atmosphere

MDPI

Article

Air Pollutant Concentration Prediction Based on a CEEMDAN-FE-BiLSTM Model

Xuchu Jiang [1], Peiyao Wei [1], Yiwen Luo [1] and Ying Li [2,*]

[1] School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan 430073, China; xuchujiang@zuel.edu.cn (X.J.); z0004994@zuel.edu.cn (P.W.); z0004364@zuel.edu.cn (Y.L.)
[2] Department of Scientific Research, Zhongnan University of Economics and Law, Wuhan 430073, China
* Correspondence: yingli@zuel.edu.cn

Abstract: The concentration series of $PM_{2.5}$ (particulate matter $\leq$ 2.5 μm) is nonlinear, nonstationary, and noisy, making it difficult to predict accurately. This paper presents a new $PM_{2.5}$ concentration prediction method based on a hybrid model of complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and bi-directional long short-term memory (BiLSTM). The new method was applied to predict the same kind of particulate pollutant PM_{10} and heterogeneous gas pollutant O_3, proving that the prediction method has strong generalization ability. First, CEEMDAN was used to decompose $PM_{2.5}$ concentrations at different frequencies. Then, the fuzzy entropy (FE) value of each decomposed wave was calculated, and the near waves were combined by K-means clustering to generate the input sequence. Finally, the combined sequences were put into the BiLSTM model with multiple hidden layers for training. We predicted the $PM_{2.5}$ concentrations of Seoul Station 116 by the hour, with values of the root mean square error (*RMSE*), the mean absolute error (*MAE*), and the symmetric mean absolute percentage error (*SMAPE*) as low as 2.74, 1.90, and 13.59%, respectively, and an R^2 value as high as 96.34%. The "CEEMDAN-FE" decomposition-merging technology proposed in this paper can effectively reduce the instability and high volatility of the original data, overcome data noise, and significantly improve the model's performance in predicting the real-time concentrations of $PM_{2.5}$.

Keywords: $PM_{2.5}$; PM_{10}; O_3; CEEMDAN; FE; BiLSTM; hourly forecast

Citation: Jiang, X.; Wei, P.; Luo, Y.; Li, Y. Air Pollutant Concentration Prediction Based on a CEEMDAN-FE-BiLSTM Model. *Atmosphere* **2021**, *12*, 1452. https://doi.org/10.3390/atmos12111452

Academic Editors: Esther Hontañón and Bernardete Ribeiro

Received: 28 August 2021
Accepted: 1 November 2021
Published: 3 November 2021

1. Introduction

With the acceleration of industrialization and urbanization, the discharge of pollutants, such as exhaust gas, wastewater, and waste, has greatly increased. In addition, air pollution has become an issue of close concern to all countries. The main causes of serious atmospheric pollution are aerosols ($PM_{2.5}$, PM_{10}, etc.) and gases (O_3, etc.). When PM particles exceed the standard, the atmosphere is in a turbid state. When the O_3 concentration exceeds the standard, it causes pollution, such as ash and photochemical smog [1]. In addition, $PM_{2.5}$ blocks the transmission of solar radiation, causing air convection to stagnate, which is not conducive to the diffusion of air pollutants. Rising $PM_{2.5}$ concentrations can greatly reduce visibility, affect people's normal travel and traffic order, and easily cause large-scale car accidents [2]. Therefore, the accurate predictions of real-time $PM_{2.5}$, PM_{10}, and O_3 are of great practical significance to the governments of various countries for implementing air pollution improvement policies, protecting human health, and ensuring normal production and life activities.

At present, the methods of predicting $PM_{2.5}$ mainly include the numerical model method [3], the statistical modeling method [4], the machine learning method [5], and the deep learning method [6]. The numerical model method is mainly based on the aerodynamic theory and the physical and chemical change process, and it uses mathematical methods to establish the dilution and diffusion model of the air pollution concentration to dynamically predict the air quality and the concentration changes of the main pollutants.

Most experts and scholars conduct research on the latter three methods considering factors that may affect $PM_{2.5}$ concentration and establish a single or combined model based on historical $PM_{2.5}$ concentrations to predict the concentrations of $PM_{2.5}$ or other pollutants. To a certain extent, it can make up for the uncertainty of a single numerical model prediction. In terms of statistical modeling, commonly used models include the autoregressive moving average model (ARMA) [7], the autoregressive integrated moving average model (ARIMA) [8], and multiple linear regression (MLR) [9]. As $PM_{2.5}$ concentration is affected by many factors, showing instability and nonlinearity, the statistical modeling methods mentioned above are not accurate in processing nonlinear time series data. Scholars have further studied machine learning methods for $PM_{2.5}$. The common applications of concentration prediction are support vector machines (SVM) [10], random forest, and BP neural networks. In recent years, as deep learning has achieved significant results in different fields, more and more scholars have begun to use deep learning models to predict $PM_{2.5}$ concentrations. The common ones are the recurrent neural network (RNN) [11], long short-term memory (LSTM) [12], and gated recurrent unit (GRU) [13] models. Hybrid models are often more robust than single models, and most of the models studied today are hybrid models. Xiao F et al. established a weighted long short-term memory neural network extended model (WLSTME) model to predict $PM_{2.5}$ concentrations, and proved that the prediction accuracy and reliability of WLSTME are higher than the space–time support vector regression model (STSVR), the long short-term memory neural network extended model (LSTME), and the geographically weighted regression (GWR) [14]. Al-Qaness MAA et al. proposed an improved version of the adaptive neuro-fuzzy inference system (ANFIS) for forecasting the air quality index in Wuhan City [15].

Many scholars have considered adding data decomposition technology to decompose the original data column in order to highlight the time series characteristics of the data and enhance the data characteristics. Singh S. et al. [16] proposed the combination of wavelet decomposition (WD) and ARIMA, Liu S. et al. [17] proposed the combination of WD and least squares support vector regression (LSSVR), and Zheng H. et al. [18] proposed the EMD-LSTM algorithm. These combined models prove that decomposition technology can effectively improve prediction accuracy. Niu M. F. et al. [19] proposed a hybrid model of ensemble empirical mode decomposition (EEMD) and LSSVR, which effectively suppresses modal aliasing caused by traditional decomposition methods when decomposing time series. Weng K. et al. [20] introduced the TPE-XGBoost model and the LASSO–LARS model to high-frequency data and low-frequency data, respectively. The researchers combined air quality factors and meteorological factors to reflect the change trend of decomposition characteristics and predicted the $PM_{2.5}$ concentrations. Sun W. et al. [21] used fast ensemble empirical model decomposition (FEEMD) to decompose the original $PM_{2.5}$ concentration sequence, reorganized the decomposed sequence based on the sample entropy, and then used a general regression neural network (GRNN) and an extreme learning machine (ELM) to predict the recombination sequence, respectively.

At present, no scholar has used complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)-fuzzy entropy (FE) decomposition-merging technology for $PM_{2.5}$ concentration prediction. In this paper, we propose using CEEMDAN to decompose the original concentration sequence to reduce data noise and enhance the periodicity of data changes. Then, according to the FE value, K-means clustering combines the decomposition waves with similar entropy values to further reduce the amount of calculation. Finally, the recombined sequence is input into the bi-directional long short-term memory (BiLSTM) model for prediction. The proposed CEEMDAN-FE-BiLSTM hybrid model predicts the $PM_{2.5}$ concentration of Seoul Station 116 included on the Kaggle website hour by hour and compares it with the prediction effects of other models.

2. Research Methods

2.1. CEEMDAN

CEEMDAN [22] was developed based on the EMD algorithm. EMD [23] is an adaptive orthogonal basis time–frequency signal processing method for unknown nonlinear signals. It decomposes the signal into eigenmode components (*IMF*) with different frequencies. Due to the existence of data noise, traditional EMD decomposition has the phenomenon of modal aliasing. Wu Z. H. [24] proposed an EEMD, which adds zero mean to the original signal every time the signal is decomposed. White noise with fixed variance can effectively improve the modal aliasing phenomenon, but the added Gaussian white noise is difficult to eliminate, and there is a problem of reconstruction error. Based on EEMD, Torres et al. proposed CEEMDAN, adaptively adding and eliminating white noise. This model not only effectively overcomes modal aliasing, but also reduces reconstruction errors, iterations, and calculation costs.

Assuming that the original time series is $X(t)$, the steps of CEEMDAN's decomposition are as follows:

(1) We add n times the Gaussian white noise $X(t)$ of the same length to the original time series data, where $i = 1, 2, \cdots, n$, and ξ_0 are adaptive coefficients. The first modal component IMF_1 is obtained through EMD decomposition; m-many IMF_1 are obtained by m-many experiments and the average value of $\overline{IMF_1}$ is obtained. The process is shown in Equations (1) and (2):

$$X(t) + \xi_0 n^i(t) = IMF_1^i(t) + r_1^i(t) \tag{1}$$

$$\overline{IMF_1} = \frac{1}{m}\sum_{i=1}^{m} IMF_1^i(t) \tag{2}$$

(2) IMF_1 is removed from the original sequence $X(t)$. The remaining time series is marked as $r_1(t)$. The adaptive signal $E_1(n^i(t))$ is calculated by EMD and added to the remaining time series $r_1(t)$. Then, at each new round of EMD decomposition, we repeat m times the average value to obtain $\overline{IMF_2}$. The process is shown in Equations (3)–(5):

$$r_1(t) = X(t) - \overline{IMF_1} \tag{3}$$

$$r_1(t) + \xi_1 E_1(n^i(t)) = {IMF_2}^i(t) + {r_2}^i(t) \tag{4}$$

$$\overline{IMF_2} = \frac{1}{m}\sum_{i=1}^{m} IMF_2^i(t) \tag{5}$$

(3) For the kth component (k = 2, 3, … , n), similar to Step (2), we obtain the kth component $\overline{IMF_k}$. The process is shown in Equations (3)–(5).

$$r_{k-1}(t) = X(t) - \overline{IMF_{k-1}} \tag{6}$$

$$r_{k-1}(t) + \xi_1 E_{k-1}(n^i(t)) = {IMF_k}^i(t) + {r_k}^i(t) \tag{7}$$

$$\overline{IMF_k} = \frac{1}{m}\sum_{i=1}^{m} IMF_k^i(t) \tag{8}$$

(4) We repeat the above steps until the residual component is not suitable to decompose again, and then we stop decomposing. At this time, all $\overline{IMF_s}$ that meet the conditions are extracted, and the trend term is $r_n(t)$:

$$X(t) = \sum_{i=1}^{n} \overline{IMF_i} + r_n(t) \tag{9}$$

2.2. FE

FE [25] is an improvement of sample entropy (SE) [26] and approximate entropy (AE) [27], which is used to measure the complexity of time series. FE introduces the

concept of fuzzy sets, using exponential functions as fuzzy functions to calculate the similarity of vectors. Not only does FE integrate the advantages of sample entropy not dependent on data length, consistency, approximate entropy, strong anti-noise, anti-outlier ability, and so forth, but the introduced fuzzy function also enables FE to solve the problem of sample entropy breakpoints in the calculation and make the value change more stable.

2.3. BiLSTM

RNN can process time series data using neurons with self-feedback. However, as the time series grows, the residual error that RNN needs to return decreases exponentially, resulting in a slow update of the network weights and the problem of gradient disappearance or gradient explosion. Hochreiter S et al. [28] proposed the most primitive LSTM, and Gers et al. [29] proposed adding a forget gate. The information of the preceding and following time steps can be filtered without all steps going through the fully connected layer to form the basic framework of LSTM that is commonly used today. The long-term and short-term neural networks replace the traditional hidden layer with the LSTM layer. It can obtain two kinds of information of the cell state and the hidden layer state from the previous moment. It adopts a control gate mechanism and is composed of memory cells, input gates, output gates, and forgetting gates. Schuster M et al. [30] inherited the construction ideas of LSTM and bidirectional recurrent neural network (BRNN), and constructed BiLSTM. The internal structure of the unit is the same as that of LSTM. The overall network structure of BiLSTM is shown in Figure 1.

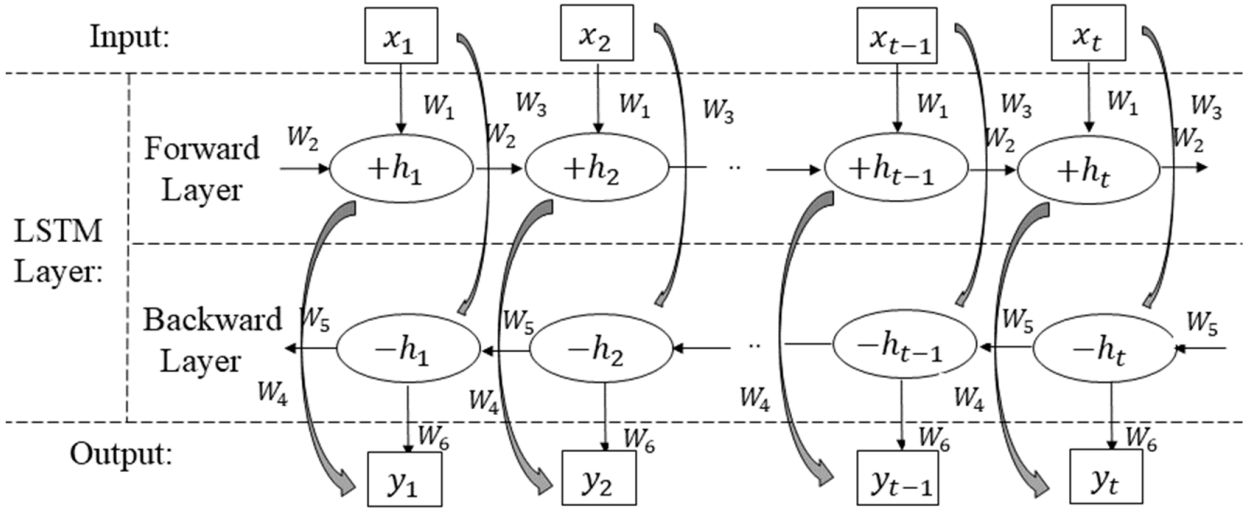

Figure 1. BiLSTM unit structure.

On the basis of the forward layer, where the original information propagates forward from the initial time to time t, the backward layer is added, where the information is propagated back from time t to the initial time. Both layers determine the output at the same time.

2.4. CEEMDAN-FE-BiLSTM

The CEEMDAN prediction model was divided into three parts. The first part is the decomposition part, which uses the CEEMDAN model to decompose the hourly $PM_{2.5}$ concentration to form K-many *IMF* components. The second part is the *IMF* component merging part where (1) the concept of FE was introduced to measure the similarity between *IMF*, (2) K-many FE values were obtained, and (3) K-means clustering was used to add and merge the similar *IMF* sequences to obtain m-many components, which were recorded as $Feat_i$ ($I = 1, 2, \cdots, m$). The third part was the prediction part of the BiLSTM model. BiLSTM is composed of forward LSTM and backward LSTM. The former is responsible for forward feature extraction, and the latter is responsible for reverse feature extraction. The feature information

propagated in these two directions was fused, and the final feature was output to obtain the predicted value of the $PM_{2.5}$ concentration. The modeling process is shown in Figure 2.

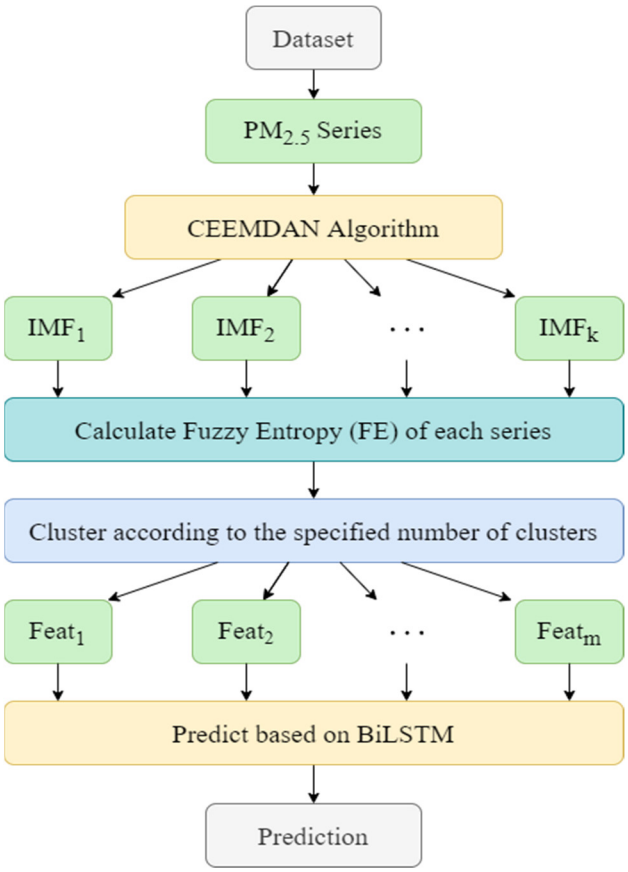

Figure 2. CEEMDAN-FE-BiLSTM model flow chart.

3. Experimental Analysis

3.1. Data Sources

The data in this paper came from the air pollution values recorded in the Seoul dataset from the Kaggle website. The hourly $PM_{2.5}$ concentrations of Station 116 were selected as the research objects. The time range was from 0:00 on 1 January 2017 to 23:00 on 31 December 2019, providing a total of 25,906 data points.

3.2. Evaluation Criteria

To quantitatively evaluate the prediction performance of the model, we selected the root mean square error (*RMSE*), the mean absolute error (*MAE*), the symmetric mean absolute percentage error (*SMAPE*), and R^2 to measure the prediction accuracy and generalization ability of different models. Let y_i be the real value and $\hat{y}_i$ be the model prediction value, where $I = 1,2, \cdots, n$ (n is the number of samples). The expressions of the above evaluation indexes are shown in Equations (10)–(13).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2} \tag{10}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{11}$$

$$SMAPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|)/2} \tag{12}$$

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{13}$$

It can be seen from the definition of the three evaluation indicators that the smaller the *RMSE* and *MAE* values are, the closer the *SMAPE* value is to 0, and the closer the R^2 value is to 1, which means that the prediction error of the model is smaller and the generalization ability is stronger.

3.3. Experimental Setup

In this experiment, we used the hourly $PM_{2.5}$ concentrations over 3 years, with 374 h of data missing in the middle. There were a few missing values, but these missing values had little impact on the model effect. Therefore, no filling was performed, because the $PM_{2.5}$ concentration changed greatly in different time periods. To retain the original information of the data, no discrete values were processed. The first 80% of the data were divided into the training set, and the last 20% were divided into the test set. The training set was used for the key parameter selection and model establishment, and the test set was used for the model prediction effect evaluation.

(1) CEEMDAN parameter setting: We used the CEEMDAN algorithm in the PyEMD package to set different modal numbers, and to decompose and test the training set data. When the modal number was set to 14, the score of each decomposed wave was the most stable.

(2) FE parameter setting: CEEMDAN has more *IMF* components after decomposing the original sequence. We considered combining the components to reduce the amount of calculation for subsequent predictions. The FE value of each *IMF* component was calculated by the concept of FE, and the FE value was similar. The subsequence was reconstructed into a new sequence. According to previous experience [31], we set the embedding dimension m to 2 and the function boundary width r to 0.15, and we calculated the FE value of each IMF_{i} ($I = 1, 2, \cdots, 14$).

(3) BiLSTM prediction: To merge similar sequences more objectively, K-means clustering is used to merge the IMF sequences according to each FE value to form the input sequence of BiLSTM. The parameter settings of BiLSTM are shown in Table 1 after the investigation and case analysis.

Table 1. Key parameters of BiLSTM.

Main Hyperparameter	Set Value
Batch size	12
Number of hidden layer units	32
Hidden layers	2
Learning rate	5×10^{-3}
Max epoch	30
Optimizer	Adam
Loss function	MSE

We input 12 samples to BiLSTM each time, that is, a time window of 12 h, and predicted the $PM_{2.5}$ concentration value of the next hour based on the historical data of the previous 12 h. The learning rate was 5×10^{-3}, and the learning passed through two hidden layers. The number of iterations was set to 30. For the optimization algorithm in the model, compared to stochastic gradient descent, the Adam algorithm used in this study had a faster and more stable convergence rate, and the common MSE was used for the loss function. We selected the optimal number of clusters by searching, set the number of clusters of K-means clustering to $\{1, 2, \cdots, 14\}$, and we examined the new input sequence on the training set under different clustering conditions. In the performance situation, the input sequence corresponding to the minimum MSE was the input of the final BiLSTM prediction, and then the test set was predicted.

3.4. Experimental Results and Analysis

3.4.1. CEEMDAN Modal Decomposition

According to the CEEMDAN modal decomposition method, the original sequence of the $PM_{2.5}$ concentration was decomposed into 14 groups of decomposed waves, as shown in Figure 3.

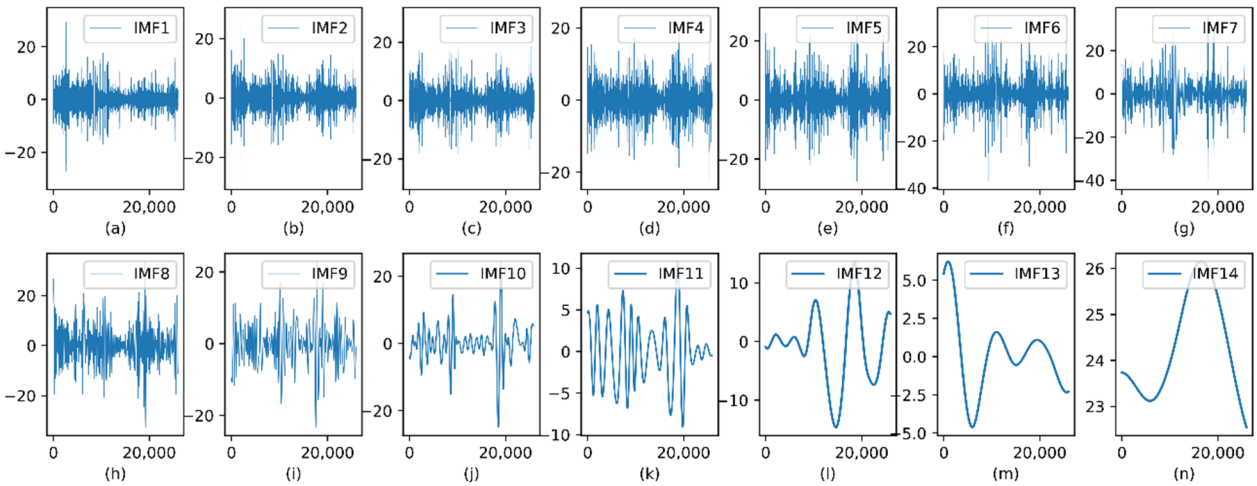

Figure 3. CEEMDAN modal decomposition wave *IMFi*: (**a**–**n**) represent IMF_1–IMF_{14}.

From the IMF_i of Figure 3a–n, these subsequences decomposed by CEEMDAN showed a trend of decreasing frequency, decreasing amplitude, and increasing wavelength from IMF_1 to IMF_{14}. The subsequences showed a certain change rule and period, indicating that the complex sequence of $PM_{2.5}$ concentration was decomposed into subsequences containing information of different scales and gradually reducing noise.

3.4.2. FE Calculation Results

According to the set dimension and function boundary width, the sample entropy of the decomposition wave IMF_i was calculated, which was used to evaluate the degree of confusion between the wavefront parts, that is, the frequency of the wave, to provide a basis for the next step of merging and reorganizing the *IMF* components. The sample entropy of the decomposition wave IMF_i is shown in Table 2.

Table 2. Sample entropy of decomposed wave IMF_i.

Decomposition Sequence IMF_i	FE Value	Decomposition Sequence IMF_i	FE Value
IMF_1	2.610	IMF_8	0.424
IMF_2	2.463	IMF_9	0.160
IMF_3	1.884	IMF_{10}	0.038
IMF_4	1.275	IMF_{11}	0.006
IMF_5	0.915	IMF_{12}	0.001
IMF_6	0.704	IMF_{13}	7.40×10^{-5}
IMF_7	0.578	IMF_{14}	4.82×10^{-6}

The smaller the value of FE is, the more structured the signal's pattern is, and the larger the value is, the more random or unpredictable the signal is. It can be seen from Table 2 that from IMF_1 to IMF_{14}, the FE gradually decreased, which once again shows that the subsequence noise obtained by CEEMDAN decomposition gradually decreased.

3.4.3. BiLSTM Experiment Results

According to the key parameter settings of BiLSTM, the performance of BiLSTM on the training set, when the number of clusters of K-means clustering increased, is shown in Figure 4.

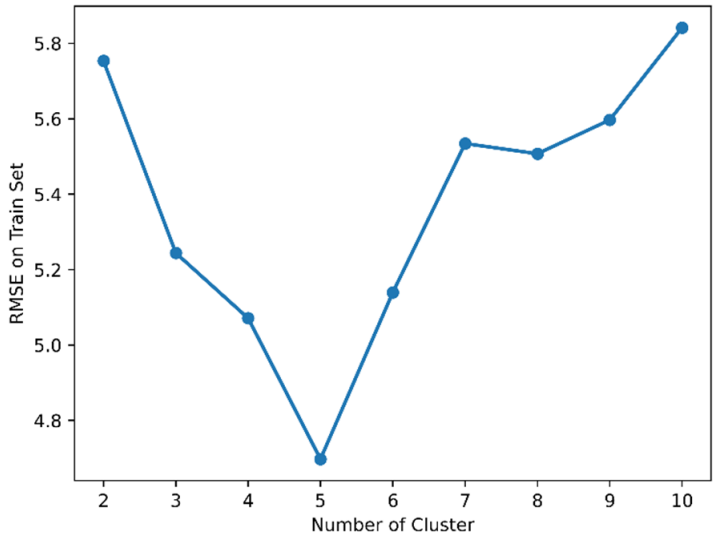

Figure 4. The performance of BiLSTM on the training set.

When the number of K clusters increased, the training error of BiLSTM first decreased and then increased. When there were five K clusters, the training set *RMSE* value was the smallest (4.70). Therefore, the IMF components were combined and reorganized into five reconstruction sequences $Feat_i$ ($I = 1, 2, \cdots, 5$). To reflect the changes more intuitively in the reconstruction sequence, the results of the partial component reconstruction are shown in Figure 5.

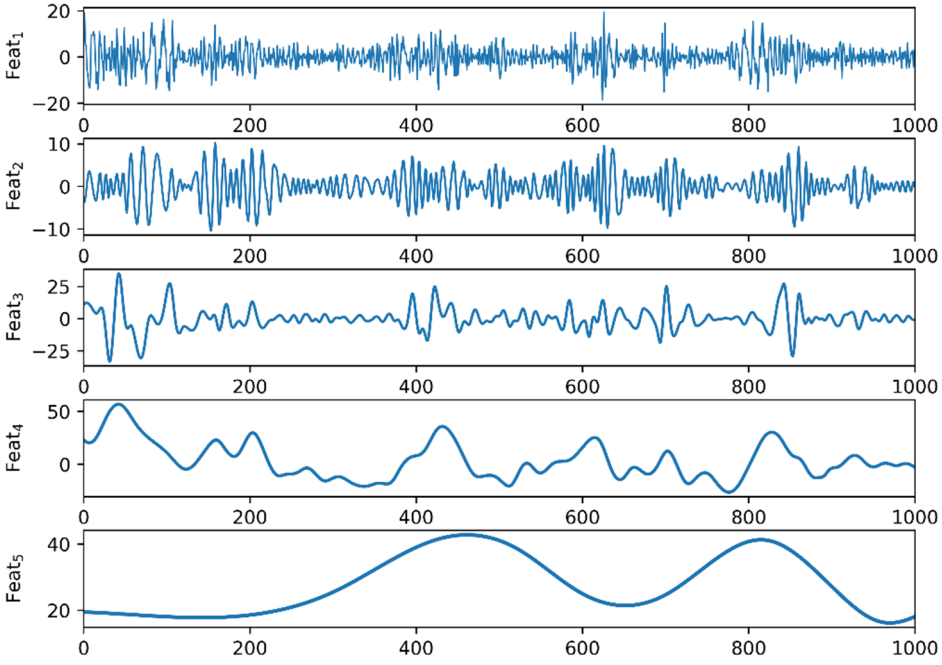

Figure 5. Subsequence after decomposition and reconstruction of CEEMDAN-FE (partial).

As seen in Figure 5, the new sequences after merging and recombination all show a certain periodicity, which indicates that the noise in the sequence after merging and recombining was less than that before being decomposed. Next, we put these five *Feat* components into the BiLSTM for training. As shown in Table 1, we set the following parameters: window_size = 12, batch_size = 12, and max_epoch = 30. The flow of the decomposed-combined *Feat* sequence in BiLSTM is shown in Figure 6.

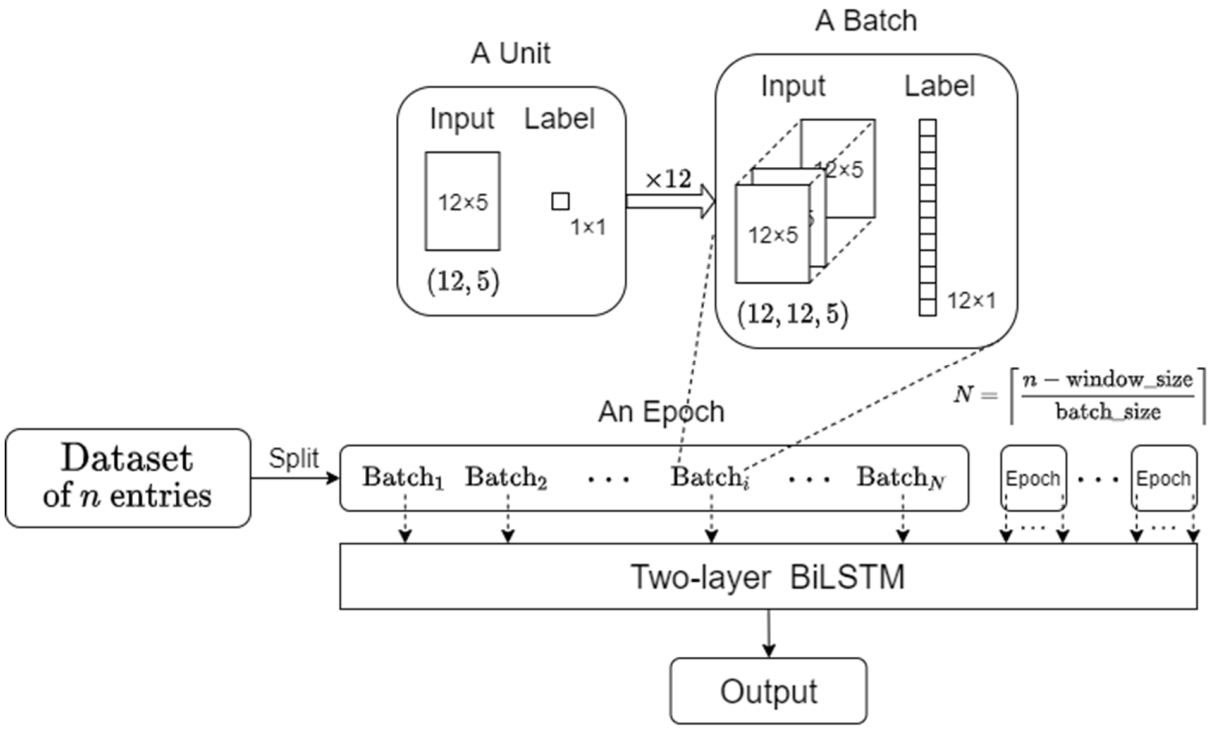

Figure 6. BiLSTM work flow chart.

After the training was completed, we performed a test on the test set to obtain the predicted values of the $PM_{2.5}$ concentrations per hour, compared it with the real values, and calculated the *RMSE*, *MAE*, *SMAP*, and R^2 values of the test set to be 2.75, 1.94, 14.02% and 0.96, respectively. To highlight the effectiveness of the CEEMDAN-FE-BiLSTM model proposed in this paper for $PM_{2.5}$ concentration prediction, we compared the CEEMDAN-FE-BiLSTM model with the other models, and the results are presented in Section 3.5.

3.5. Model Comparison Analysis

We selected CEEMD-BiLSTM, CEEMD-SE-BiLSTM, and CEEMD-AE-BiLSTM as the comparison models, and used the evaluation indicators *RMSE*, *MAE*, *SMAPE*, and R^2 to evaluate the performance of all prediction models.

Table 3 shows the performance evaluation of each hour of the $PM_{2.5}$ concentration predictions for each model on the test set. Figures 7a–c and 8a–c, from the horizontal direction (*RMSE*, *MAE*, *SMAPE*) and the vertical direction (R^2), respectively, compare the prediction effects of each model intuitively.

Table 3. Forecast errors of different models.

Models	*RMSE*	*MAE*	*SMAPE*	R^2
BiLSTM	4.09	2.74	17.49%	91.84%
EMD-BiLSTM	3.37	2.28	16.35%	94.44%
EMD-SE-BiLSTM	3.67	2.62	18.71%	93.43%
EMD-AE-BiLSTM	3.55	2.45	17.18%	93.85%
EMD-FE-BiLSTM	2.97	2.09	15.47%	95.71%
EEMD-BiLSTM	5.08	3.71	22.58%	87.38%
EEMD-SE-BiLSTM *	3.41	2.57	18.64%	94.32%
EEMD-AE-BiLSTM *	3.41	2.57	18.64%	94.32%
EEMD-FE-BiLSTM	3.26	2.45	17.96%	94.81%
CEEMDAN-BiLSTM	3.93	2.92	19.46%	92.47%
CEEMDAN-SE-BiLSTM	3.38	2.30	16.53%	94.42%
CEEMDAN-AE-BiLSTM	3.12	2.18	15.60%	95.24%
CEEMDAN-FE-BiLSTM	2.74	1.90	13.59%	96.34%

* Means: The clustering results of EEMD-AE-BiLSTM and EEMD-SE-BiLSTM are the same. Therefore, the final result is the same.

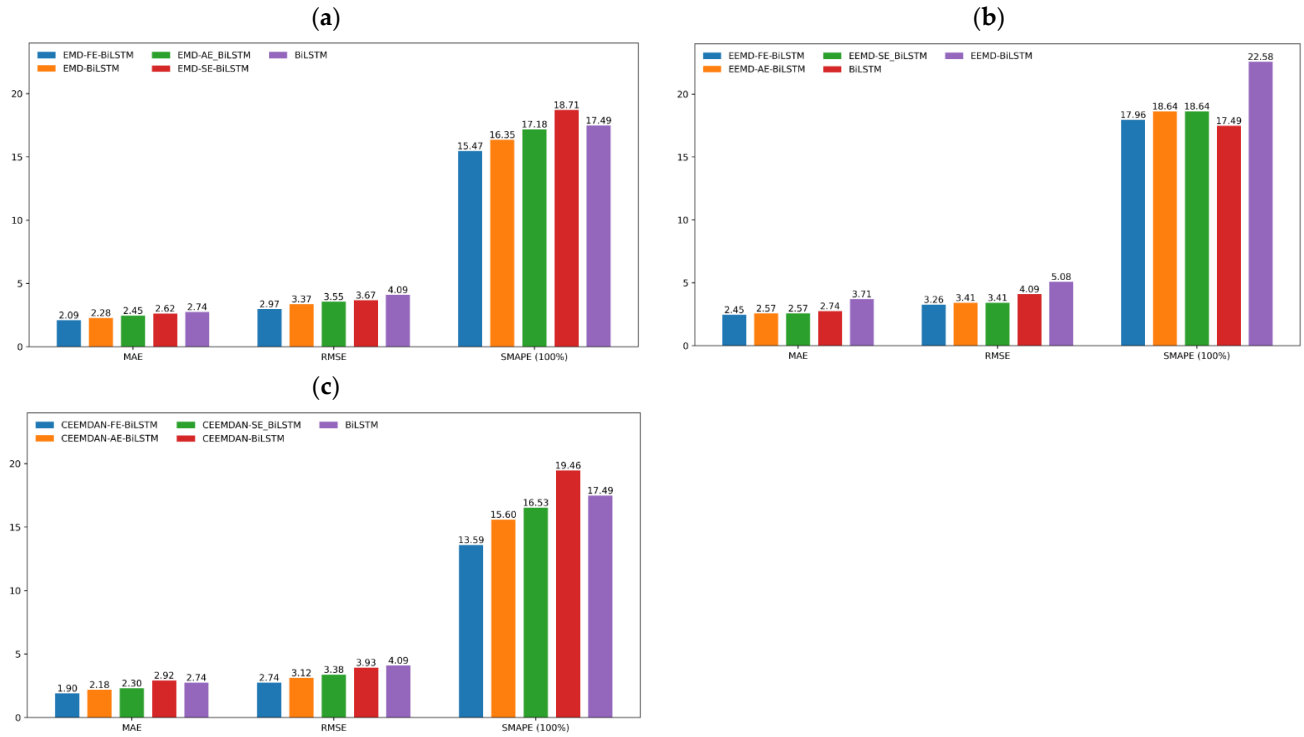

Figure 7. Comparison of horizontal errors of prediction models: (**a**) EMD decomposition; (**b**) EEMD decomposition; (**c**) CEEMDAN decomposition.

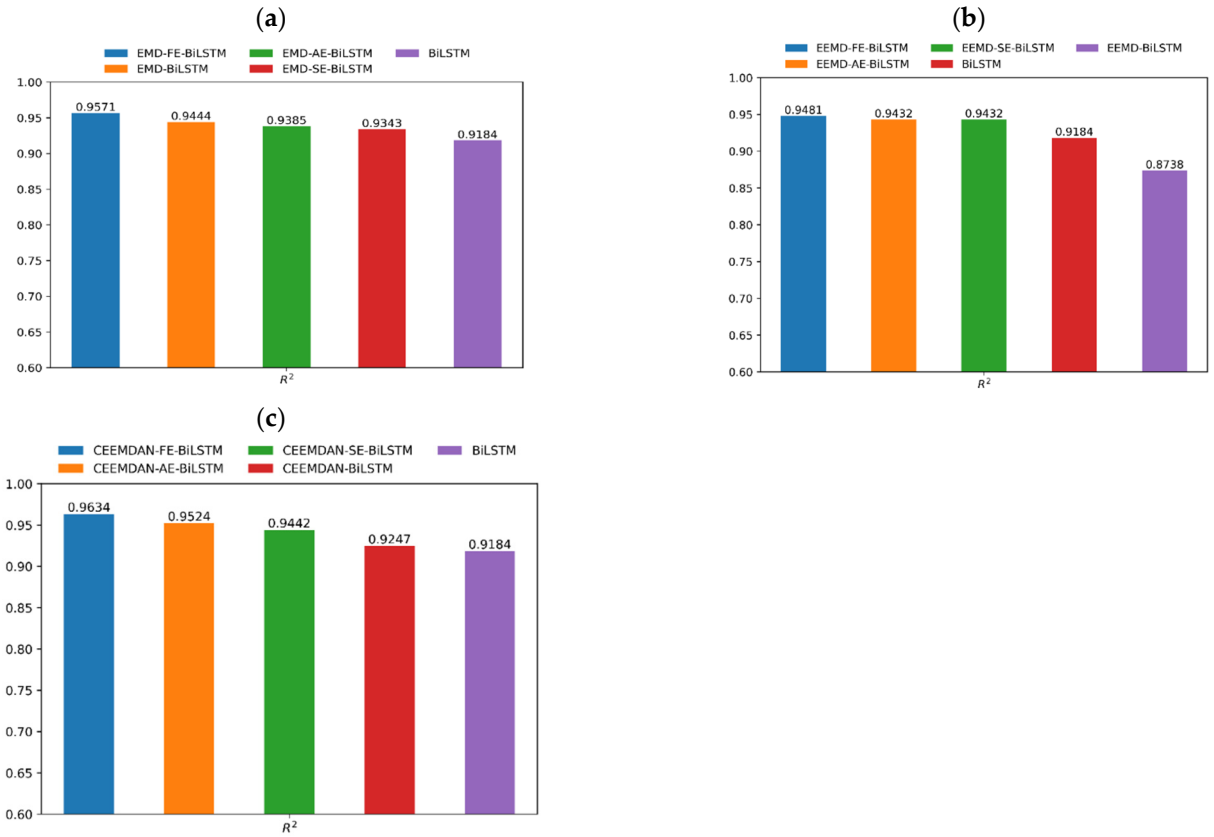

Figure 8. Comparison of goodness of fit of prediction models: (**a**) EMD decomposition; (**b**) EEMD decomposition; (**c**) CEEMDAN decomposition.

The results show the following: (1) The models that used the decomposition-merging technology have significant effects. Models that did not use decomposition technology or merging technology showed that decomposition-merging technology could effectively overcome the non-linearity, large fluctuations, and noise of the $PM_{2.5}$ concentration series. The impact of accuracy significantly improved the predictive ability of the model. (2) In the hybrid model, the CEEMDAN decomposition method was more suitable for the decomposition of the $PM_{2.5}$ concentration sequence than EMD and EEMD. In terms of the decomposition effect, CEEMDAN > EEMD > EMD. (3) FE presented the best merging effect after decomposing the sequence. Compared with SE and AE, the *SMAPE* of the hybrid model using FE was reduced by 12.88% and 17.79%, respectively, and the SE and AE methods had similar effects. In the case of EMD and CEEMDAN decomposing the sequence, the AE merging effect was better. In the EEMD decomposition-merging, SE and AE had the same entropy clustering effect. The result of the merging of 14 identical decomposition sequences was the same. Therefore, the final result was the same. In terms of the merging effect, FE > AE $\geq$ SE. (4) The prediction effect of CEEMDAN-FE-BiLSTM was better than the other models in terms of the horizontal accuracy and goodness of fit. The *RMSE*, *MAE*, and SMPE values were as low as 2.74, 1.90, and 13.59%, respectively, and the R^2 value was as high as 96.34%. (5) Compared with the single-model BiLSTM, the horizontal prediction error *RMSE*, *MAE*, and *SMAPE* values of the CEEMDAN-FE-BiLSTM model were reduced by 33.01%, 30.66%, and 22.30%, respectively, and the goodness of fit was improved by 4.90%. Decomposing the unmerged CEEMDAN-BiLSTM, the CEEMDAN-FE-BiLSTM model's horizontal prediction errors were reduced by 30.28%, 34.93%, and 30.16%, respectively, and the goodness of fit was increased by 4.19%, indicating that the combination of FE values can significantly improve the model's prediction accuracy.

4. Extension Analysis

This section presents our exploration into the general applicability of the CEEMDAN-FE-BiLSTM model. We tested the stability of the hybrid model based on PM_{10} and O_3 concentration sets and compared them with the other models. The data used in this section are the same as those in Section 3.1.

4.1. Predictive Analysis of PM_{10}

The PM_{10} concentrations used in this section were monitored together with $PM_{2.5}$, and the evaluation indicators remained unchanged from the abovementioned *RMSE* and *MAE*. Table 4 shows the performance evaluation of each model on the test set for each hour of the PM_{10} concentration prediction, and Figure 9 presents the corresponding histograms.

Table 4. Prediction error of PM_{10} concentration prediction model.

Model	*RMSE*	*MAE*	*SMAPE*	R^2
CEEMDAN-BiLSTM	8.01	5.72	21.94%	89.87%
CEEMDAN-SE-BiLSTM	7.78	4.97	18.55%	90.44%
CEEMDAN-AE-BiLSTM	7.14	4.37	16.27%	91.96%
CEEMDAN-FE-BiLSTM	5.64	3.57	14.05%	94.98%

From Figure 9, compared to the prediction model of $PM_{2.5}$ concentrations, although the prediction accuracy of the model when predicting PM_{10} was reduced, the difference between the *MAE* and *RMSE* was less than 4. Regardless of the level of accuracy or the goodness of fit, the CEEMDAN-FE-BiLSTM model remains the model with the best predictive effect, with *RMSE*, *MAE*, and *SMAPE* values as low as 5.64, 3.57, and 14.05%, and an R^2 value as high as 94.98%. The hybrid model that does not use the entropy value to merge the decomposition sequence had the worst effect, and the model effect of using the entropy value to merge the decomposition sequence FE > AE > SE once again proved the effectiveness of the FE model merging. In summary, the same model had the same

effect on the predictions of PM_{10} and $PM_{2.5}$, which proves the effectiveness and accuracy of CEEMDAN-FE-BiLSTM in predicting similar particles.

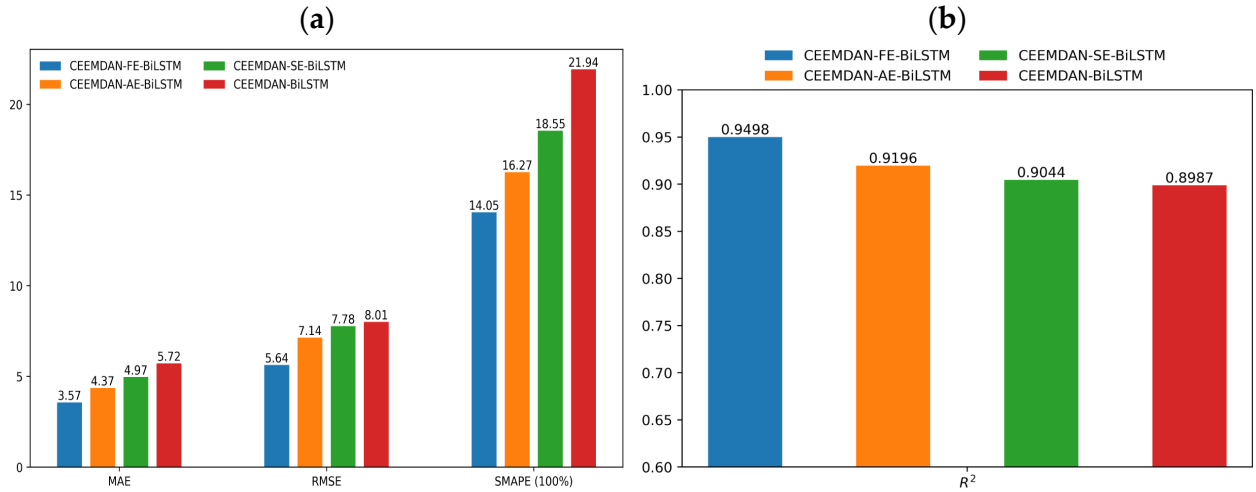

Figure 9. Comparison of PM_{10} concentration prediction model: (**a**) horizontal error; (**b**) goodness of fit.

4.2. Predictive Analysis of O_3

Section 4.1 proves the applicability of the CEEMDAN-FE-BiLSTM hybrid model in PM particulate matter prediction. As described in this section, we selected a gas O_3 dataset that is different from $PM_{2.5}$, used the same model in Section 4.1 to predict the hourly concentration of O_3, and used the same evaluation index for evaluation. In addition, the stability of model prediction was further explored.

Table 5 shows the performance evaluation of each model on the test set for each hour of O_3 concentration prediction, and Figures 10 and 11 present the corresponding histograms. Continuing the performance of predicting the concentrations of $PM_{2.5}$ and PM_{10}, out of the four evaluation indicators used, CEEMDAN-FE-BiLSTM was significantly better than the other models. Moreover, CEEMDAN-FE-BiLSTM was the best model for predicting the hourly concentration of O_3, except for the value of *SMAPE*. In addition to the obvious increase, the *RMSE* and *MAE* values were still very low, as low as 0.0044 and 0.0036, respectively, and the R^2 value was as high as 95.61%. The model that did not use entropy decomposition had the worst effect. Unlike for predicting PM particles, the effect of using SE value decomposition was significantly better than that of the AE decomposition. Compared to CEEMDAN-AE-BiLSTM, the CEEMDAN-SE-BiLSTM model showed that horizontal prediction error *RMSE*, *MAE*, and *SMAPE* values decreased by 40.71%, 42.62%, and 25.79%, respectively, and the goodness of fit increased by 37.47%. Compared to the CEEMDAN-SE-BiLSTM model with the second-best prediction effect, the horizontal prediction error *RMSE*, *MAE*, and *SMAPE* values of the CEEMDAN-FE-BiLSTM model were reduced by 46.99%, 37.69%, and 12.10%, respectively. Therefore, in the gas prediction, the FE decomposition also plays a decisive role in the accuracy of the model. In summary, for O_3 prediction, the prediction accuracy of CEEMDAN-FE-BiLSTM remained the highest, and the goodness of fit was much better than the other three models. The other three models presented large fluctuations, which proves that the CEEMDAN-FE-BiLSTM hybrid demonstrated the best stability in its model predictions.

Table 5. Prediction error of O_3 concentration prediction model.

Model	*RMSE*	*MAE*	*SMAPE*	R^2
CEEMDAN-BiLSTM	0.0161	0.0140	63.67%	40.44%
CEEMDAN-SE-BiLSTM	0.0083	0.0070	43.99%	84.04%
CEEMDAN-AE-BiLSTM	0.0140	0.0122	59.28%	55.05%
CEEMDAN-FE-BiLSTM	0.0044	0.0036	27.41%	95.61%

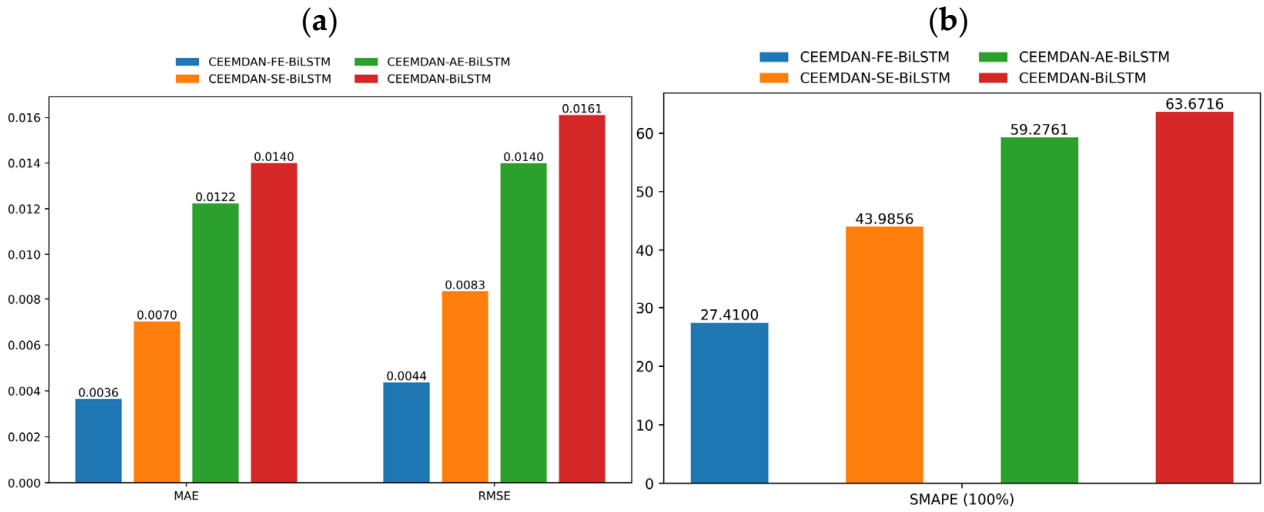

Figure 10. Comparison of horizontal error of O_3 concentration prediction model: (**a**) *MAE* and *RMSE*; (**b**) *SMAPE*.

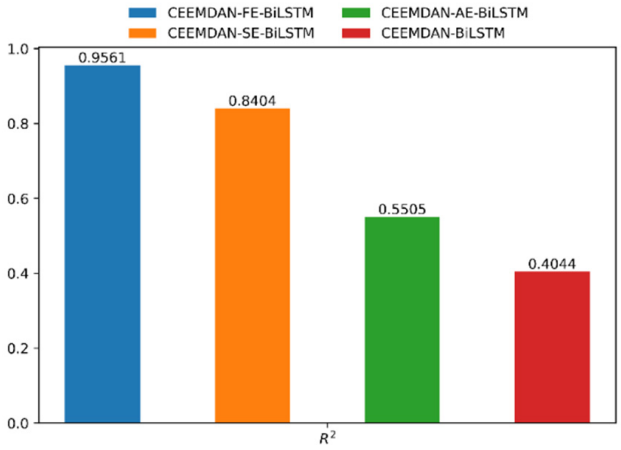

Figure 11. Comparison of the goodness of fit of the O_3 concentration prediction model.

5. Conclusions

The $PM_{2.5}$ concentration sequence has the characteristics of non-linearity, non-stationary, and a lot of noise. We propose a CEEMDAN-FE-BiLSTM hybrid model based on the decomposition-merge technology to predict $PM_{2.5}$ concentrations hour by hour. First, the CEEMDAN algorithm is used to decompose the original $PM_{2.5}$ concentration sequence to obtain 14 *IMF* components; then, the FE values of the 14 *IMF* components are calculated according to the FE definition, the number of clusters is set, and the K-means clustering is based on the FE value. *IMF* components are merged to obtain a new component *Feat*, which is input into BiLSTM, and the optimal number of clusters is five, according to the training effect. Finally, five new components are input into the BiLSTM model to predict $PM_{2.5}$ concentrations hour by hour. In order to prove the validity and stability of the CEEMDAN-FE-BiLSTM model, it was used to predict PM_{10} and O_3. The experimental results show the following: (1) Decomposing the original sequence using the CEEMDAN algorithm can effectively remove noise and extract timing information. (2) Using entropy values to recombine *IMF* sequence can significantly improve the prediction performance of the BiLSTM model. In different cases, SE and AE had different effects on the combination of sequences. Whether for PM particles or gas particles, the prediction effect of BiLSTM after using the FE value to recombine the *IMF* sequence was significantly better than the first two, that is, FE combination plays a decisive role in improving the goodness-of-fit of the model. (3) Regardless of whether it is. predicting similar or heterogeneous pollutants, the CEEMDAN-FE-BiLSTM model is significantly better than the other models in terms of

the horizontal accuracy and goodness of fit, with little fluctuation and a stable prediction effect. The "CEEMDAN-FE" decomposition-merging technology proposed in this paper can effectively reduce the instability and high volatility of the original data, overcome data noise, and significantly improve the model's performance in predicting the real-time concentrations of $PM_{2.5}$.

Although the proposed CEEMDAN-FE-BiLSTM hybrid model can solve the irregular and unstable characteristics of $PM_{2.5}$ concentration sequences and improve the prediction accuracy of a $PM_{2.5}$ concentration sequence, there are still many problems to be solved. First of all, in this study, we only considered hourly forecasting. Next, we can explore the prediction accuracy of the model for the next 12 h and 24 h to further enhance the broadness of the model's applicability. Secondly, we only considered the single series prediction without considering the factors affecting its change, such as wind speed, temperature, precipitation, and so on. If influencing factors are added, the prediction accuracy of the model may be improved again.

Author Contributions: Conceptualization, X.J. and Y.L. (Yiwen Luo); methodology, P.W. and Y.L. (Ying Li); formal analysis, Y.L. (Yiwen Luo) and P.W.; data curation, P.W.; supervision, X.J.; writing—original draft preparation, Y.L. (Ying Li) and P.W.; writing—review and editing, X.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data and methods used in the research have been presented in sufficient detail in the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Guérette, E.A.; Chang, L.T.C.; Cope, M.E.; Duc, H.N.; Emmerson, K.M.; Monk, K.; Rayner, P.J.; Scorgie, Y.; Silver, J.D.; Simmons, J. Evaluation of Regional Air Quality Models over Sydney, Australia: Part 2, Comparison of $PM_{2.5}$ and Ozone. *Atmosphere* **2020**, *11*, 233. [CrossRef]
2. Olukanni, D.; Enetomhe, D.; Bamigboye, G.; Bassey, D. A Time-Based Assessment of Particulate Matter ($PM_{2.5}$) Levels at a Highly Trafficked Intersection: Case Study of Sango-Ota, Nigeria. *Atmosphere* **2021**, *12*, 532. [CrossRef]
3. Yin, X.; Huang, Z.; Zheng, J.; Yuan, Z.; Zhu, W.; Huang, X.; Chen, D. Source contributions to $PM_{2.5}$ in Guangdong province, China by numerical modeling: Results and implications. *Atmos. Res.* **2017**, *186*, 63–71. [CrossRef]
4. Shoji, H.; Tsukatani, T. Statistical model of air pollutant concentration and its application to the air quality standards. *Atmos. Environ.* **1973**, *7*, 485–501. [CrossRef]
5. Liu, J.; Weng, F.; Li, Z. Satellite-based $PM_{2.5}$ estimation directly from reflectance at the top of the atmosphere using a machine learning algorithm. *Atmos. Environ.* **2019**, *208*, 113–122. [CrossRef]
6. Altikat, S. Prediction of CO_2 emission from greenhouse to atmosphere with artificial neural networks and deep learning neural networks. *Int. J. Environ. Sci. Technol.* **2021**, *18*, 3169–3178. [CrossRef]
7. Choi, B.S. *ARMA Model Identification*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
8. Zhang, L.; Lin, J.; Qiu, R.; Hu, X.; Zhang, H.; Chen, Q.; Tan, H.; Lin, D.; Wang, J. Trend Analysis and Forecast of $PM_{2.5}$ in Fuzhou, China Using the ARIMA Model. *Ecol. Indic.* **2018**, *95*, 702–710. [CrossRef]
9. Venkataraman, V.; Usmanulla, S.; Sonnappa, A.; Sadashiv, P.; Mohammed, S.S.; Narayanan, S.S. Wavelet and multiple linear regression analysis for identifying factors affecting particulate matter $PM_{2.5}$ in Mumbai City, India. *Int. J. Qual. Reliab. Manag.* **2019**, *36*, 1750–1783. [CrossRef]
10. Zhou, Y.; Chang, F.J.; Chang, L.C.; Kao, I.F.; Wang, Y.S.; Kang, C.C. Multi-output Support Vector Machine for Regional Multi Step-ahead $PM_{2.5}$ Forecasting. *Sci. Total Environ.* **2019**, *651*, 230–240. [CrossRef] [PubMed]
11. Zaremba, W.; Sutskever, I.; Vinyals, O. Recurrent neural network regularization. *arXiv* **2014**, arXiv:1409.2329.
12. Graves, A. Long short-term memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 37–45.
13. Dey, R.; Salem, F.M. Gate-variants of gated recurrent unit (GRU) neural networks. In Proceedings of the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), Boston, MA, USA, 6–9 August 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1597–1600.

14. Xiao, F.; Yang, M.; Fan, H.; Fan, G.; Al-Qaness, M.A. An improved deep learning model for predicting daily $PM_{2.5}$ concentration. *Sci. Rep.* **2020**, *10*, 20988. [CrossRef]
15. Al-Qaness, M.A.; Fan, H.; Ewees, A.A.; Yousri, D.; Abd Elaziz, M. Improved ANFIS model for forecasting Wuhan City air quality and analysis COVID-19 lockdown impacts on air quality. *Environ. Res.* **2021**, *194*, 110607. [CrossRef] [PubMed]
16. Singh, S.; Parmar, K.S.; Kumar, J.; Makkhan, S.J.S. Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19. *Chaos Soliton Fract.* **2020**, *135*, 109866. [CrossRef] [PubMed]
17. Singh, S.; Parmar, K.S.; Kumar, J.; Makkhan, S.J.S. A hybrid WA–CPSO-LSSVR model for dissolved oxygen content prediction in crab culture. *Eng. Appl. Artif. Intell.* **2014**, *29*, 114–124.
18. Zheng, H.; Yuan, J.; Chen, L. Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. *Energies* **2017**, *10*, 1168. [CrossRef]
19. Niu, M.F.; Gan, K.; Sun, S.L.; Li, F.Y. Application of decomposition-ensemble learning paradigm with phase space reconstruction for day-ahead $PM_{2.5}$ concentration forecasting. *J. Environ. Manag.* **2017**, *196*, 110–118. [CrossRef]
20. Kerui, W.; Miao, L.; Qian, L. An integrated prediction model of $PM_{2.5}$ concentration based on TPE-XGBOOST and LassoLars. *Syst. Eng. Theory Pract.* **2020**, *40*, 748–760.
21. Sun, W.; Li, Z. Hourly $PM_{2.5}$ Concentration Forecasting Based on Mode Decomposition Recombination Technique and Ensemble Learning Approach in Severe Haze Episodes of China. *J. Clean. Prod.* **2020**, *263*, 121442. [CrossRef]
22. Cao, J.; Li, Z.; Li, J. Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A* **2019**, *519*, 127–139. [CrossRef]
23. Rilling, G.; Flandrin, P.; Gonçalves, P.; Lilly, J.M. Bivariate empirical mode decomposition. *IEEE Signal. Proc. Lett.* **2007**, *14*, 936–939. [CrossRef]
24. Wu, Z.H.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [CrossRef]
25. Tran, D.; Wagner, M. Fuzzy entropy clustering. In Proceedings of the Ninth IEEE International Conference on Fuzzy Systems, San Antonio, TX, USA, 7–10 May 2000; FUZZ-IEEE 2000 (Cat. No. 00CH37063). IEEE: Piscataway, NJ, USA, 2000; Volume 1, pp. 152–157.
26. Yentes, J.M.; Hunt, N.; Schmid, K.K.; Kaipust, J.P.; McGrath, D.; Stergiou, N. The appropriate use of approximate entropy and sample entropy with short data sets. *Ann. Biomed. Eng.* **2013**, *41*, 349–365. [CrossRef] [PubMed]
27. Wang, D.; Zhong, J.; Shen, C.; Pan, E.; Peng, Z.; Li, C. Correlation dimension and approximate entropy for machine condition monitoring: Revisited. *Mech. Syst. Signal. Process.* **2021**, *152*, 107497. [CrossRef]
28. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
29. Gers, F.A.; Jürgen, S.; Cummins, F. Learning to Forget: Continual Prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [CrossRef]
30. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal. Process.* **1997**, *45*, 2673–2681. [CrossRef]
31. Li, J.; Li, D.C. Short-term wind power forecasting based on CEEMDAN-FE-KELM method. *Inf. Control* **2016**, *45*, 135–141.