Atmospheric PM2.5 Concentration Prediction Based on the GV-MCBiLSTM Hybrid Model

Boyu Dong, Boqun Li, Xi Chen

***Abstract*—In the context of exacerbated atmospheric compound pollution, accurate prediction of PM2.5 concentrations is crucial for pollution prevention, control, and public health management. To address the limitations of existing methods in handling non-stationary signals and integrating spatiotemporal characteristics, this study proposes a novel hybrid hourly PM2.5 prediction model, namely GWO-VMD-MultiscaleConv-BiLSTM (GV-MCBiLSTM). Firstly, the Grey Wolf Optimizer (GWO) algorithm is employed to dynamically optimize the key parameters of Variational Mode Decomposition (VMD), thereby enhancing the accuracy of the decomposition process. Secondly, the original PM2.5 concentration time series is decomposed into multiple Intrinsic Mode Functions (IMFs), which effectively separate the aliased modal features, yielding clearer and more interpretable signal components for subsequent modeling. Finally, a Multi-scale Convolutional Bidirectional Long Short-Term Memory Network (MCBiLSTM) is constructed, integrating various influencing factors such as air pollutants and meteorological variables, to capture the spatiotemporal dependencies of PM2.5 concentrations. Experimental results demonstrate that the proposed model outperforms existing methods in PM2.5 prediction in Shenyang, achieving an RMSE of 1.179 μg/m³, an MAE of 0.746 μg/m³, and an R^2 value of 0.997, thus validating its effectiveness and accuracy.**

***Index Terms*—PM2.5 prediction, GWO-VMD, BiLSTM, Multi-scale Convolution**

I. Introduction

UNDER the background of rapid global economic development and accelerated industrialization, the increase in environmental pollution sources such as vehicle exhaust emissions and the combustion of chemical fuels has led to the worsening of air pollution. In recent years, the frequent occurrence of air pollution events and extreme weather has posed a tremendous threat to the ecological environment and public health [1]. Among these pollutants, PM2.5 (fine particulate matter with a diameter of less than or equal to 2.5 μm) is regarded as a crucial indicator for assessing air quality. Due to the extremely small size of PM2.5 particles, they can remain suspended in the air for an extended period and carry toxic and harmful substances, thus posing a significant threat to human health [2–4]. Therefore, it is particularly essential to propose a reliable and accurate early warning method for predicting the changes in PM2.5 concentration, which can help reduce exposure risks and provide a scientific basis for environmental governance[5,6].

Manuscript received February 18, 2025; revised May 31, 2025.

Boyu Dong is postgraduate student of School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, Liaoning, 114051, China (e-mail:849052973@qq.com).

Boqun Li is professor of the School of Electronic and Information Engineering,University of Science and Technology Liaoning, Anshan, Liaoning,114051, China (Corresponding author, e-mail: Lbqhylyxab@163.com).

Xi Chen is postgraduate student of School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, Liaoning, 114051, China (e-mail:2100914719@qq.com.).

In the domain of PM2.5 concentration prediction research, scholars have proposed various prediction models, mainly encompassing four categories: deterministic mechanism models, traditional statistical models, artificial intelligence models, and hybrid models. Deterministic mechanism models predict PM2.5 concentration by studying the physical and chemical generation mechanisms of pollutants and simulating their diffusion and transfer processes. Common models such as WRF [7] and WRF-CMAQ [8] are employed. Nevertheless, these models are complex and require a substantial amount of detailed data, especially the difficulty in obtaining ground emission data, which leads to significant errors in simulation results. Traditional statistical models like the grey model (GM) [9], multivariate linear regression model (MLRM) [10] autoregressive integrated moving average model (ARIMA) [11], generalized additive model (GAM) [12], and land use regression (LUR) [13] can make predictions using historical data, but they are often suitable for linear relationships and have difficulty in handling the nonlinearity and non-stationarity of PM2.5 sequences, resulting in low prediction accuracy. With the rapid advancement of artificial intelligence, deep learning models have been widely utilized in PM2.5 prediction due to their excellent nonlinear modeling capabilities. Models such as backpropagation neural network (BPNN) [14], support vector regression (SVR) [15], and long short-term memory network (LSTM) [16] can effectively capture long-term dependencies in time series, but they are weak in extracting spatial features. Especially when dealing with complex environmental data, their performance is often unsatisfactory. To address this issue, Ning Zhou proposed the CNN-LSTM model [17], which combines traditional LSTM with convolutional neural networks CNN to enhance the ability to capture spatial features and has made certain progress. CNN-LSTM has to some extent addressed the problem of local spatial dependence, however,it still has difficulty in effectively capturing remote spatial dependence and has limitations in complex spatiotemporal data modeling. To further overcome these deficiencies, this paper proposes a neural network model based on the MultiscaleConv-BiLSTM (MCBILSTM). Compared with traditional models, the MCBILSTM effectively enhances the model's ability to capture spatial features and long-term spatial dependencies by constructing a multi-branch network structure and integrating the BiLSTM unit. This method, through the fusion of multi-scale convolution and BiLSTM, enables the model to understand the dynamic changes of spatio-temporal features at different scales.

However, due to the complexity and non-stationarity of the PM2.5 time series, prediction errors are still inevitable. To address this, researchers have introduced signal processing techniques to reduce the interference of nonlinearity and non-stationarity on prediction results [18]. For example, Huang

combined Empirical Mode Decomposition (EMD) with the Gated Recurrent Neural Network (GRNN), achieving certain success. However, EMD has issues of mode aliasing and an insufficiently rigorous mathematical foundation. To overcome this limitation, this paper selects the VMD method, which has a more rigorous mathematical theoretical basis. VMD decomposes the signal into modes with clear frequency band limitations by minimizing frequency deviation, effectively avoiding the mode aliasing problems that may occur in EMD and Ensemble Empirical Mode Decomposition (EEMD), and does not rely on noise addition, thereby achieving more accurate and stable signal decomposition [19,20]. Although VMD performs well in decomposition, its performance still depends on the selection of decomposition layers and penalty factors, which usually require manual adjustment. To solve this problem, Liu proposed using Particle Swarm Optimization (PSO) to automatically optimize VMD parameters, thereby improving decomposition accuracy [21]. However, PSO is prone to getting stuck in local optima and cannot guarantee the acquisition of the global optimum. To overcome this limitation, this study adopts the GWO algorithm to optimize the key parameters of VMD. GWO effectively balances local and global search through adaptive convergence factors and information feedback mechanisms, significantly enhancing the parameter optimization effect [22].

II. Method

A. Variational modal decomposition

VMD algorithm employs a completely non-recursive mode decomposition method, effectively avoiding the problem of mode mixing [23]. The core idea of VMD is to construct and solve the following variational problem:

(1) Construct the variational problem

$$\begin{aligned}&\min_{\{u_k\},\{\omega_k\}}\left\{\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-j\omega_k t}\right\|_2^2\right\}\\&\text{s.t.}\quad\sum_{k=1}^{K}u_k=f\end{aligned}\tag{1}$$

where $u_k(t)$ constitutes the subsequence resulting from VMD decomposition, $\omega_k(t)$ is the frequency, $\delta(t)$ is the Dirac delta function, and f is the signal to be decomposed.

(2) By introducing a quadratic penalty factor α and Lagrange multiplier λ into the Lagrange value function, the constrained variational problem is transformed into an unconstrained variational problem.

$$\begin{aligned}\underset{\{u_k\},\{\omega_k\},\lambda}{L}&=\alpha\sum_{k=1}^{K}\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)u_k(t)\right]e^{-j\omega_k t}\right\|_2^2\\&+\left\|f(t)-\sum_{k=1}^{K}u_k(t)\right\|_2^2+\left\langle\lambda(t),f(t)-\sum_{k=1}^{K}u_k(t)\right\rangle\end{aligned}\tag{2}$$

(3) By using the alternating direction method of multipliers in combination with the Fourier equidistant transformation and other methods, the individual components and their central frequencies are continuously updated. Eventually, the optimal solution to the original constrained variational problem is obtained. Update all u_k and ω_k when $k \geq 0$.

$$\widehat{u}_k^{n+1}(\omega)=\frac{f(\omega)-\sum_{i=1,i<k}^{k}\widehat{u}_i^n(\omega)+\frac{\widehat{\lambda}^n(\omega)}{2}}{1+2a\left(\omega-\omega_k^n\right)^2}\tag{3}$$

$$\omega_k^{n+1}=\frac{\int_0^\infty\omega\left|\widehat{u}_k^{n+1}(\omega)\right|^2d\omega}{\int_0^\infty\left|\widehat{u}_k^{n+1}(\omega)\right|^2d\omega}\tag{4}$$

The updated λ with double improvements, where γ is the update rate.

$$\widehat{\lambda}^{n+1}(\omega)=\widehat{\lambda}^n(\omega)+\gamma\left[\widehat{f}(\omega)-\sum_{k=1}^{K}\widehat{u_k}^{n+1}(\omega)\right]\tag{5}$$

(4) When the iterative constraint of the following formula is satisfied, stop updating the functional to obtain the solution of the constrained variational problem. In the formula, ε is the discrimination accuracy parameter.

$$\sum_{k=1}^{K}\left(\frac{\left\|\widehat{u}_k^{n+1}-\widehat{u}_k^n\right\|_2^2}{\left\|\widehat{u}_k^n\right\|_2^2}\right)<\varepsilon\tag{6}$$

The overall execution process of the VMD algorithm is:

Step 1. Initialize u_1^k, ω_1^k, λ_1^k, $n=0$, $k=1$.

Step 2. According to the VMD algorithm formulas (3) and (4), update $\widehat{u}_k^{n+1}$ and $\widehat{\omega}_k^{n+1}$ until the decomposition times reach k, then stop the inner loop and update the Lagrange multiplier $\widehat{\lambda}^{n+1}$ according to formula (5).

Step 3. Continue the loop until the iteration constraint formula (6) is satisfied; otherwise, proceed to Step 2 and continue the iteration.

B. Parameter optimization of variational mode decomposition using grey wolf optimization algorithm

The modal decomposition accuracy of the VMD algorithm is predominantly governed by two crucial parameters: the mode number K and the bandwidth constraint α. Selecting the optimal parameters through manual configuration is challenging. Hence, in this paper, the GWO algorithm proposed by Mirjalili et al. is employed to optimize the parameters K and α. This algorithm strikes a balance between local search optimization and global search exploration by emulating the four-level social structure ($\alpha > \beta > \delta > \omega$) of the grey wolf pack and the group hunting strategy [24]. The principal processes of GWO encompass establishing the wolf hierarchy, encircling the prey, hunting the prey, and attacking the prey. Eventually, the position vector X_α of the wolf is regarded as the optimal solution of the objective function. The main steps are as follows:

Step 1: Initialize the grey wolf population X_i ($i = 1, 2, \ldots, n$) and other parameters.

Step 2: Determine the adaptive function and calculate the fitness value of each grey wolf individual.

Step 3: Identify α, β, δ. α is the grey wolf individual with the highest fitness value, β is the one with the second highest fitness value, and δ is the one with the third highest fitness value. X_α, X_β, X_δ represent the positions of α, β, and δ.

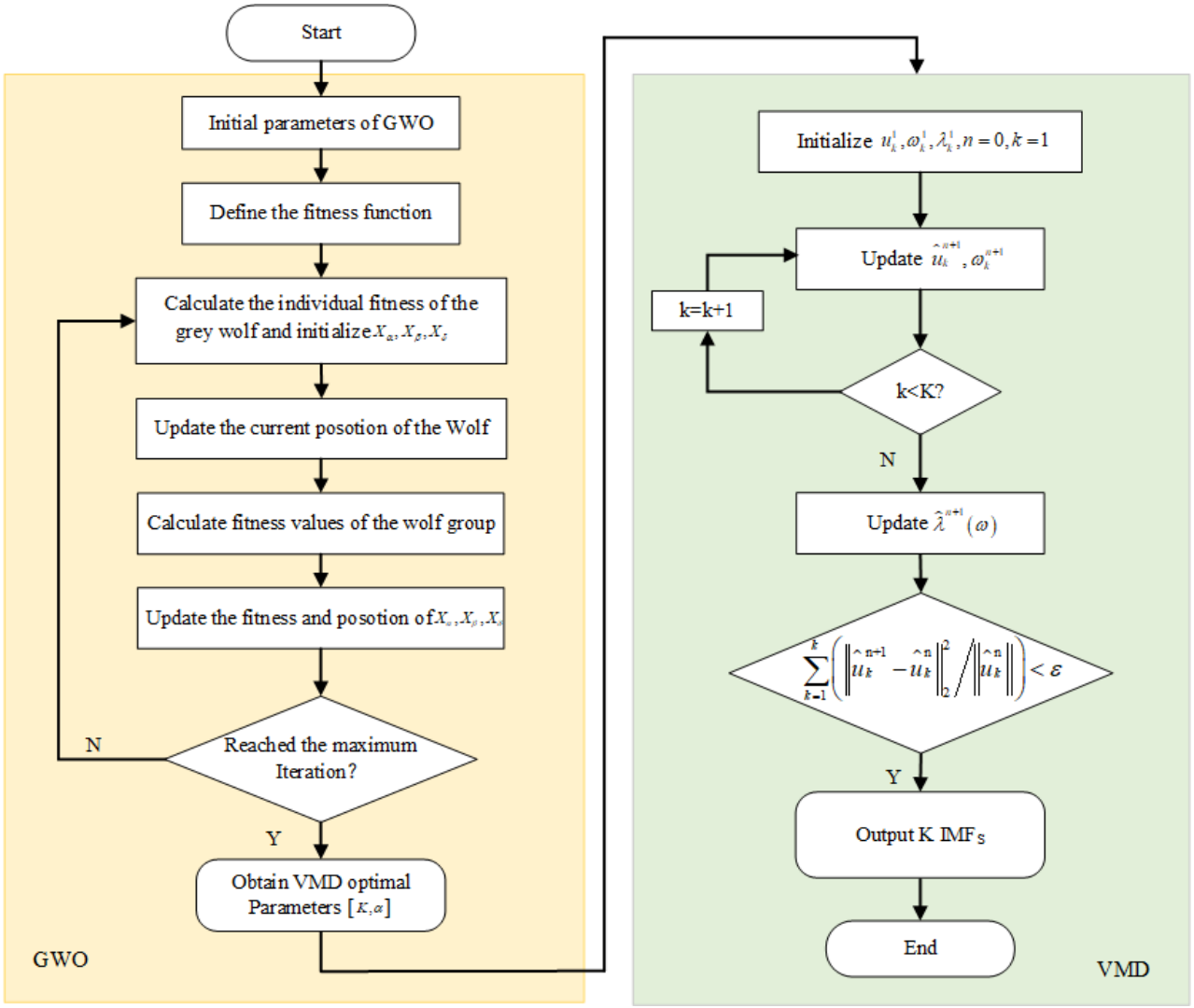

Fig. 1. Flowchart of GWO-VMD

Step 4: Update the position of the current grey wolf individual$X_\omega = \frac{X_1+X_2+X_3}{3}$, where X_1, X_2, and X_3 are influenced by X_α, X_β, X_δ, and X_ω represents the adjusted position.

Step 5: Calculate the fitness value of all grey wolf individuals and update X_α, X_β, and X_δ.

Step 6: $t = t+1$, repeat steps 4 to 6 until the maximum number of iterations is reached.

Step 7: Return the optimal solution X_α as the expected result.

After obtaining the optimal parameters $[K, \alpha]$, VMD is used to decompose the subsequences. The overall workflow diagram of GWO-VMD is presented, as shown in Fig.1.

C. Multi-scale extraction module

One-dimensional convolutional networks have demonstrated significant advantages in extracting temporal features [25], and their performance is closely related to the scale sensitivity of the receptive field [26]. Fixed-scale convolutional kernels have inherent limitations when dealing with the multimodal feature coupling problem in PM2.5 concentration sequences, resulting in a mismatch of feature resolution and insufficient feature separation across different time scales. This study proposes a multi-scale convolutional architecture, which builds a temporal feature pyramid through four sets of heterogeneous convolutional kernels of 1×5, 1×7, 1×9, and 1×11. Through multi-scale convolution, it can effectively model temporal features at different levels and capture subtle fluctuations and abrupt changes. The mathematical expression is as follows :

$$\begin{cases} A_1(t) = \sum_{k=0}^{4} x_A(t-k)\omega_A(k) + b_A \\ A_2(t) = \sum_{k=0}^{6} x_A(t-k)\omega_A(k) + b_A \\ A_3(t) = \sum_{k=0}^{8} x_A(t-k)\omega_A(k) + b_A \\ A_4(t) = \sum_{k=0}^{10} x_A(t-k)\omega_A(k) + b_A \end{cases} \tag{7}$$

Where $A_i(t)$ represents the output feature of the i-th convolutional layer at time step t, $x_d(t-k)$ is the input signal, $\omega_A(k)$ is the convolution kernel weight, and b_A is the bias term.

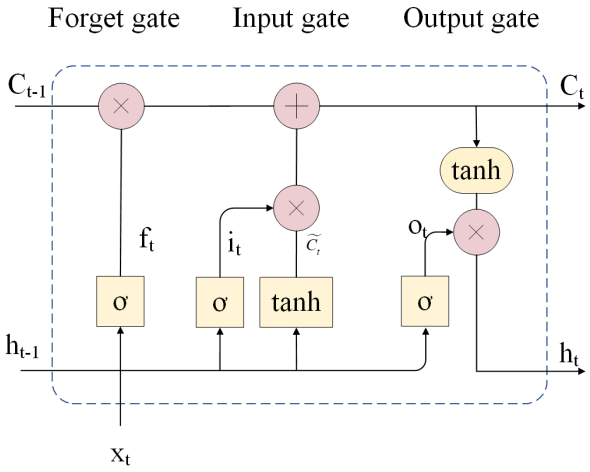

Fig. 2. Structure diagram of the LSTM

D. Bidirectional long short-term memor

The Long Short-Term Memory (LSTM) network was initially proposed by Hochreiter and Schmidhuber in 1997, resolving the issues of gradient explosion and vanishing in traditional Recurrent Neural Networks (RNN) through the introduction of an internal gating mechanism [27,28].The core structure of LSTM consists of three key gated units: the Forget Gate, the Input Gate, and the Output Gate, as shown in Fig.2. Among them, the Forget Gate is responsible for screening the historical information that needs to be retained, the Input Gate controls the update of new information, and the Output Gate determines the output of the hidden state at the current moment.The detailed process of LSTM neural unit update is as follows:

(1) The output of the previous moment and the input of the current moment are input into the forget gate, and the output of the forget gate is obtained after calculation, as shown in the following figure:

$$f_t = \text{sigmoid}\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{8}$$

Where the value range of f_t is $(0, 1)$, W_f represents the weights of the forget layer, and b_f represents the bias of the forget layer. h_{t-1} indicates the hidden state of the previous time step, and x_t represents the input information of the current time step.

(2) Feed the output of the previous moment and the input of the current moment into the input gate, and calculate to obtain the output of the input gate and the state of the candidate cell, as shown below:

$$i_t = \text{sigmoid}\left(U_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{9}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{10}$$

Where it lies within the range of $(0, 1)$, where W_i denotes the weight of the input gate, b_i indicates the bias of the input gate, W_c represents the weight of the candidate input gate, and b_c represents the bias of the candidate input gate.

(3) Update the current cell state, calculated as follows:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C_t \tag{11}$$

where C_t lies within the range $(-1, 1)$.

(4) Feed the output of the previous moment and the input of the current moment into the output gate, and calculate to obtain the output value, as follows:

$$o_t = \text{sigmoid}\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{12}$$

where o_t lies within the range $(0, 1)$, W_o represents the weight of the output gate, and b_o denotes the bias of the output gate.

(5) The output of the output gate is calculated based on the cell state to obtain the final LSTM output, as follows:

$$h_t = o_t \cdot \tanh(C_t) \tag{13}$$

However, LSTM is capable of merely exploiting past information. On the foundation of LSTM, BiLSTM incorporates an additional backward LSTM layer, integrating the forward and backward LSTMs to form a BiLSTM network, as illustrated in Fig.3. The forward LSTM is capable of extracting the past data information of the input sequence, whereas the backward LSTM can acquire the future data information of the input sequence. The computations of the two LSTMs ameliorate the learning of long-term dependencies, thereby enhancing the accuracy of the model[29]. Consequently, for the purpose of better capturing the changing trends of indoor environmental variables over time, this paper will employ BiLSTM as the underlying architecture of the deep neural network.

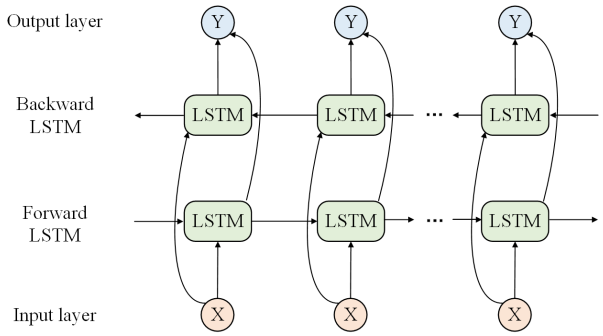

Fig. 3. Structure diagram of the biLSTM

At time t, the output value of the BiLSTM hidden layer, denoted as h_t, is composed of the forward hidden state, denoted as h_t^{forward}, and the backward hidden state, denoted as h_t^{backward}.

$$\overrightarrow{h_t} = \overrightarrow{\text{LSTM}}\left(h_{t-1}, x_t, C_{t-1}\right) \tag{14}$$

$$\overleftarrow{h_t} = \overleftarrow{\text{LSTM}}\left(h_{t+1}, x_t, C_{t+1}\right) \tag{15}$$

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{16}$$

Wherer $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ respectively represent the forward and backward hidden states, x_t represents the input at time t, and C_{t+1} and C_{t-1} respectively represent the cell states of the network at the previous and next moments.

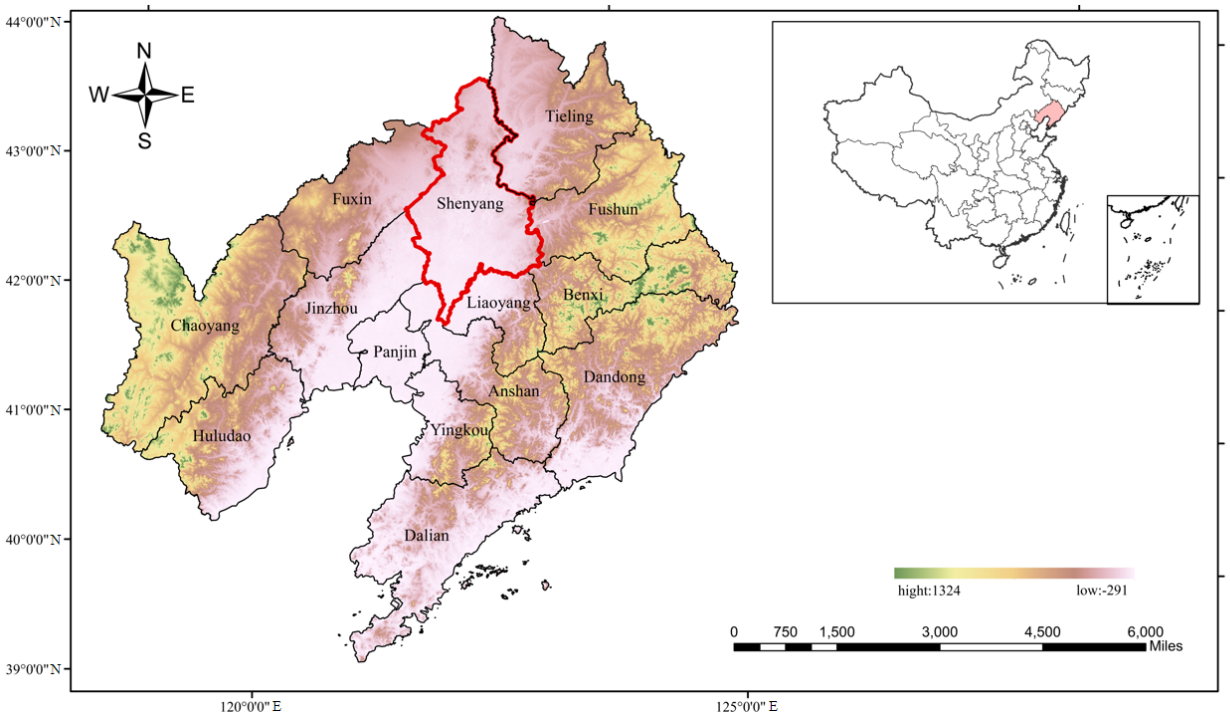

Fig. 4. Air Pollution Research Area of Shenyang City

III. Results

In this chapter, the basic situation of the regional data used and the evaluation indicators are first introduced, and comparative experiments and decomposition effectiveness verification of the proposed model are conducted. All experiments are implemented in the PyTorch 2.2.1 environment, and the hardware configuration of the computing platform is Intel Core i7-12700 and NVIDIA GeForce RTX 3060.

A. Research area

Shenyang is situated in the middle of Liaoning Province, China, with geographical coordinates approximately at 41.8°N latitude and 123.4°E longitude. It is a typical city featuring a temperate continental monsoon climate, with well-defined four seasons and remarkable climate variations, as depicted in Fig.4. As an important old industrial base in China, certain areas of Shenyang have long been concentrated with heavy industrial enterprises and outdated industrial facilities. Its air quality is profoundly influenced by multiple factors such as industrial emissions, traffic pollution, and meteorological conditions. In recent years, along with the rapid economic growth and the continuous acceleration of the urbanization process, the issues of industrial emissions and traffic pollution have become increasingly prominent, resulting in a continuous increase in PM2.5 concentrations and posing a distinct threat to the health of residents. Therefore, this research selects Shenyang as the study area. All the data utilized are derived from the China Air Quality Online Monitoring and Analysis Platform (https://www.aqistudy.cn/historydata/), with a data time span ranging from January 1, 2017 to January 1, 2024. The data encompass concentrations of major pollutants such as PM2.5, PM_{10}, SO_2, NO_2, CO, and O_3 , as well as crucial meteorological parameters like relative humidity (RH), temperature (TEMP), wind direction (WD), and wind speed (WS), providing a solid data foundation for an in-depth analysis of air quality variations and air pollution prevention and control measures in the Shenyang region.

B. Evaluation criteria

In order to evaluate the performance of the model more accurately, four statistical metrics were employed to compare the predicted sequence with the actual PM2.5 sequence, namely root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination R^2.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2} \tag{17}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \widehat{y_i}| \tag{18}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \widehat{y_i}}{y_i}\right| \tag{19}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{20}$$

Where i is the sample serial number, n is the total number of samples, y_i is the measured data of sample i, $\widehat{y_i}$ is the model-predicted data of sample i, and $\overline{y}$ is the average of the actual values of sample i.

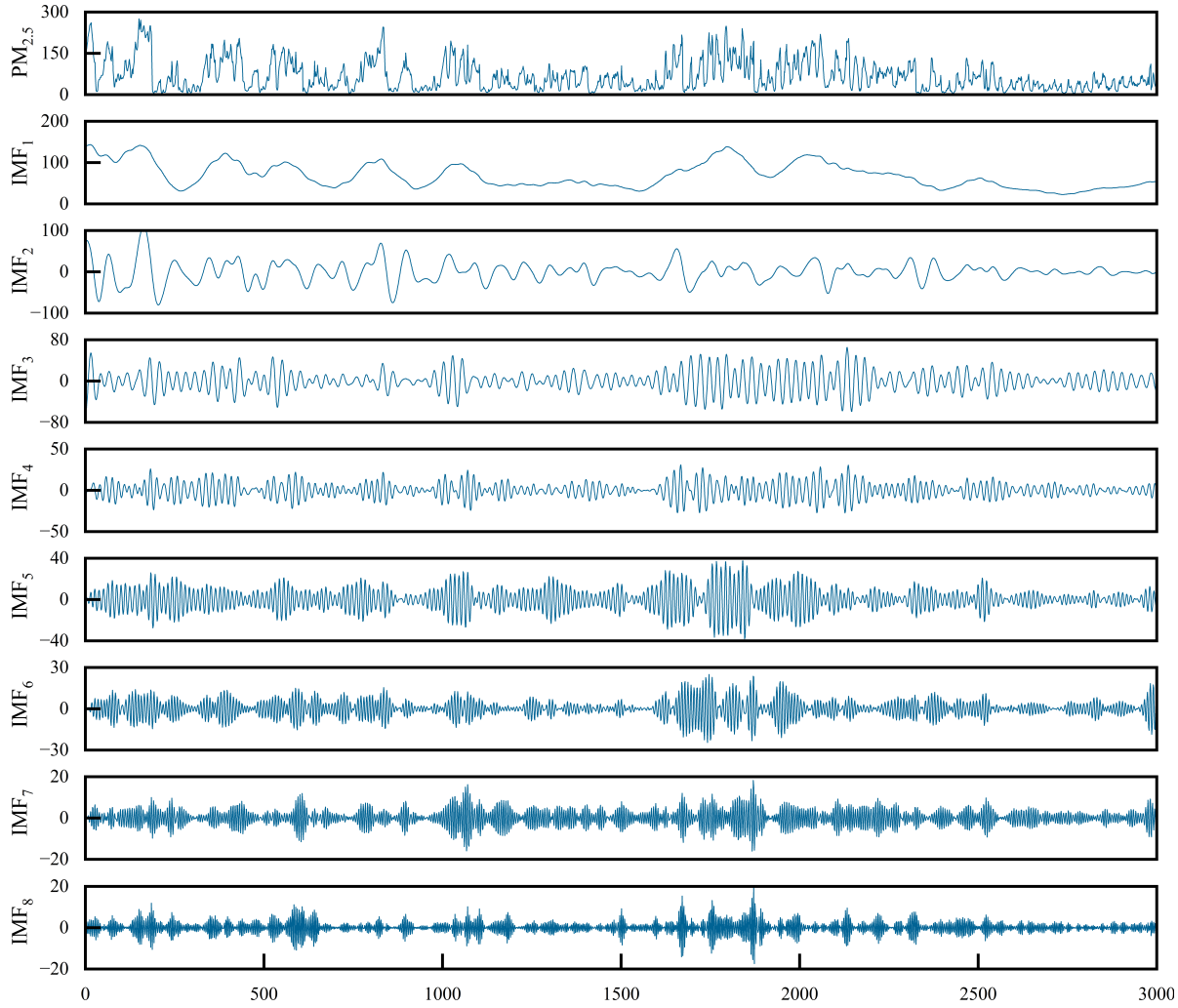

Fig. 5. VMD Decomposition Results of the Original PM2.5 Time Series

C. Evaluation of model performance

To capture the dynamic variations of time series data more effectively, this research employed the sliding window technique. The sliding window divides the time series data into subsequences of fixed length, allowing the model to gradually learn and predict the changing trends of PM2.5 concentrations. Specifically, we set the window length at 30 and the step size at 1, thereby ensuring that the model can fully leverage the information of historical data for prediction. The introduction of this approach further enhances the model's capacity for modeling time dependencies.

This study establishes a comprehensive validation system with 8 sets of control experiments, ensuring a thorough evaluation of the model architectures. The correspondence between the model architectures and their respective numbers is shown in Table 1.

To further enhance the efficacy of VMD, this paper introduces the GWO algorithm for the dynamic optimization of the key parameters of VMD. Based on the principle of swarm intelligence optimization, by minimizing the objective function to reduce the degree of mode aliasing, GWO conducts global optimization within the preset parameter space (the decomposition layer number $K \in [3, 10]$, the penalty factor $\alpha \in [1000, 4000]$). Fig.5 demonstrates that

TABLE I
MODEL NAME REFERENCE

Model ID	Model Name
M1	LSTM
M2	BiLSTM
M3	CNN-BiLSTM
M4	MultiScaleConv-BiLSTM
M5	VMD-GWO-LSTM
M6	VMD-GWO-BiLSTM
M7	VMD-GWO-CNN-BiLSTM
M8	VMD-GWO-MultiScaleConv-BiLSTM

the optimized VMD successfully realizes the intrinsic mode decomposition of the PM2.5 time series signal, effectively separating the mode components with different time scales. To comprehensively assess the performance of each model, we analyzed the performance of different models in PM2.5 concentration prediction by comparing the RMSE , MAE , R^2 and MAPE of the prediction results. The comparative experiments in Table 2 validate the effectiveness of VMD in improving the prediction accuracy. The experimental results indicate that the models applying VMD decomposition outperform those without VMD in different prediction periods (T+1, T+6, T+12). This might stem from the fact that VMD constrains the variational model, suppressing the endpoint

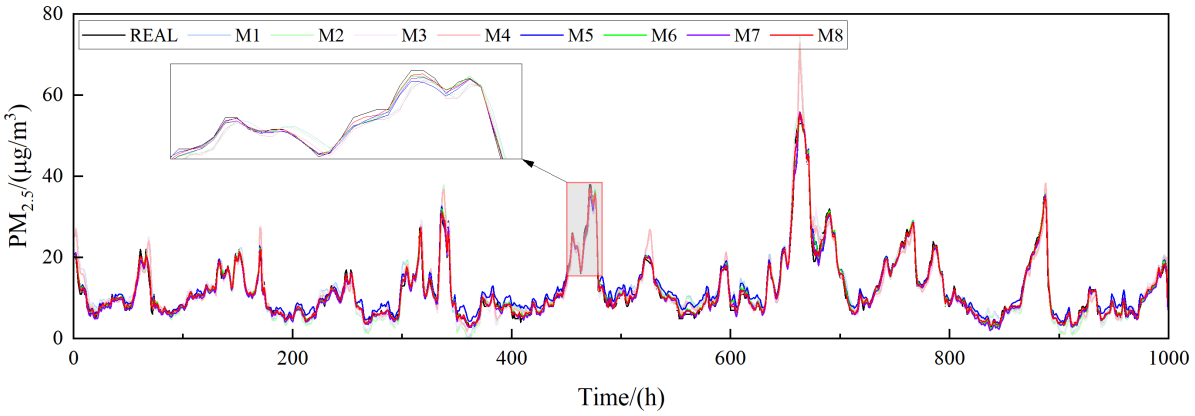

Fig. 6. Comparison of Prediction Results from Different Models within the 1-Hour Prediction Interval

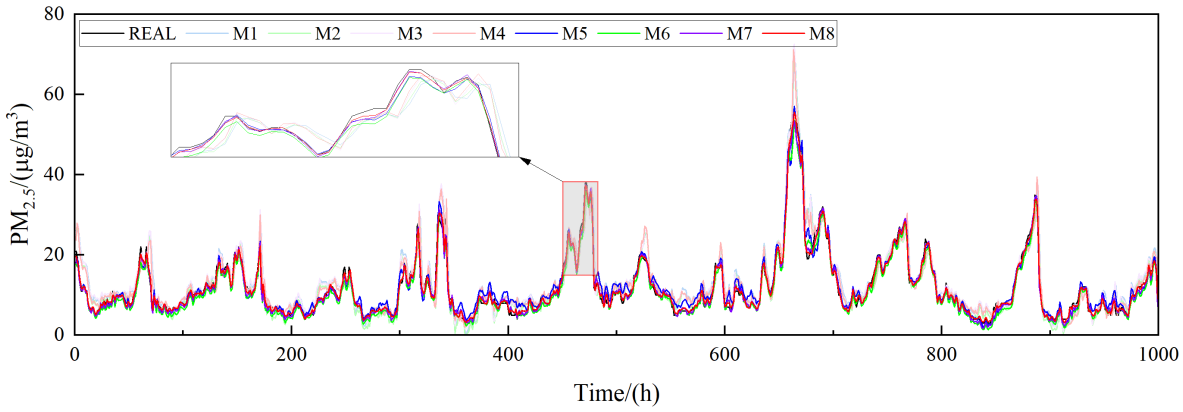

Fig. 7. Comparison of Prediction Results from Different Models within the 6-Hour Prediction Interval

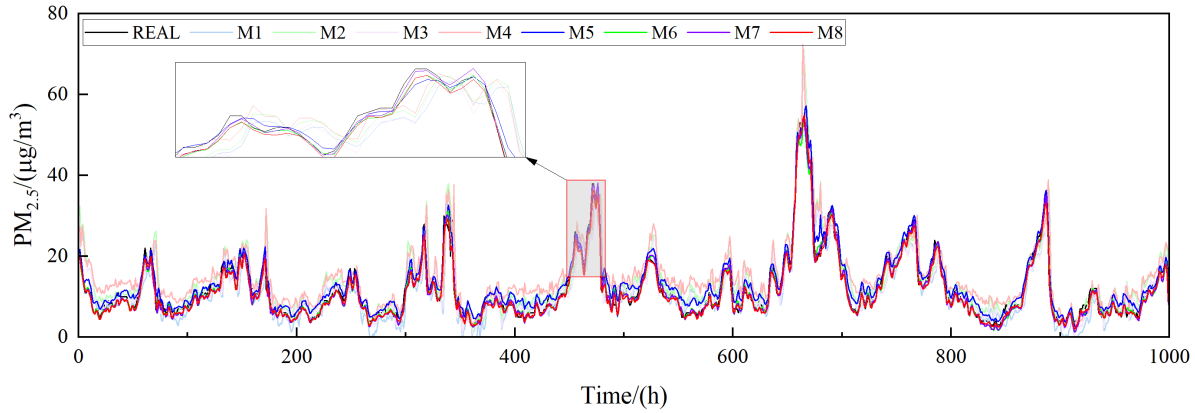

Fig. 8. Comparison of Prediction Results from Different Models within the 12-Hour Prediction Interval

effect and mode aliasing, thereby significantly enhancing the prediction accuracy of the model. This suggests that VMD technology effectively strengthens the model's ability to capture signals and reduces the prediction errors, thereby proving its significant role in PM2.5 concentration prediction.Further analyses reveal that the M7 and the M8 models proposed in this study surpass M5 and M6 in predicting PM2.5 levels. While M5 and M6 can capture the dependencies within the time series, they have limitations in addressing spatial correlations. In contrast, M7 possesses the capability to extract spatial correlations and demonstrates superiority in spatio-temporal data modeling. Nevertheless, the proposed M8 model exhibits even more remarkable performance in capturing spatio-temporal relationships, emphasizing the significance of spatial correlations in PM2.5 prediction. Specifically, the core performance indicators of the M8 model constructed in this study are as follows: RMSE = 1.179 μg/m³, MAE = 0.746 μg/m³, MAPE = 0.043, and R^2 = 0.997, demonstrating significant advantages over the M7 model. During the T+1 period, RMSE improved by approximately 31.8% and MAE by approximately 30.5%; in the T+6 period, RMSE improved by approximately 27.5% and MAE by approximately 13.7%; in the T+12 period, RMSE improved by approximately 22.0% and MAE

TABLE II
THE PREDICTION PERFORMANCE INDICES OF DIFFERENT MODELS IN DIVERSE TIME INTERVALS

Model	T+1				T+6				T+12			
	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2
M1	4.603	2.833	0.153	0.972	6.594	4.054	0.211	0.931	9.021	5.534	0.281	0.872
M2	4.184	2.605	0.144	0.972	5.963	3.670	0.190	0.944	8.469	5.180	0.278	0.887
M3	3.968	2.435	0.133	0.975	5.584	3.444	0.181	0.950	8.072	4.886	0.237	0.897
M4	3.841	2.214	0.115	0.976	5.441	3.259	0.165	0.953	7.919	4.994	0.293	0.901
M5	2.629	1.632	0.109	0.989	2.997	1.827	0.109	0.985	3.755	2.400	0.157	0.977
M6	1.948	1.151	0.065	0.994	2.698	1.693	0.090	0.988	2.902	1.713	0.098	0.986
M7	1.728	1.073	0.056	0.995	2.071	1.091	0.061	0.993	2.628	1.511	0.087	0.989
M8	1.179	0.746	0.043	0.997	1.502	0.942	0.053	0.996	2.050	1.389	0.080	0.993

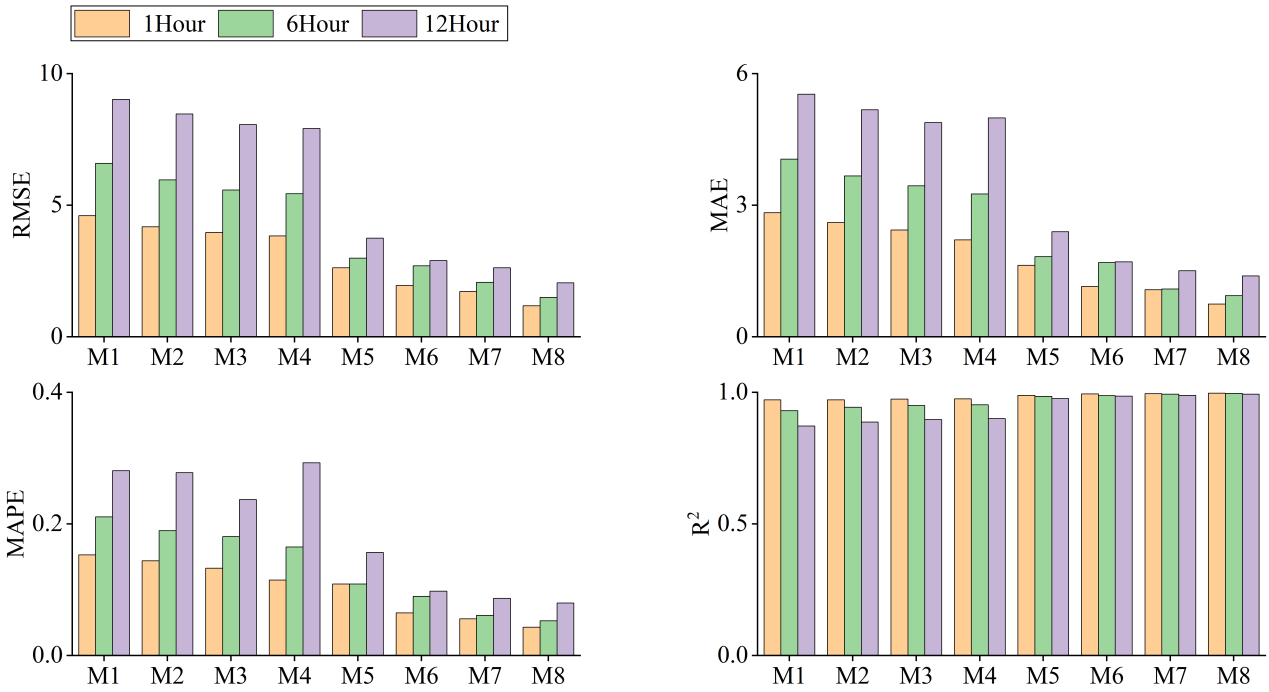

Fig. 9. Comparison of Model Metrics

by approximately 8.1%. These enhancements can mainly be attributed to the advantages of the proposed model in extracting spatio-temporal features, particularly the multi-scale convolution that enables the collaborative capture of cross-scale spatial features, thereby enhancing the prediction accuracy.

To present the research findings more visually and intuitively, we conducted a visualization of the data. The black curve represents the actual data, while the light blue, light green, light purple, light red, blue, green, purple, and red respectively correspond to the predicted values of the different models employed in this paper. The data visualization clearly indicates that, as depicted from Fig.6-8, within the short-term prediction range, the prediction results of all models are relatively consistent with the actual data. However, the proposed model demonstrates superior performance in terms of prediction accuracy. Additionally, the proposed model outperforms the others in peak prediction. As the prediction range extends, the prediction accuracy of all models for peaks and future trends gradually declines. By contrast, the model proposed in this study exhibits superior performance in the accuracy of the fit between the prediction results and the actual data, as shown in Fig.9. To sum up, this model can track the temporal evolution law of PM2.5 concentration more accurately.

IV. CONCLUSION

In this paper, a GV-MCBiLSTM fusion model is proposed and verified in the Shenyang region. The experimental results indicate that the proposed model achieves outstanding outcomes. Through the introduction of the VMD technique, a frequency-domain feature decoupling framework with adaptive bandwidth constraints is constructed, effectively mitigating the issues of noise interference and mode aliasing. Furthermore, the designed MCBILSTM network architecture possesses distinctive advantages in capturing the spatio-temporal evolution patterns of pollution events. The empirical results demonstrate that this model exhibits excellent accuracy and robustness in the task of predicting $PM_{2.5}$ concentration in the next hour, offering a novel technical approach for complex environmental time series prediction. Future studies will mainly explore multi-source data fusion and adaptive optimization mechanisms to further enhance the environmental generalization ability of the model.

REFERENCES

[1] H. Yang, Z. Liu, and G. Li, "A new hybrid optimization prediction model for $PM_{2.5}$ concentration considering other air pollutants and meteorological conditions," *Chemosphere*, vol. 307, no. 3, pp. 135701–135810, Jul. 2022.

[2] I. Seeni, A. Williams, C. Nobles, Z. Chen, S. Sherman, and P. Mendola, "Acute air pollution exposure and NICU admission: a case-crossover analysis," *Annals of Epidemiology*, vol. 37, pp. 64–70, Jul. 2019.

[3] H. Yang, Z. Liu, and G. Li, "A new hybrid optimization prediction model for $PM_{2.5}$ concentration considering other air pollutants and meteorological conditions," *Chemosphere*, vol. 307, no. 3, pp. 135701–135810, Jul. 2022.

[4] M. Kampa and E. Castanas, "Human health effects of air pollution," *Environmental Pollution*, vol. 151, no. 2, pp. 362–367, Feb. 2008.

[5] P. Yin, M. Brauer, A. Cohen, R. T. Burnett, J. Liu, Y. Liu, et al., "Long-term fine particulate matter exposure and nonaccidental and cause-specific mortality in a large national cohort of Chinese men," *Environmental Health Perspectives*, vol. 125, no. 11, p. 117002, Nov. 2017.

[6] B. Ostro, J. Hu, D. Goldberg, P. Reynolds, A. Hertz, L. Bernstein, and M. J. Kleeman, "Associations of mortality with long-term exposures to fine and ultrafine particles, species and sources: results from the California Teachers Study Cohort," *Environmental Health Perspectives*, vol. 123, no. 6, pp. 549–556, Jun. 2015.

[7] Q. Cao, L. Shen, S. C. Chen, and D. Y. Pui, "WRF modeling of $PM_{2.5}$ remediation by SALSCS and its clean air flow over Beijing terrain," *Science of the Total Environment*, vol. 626, pp. 134–146, Jun. 2018.

[8] F. Y. Cheng, C. Y. Feng, Z. M. Yang, C. H. Hsu, K. W. Chan, C. Y. Lee, and S. C. Chang, "Evaluation of real-time $PM_{2.5}$ forecasts with the WRF-CMAQ modeling system and weather-pattern-dependent bias-adjusted $PM_{2.5}$ forecasts in Taiwan," *Atmospheric Environment*, vol. 244, p. 117909, Jan. 2021.

[9] W. Zhou, X. Wu, S. Ding, X. Ji, and W. Pan, "Predictions and mitigation strategies of $PM_{2.5}$ concentration in the Yangtze River Delta of China based on a novel nonlinear seasonal grey model," *Environmental Pollution*, vol. 276, p. 116668, Jan. 2021.

[10] Z. Darynova, M. Malekipirbazari, D. Shabdirov, H. A. Khwaja, and M. A. Torkmahalleh, "Reliability and Stability of a Statistical Model to Predict Ground-Based $PM_{2.5}$ over 10 Years in Karachi, Pakistan, Using Satellite Observations," *Air Quality, Atmosphere & Health*, vol. 16, no. 5, pp. 1023–1035, May 2023.

[11] L. Zhang, J. Lin, R. Qiu, X. Hu, H. Zhang, Q. Chen, H. Tan, D. Lin, and J. Wang, "Trend analysis and forecast of $PM_{2.5}$ in Fuzhou, China using the ARIMA model," *Ecological Indicators*, vol. 95, pp. 702–710, Dec. 2018.

[12] Y. Zhang, R. Zhou, D. Hu, J. Chen, and L. Xu, "Modelling driving factors of $PM_{2.5}$ concentrations in port cities of the Yangtze River Delta," *Marine Pollution Bulletin*, vol. 184, p. 114131, Nov. 2022.

[13] C.-S. Huang, T.-H. Lin, H. Hung, C.-P. Kuo, C.-C. Ho, Y.-L. Guo, K.-C. Chen, and C.-F. Wu, "Incorporating satellite-derived data with annual and monthly land use regression models for estimating spatial distribution of air pollution," *Environmental Modelling & Software*, vol. 114, pp. 181–187, Mar. 2019.

[14] X. Wang, J. Yuan, and B. Wang, "Prediction and analysis of $PM_{2.5}$ in Fuling District of Chongqing by artificial neural network," *Neural Computing and Applications*, vol. 33, pp. 517–524, May 2021.

[15] W. Yang, M. Deng, F. Xu, and H. Wang, "Prediction of hourly $PM_{2.5}$ using a spacetime support vector regression model," *Atmospheric Environment*, vol. 181, pp. 12–19, Jun. 2018.

[16] M. Chen, P. Xu, Z. Liu, F. Liu, H. Zhang, and S. Miao, "Air pollution prediction based on optimized deep learning neural networks: PSO-LSTM," *Atmospheric Pollution Research*, vol. 16, no. 3, p. 102413, Mar. 2025.

[17] N. Zhou, B. Shang, M. Xu, L. Peng, and G. Feng, "Enhancing photovoltaic power prediction using a CNN-LSTM-attention hybrid model with Bayesian hyperparameter optimization," *Global Energy Interconnection*, vol. 7, no. 5, pp. 667–681, May 2024.

[18] G. Huang, X. Li, B. Zhang, and J. Ren, "$PM_{2.5}$ concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition," *Science of the Total Environment*, vol. 768, p. 145678, Apr. 2021.

[19] K. Dragomiretskiy and D. Zosso, "Variational Mode Decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, Mar. 2013.

[20] Y. Zhao, X. Peng, T. Tu, Z. Li, P. Yan, and C. Li, "WOA-VMDSCINet: Hybrid model for accurate prediction of ultra-short-term photovoltaic generation power considering seasonal variations," *Energy Reports*, vol. 12, pp. 3470–3487, Sep. 2024.

[21] S. Liu, Y. Chen, C. Luo, H. Jiang, H. Li, H. Li, and Q. Lu, "Particle Swarm Optimization-Based Variational Mode Decomposition for Ground Penetrating Radar Data Denoising," *Remote Sensing*, vol. 14, no. 13, p. 2973, Jul. 2022.

[22] S. Saremi, S. Z. Mirjalili, and S. M. Mirjalili, "Evolutionary population dynamics and grey wolf optimizer," *Neural Computing and Applications*, vol. 26, no. 5, pp. 1257–1274, May 2015.

[23] F. Wu, P. Min, Y. Jin, K. Zhang, H. Liu, J. Zhao, and D. Li, "A novel hybrid model for hourly PM2.5 prediction considering air pollution factors, meteorological parameters and GNSS-ZTD," *Environmental Modelling & Software*, vol. 167, p. 105780, Jul. 2023.

[24] L. Liu, X. Chen, X. Kang, X. Ma, J. Li, S. Liu, B. Zhao, Y. Zhang, and X. Liu, "A highly accurate double-ended traveling wave fault location method using GWO-VMD and the Hilbert transform for offshore MMC-MTDC grids," *Smart Power & Energy Security*, vol. 1, p. 1, Oct. 2024.

[25] D. X. Niu, M. Yu, L. J. Sun, T. Gao, and K. K. Wang, "Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism," *Applied Energy*, vol. 313, p. 118801, Aug. 2022.

[26] W. Luo, Z. Li, R. Zeng, and J. Wang, "Understanding the effective receptive field in deep convolutional neural networks," *Advances in Neural Information Processing Systems*, pp. 2924–2932, Dec. 2016.

[27] Y. Huang, J. Li, Y. Li, R. Lin, J. Wu, L. Wang, and R. Chen, "An Improved Hybrid CNN-LSTM-Attention Model with Kepler Optimization Algorithm for Wind Speed Prediction," *Engineering Letters*, vol. 32, no. 10, pp. 1957–1965, 2024.

[28] X. Li, Q. Ai, and M. Xu, "RUL Prediction of Lithium-ion Batteries Based on TimeGAN-Pyraformer-BiLSTM," *Engineering Letters*, vol. 32, no. 8, pp. 1675–1689, 2024.

[29] L. Zhu, X. Yang, D. Wang, and X. Huo, "Short-Time Inbound Passenger Flow Prediction of Urban Rail Transit Based on STL-HEOA-BiLSTM," *Engineering Letters*, vol. 32, no. 7, pp. 1340–1352, 2024.