Transactions on Environment, Energy and Earth Sciences
ISSN: 2960-1924, eISSN: 2960-2300 | Volume 3
GRET 2024

Prediction of pm2.5 concentration based on VMD-CNN-Transformer hybrid model

Xinjie Wang [1, *], Changsheng Zheng [1], Ziyang Zheng [2]

[1] School of Cyber Science and Engineering, Qufu Normal University, Jining, China, 273165

[2] School of Computer Science, Qufu Normal University, Rizhao, China,276826

* Corresponding author: engineer_cradlew@163.com

Abstract. In recent years, PM2.5 pollution has become increasingly serious, seriously affecting people's health and the world's ecological environment, and it has important research significance for the accurate prediction of PM2.5 concentration. In this paper, a VMD-CNN-Transformer hybrid model is proposed to predict future pm2.5 concentration. Firstly, the sliding window method is used to reconstruct the data set, and the required historical data range is set according to the window size. The research shows that the model has the best performance when the window size is 6. Secondly, the ablation experiment shows that the VMD module can enhance the predictive performance of the model. The test set is used to test the VMD-CNN-Transformer hybrid model, and the comparison experiment is conducted with the comparison model (LSTM, random forest). The results show that in terms of MAE, RMSE and determination coefficient, the Transformer model can be improved. The VMD-CNN-Transformer hybrid model can predict pm2.5 concentration better than the comparison model, with a determination coefficient of 0.97. Therefore, this study aims to improve the prediction of PM2.5 concentration, which is of great significance for controlling air pollution and solving health problems, and provides a scientific basis for related policies.

Keywords: Prediction of pm2.5 concentration; VMD-CNN-Transformer; Comparison Experiment.

1. Introduction

With the rapid development of industrialization and urbanization, PM2.5 pollution has become a widespread phenomenon around the world. PM2.5 not only causes direct harm to human health, such as respiratory system diseases and cardiovascular diseases, but also has a serious impact on environmental quality [1]. High concentrations of PM2.5 can reduce air visibility and increase the risk of traffic accidents, as well as harm plant growth and aquatic life, disrupting the ecological balance. Therefore, the establishment of accurate PM2.5 concentration model is of great significance for air pollution prevention and control.

In the field of PM2.5 concentration prediction, researchers have used time series prediction statistical methods such as ARIMA [2], which usually require high data stability. Subsequently, the use of historical meteorological data or historical pollution data, combined with machine learning and deep learning models to predict PM2.5 concentration has become a research hotspot. These methods include random forest (RF), XGBoost [3], and various deep learning neural network models such as BP neural network, LSTM, CNN, etc. [4]. A single machine learning model can learn single-frequency time series very well, but it is not good at learning multi-frequency series. Recently, Xiaodi Xu et al. [5] proposed a 24-hour time series prediction method for PM2.5 concentration based on LSTM, regression tree and BP fully connected neural network. Jinsong Zhang [6] et al. also proposed the CNN-BiLSTM-Attention mixed model to predict PM2.5 concentration. These hybrid models have better performance. However, the ability to predict the abrupt change point of PM2.5 concentration is limited, and some extreme cases will be ignored.

Based on the existing problems in the current research, this paper uses VMD to decompose multi-frequency sequences of target time series into a set of inherent mode functions (IMFs), proposes a VMD-CNN-Transformer model, and conducts comparison experiments with LSTM and random

forest models. This will be of great significance for improving the prediction accuracy of PM2.5 concentration, which is expected to help improve the understanding of the dynamic change of atmospheric environmental pollution, and provide a scientific basis for relevant decision-making, and provide a reference for controlling air pollution and solving health problems.

2. Data and methods

2.1. Data Set

The data used in this paper are the historical data of air quality monitoring stations and historical meteorological data from March 1, 2013 to February 28, 2017. The air quality data comes from the Beijing Municipal Environmental Monitoring Center, and the meteorological data of each air quality monitoring site is matched with the nearest meteorological station to the China Meteorological Administration. It contains air quality data (SO_2、NO_2、CO、O_3、PM_{10}、$PM_{2.5}$) related to 12 air quality monitoring stations in Beijing and meteorological data (such as temperature, humidity, pressure, precipitation, wind speed, etc.).

2.2. MD-CNN-Transformer method

2.2.1. VMD variational mode decomposition.

VMD variational mode decomposition [7] decomposes the signal into multiple intrinsic modes (IMFs) based on the variational principle by minimizing the complexity of the signal and the interaction between different components.

VMD constructs an optimization problem, which is solved to obtain the optimal modal function decomposition. This method avoids the problem of end effect and false component in the iterative process, and can deal with nonlinear and non-stationary signals effectively. The decomposition process of VMD is to decompose the original signal into a specified number of IMF components by constructing and solving the constrained variational problem. In the process of solving, the optimization algorithm will try to find the best parameter combination, reduce the mutual interference between different mode functions, and further improve the anti-mode aliasing ability.

2.2.2. CNN-Transformer Network Architecture.

The CNN-Transformer model first extracts local features through two one-dimensional convolution layers and pooling layers, then uses Transformer with multi-head attention mechanism to encode global features, and then generates prediction results through feedforward neural network and normalization layer. As shown in Figure 1 and Figure 2, CNN-Transformer model can capture both local features and global context information in the input sequence [8], and output accurate prediction of PM2.5 concentration at the next moment through training and optimization of multi-layer neural network.

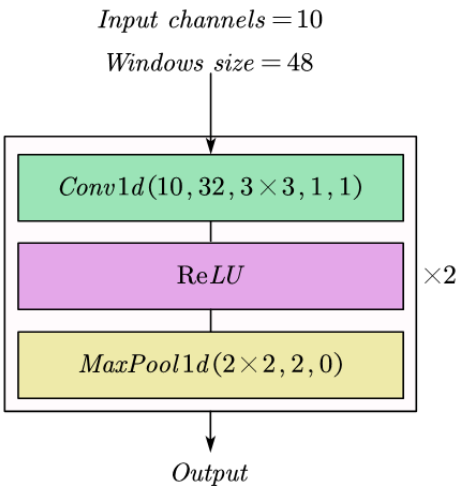


Figure 1. CNN network structure

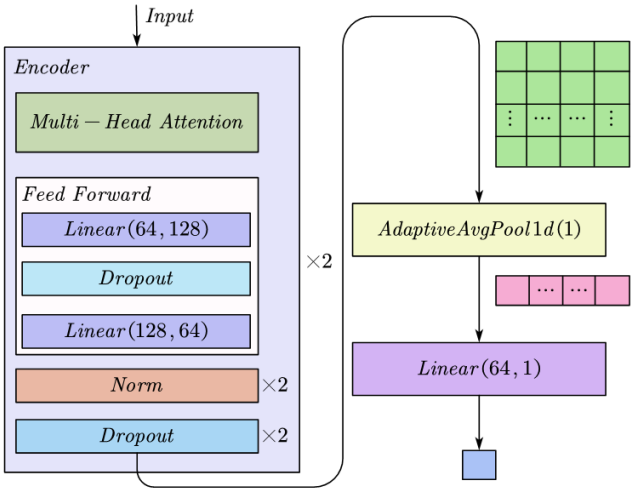


Figure 2. Transformer network structure

2.2.3. Overall structure of the VMD-CNN-Transformer model.

Figure 3 shows the overall structure of a VMD-CNN-Transformer model for PM2.5 prediction. Firstly, data preprocessing is carried out, including data cleaning, normalization and data set partitioning. Then the original signal is decomposed into intrinsic mode function (IMFs) by VMD module. Then, the features are extracted using CNN and encoded by Transformer. The prediction results are obtained through pooling and linear layer, and the test set is used for model evaluation. The optimal VMD-CNN-Transformer hybrid model is selected for the prediction of PM2.5.

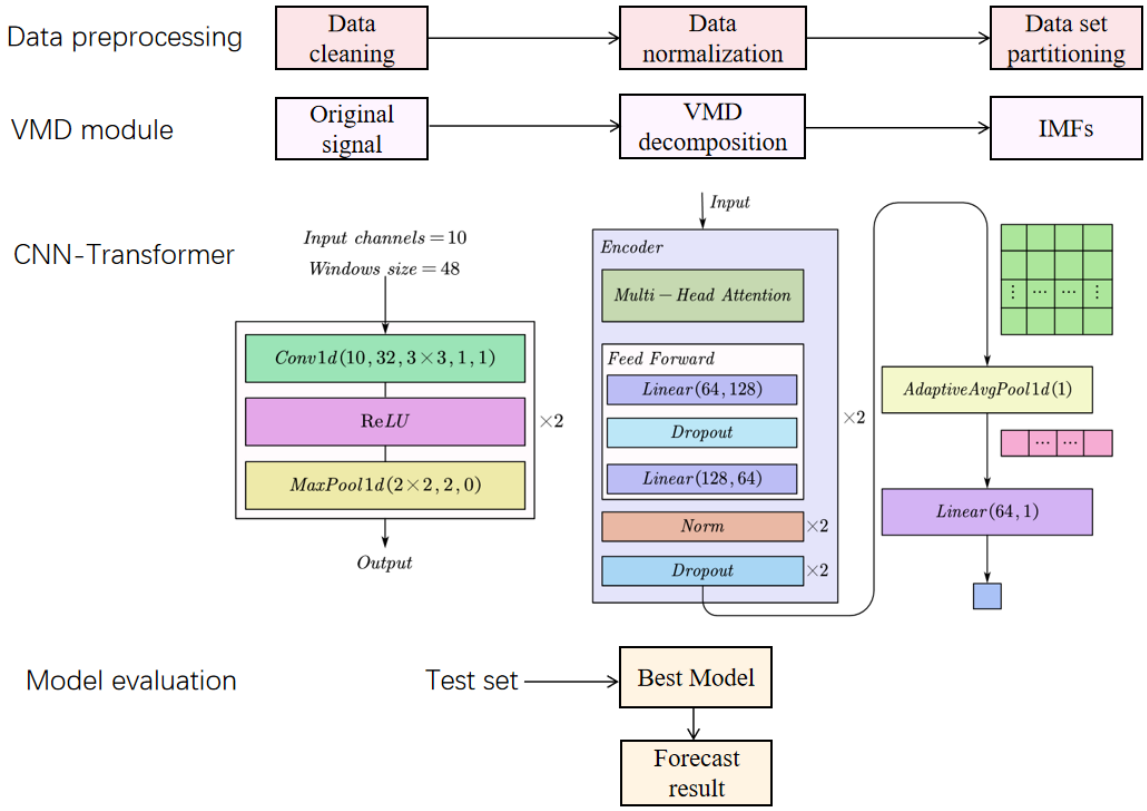


Figure 3. VMD-CNN-Transformer network structure

2.3. Long Short-Term Memory Network LSTM

As shown in Figure 4, LSTM [9], a short-duration memory network, performs well in processing sequence data with long-term dependencies. The core idea is to capture and remember long-term dependencies in the sequence by introducing cell states and three gate structures (input gate, output

gate and forget gate). The LSTM model can combine the input at the current moment and the hidden state at the previous moment, and generate a value between 0 and 1 through the sigmoid activation function. To decide what information to update, forget, or output.

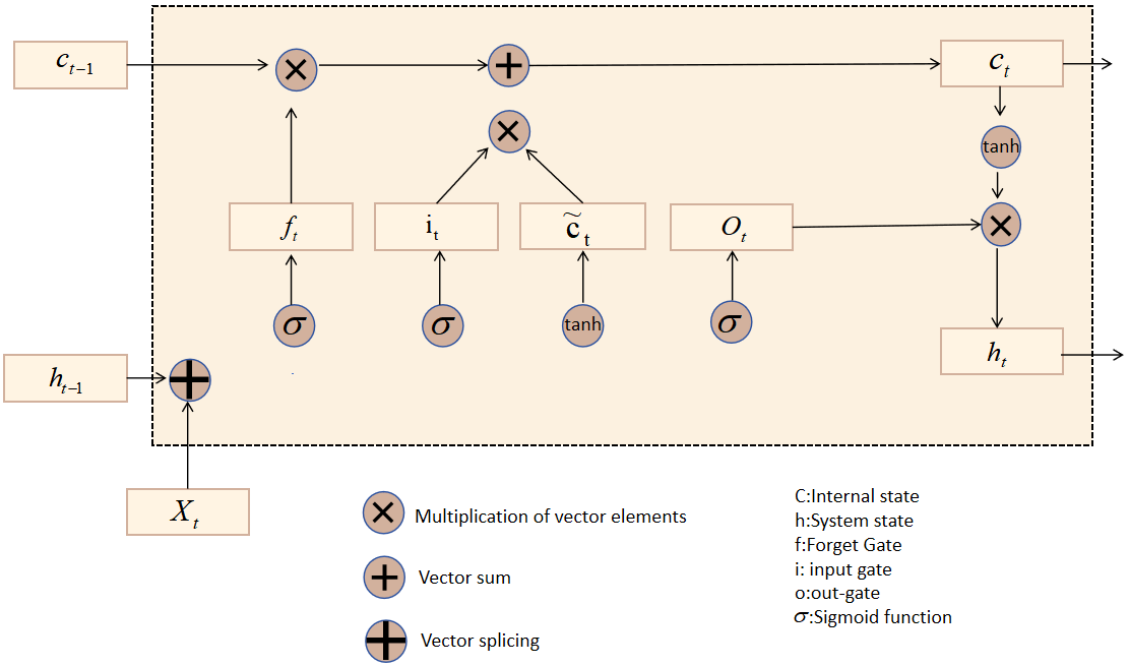


Figure 4. LSTM network structure

The picture f_t representative forgetting gate

$$f_t = \sigma(W_f \bullet |h_{t-1}, x_i| + b_f) \tag{1}$$

The picture i_t representative input gate

$$i_t = \sigma(W_i \bullet [h_{t-1}, x_i] + b_i) \tag{2}$$

The picture $\tilde{C}$ represents a computational state in an internal system

$$\tilde{C} = \tanh(W_C \bullet [h_{t-1}, x_t] + b_c) \tag{3}$$

The picture C is memory cell, stored memory information, C_t represents the current moment of memory information, C_{t-1} represents the memory information of the last moment

$$C_{t=} f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4}$$

The picture O_t representative output gate, used to control whether the information in the memory unit is output

$$O_t = \sigma(W_o [h_{t-1}, x_i] + b_o) \tag{5}$$

h Is the output of the LSTM unit, h_{t-1} is the output from a moment ago

$$h_t = O_t * \tanh(C_t) \tag{6}$$

2.4. Random Forest model

As shown in Figure 5 and Figure 6, random forest method[10] is used as an integrated learning algorithm. Random forest methods improve the performance of a single decision tree by building multiple decision trees and averaging or voting on their results. This method performs well in both classification and regression tasks, and has the ability to handle complex situations such as high-dimensional data, missing values, outliers, and so on.

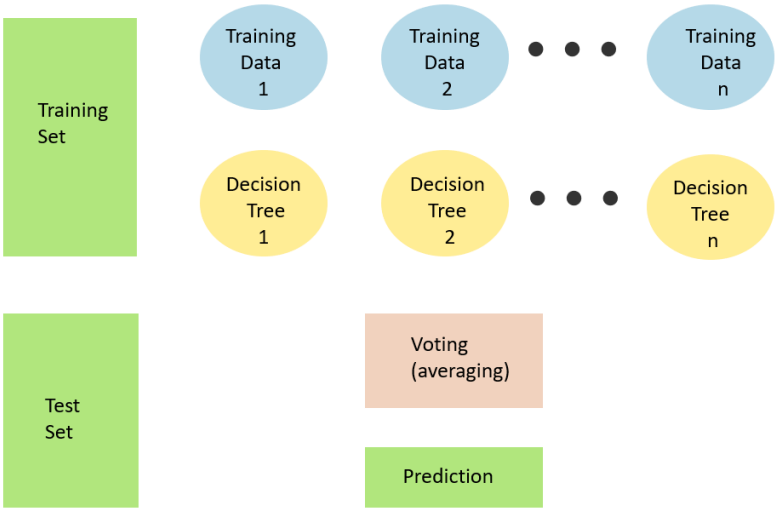


Figure 5. Random forest model

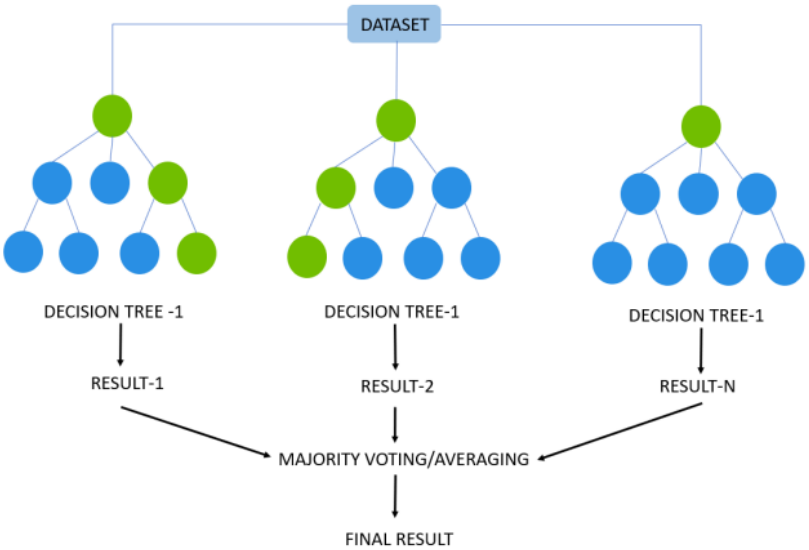


Figure 6. Random forest model

3. Results

3.1. VMD variational mode decomposition

3.1.1. Original signal.

Visualize the original time series signal, as shown in Figure 7, where the horizontal axis represents the time line and the vertical axis represents the concentration of PM2.5. As can be seen from Figure 7, the signal of the series fluctuates greatly and irregularly. If the signal of this time series is directly

predicted, it will be difficult to obtain a high accuracy. In this paper, VMD algorithm is used to decompose the signal of time series data, and predict the single component signal after decomposition, so as to realize the prediction of the total signal indirectly.

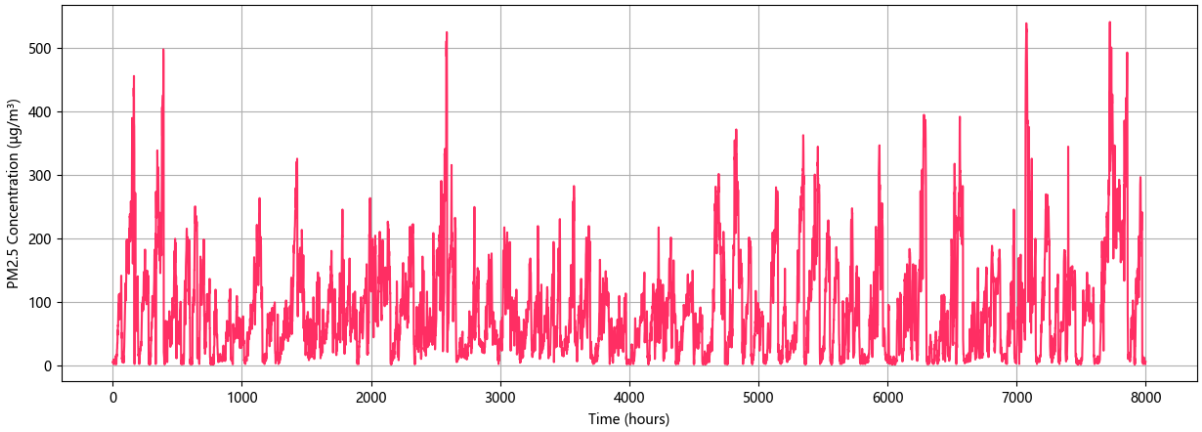


Figure 7. Raw signal

3.1.2. Select K value.

In this paper, the value of K is determined by the center frequency. First, 6 K values are set: 3, 4, 5, 6, 7, 8. VMD decomposition is performed for each K value respectively, and the most appropriate K value is determined by the final distribution of their respective center frequencies.

Figure 8 shows the change of the center frequency of each component in the iterative process under different K values. Criteria for selecting the value of K: there is a certain interval between the respective central frequencies of the components, and this interval must reach more than 1 times to meet the requirements. Combining Figure 8, it can be determined that the most appropriate value of K is 3. The center frequencies of these three components are 0.000381, 0.013512 and 0.043901, respectively.

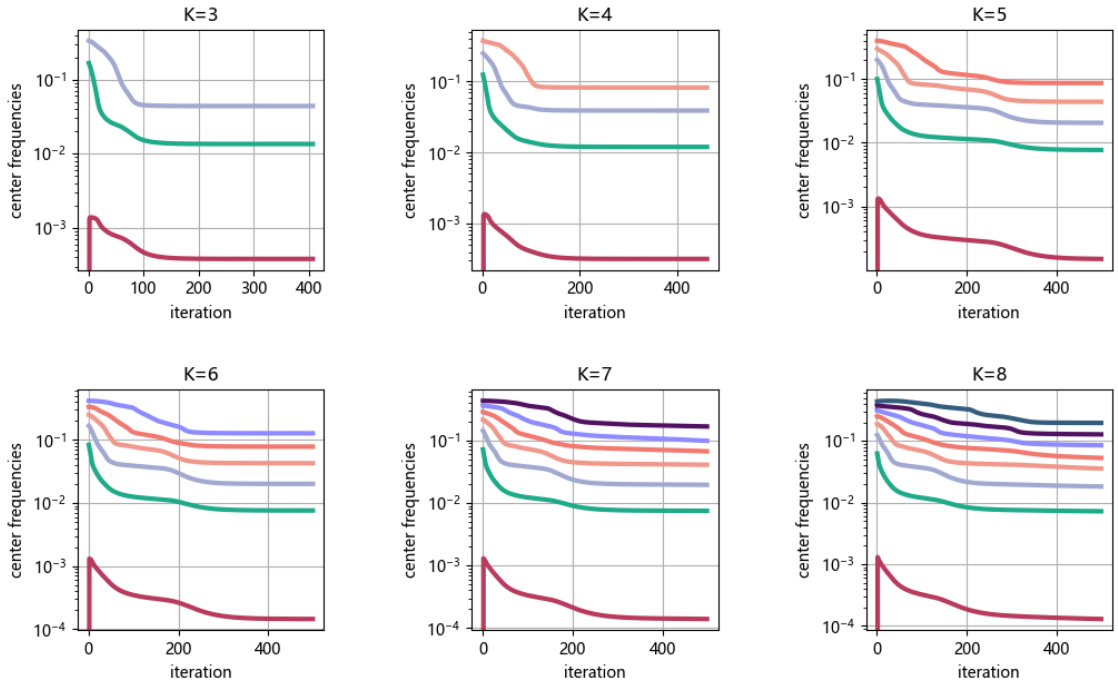


Figure 8. Center frequency iteration under different K values

Figure 9 shows the spectrum of three different components. The horizontal axis represents the frequency and the vertical axis represents the spectrum amplitude, both of which are logarithmic scale. The black dashed lines are raw signals that have higher energies across the entire spectrum and contain multiple frequency components from low to high frequencies. The VMD successfully

decomposed the original signal into three distinct modal components, each of which has a spectrum concentrated in a different frequency range. From the 3 circled regions, it can be seen that the frequency distribution of the 3 components presents three frequency levels. The solid red line is the first modal component, capturing the low-frequency component of the signal and rapidly decreasing as the frequency increases. The solid green line is the second modal component that captures the intermediate frequency component of the signal. The solid blue line is the third modal component that captures the high frequency component of the signal, which is higher in energy at high frequency bands and lower in energy at low frequency bands.

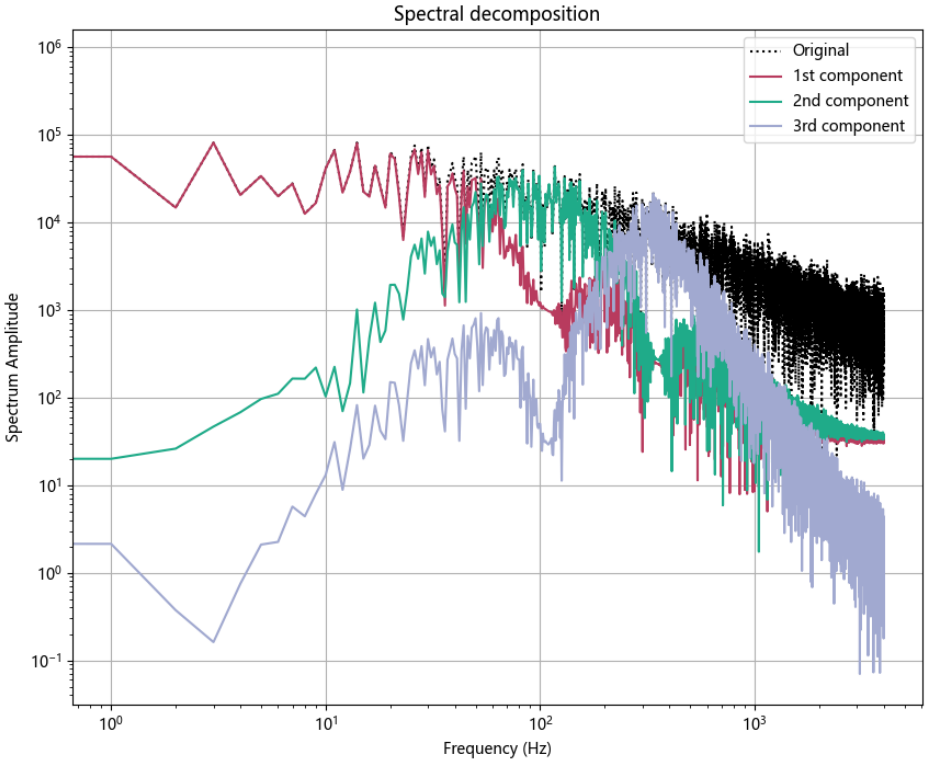


Figure 9. Spectral log-log images of different components for K=3

3.1.3. Variational mode decomposition.

Select K value as 3 for VMD decomposition of the original signal, and the results of decomposition components are shown in Figure 10. IMF1 is the first mode, which represents the main trend component in the signal and is able to describe the overall change trend of PM2.5 concentration. This trend component is useful for understanding overall behavior and making long-term predictions. IMF2 is the second mode that depicts the mid-frequency component in the signal, which may correspond to seasonal variations or other cyclical factors. IMF3 is the third mode, which represents high frequency fluctuations in the data that may be related to short-term environmental changes or noise. These three modes are modeled and then the results are combined to get the overall prediction.

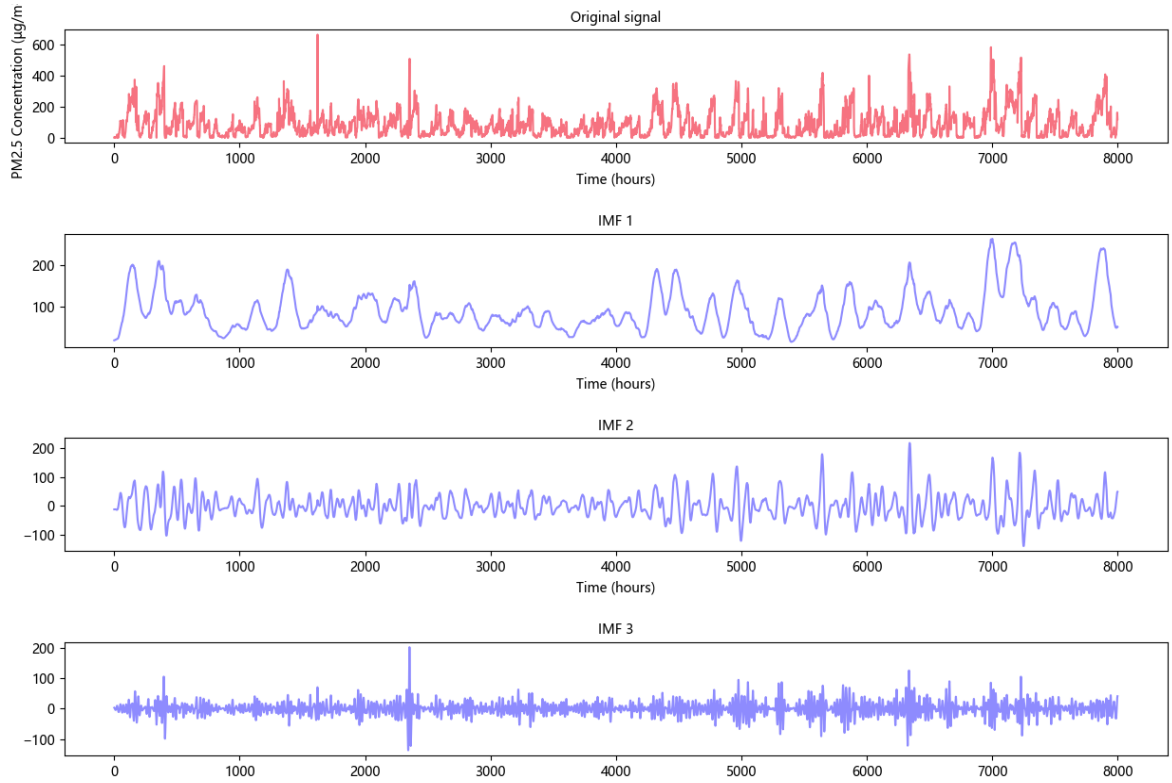


Figure 10. Variational mode decomposition

3.2. Prediction result analysis

The training set is iterated with epochs=50 rounds, and the optimal model weight is found as the final model weight. The evaluation index R2 of the optimal model on the test set is 0.96, and its mean square error MSE is 408.86. The model can correctly interpret 96% of the data in the test set, and the prediction accuracy is high. Figure 11 shows the iterative process of the model's loss on the training set and the test set respectively during the training process. In the last period of training, the loss value on the training set and the test set is relatively ideal, and the loss of the test set gradually approaches the loss on the training set. Figure 12 shows the model's prediction of the original signal on the test set. It can be seen that the difference between the predicted value and the real value is very small, which shows the advantages of the VMD-CNN-Transformer model.

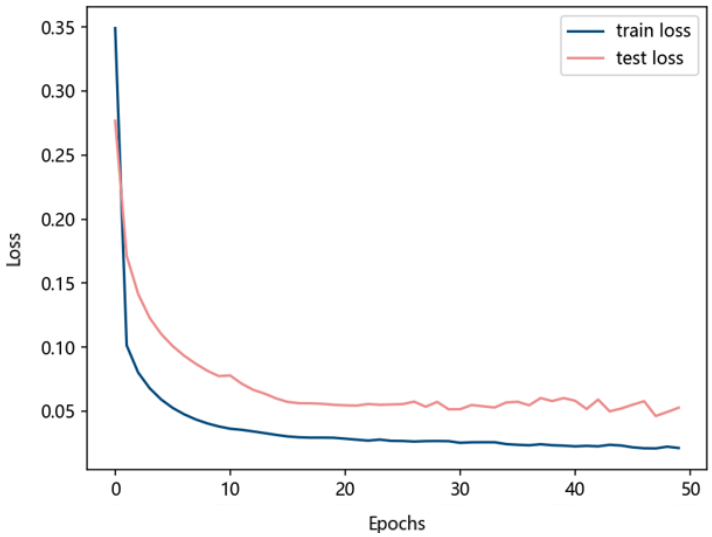


Figure 11. Loss of VMD-CNN-Transformer model

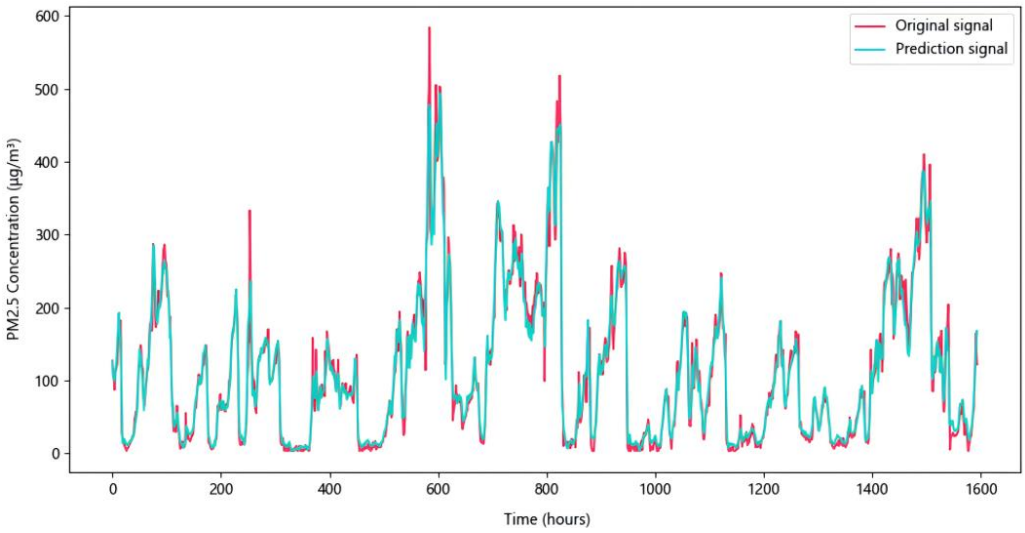


Figure 12. Model prediction results

3.3. Ablation experiment

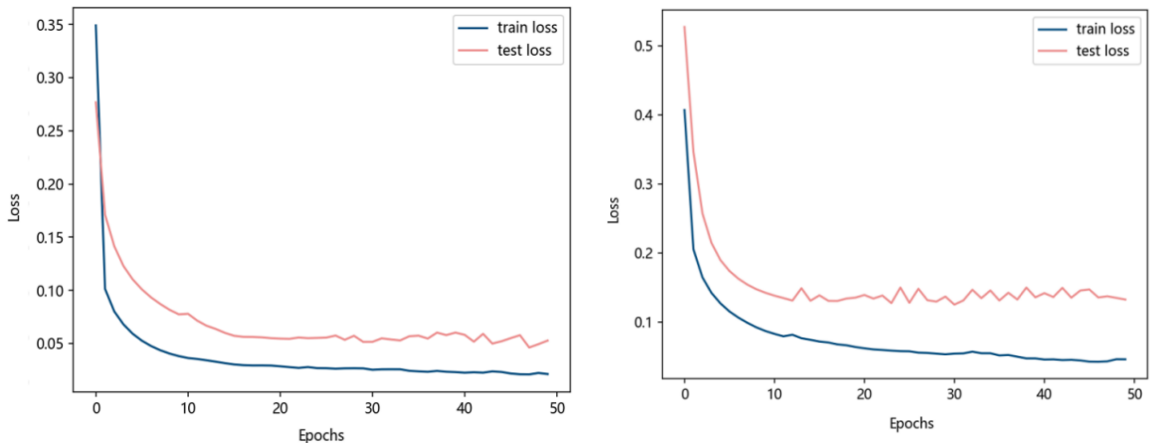
3.3.1. Analysis of VMD's contribution to the model.

In order to explore the positive contribution of VMD module in deep learning model, this paper conducts VMD ablation experiment. As shown in Table.1, the window size is set to 48, that is, the PM2.5 concentration in the next moment is predicted according to the indicators of the first 48 hours of historical experience, and the contribution of VMD module to the models is analyzed respectively for CNN-Transformer and LSTM deep learning models.

Table 1. Evaluation indicators with and without VMD

	CNN-Transformer				LSTM			
	MSE	RMSE	MAE	R2	MSE	RMSE	MAE	R2
VMD	408.86	20.22	12.25	0.96	1470.47	38.35	29.06	0.87
No VMD	875.51	29.59	18.49	0.92	1720.20	41.48	30.43	0.84

Table.1 shows that both CNN-Transformer model and LSTM model have much better prediction effect when using VMD module than when not using VMD module. As shown in Figure 13 (b), there is a slight upward trend in losses on the test set late in the iteration, which is not an ideal result. In addition, the final loss value in Figure 13 (b) is greater than the corresponding position loss value in Figure 13 (a). Therefore, the VMD module has a positive contribution to the deep learning model and can promote the accurate prediction of the model.



(a) Use the VMD module (b) Not using the VMD module

Figure 13. Whether CNN-Transformer has VMD training loss

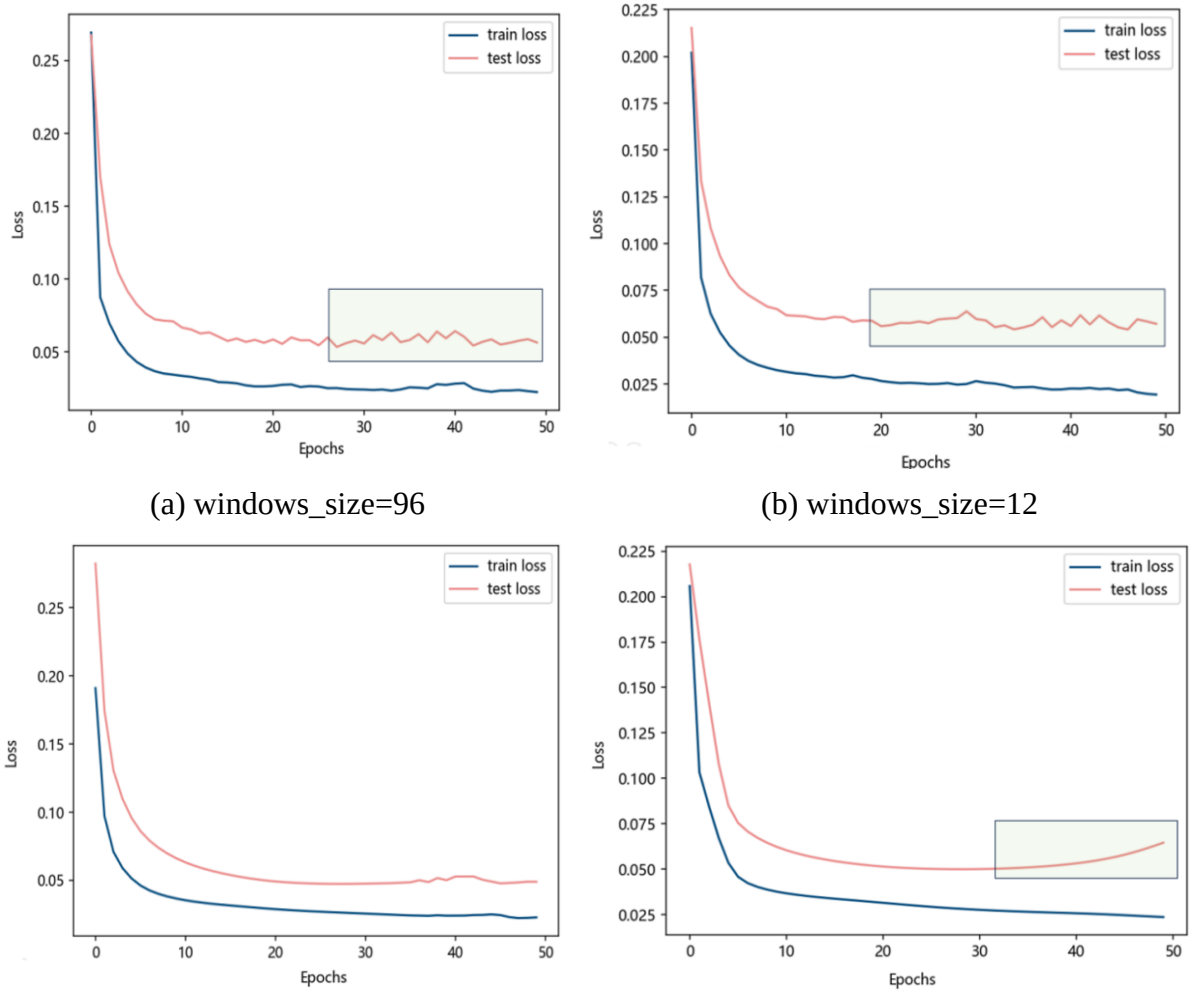
3.3.2. Influence analysis of different window sizes.

The model evaluation metrics corresponding to different window sizes can be seen in Table.2. When the window size is 6, it can be regarded as the best window size, and the prediction accuracy can reach 97%. Therefore, a window size of 6 allows the model to achieve optimal performance release on this training set.

Table 2. Evaluation indicators corresponding to different window sizes

window size	MSE	RMSE	MAE	R2
96	373.96	19.34	11.62	0.9664
48	408.86	20.22	12.25	0.9627
24	421.76	20.54	12.12	0.9611
12	378.50	19.45	12.00	0.9649
6	324.27	18.01	11.39	0.9701
3	427.99	20.69	11.67	0.9601

Figure 14 shows the record of the iterative process of loss on the training set and the test set during the model training process with different window sizes. It can be found that Figure 14 (b) and Figure 14 (d) show a slightly increasing trend of model loss on the test set in the later period of iterative training, and Figure 14 (a) shows the phenomenon of unstable loss on the test set in the later period of iterative training.



(a) windows_size=96 (b) windows_size=12

(c) windows_size=6 (d) windows_size=3

Figure 14. Loss iteration of the model under different window sizes

3.4. Comparison with other predictive models

Comparison experiments between VMD-CNN-Transformer model and other models (LSTM, random forest) are conducted to evaluate the advantages and disadvantages of the model. As shown in Table.3, the VMD-CNN-Transformer model proposed in this paper is compared with LSTM, a deep learning method, and random forest, a machine learning method, respectively. It is found that the model proposed in this paper has more advantages in the training set.

Table 3. Comparison of prediction accuracy of different models

Model name	MSE	RMSE	MAE	R2
VMD-CNN-Transforemer	408.86	20.22	12.25	0.9627
LSTM	708.86	26.62	17.61	0.93
random forest	1170.47	32.35	22.06	0.90

As shown in Figure 15/16/17, the prediction results of the VMD-CNN-Transformer model, LSTM and random forest model in the test set are respectively shown. The original signals of the VMD-CNN-Transformer model are almost completely identical with the prediction signals, indicating that the model has high prediction accuracy. And it performs well at the peaks and troughs of the signal. LSTM can follow the trend of the original signal in most cases, but performs poorly at the inflection point of the signal. However, the predicted signal of random forest model has the lowest coincidence degree with the original signal, especially in the region with large fluctuations. At multiple signal inflection points, the random forest model cannot accurately capture the trend of signal change. Therefore, the VMD-CNN-Transformer model proposed in this paper has the most advantages and the highest prediction performance.

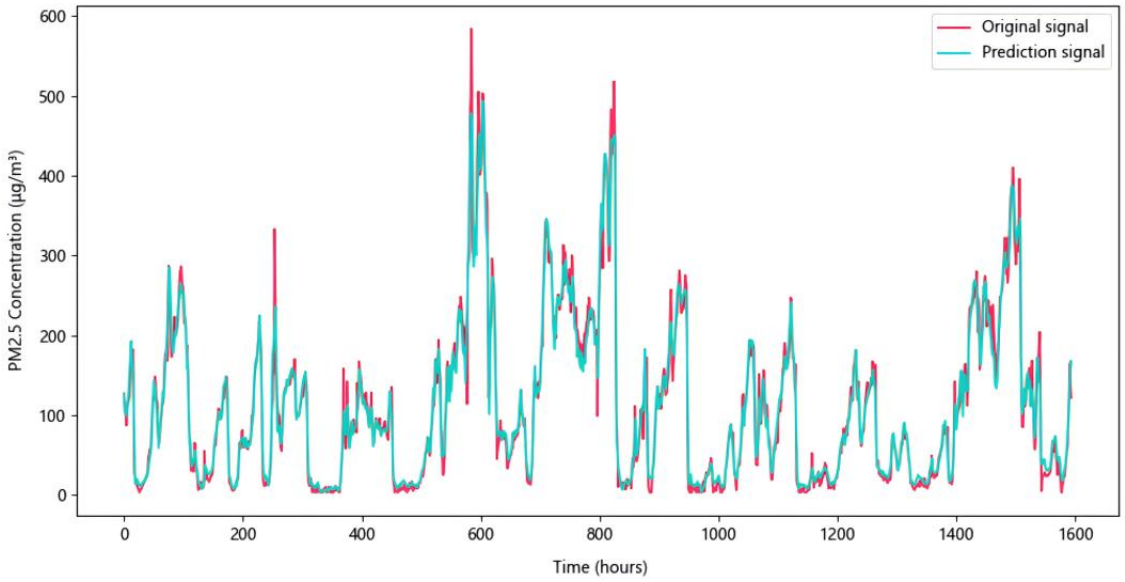


Figure 15. Forecast results from VMD-CNN-Transformer

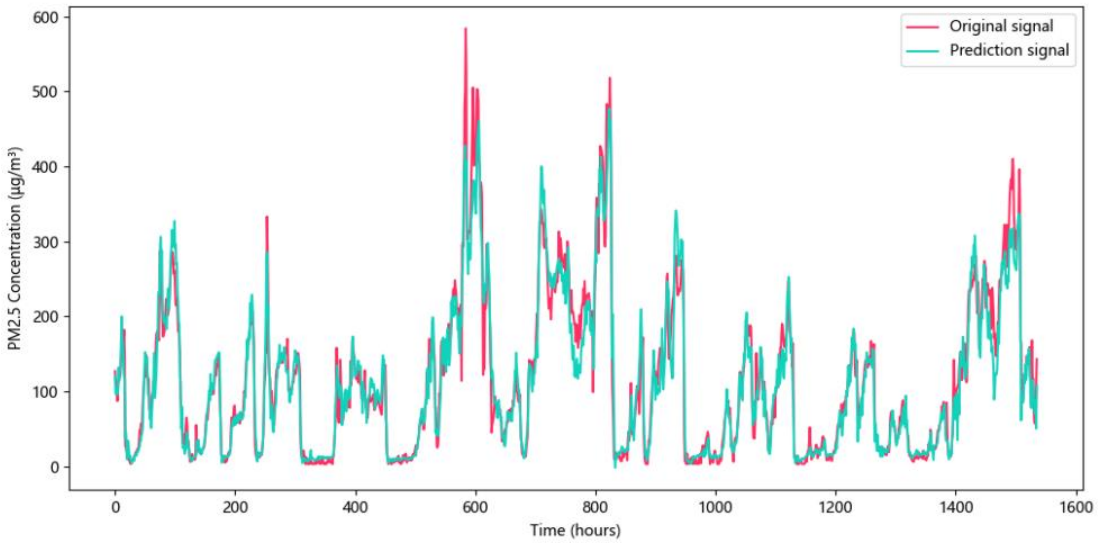


Figure 16. LSTM forecast results

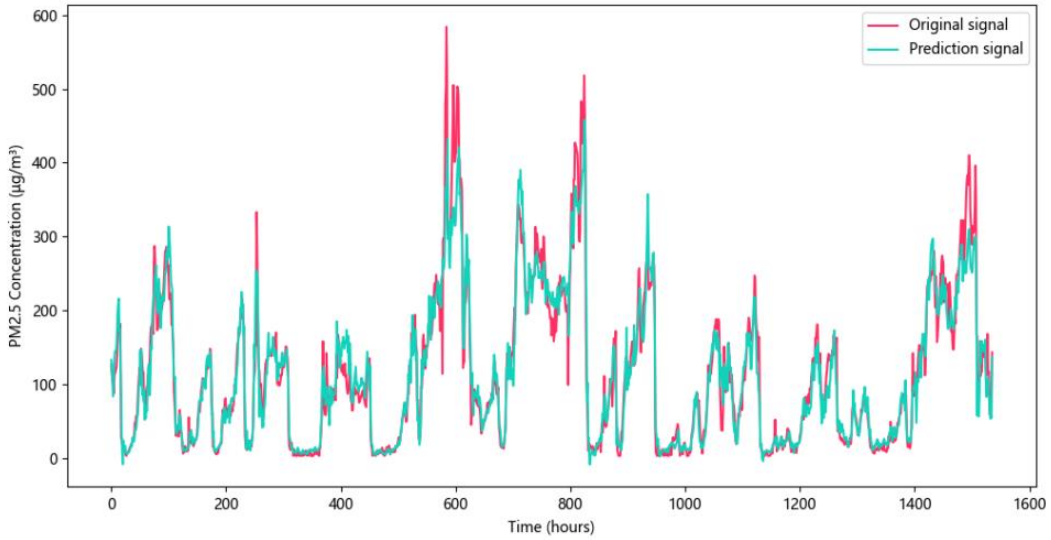


Figure 17. Random forest prediction results

4. Conclusion

The prediction of PM2.5 concentration faces many challenges due to the irregularity and unpredictability of its time series. In this paper, a new model combining variational modal decomposition (VMD), convolutional neural network (CNN) and Transformer network is proposed to improve the prediction accuracy of PM2.5 concentration. The model successfully predicted the PM2.5 concentration at the next moment, and the study showed that the VMD module could significantly improve the prediction accuracy. The model performs best when the window size is 6, and the prediction accuracy reaches 97%. Compared with LSTM and Random Forest models, the VMD-CNN-Transformer model has a higher coefficient of determination and better performance. Future research can further optimize the model parameters, expand the application range, and improve the prediction ability and application value.

References

[1] Wang S, Kaur M, Li T, et al. Effect of different pollution parameters and chemical components of PM2. 5 on health of residents of **nxiang City, China [J]. International Journal of Environmental Research and Public Health, 2021, 18 (13): 6821.

[2] Zhang L, Lin J, Qiu R, et al. Trend analysis and forecast of PM2. 5 in Fuzhou, China using the ARIMA model [J]. Ecological indicators, 2018, 95: 702 - 710.

[3] Kang J, Zou X, Tan J, et al. Short-Term PM2. 5 concentration changes prediction: a comparison of meteorological and historical data [J]. Sustainability, 2023, 15 (14): 11408.

[4] Liang Y, Ma J, Tang C, et al. Hourly forecasting on PM2. 5 concentrations using a deep neural network with meteorology inputs [J]. Environmental Monitoring and Assessment, 2023, 195 (12): 1510.

[5] Xu X, Tong T, Zhang W, et al. Fine-grained prediction of PM2. 5 concentrations based on multisource data and deep learning [J]. Atmospheric Pollution Research, 2020, 11 (10): 1728 - 1737.

[6] Zhang J, Peng Y, Ren B, et al. Pm2. 5 concentration prediction based on cnn-bilstm and attention mechanism[J]. Algorithms, 2021, 14 (7): 208.

[7] Qin C, Huang G, Yu H, et al. Adaptive VMD and multi-stage stabilized transformer-based long-distance forecasting for multiple shield machine tunneling parameters [J]. Automation in Construction, 2024, 165: 105563.

[8] Zhang Z, Zhang S. Modeling air quality PM2. 5 forecasting using deep sparse attention-based transformer networks [J]. International Journal of Environmental Science and Technology, 2023, 20 (12): 13535 - 13550.

[9] Kristiani E, Lin H, Lin J R, et al. Short-term prediction of PM2. 5 using LSTM deep learning methods [J]. Sustainability, 2022, 14 (4): 2068.

[10] Chen M, Bai J, Zhu S, et al. The influence of neighborhood-level urban morphology on PM2. 5 variations based on random forest regression [J]. Atmospheric Pollution Research, 2021, 12 (8): 101147.