasting methods, and shows a high prediction precision, and good universality, which provided effective technical support for pre-control of air pollution.

Index Terms—PM2.5, concentration prediction, ensemble empirical mode decomposition, permutation entropy, time series.

I. INTRODUCTION

With the acceleration of economic development worldwide, environmental issues are becoming more and more important in various countries, and the causes of environmental problems in different regions are becoming more and more complicated. Among them, the problem of air pollution has become more serious. Particulate matter (PM) in the air is an important evaluation index of air quality. Generally speaking, PM10 and PM2.5 mean that the aerodynamic equivalent diameter in air is less than or equal to 10 μ m and less than or equal to 2.5 μ m. Atmospheric particulate matter is the primary atmospheric pollutant in most cities in northern China. Compared with PM10, PM2.5 is small in size and light in weight, and it is easier to absorb some harmful substances, which not only endangers human health [1], but also affects air visibility. Therefore, timely and accurate prediction of PM2.5 concentration is beneficial for people to prepare for and control in advance to achieve the

purpose of protecting people's health and even life.

At present, the methods for predicting PM2.5 concentration mostly use the unary linear regression method [2], multiple linear regression method [3], and support vector machine method [4]. With the development of artificial intelligence, artificial neural network (ANN) has achieved good performance in PM2.5 concentration prediction [5], [6]. However, artificial neural networks tend to converge to local optimum and are not suitable for small samples, which seriously affects prediction accuracy. In recent years, deep learning as a new research direction in the field of machine learning has been developing rapidly. The deep learning model has multiple non-linear mapping levels, which can grasp the deep potential laws contained in massive data through layer-by-layer learning. In many deep learning models, the recurrent neural network (RNN) introduces the concept of time step and shows greater adaptability in the analysis of time series data. The long- and short-term memory neural network improves the structure of the cyclic neural network. It has a special memory structure and gate structure to enhance the long-term memory ability, make up for the gradient explosion of RNN and easily ignore the long-term dependence within the time series [7]. Therefore, using the LSTM network as a prediction algorithm to construct a PM2.5 concentration prediction model has higher prediction accuracy.

The PM2.5 concentration data generates noise due to interference from many factors during the acquisition process. Therefore, there is a large error in establishing a prediction model directly using the original wind speed time series. In the literature [8], the EMD method is used to decompose the time series, but the modal aliasing phenomenon is easy to occur when using the EMD method. In [9], the ensemble empirical mode decomposition is used to decompose the PM2.5 time series and then use the support vector regression prediction to effectively improve the prediction accuracy. EEMD inherits the advantages of multivariate analysis of wavelet transform and improves the modal aliasing of EMD. CEEMDAN has been improved on the basis of EEMD, which not only overcomes the problem of low decomposition efficiency, but also has a reconstruction error of almost zero, which is more suitable for the decomposition of nonlinear time series.

In view of the advantages of CEEMDAN and LSTM, a model based on complete ensemble empirical mode decomposition with adaptive noise- permutation entropy and long short-term memory neural network is proposed. The CEEMDAN is used to decompose the non-stationary PM2.5 concentration time series into a series of IMF components

Manuscript received October 15, 2019; revised July 13, 2020. This work was supported in part by the Hebei Province Science and Technology Plan Project (16214510D, 1721430 4D, 19210108D) and the support of development project of Shijiazhuang science and technology bureau Plan Project (181060481a).

Tao Liang and Gaofeng Xie are with the Hebei University of Technology, college of Artificial Intelligence, Tianjin, 300130, China (Corresponding author; e-mail: LiangTao@hebut.edu.cn, xiegaofeng821@163.com).

Dabin Mi, Wen Jiang, and Guilin Xu are with the Jointo Energy Investment Co., Ltd. Hebei, Shijiazhuang, 050000, China (e-mail: 2017919074@qq.com, 2743247984@qq.com, 1129215323@qq.com).

with different feature scales. Then, considering the high sensitivity and high computational efficiency of permutation entropy for time series variation, the remaining entropy is determined based on permutation entropy. The complexity of the IMF component combines and recombines similar IMF components to reduce the computational scale of predicting each IMF component separately. The LSTM neural network prediction model with appropriate parameter space is established by adding meteorological factors to each recombination subsequence. Finally, the predicted values are superimposed to obtain the final prediction value.

II. RELATED METHODOLOGY

A. CEEMDAN

The EMD method decomposes the original sequence according to different fluctuation scales to obtain IMF components of different amplitudes [10]. The CEEMDAN method calculates the unique margin to obtain the IMF component by adding adaptive white noise at each decomposition stage, and the decomposition process is complete.

Record $s(n)$ as the original PM2.5 concentration time series, $v^i(n)$ is the Gaussian white noise sequence added in the i -th experiment, and the i -th decomposition PM2.5 concentration time series can be expressed as $s^i(n) = s(n) + v^i(n)$. Assume that the k -th modal component produced by EMD and CEEMDAN is denoted as $E_k(\cdot)$ and IMF_k , respectively. The specific steps of the CEEMDAN algorithm are as follows:

- 1) Like the EEMD method, CEEMDAN performs I-time decomposition for the PM2.5 concentration time series $s^i(n) = s(n) + v^i(n)$. The first modal component is calculated by the EMD method:

$$IMF_1(n) = \frac{1}{I} \sum_{i=1}^I IMF_1^i(n) = \overline{IMF_1}(n) \quad (1)$$

- 2) Calculate the first unique margin time series decomposed by CEEMDAN as:

$$r_1(n) = s(n) - IMF_1(n) \quad (2)$$

- 3) Perform i -time decomposition on $r_1(n) + \varepsilon_1 E_1(v^i(n))$. Calculate the second modal component as:

$$IMF_2(n) = \frac{1}{I} \sum_{i=1}^I E_1(r_1(n) + \varepsilon_1 E_1(v^i(n))) \quad (3)$$

- 4) Similarly, for each of the remaining stages, calculate the k th residual sequence; then, according to step (3), calculate the $(k+1)$ th modal component, namely:

$$r_k(n) = r_{k-1}(n) - IMF_k(n) \quad (4)$$

$$IMF_{k+1}(n) = \frac{1}{I} \sum_{i=1}^I E_1(r_k(n) + \varepsilon_k E_k(v^i(n))) \quad (5)$$

- 5) Execution step (4), when the residual sequence can no longer be decomposed, that is, the maximum number of extreme points of the residual signal does not exceed two, the algorithm terminates. At this point, K modal components are obtained, and the final result of the remainder sequence is:

$$R(n) = s(n) - \sum_{k=1}^K IMF_k \quad (6)$$

At this time, the original PM2.5 concentration time series $s(n)$ is finally decomposed into $\sum_{k=1}^K IMF_k$ and $R(n)$, that is

$$s(n) = \sum_{k=1}^K IMF_k + R(n) \quad (7)$$

B. Permutation Entropy

Permutation Entropy (PE) [11] is a measure of time series complexity, with fast calculation speed and strong anti-interference ability, especially suitable for nonlinear data. It is highly sensitive to time series changes and has good robustness and has been widely used in various time series. The PM2.5 concentration data has certain randomness and non-stationarity, which makes the IMF component obtained by CEEMDAN decomposition more. Therefore, in order to reduce the computational scale, this paper uses PE algorithm to carry out IMF reorganization, the steps are as follows:

Step 1: First, phase space reconstruction is performed on each IMF traffic sequence $\{X(i), i = 1, 2, \dots, N\}$ obtained by CEEMDAN to obtain a phase space matrix:

$$Y = \begin{bmatrix} x(1) & x(1+\tau) & \cdots & x(1+(m-1)\tau) \\ & & \vdots & \\ x(j) & x(j+\tau) & \cdots & x(j+(m-1)\tau) \\ & & \vdots & \\ x(K) & x(K+\tau) & \cdots & x(K+(m-1)\tau) \end{bmatrix}$$

where m and τ represent the embedding dimension and delay time. $j = 1, 2, \dots, K$

Step 2: Treat each row of the above matrix as a component, then there are K components, the existence relationship $K = n - (m-1)\tau$. Taking the j -th component $(x(j), x(j+\tau), \dots, x(j+(m-1)\tau))$ as an example, in ascending order according to the size of the element, that is

$$x(i + (j_1 - 1)\tau) \leq x(i + (j_2 - 1)\tau) \leq \cdots \leq x(i + (j_m - 1)\tau) \quad (8)$$

where j_1, j_2, \dots, j_m represents the serial number of each component element. If $x(i + (j_1 - 1)\tau) = x(i + (j_2 - 1)\tau)$ is present, it is sorted by j_1 and j_2 values. In summary, for each row vector of the phase space reconstruction matrix of any concentration time series $X(i)$, a set of sequence $S(l)$ can be obtained.

$$S(l) = \{j_1, j_2, \dots, j_m\} \quad (9)$$

where $l=1,2,\dots,K$, Then there are a total of $k \leq m!$ different mapping symbol sequences.

Step 3: Calculate the probability p_1, p_2, \dots, p_k of occurrence of each symbol sequence $S(l)$. According to the form of Shannon entropy, the permutation entropy $H_p(m)$ of the k th different symbol sequence of the flow time series $X(i)$ can be defined as:

$$H_p(m) = -\sum_{j=1}^k P_j \ln P_j \quad (10)$$

It can be seen from the above formula that when $P_j = 1/m!$, $H_p(m)$ reaches the maximum value $\ln(m!)$. For convenience, $H_p(m)$ is typically normalized with $\ln(m!)$, namely:

$$H_{PE}(m) = H_p(m) / \ln(m!) \quad (11)$$

where $0 \leq H_{PE}(m) \leq 1$, the size of $H_{PE}(m)$ represents the degree of randomness of the time series. The larger the value, the stronger the randomness; on the contrary, the weaker the randomness. When calculating PE, the embedding dimension m and the delay τ need to be predetermined.

C. Long Short-Term Memory Network

As an improvement of RNN, LSTM not only adds a new hidden cell state, but also subtly designs various door structures to control the input data through the door. This way of data transfer makes the weight of the self-loop no longer fixed [12]. The basic unit of the LSTM neural network is shown in Fig. 1.

Suppose x_t is the input vector at time t , and the output at time $t-1$ is h_{t-1} . $W^{(f)}$, $W^{(i)}$, $W^{(c)}$, $W^{(o)}$, $U^{(f)}$, $U^{(i)}$, $U^{(c)}$, $U^{(o)}$ and b_f , b_i , b_c , b_o represent the weight matrix and the offset vector, respectively, and the

calculation process is as follows:

First, forget the door and forget the useless historical information:

$$F_t = \sigma(W^{(f)}x_t + U^{(f)}h_{t-1} + b_f) \quad (12)$$

The input gate then updates the status based on the input data and historical information:

$$I_t = \sigma(W^{(i)}x_t + U^{(i)}h_{t-1} + b_i) \quad (13)$$

$$\tilde{C}_t = \tanh(W^{(c)}x_t + U^{(c)}h_{t-1} + b_c) \quad (14)$$

$$C_t = F_t * C_{t-1} + I_t * \tilde{C}_t \quad (15)$$

Finally, the final output of the LSTM is determined by the output gate and the unit state:

$$O_t = \sigma(W^{(o)}x_t + U^{(o)}h_{t-1} + b_o) \quad (16)$$

$$h_t = O_t * \tanh(C_t) \quad (17)$$

where: σ represents the sigmoid activation function; F_t , I_t and O_t respectively indicate the output states of the forgetting gate, the input gate and the output gate at time t ; \tilde{C}_t represents the unit state input at time t ; \tanh represents the tanh activation function.

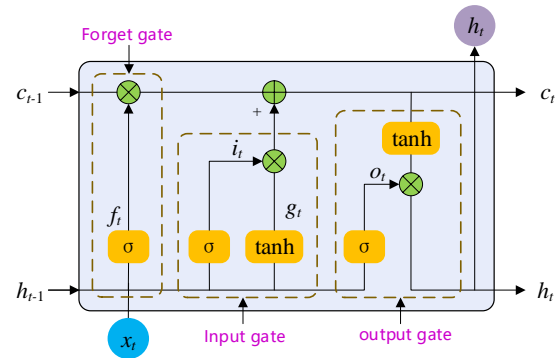


Fig. 1. Basic unit of LSTM network.

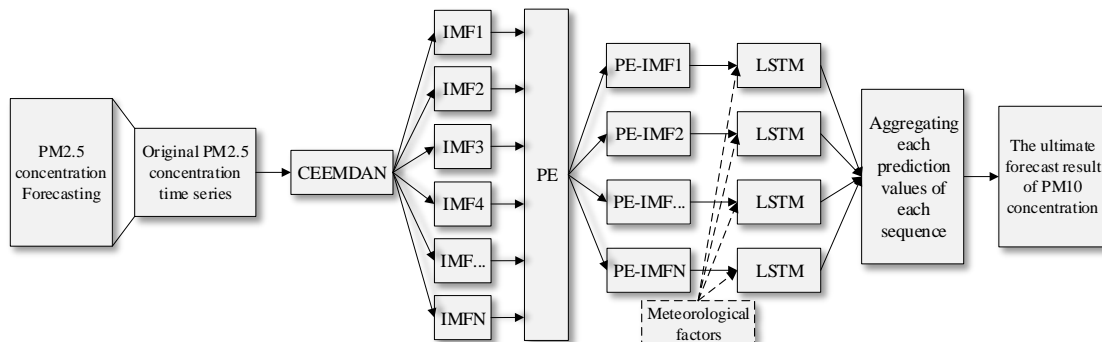


Fig. 2. Structure of CEEMDAN-PE-LSTM forecasting model.

D. PM2.5 Concentration Prediction Model (CEEMDAN-PE-LSTM)

Construct a CEEMDAN-PE-LSTM combined prediction model with PM2.5 concentration. The flowchart in Fig. 2 illustrates its operation. The specific steps are as follows:

Step1: Information extraction. Using CEEMDAN to

decompose the original PM2.5 concentration sequence to obtain several IMF components and one trend component.

Step2: Data recombining. Calculate the sample entropy of each IMF component separately, and recombine the sub-sequences into obtain new subsequences (PE-IMF) with obvious complexity differences based on their approximate SE values.

Step3: Data normalization. In order to improve the prediction accuracy of the model, the new subsequence and meteorological parameters are normalized before the data is sent to the model.

$$X'(i) = \frac{(X(i) - \min\{X(i)\})}{(\max\{X(i)\} - \min\{X(i)\})}$$

where $X(i)$ is the original data and $X'(i)$ is the normalized data.

Step4: Training and validation of the model. Add the meteorological factors to each new subsequence to establish the LSTM neural network prediction model of the corresponding parameter space, and output the predicted values of each model.

Step5: PM2.5 concentration forecasting. The n predicted values obtained in step 4 are inversely normalized and superimposed to obtain a PM2.5 concentration prediction result.

Step6: Evaluation of prediction results. Calculate the error index of the error analysis by comparing the predicted result with the actual PM2.5 concentration data.

III. EXPERIMENTAL RESULTS

A. Dataset

The PM2.5 concentration data used in this paper was obtained from the measured PM2.5 concentrations of four air quality monitoring stations (S1-S4) in Tangshan City, Hebei Province. The sample time span is 31 days from March 1, 2018 to March 31, 2018, and data is collected every hour. The values of the day before, the first two days and the first three days of the predicted time point of the recombination subsequence and the meteorological parameters of the predicted time point: temperature (°C), humidity (%), wind direction, wind speed (m s⁻¹), atmospheric pressure (mmHg) As an input parameter of the LSTM network, the value of the recombination subsequence corresponding to the predicted time point is used as the output of the model.

A total of 696 data sets were used as model training samples from March 1st to March 29th, and a total of 48 sets of data were used as model test samples from March 30th to March 31st. The simulation result of the measured data of the monitoring station S1 is selected as the display object. The original time series of the measured PM2.5 concentration of the monitoring station S1 is shown in Fig. 3.

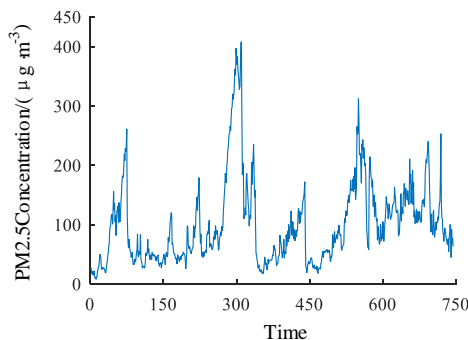


Fig. 3. Time series of PM2.5 concentration.

B. CEEMDAN-PE Decomposition and Recombination of PM2.5 Concentration Time Series

CEEMDAN decomposition was performed on the PM2.5 concentration time series, and $I=500$ groups of white noise signals with a non-standard deviation of 0.2 were added to obtain 10 IMF components. According to the foregoing, the permutation entropy values of the IMF components are calculated, and the results are shown in Table I. In order to avoid excessive decomposition and reduce the calculation scale, the IMF components with similar entropy values are superimposed to obtain the recombination component: PE-IMF. The difference in PE between IMF1, IMF2 and IMF3 is less than 0.05, and the difference is small, so they are combined. Although IMF4 and IMF5 are adjacent, the PE value and other adjacent components are larger, so they are respectively used as separate components. The PE values between IMF 6, IMF 7, IMF 8 and IMF 9 were all less than 0.1 and were combined. The difference between the PE values of IMF9 and IMF10 is 0.164, which is quite different. To ensure accuracy, no merging is performed. The specific combined results are shown in Table I. The recombination sequence is shown in Fig. 4.

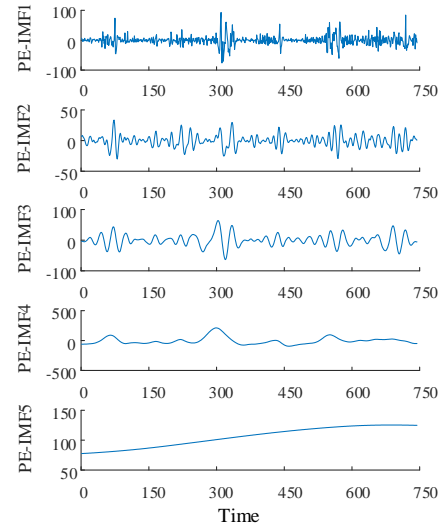


Fig. 4. PM2.5 concentration subsequences processed by CEEMDAN-PE.

TABLE I: NEW SUBSEQUENCES WITH MERGED IMF COMPONENTS

IMFn	$H_{PE}(n)$	Recombination	New IMF
IMF1	0.992	IMF1&IMF2&IMF3	PE-IMF1
IMF2	0.973		
IMF3	0.956		
IMF4	0.789	IMF4	PE-IMF2
IMF5	0.617	IMF5	PE-IMF3
IMF6	0.381	IMF6&IMF7&IMF8&IMF9	PE-IMF4
IMF7	0.313		
IMF8	0.285		
IMF9	0.258		
IMF10	0.094	IMF10	PE-IMF5

C. Evaluation Matrix

The MAE, RMSE and MAPE are selected as evaluation indicators to analyze the prediction accuracy of the model. The formula for each indicator is as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (18)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (19)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right| \quad (20)$$

where n is the number of test data, y_i and \hat{y}_i are the predicted values of PM2.5 and the actual monitored values, respectively.

D. Forecast Results and Contrast Analysis

It is known from Table I that five LSTM neural network prediction models need to be established. It is found through experiments that increasing the model depth of the LSTM network can effectively improve the prediction accuracy of the model. Therefore, this paper experiments on different models of LSTM network layers. At the same time, an appropriate increase in the number of LSTM network layers can also improve the prediction accuracy of the model. Finally, the LSTM network used in this paper contains four hidden layers, each with 5, 10, 20, and 40 neurons. The Drop Out layer is set to randomly disconnect 20% of neurons to

prevent over-fitting.

The rolling single-step prediction was used to construct the CEEMDAN-PE-LSTM, EEMD-PE-LSTM LSTM and ELMAN models to compare the prediction effects. The prediction results of each model and the actual PM2.5 concentration values are shown in Fig. 5. The results of the prediction error indicators for different models are shown in Table II.

It can be seen from Table II and Fig. 5 that compared with other prediction models, the predicted values of the proposed CEEMDAN-PE-LSTM model are the closest to the true values. The mean absolute error, root mean square error and mean absolute percentage error are $8.1026 \mu\text{g}\cdot\text{m}^{-3}$, $10.8694 \mu\text{g}\cdot\text{m}^{-3}$, 7.1644% , which are superior to EEMD-PE-LSTM, LSTM and ELMAN prediction models. The results show that the CEEMDAN-PE-LSTM model has the least deviation from the true value, the prediction error is the smallest, and the prediction accuracy is the highest. In summary, the CEEMDAN-PE-LSTM based PM2.5 concentration prediction model performs well and can meet the forecasting requirements.

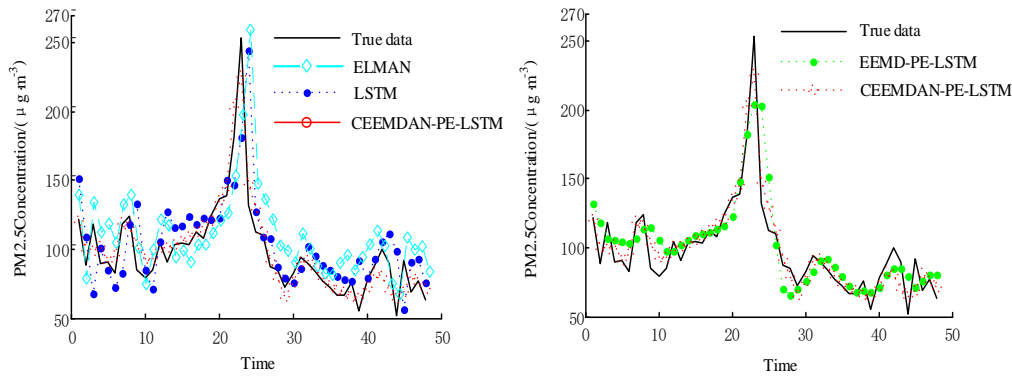


Fig. 5. Prediction results of each model.

TABLE II: COMPARISONS OF PREDICTION PERFORMANCE INDEXES FOR EACH PREDICTION MODEL

Predictive model	MAE	RMSEE	MAPE
ELMAN	16.325	24.0132	15.5362
LSTM	15.8948	22.8832	14.6887
EEMD-PE-LSTM	13.5092	18.0745	11.3949
CEEMDAN-PE-LSTM	8.1026	10.8694	7.1644

TABLE III: COMPARISONS OF PREDICTION PERFORMANCE INDEXES OF EACH PREDICTION MODEL FOR DIFFERENT SITE DATA

Predictive model	Monitoring station S2			Monitoring station S3			Monitoring station S4		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
ELMAN	15.8864	23.5632	14.2356	13.6879	24.8653	15.3256	16.1235	23.9563	14.9586
LSTM	14.9945	21.8754	13.8763	15.6734	22.7633	14.2156	15.0025	21.8452	13.8857
EEMD-PE-LSTM	12.9142	17.9554	10.6324	13.2135	18.0244	11.1253	12.6752	17.8746	10.9659
CEEMDAN-PE-LSTM	7.3024	10.5683	7.0524	8.0036	10.6253	7.0123	7.8026	10.0256	7.1034

In order to verify the universality of the CEEMDAN-PE-LSTM model proposed in this paper, CEEMDAN-PE-LSTM, EEMD-PE-LSTM LSTM and ELMAN models were used to predict the measured data of S2, S3 and S4 monitoring stations. As shown in Table III. It can be seen from Table III that the CEEMDAN-SE-LSTM model proposed in this paper is still optimal for PM2.5

concentration prediction at different sites, which proves the universality of the proposed model.

IV. CONCLUSION

Aiming at the obvious nonlinear and volatility characteristics of PM2.5 concentration time series, the

CEEMDAN-PE-LSTM model is proposed. Using CEEMDAN-PE to decompose the PM2.5 concentration time series, several subsequences with obvious complexity differences were obtained. Reduce the size of the calculation. The multi-layer LSTM network is used as the prediction algorithm to describe the change trend of PM2.5 concentration more accurately and improve the prediction accuracy of the model. The CEEMDAN-PE-LSTM model proposed in this paper is compared with the predicted performance of other models. The PM2.5 concentration at different sites was predicted. The results show that the CEEMDAN-PE-LSTM model is better than other models, which can effectively fit the nonlinear variation of PM2.5 concentration time series, with good prediction accuracy and universality.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Dabin Mi concuted the research; Gaofeng Xie and Tao Liang analyzed the data; Gaofeng Xie and Tao Liang wrote the paper; Dabin Mi and Wen Jiang correction article. All authors had approved the final version.

ACKNOWLEDGMENTS

This work obtains the support of Hebei Province Science and Technology Plan Project (16214510D, 17214304D, 19210108D) and the support of development project of Shijiazhuang science and technology bureauPlan Project (181060481a).

REFERENCES

- [1] Y. O. Khaniabadi, G. Goudarzi, S. M. Daryanoosh, A. Borgini, A. Tittarelli, and A. De Marco, "Exposure to pm10, no2, and o3 and impacts on human health," *Environmental Science and Pollution Research*, vol. 24, no. 3, pp. 2781-2789, January 2017.
- [2] Y. Li, Q. Chen, H. Zhao, L. Wang, and R. Tao, "Variations in PM10, PM2.5 and PM10 in an urban area of the Sichuan basin and their relation to meteorological factors," *Atmosphere*, vol. 6, no. 3 pp. 150-163, March 2015.
- [3] A. Vlachogianni, P. Kassomenos, A. Karppinen, S. Karakitsios, and J. Kukkonen, "Evaluation of a multiple regression model for the forecasting of the concentrations of NOx and PM 10 in Athens and Helsinki," *Science of the Total Environment*, vol. 409, no. 8 pp. 1559-1571, March 2011.
- [4] W. Li, D. Kong, and J. Wu, "A new hybrid model FPA-SVM considering cointegration for particular matter concentration forecasting: a case study of Kunming and Yuxi, China," *Computational Intelligence and Neuroscience*, vol. 2017, pp. 1-11, 2017.
- [5] M. M. Dedovic, S. Avdakovic, I. Turkovic, N. Dautbasic, and T. Konjic, "Forecasting PM10 concentrations using neural networks and system for improving air quality," in *Proc. XI International Symposium on Telecommunications*, IEEE, 2016.
- [6] X. Qu, X. Kang, C. Zhang, J. Shuai, and X. Ma, "Short-term prediction of wind power based on deep long short-term memory," in *Proc. Asia-Pacific Power and Energy Engineering Conference*, Dec. 2016, pp. 1148-1152.
- [7] H. Chiou-Jye and K. Ping-Huan, "A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities," *Sensors*, vol. 18, no. 7, 2018.

- [8] A. Deng and M. Jin, "Time-scale feature extraction based on EMD and its application on short-term power load forecasting," *Application Research of Computers*, vol. 35, no. 10, pp. 2952-2955, 2017.
- [9] X. Qin, Y. Liu, X. Wang, and X. Dong, "PM2.5 prediction of Beijing city based on ensemble empirical mode decomposition and support vector regression," *Journal of Jilin University (Earth Science Edition)*, vol. 46, no. 2, pp. 563-568, 2016.
- [10] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1-41, 2009.
- [11] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, pp. 278, no. 6, pp. 2039-2049, June 2000.
- [12] S. Hochreite and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, March 1997.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).



Tao Liang received his doctorate degree in power electronics and power transmission from Hebei University of Technology in 2007. He is currently a professor at the School of Artificial Intelligence, Hebei University of Technology.

His current research interests include new energy power generation, fault prediction and fault diagnosis based on big data and cloud computing.

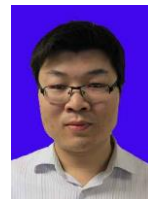


Gaofeng Xie received the B.S. degree from the School of physics and electrical engineering, Anyang Normal University, China, in 2017. He is currently pursuing the degree with the School of Artificial Intelligence, Hebei University of Technology.

His main research interests include data mining, machine learning and the application of deep learning in the energy.



Dabin Mi holds master's degree and he is a senior engineer. His main research interests include artificial intelligence control of thermal power generation enterprises and optimization and application of intelligent blending of coal-fired boilers in thermal power plants based on artificial intelligence technology.



Wen Jiang holds master's degree and he is an engineer. His main research interests include fault diagnosis method based on artificial neural network automatic control system intelligent fault detection.



Guilin Xu holds master's degree, senior engineer. His main research interests include artificial intelligence control of thermal power generation enterprises and optimization and application of intelligent blending of coal-fired boilers in thermal power plants based on artificial intelligence technology.