

Article

# Prediction of PM<sub>2.5</sub> Concentration on the Basis of Multi-Time Scale Fusion

Jianfei Zhang \* and Wangui Xia

School of Computer & Information Engineering, Heilongjiang University of Science & Technology, Harbin 150027, China; xiaofangdui@163.com

\* Correspondence: zjfnfu2008@163.com

**Abstract:** Long-term prediction of hour-concentration of PM<sub>2.5</sub> (particles in atmospheric suspension with effective dimensions equal or lower than 2.5 microns) is of great significance for environmental protection and people's health. At present, the prediction of hour-concentration of PM<sub>2.5</sub> is mostly single-step prediction, which is to predict PM<sub>2.5</sub> concentration at a future time point based on a period of historical data. In this paper, a model based on multi-time scale fusion is proposed to study single-step prediction and multi-step prediction, respectively. Experimental results show that the proposed model is better than stacked LSTM and CNN-LSTM in predicting PM<sub>2.5</sub> hour-concentration.

**Keywords:** PM<sub>2.5</sub> concentration prediction; multi-time scale fusion; time series



**Citation:** Zhang, J.; Xia, W. Prediction of PM<sub>2.5</sub> Concentration on the Basis of Multi-Time Scale Fusion. *Processes* **2022**, *10*, 171. <https://doi.org/10.3390/pr10010171>

Academic Editors: Daniele Sofia and Paolo Trucillo

Received: 20 December 2021

Accepted: 11 January 2022

Published: 17 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

At present, the traditional methods in the field of PM<sub>2.5</sub> concentration prediction research mainly combine four conventional methods formed by meteorology, environmental science, mathematics, and computational science. That is, empirical model prediction based on historical data and statistical methods, probability model prediction based on statistical and mathematical methods or models, prediction based on synthetic methods, and prediction based on conventional machine learning models. With the rapid development of deep learning, Fan et al. used the recurrent neural network model for predicting PM<sub>2.5</sub> concentration in the future 1 hour based on air quality and meteorological data in the past 48 h [1]. Qi et al. proposed the model GCN-LSTM (A model based on Graph Convolutional Network and Long Short-Term Memory), which proved that the model was superior to CNN (Convolutional Neural Networks) and LSTM in predicting the air quality in the future one hour [2]. He et al. combined the wavelet transform with the LSTM (Long Short-Term Memory) model and took the daily average concentration as the input to predict the pollutant concentration of the next day, and proved that the proposed model was superior to MLR (Mixed Logistic Regression), LSTM and WT-MLR (Mixed Logistic Regression based on Wavelet Transform) [3]. Huang and Kuo constructed the model APNet (Attention-based Parallel Networks) and proved through experiments that the model was superior to CNN and LSTM in predicting PM<sub>2.5</sub> concentration in the future one hour [4].

However, the above studies only focus on PM<sub>2.5</sub> concentration of single-step prediction, did not predict the PM<sub>2.5</sub> concentration for a period of time in the future, that is, did not make a multi-step prediction. To solve this problem, this article builds a CNN-LSTM network on the basis of the combination of attention mechanism, the multi-time scale fusion model of multi-time scale features is integrated. The aim is to accurately predict the value of PM<sub>2.5</sub> corresponding to each hour in a continuous period of time in the future. Through experiments, the validity and superiority of the method proposed in this paper are verified.

## 2. PM2.5 Prediction Model Based on Multi-Time Scale Fusion

### 2.1. LSTM (Long Short-Term Memory)

LSTM is a variant produced to solve the long-term dependence problem that RNN (Recurrent Neural Network) cannot solve, which effectively alleviates the gradient explosion problem that RNN cannot avoid and can better predict the time series [5]. The architecture of an LSTM memory cell is shown in Figure 1, where each cell has three “gate” structures, include, the input gate, the forget gate, and the output gate. A chain of repeating cells forms the LSTM layer. The calculation process of the spatiotemporal feature matrix  $X = [x_1, x_2, \dots, x_t]$  in the LSTM layer is given in Equations (1)–(6). Equation (1) represents the forget gate and it decides what information should be thrown away from the cell state. The directions are: Input  $h_{t-1}$  and  $x_t$  into the forget gate, and calculate the output value  $f_t$  of the forget gate through the sigmoid activation function. Equations (2) and (3) represent the input gate, which decides what new information should be stored in the state of cell. The directions are: Input  $h_{t-1}$  and  $x_t$  into the input gate, and get  $i_t$  and  $\tilde{c}_t$  through the sigmoid activation function and tanh activation function respectively. Equation (4) uses the output of the forget gate and the input gate to update the current cell state. Equations (5) and (6) together constitute the output of the current cell. The directions are: First, input  $h_{t-1}$  and  $x_t$  into the output gate, and calculate the output  $o_t$  of the output gate through the sigmoid activation function. Then get the current cell output  $h_t$  by calculating the output of the output gate and the state of the current cell.

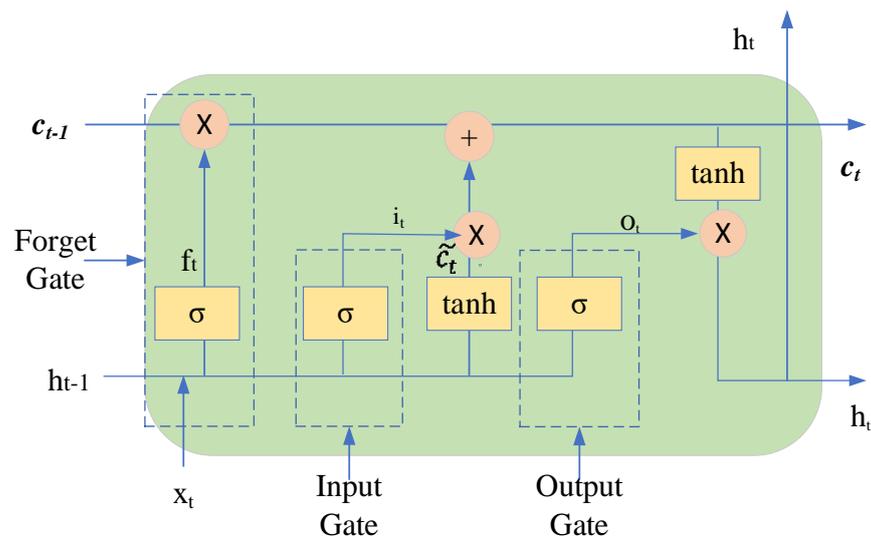


Figure 1. The structure of LSTM neurons.

The following Equations (1)–(6) describe the internal calculation process of an LSTM neural unit:

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{c}_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (4)$$

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

where  $f_t$  is the output of forget gate, the value range of  $f_t$  is (0,1);  $i_t$  is the output of input gate, the value range of  $i_t$  is (0,1);  $c_t$  is the state of the current cell;  $o_t$  is the output of output gate, the value range of  $o_t$  is (0,1);  $h_{t-1}$  is the output of the previous cell;  $h_t$  is the output of the current cell;  $w_f$ ,  $w_i$ ,  $w_c$ , and  $w_o$  are the weight matrices for input vector  $x_t$  at

time step  $t$ ;  $b_f$ ,  $b_i$ ,  $b_c$ , and  $b_o$  are the bias vectors;  $\sigma$  is sigmoid activation function;  $\tanh$  is hyperbolic sine function;  $\odot$  stands for element-wise multiplication of the matrix;  $\otimes$  stands for multiplication;  $\oplus$  stands for the sum operation;

### 2.2. Ensemble Empirical Mode Decomposition (EEMD)

As a noise-assisted signal decomposition method, EEMD adds white noise to the original signal and performs EMD decomposition on it, and finally calculates lumped average using the results of multiple decomposition [6].

The specific operation steps are as follows:

- (1) Set the overall average times  $M$ ;
- (2) Add a white noise  $n_i(t)$  with standard normal distribution to the original signal  $x(t)$  to generate a new signal:

$$x_i(t) = x(t) + n_i(t) \tag{7}$$

where  $n_i(t)$  is  $i$ -th additive white noise sequence;  $x_i(t)$  is the additional noise signal of the  $i$ -th test,  $i = 1, 2, 3, \dots, M$ .

- (3) EMD decomposition is performed on the obtained signal  $x_i(t)$  containing noise to obtain the form of their respective IMF (Intrinsic Mode Function) sum:

$$x_i(t) = \sum_{j=1}^J c_{i,j}(t) + r_{i,j}(t) \tag{8}$$

where  $c_{i,j}(t)$  is the  $J$ -th IMF obtained by decomposing after adding white noise for the  $i$ -th time.  $r_{i,j}(t)$  is the residual term represents the average trend of the signal, and  $j$  is the number of IMF;

- (4) Repeat steps (2) and (3) for  $M$  times, decompose and add white noise signals with different amplitudes each time, and the set of IMF is:  $c_{1,j}(t), c_{2,j}(t), \dots, c_{M,j}(t)$ , where  $j = 1, 2, 3, \dots, J$ ;

- (5) Based on the principle that the statistical average value of unrelated sequences is zero, the above IMF is calculated by aggregate average to obtain the final IMF, namely:

$$c_j(t) = \frac{1}{M} \sum_{i=1}^M c_{i,j}(t) \tag{9}$$

where  $c_j(t)$  is the  $j$ -th IMF,  $i = 1, 2, \dots, M$ ,  $j = 1, 2, \dots, J$ ;

### 2.3. Attention Mechanism

The attention mechanism mimics the internal process of biological observation behavior [7]. His principle is through a set of weights  $\alpha_{T_{s-e}}^{T_t} = [\alpha_{T_e}^{T_t}, \alpha_{T_{e+1}}^{T_t}, \dots, \alpha_{T_s}^{T_t}]$  to express the value of a certain time slice in the target sequence  $x_{T_t}$  and the dependent sequence  $x_{T_{s-e}} = [x_{T_e}, x_{T_{e+1}}, \dots, x_{T_s}]$  relevance. Each element in  $x_{T_t}$  and  $x_{T_{s-e}}$  has the same dimension. Map  $x_{T_t}$  and  $x_{T_{s-e}}$  to the parameter space:

$$\text{Query} = x_{T_t} W_Q \tag{10}$$

$$\text{Key} = x_{T_{s-e}} W_K \tag{11}$$

$$\text{Value} = x_{T_{s-e}} W_V \tag{12}$$

where  $W_Q$  is  $d_x \times d_q$  dimensional Query parameter matrix;  $W_K$  is  $d_x \times d_k$  dimensional Key parameter matrix;  $W_V$  is  $d_x \times d_v$  dimensional Value parameter matrix;

The attention mechanism is divided into three stages: in the first stage, the target sequence is mapped from  $x_{T_t}$  map of  $d_x$  dimension to Query of  $d_q$  dimension, and similarly transformed  $x_{T_{s-e}}$  into matrix mapping to Key matrix with  $d_k$  element dimension and Value matrix with  $d_v$  element dimension, calculating the similarity between Query

and Key; In the second stage, the original score of the first stage is normalized, and the  $\alpha_{T_s-e}^{T_t}$  weight of Value is calculated by Softmax. In the third stage, the Value is weighted and summed according to the weight coefficient to obtain the attention Value.

2.4. Multi Time Scale Fusion Model

In this paper, the multi-time scale fusion model is applied to the prediction of PM2.5 hour-concentration for the first time, and the model process is shown in Figure 2. EEMD (Ensemble Empirical Mode Decomposition) decomposition can decompose the original PM2.5 sequence into new sequences with different time scales. CNN-LSTM was employed to extract characteristic information of time series. Attention\_layer pays attention to important features and ignores non-important features through attention mechanism to improve prediction accuracy.

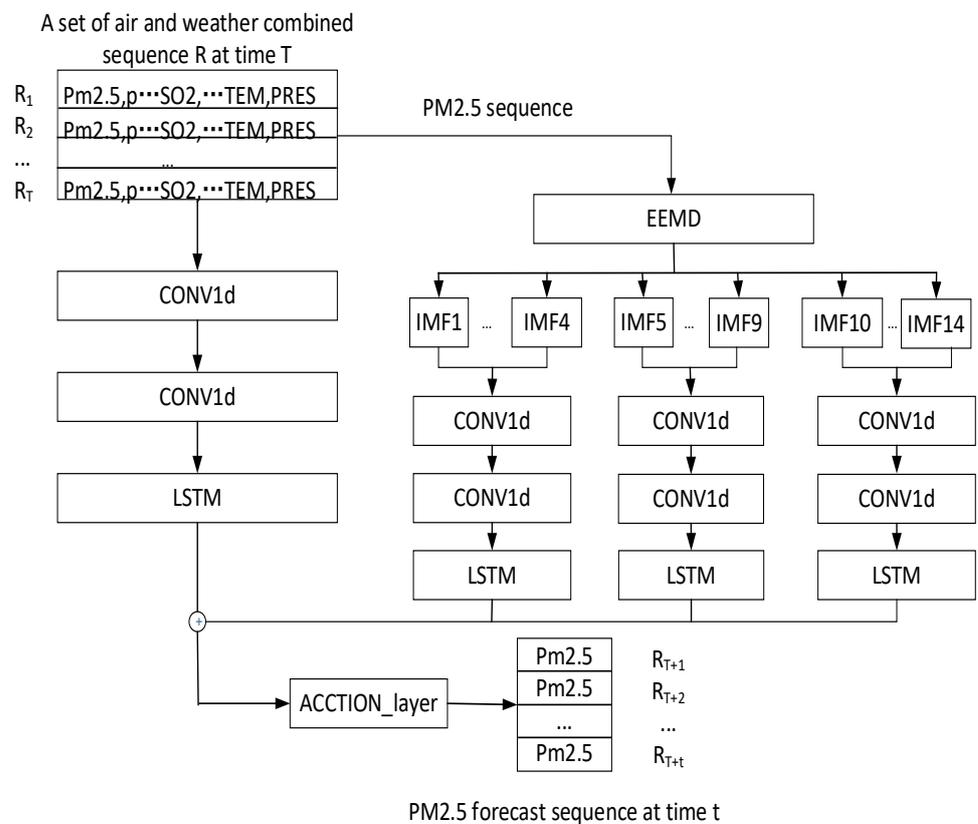


Figure 2. System flow of multi-time scale network model.

The specific steps are as follows:

(1) Input the original PM2.5 sequence into the EEMD model, and perform EEMD decomposition on the original PM2.5 concentration data. This is the first improvement made by the model in this paper on the basis of CNN-LSTM model. Compared with the original sequence, the decomposed sequence can more precisely express the period of the original sequence and better obtain information of different time scales.

(2) The original PM2.5 data sequence and the decomposed PM2.5 sequence were input into CNN-LSTM network composed of two layers of Conv1d and one layer of LSTM respectively for feature extraction. As convolutional neural network has excellent feature extraction and feature expression capabilities, LSTM has natural advantages in processing time sequence. Therefore, CNN and LSTM are used in combination in feature extraction in this paper. In this paper, the decomposed sequences are recombined into new sequences according to different time scales and used as the input of different network layers respectively with the original sequence.

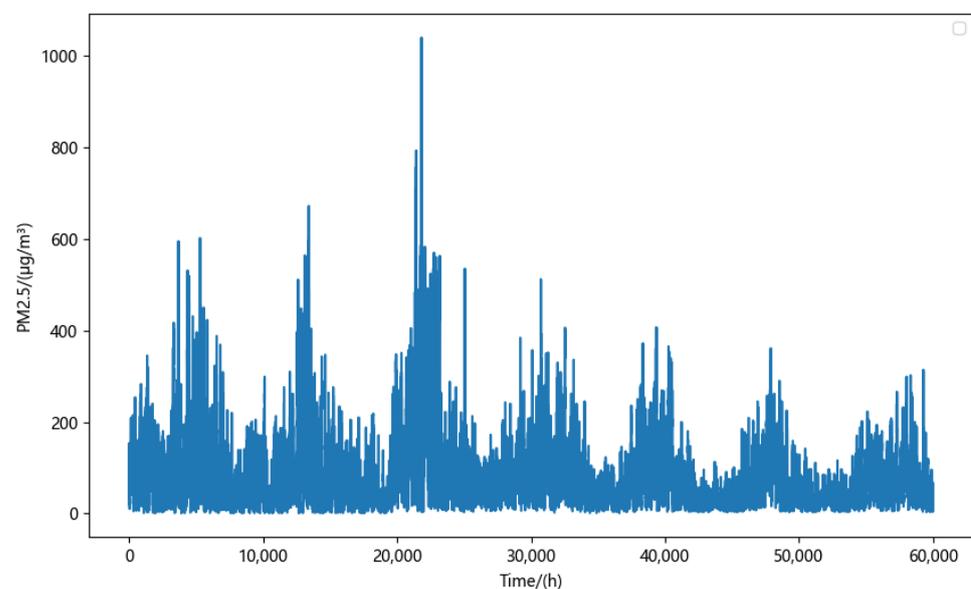
(3) The outputs of different LSTM layers output the prediction results through the attention mechanism layer. Attention mechanism is another improvement based on CNN-LSTM. Through attention mechanism, more important feature information can be paid attention to in features of different time scales to improve the accuracy of prediction.

### 3. Results and Discussion

#### 3.1. Experimental Configuration and Data Set Description

The experimental environment of this paper uses TensorFlow + Keras framework, Python 3.7 development language, the system uses Windows, with multiple Python library functions for code implementation and result analysis.

The data in this paper are the monitoring data from ground stations in Harbin, mainly including AQI, PM2.5, PM10, O<sub>3</sub>, and other data. The update frequency is one hour, and the time span is from May 2014 to April 2021. PM2.5 is shown in Figure 3.



**Figure 3.** Changes in PM2.5 concentration over time.

#### 3.2. Data Pre-Processing

In this paper, data pre-processing includes data cleaning and data normalization. During data cleaning, clear redundant data. When the pollutant data is missing, this paper uses 8 h moving average data to replace it. After processing, the short-term missing values that still exist are supplemented by simple linear interpolation of adjacent values, and the missing data that are too long are deleted.

The normalization of maximum and minimum values is used in this paper, as follows:

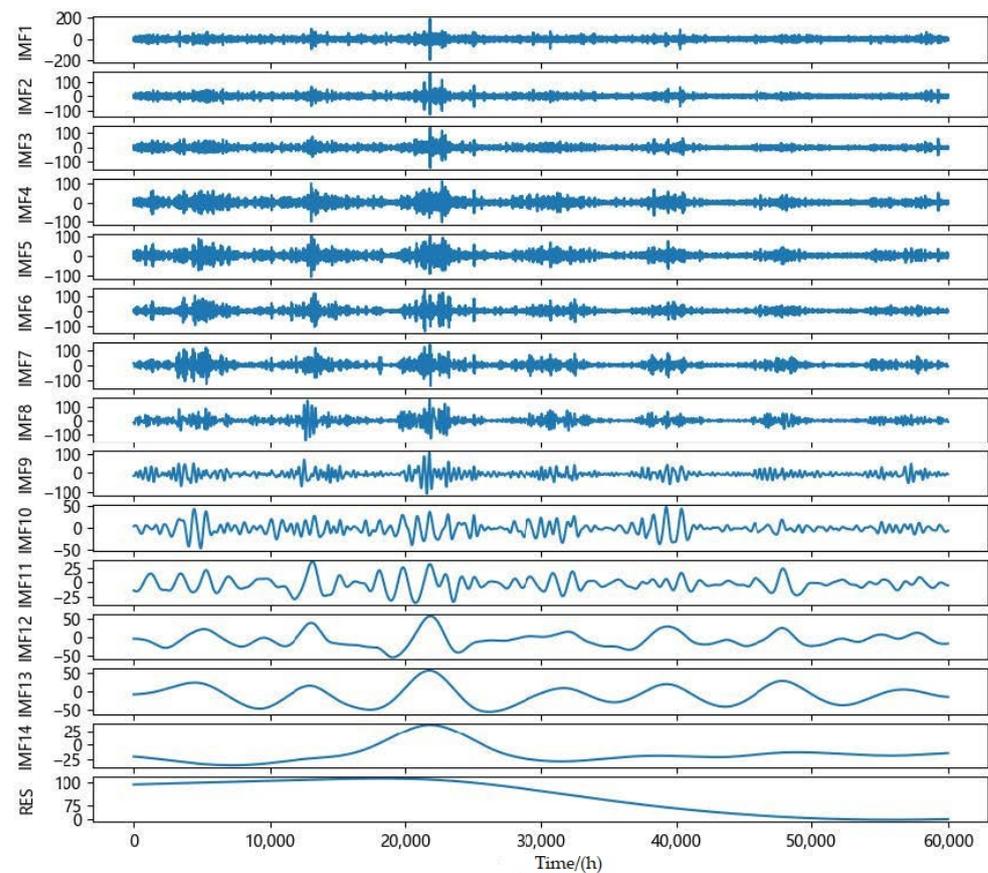
$$f^* = \frac{f - f_{\min}}{f_{\max} - f_{\min}} \quad (13)$$

where  $f_{\max}$  is the maximum value of sample data;  $f_{\min}$  is the minimum value of sample data.

#### 3.3. EEMD Decomposition of PM2.5 Concentration

In this paper, the pre-treated TIME series of PM2.5 value is decomposed into 14 IMF series and one trend item, as shown in Figure 4.

For the period calculation of IMF components, this paper uses the average period as the period of IMF components. The calculation results are shown in Table 1 below. According to the cycle calculation results, imF1-IMF4 is hour scale, IMF5-IMF9 is day scale, IMF10-IMF12 is month scale, and IMF13-IMF14 is year scale.



**Figure 4.** EEMD decomposition results of PM<sub>2.5</sub> concentration.

**Table 1.** The period of each IMF component of PM<sub>2.5</sub> concentration.

IMF Component	Period/h
IMF1	3
IMF2	5
IMF3	8
IMF4	15
IMF5	25
IMF6	46
IMF7	89
IMF8	168
IMF9	321
IMF10	659
IMF11	1395
IMF12	4000
IMF13	8572
IMF14	20,000
RES	–

### 3.4. Evaluation Index

The following indicators are selected as the evaluation criteria in this paper:

## (1) RMSE (Root Mean Square Error)

$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{m=1}^M (y_m - y'_m)^2} \quad (14)$$

where  $y_m$  is the true value in the test set;  $y'_m$  is the predicted value.

## (2) MAE (Mean Absolute Error)

$$\text{MAE} = \frac{1}{m} \sum_{m=1}^M |Y' - Y| \quad (15)$$

where  $Y'$  is predicted results;  $Y$  is true value.

(3)  $R2_{\text{adj}}$  (Adjusted R-Square)

$$R2 = 1 - \frac{\sum_{m=1}^M (y_m - \bar{y}_m)^2}{\sum_{m=1}^M (y_m - \bar{y})^2} \quad (16)$$

$$R2_{\text{adj}} = 1 - \frac{(1 - R2)(n - 1)}{n - p - 1} \quad (17)$$

where  $y_m$  is the true value in the test set;  $\bar{y}_m$  is the predicted value;  $\bar{y}$  is the average of the true values in the test set;  $R2$  is R-Square;  $n$  is the number of samples;  $p$  is the number of features;  $R2_{\text{adj}}$  offsets the impact of the number of samples on  $R2$ , so that the value of  $R2_{\text{adj}}$  is between zero and one, and the larger the value of  $R2_{\text{adj}}$ , the better the performance of the model.

## 3.5. Comparison of Experimental Results

## 3.5.1. Impact of Historical Time Windows on Model Performance

PM2.5 data is affected by a variety of related time series, but the change of each time series value does not immediately affect PM2.5 concentration value, which means that the variable value at the previous moment has a lag effect on the PM2.5 concentration value at the next moment, which may be strong in the short term and weak in the long term [8]. A smaller window size cannot guarantee sufficient long-term memory input for LSTM model, while a larger window size will increase the input of irrelevant information and increase the unnecessary computational complexity of the model [9]. In order to determine the appropriate historical time window, the historical time window in this study starts from 12 h, and every 12 h is a time interval. The prediction scale is the concentration of 1 h PM2.5 in the future. The results are shown in Table 2 below. When the historical time window is 36 h, the RMSE, MAE and  $R2$  of the model in this paper are 9.66, 6.95, and 0.95, respectively, which are the best. For LSTM model, when the history time window is 24 h, RMSE 14.0 is the best. When the historical time window is 36 h, MAE is 7.63 and  $R2$  is 0.89. For CNN-LSTM model, when the historical time window is 24 h, RMSE is 13.66, MAE is 9.88, and  $R2$  is 0.91. The model in this paper is superior to the comparison model in terms of indicators. The RMSE of the model is 31% lower than that of LSTM and 25% lower than that of CNN-LSTM. For the index MAE, it is 24% lower than LSTM and 22% lower than CNN-LSTM. For index  $R2$ , it is 5% higher than LSTM and 3% higher than CNN-LSTM.

## 3.5.2. Performance Comparison of Multi-Step Prediction

In order to test the multi-step prediction performance of the model in this paper for PM2.5 hour-concentration, experiments were carried out on the three models for 1 h, 4 h, 8 h, 12 h, and 24 h in the future, respectively, and the results are shown in Table 3. It can be seen from Table 3 that: (1) each model achieves the best effect when the prediction step

size is one hour, and the evaluation indexes of the model proposed in this paper are better. (2) With the increase of prediction step size, the accuracy of prediction decreases, but the prediction evaluation index of the model proposed in this paper is superior to LSTM and CNN-LSTM in each prediction time scale. Therefore, it indicates that the model proposed in this paper is effective in improving the long-term prediction accuracy.

**Table 2.** Performance comparison of models in different historical time windows.

Historical Window Time	LSTM			CNN-LSTM			Model of This Paper		
	RMSE	MAE	Adjusted R2	RMSE	MAE	Adjusted R2	RMSE	MAE	Adjusted R2
12 h	14.65	9.39	0.90	13.85	9.56	0.91	10.62	7.27	0.94
24 h	14.0	9.23	0.91	12.90	8.80	0.92	9.79	7.02	0.95
36 h	14.24	9.15	0.90	12.96	8.91	0.92	9.66	6.95	0.95
48 h	16.87	10.57	0.86	13.30	9.20	0.91	10.37	7.25	0.94
60 h	17.16	11.07	0.86	14.90	10.34	0.89	10.99	7.56	0.94
72 h	17.41	11.45	0.85	14.92	10.40	0.89	11.06	7.42	0.94

**Table 3.** Comparison of the performance of the three methods for different time step predictions.

Time Step (Predicted)	LSTM			CNN-LSTM			Model of This Paper		
	RMSE	MAE	Adjusted R2	RMSE	MAE	Adjusted R2	RMSE	MAE	Adjusted R2
1 h	14.25	9.15	0.90	12.96	8.91	0.92	9.96	6.95	0.95
4 h	14.95	9.72	0.89	14.51	9.70	0.90	11.68	8.10	0.93
8 h	16.88	11.16	0.86	15.20	10.26	0.89	14.25	9.69	0.90
12 h	17.65	11.27	0.84	17.60	11.32	0.85	15.00	9.85	0.89
24 h	21.21	13.48	0.78	20.80	13.63	0.79	18.48	11.78	0.83

In order to display the forecast results intuitively, the forecast data from 26 February 2021 to 18 March 2021 are selected for display, as shown in Figures 5–9 below. The blue represents the real data value, the yellow is the predicted value of the LSTM model, the green is the predicted value of the CNN-LSTM model, and the red is the predicted value of the model in this article. It can be seen from Figures 5 and 6 that when the prediction step length is short, although the prediction results of the other two models and the predicted future trend can be well consistent with the real data, the model proposed in this article has achieved better results. At the same time, the model proposed in this article is also superior to the other two models in peak prediction. It can be seen from Figures 7–9 that as the prediction duration increases, the accuracy of the peak prediction and the prediction of the future trend of each model decreases. When the prediction time step is 24 h, the prediction trend of LSTM and CNN-LSTM starts to be opposite to that of the real data, as shown in the predicted value between 400 h and 450 h in Figure 9. The prediction results and future trends of the model in this article can be better agreement with the real data. Therefore, the model in this article can better simulate the long-term forecast of PM2.5.

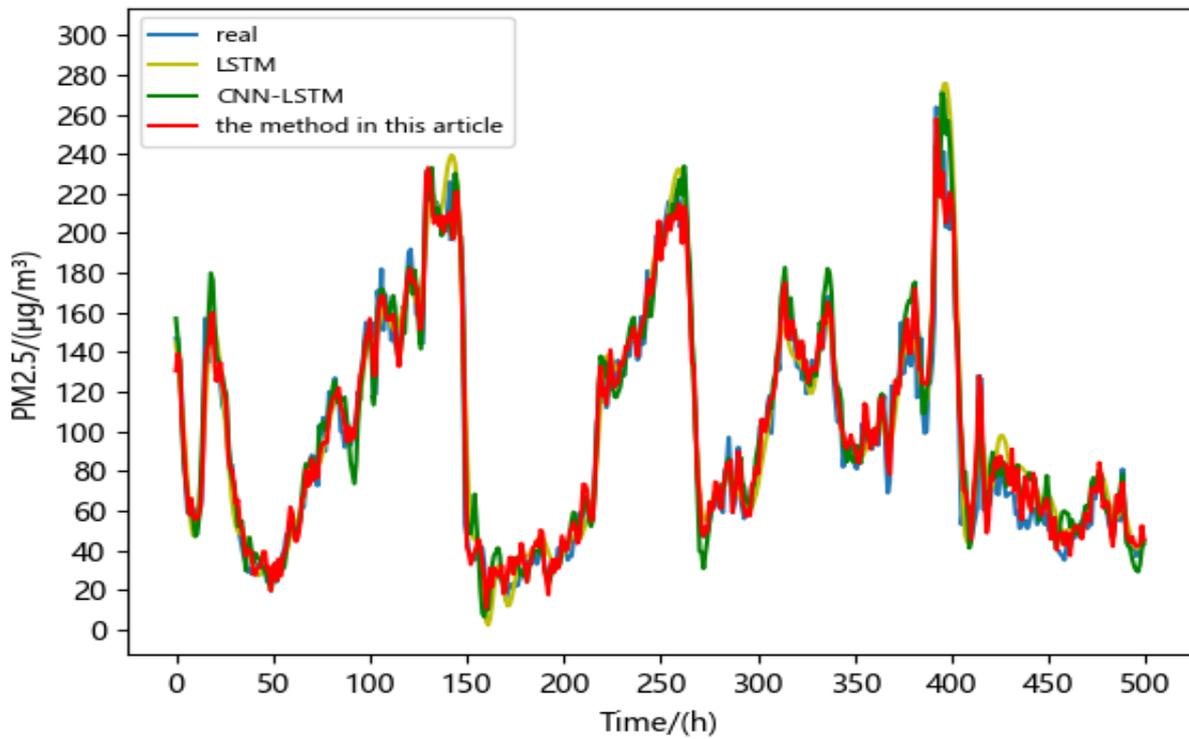


Figure 5. The prediction results of the three methods with a prediction step of 1 h.

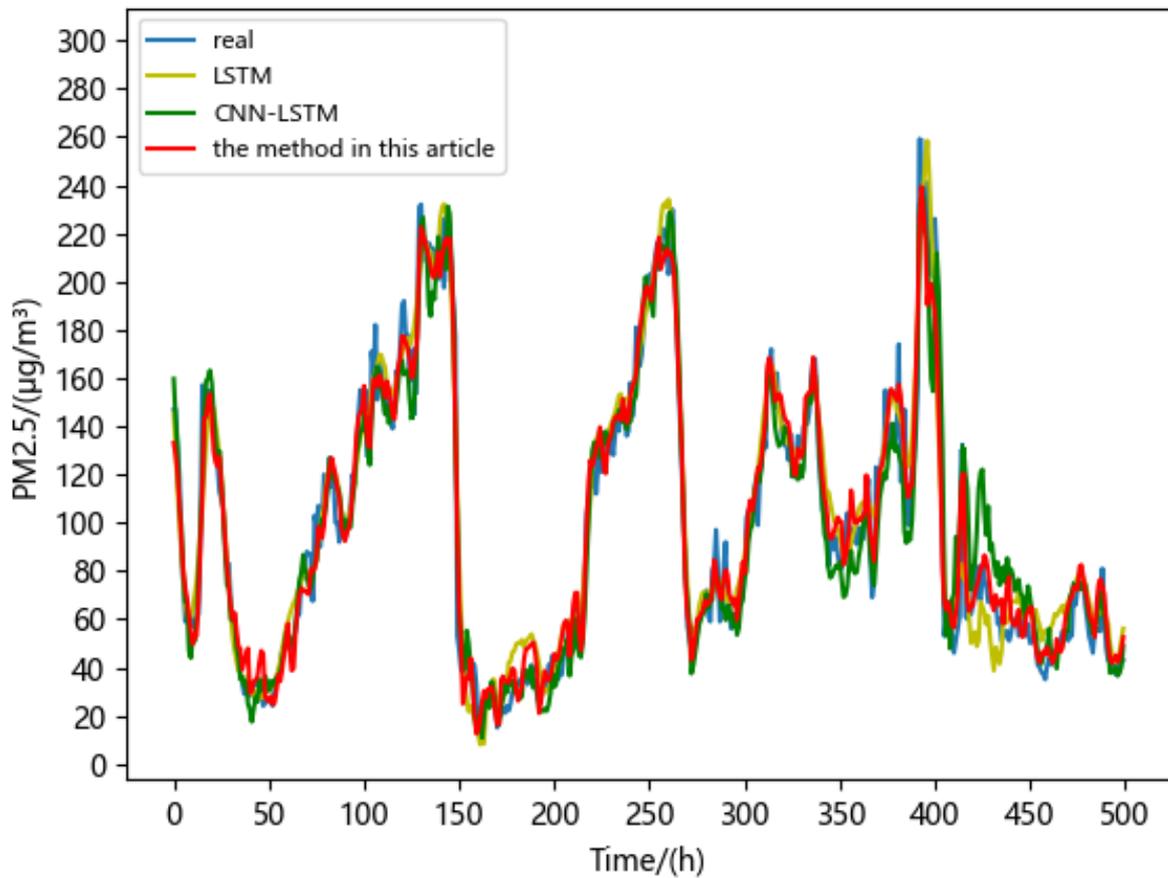


Figure 6. The prediction results of the three methods with a prediction step of 4 h.

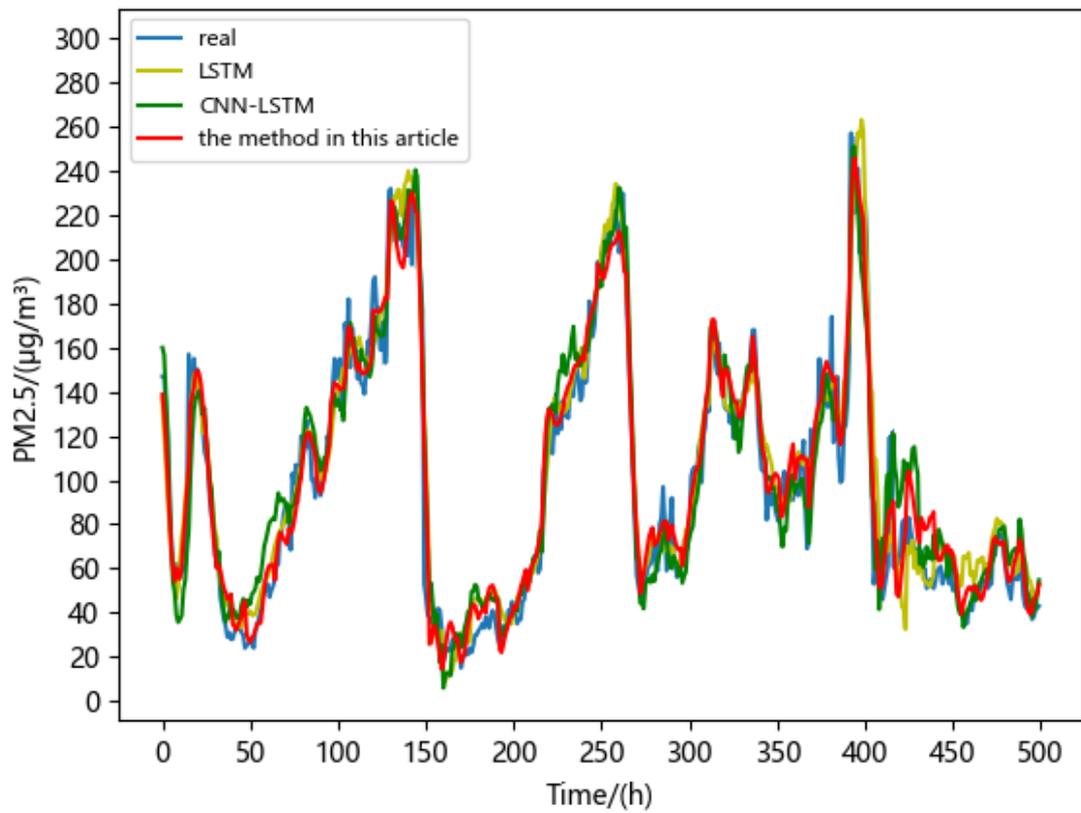


Figure 7. The prediction results of the three methods with a prediction step of 8 h.

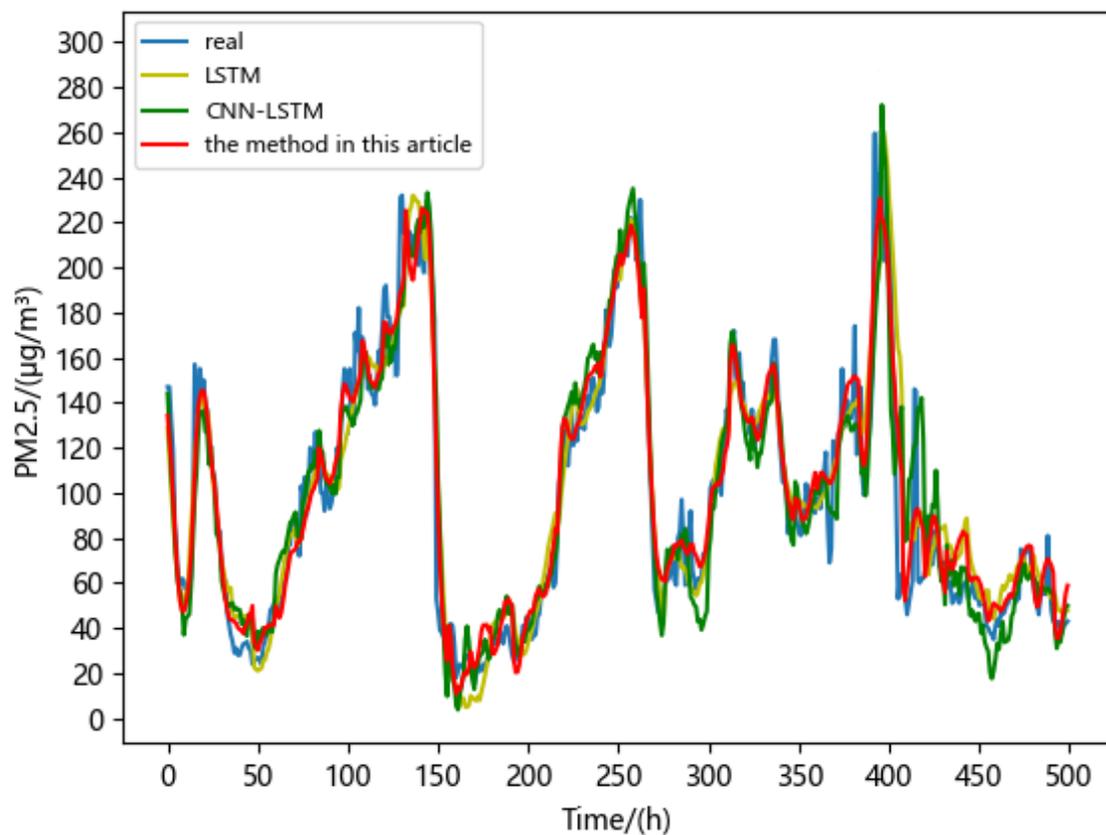
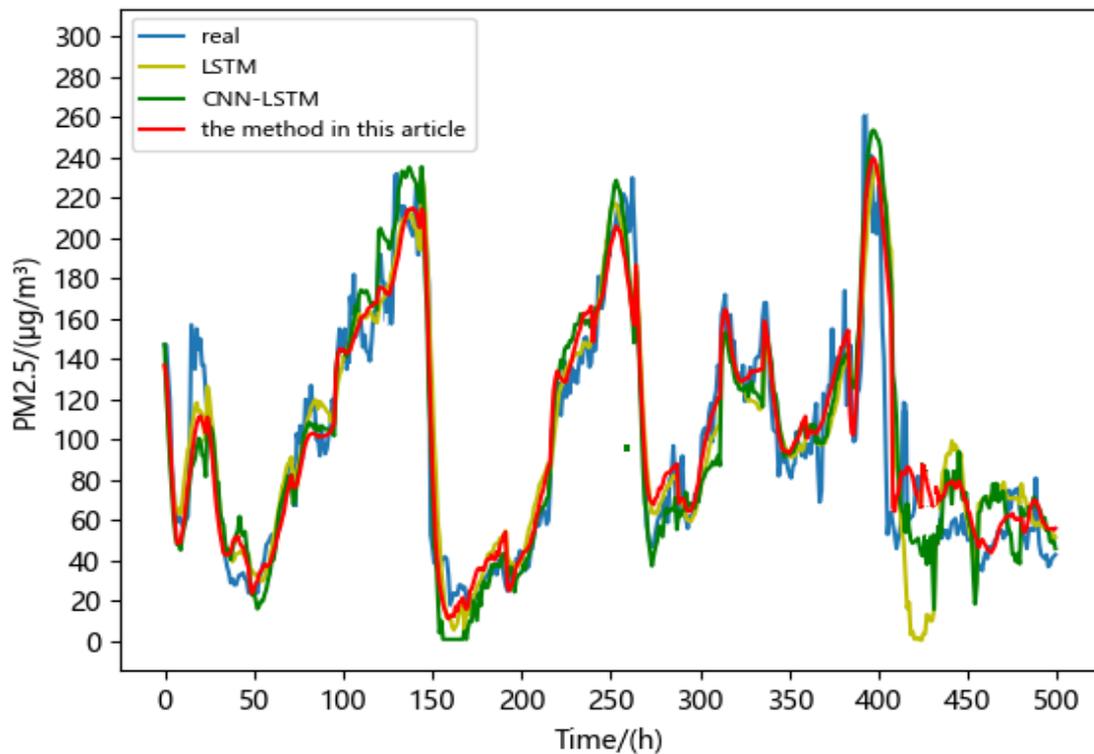


Figure 8. The prediction results of the three methods with a prediction step of 12 h.



**Figure 9.** The prediction results of the three methods with a prediction step of 24 h.

#### 4. Conclusions

The prediction of PM<sub>2.5</sub> concentration is of great significance for People's Daily life and environmental governance. Because the characteristic information of different time scales has different influence on the prediction results, a multi-time scale fusion model is proposed in this paper. The experimental results show that the proposed multi-time scale fusion model is superior to the comparison model in single and multi-step prediction, indicating that the multi-time scale fusion is effective for long-term prediction. In addition, in this paper, only the data of one site is used for the experiment, the amount of data is too small, and the influence between sites is not taken into account. In the future, PM<sub>2.5</sub> between adjacent stations will be studied and analyzed, and the accuracy of prediction will be improved by studying the spatial correlation between stations.

**Author Contributions:** Writing—original draft preparation, W.X.; writing—review and editing, J.Z.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by National Natural Science Foundation of China under Grant NSFC-61803148.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

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

#### References

1. Fan, J.X.; Li, Q.; Zhu, Y.J.; Hou, J.X.; Feng, X. Research on time and space prediction model of air pollution based on RNN. *Sci. Surv. Mapp.* **2017**, *42*, 76–83.
2. Qi, B.L.; Guo, K.P.; Yang, B.; Du, Y.M.; Liu, M.; Wang, J.N. Air quality prediction based on GCN-LSTM. *Appl. Comput. Syst.* **2021**, *30*, 208–213.

3. He, Z.X.; Li, L. A prediction model of air pollutant concentration based on wavelet transform and LSTM. *Environ. Eng.* **2021**, *39*, 111–119.
4. Huang, C.J.; Kuo, P.H. A deep CNN-LSTM model for particulate matter (PM<sub>2.5</sub>) forecasting in smart cities. *Sensors* **2018**, *18*, 2220. [[CrossRef](#)]
5. Liu, J.; Shahroudy, A.; Xu, D.; Kot, A.C.; Wang, G. Skeleton-based action recognition using spatio-temporal LSTM network with trust gates. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 3007–3021. [[CrossRef](#)]
6. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [[CrossRef](#)]
7. Lu, Y.; Yang, J.; Shao, Z.J.; Zhu, C.C. Robust prediction of PM (2.5) based on staged temporal attention network. *Environ. Eng.* **2021**, *39*, 1–9.
8. Huang, W.J.; Li, D.Y.; Huang, Y. Long-term prediction of PM<sub>2.5</sub> concentration based on deep learning. *Appl. Res. Comput.* **2021**, *38*, 1809–1814.
9. Li, X.; Peng, L.; Yao, X.J.; Cui, S.L.; Hu, Y.; You, C.Z.; Chi, T. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environ. Pollut.* **2017**, *231*, 997–1004. [[CrossRef](#)] [[PubMed](#)]