

# A Secondary Modeling Approach for Air Quality Prediction Based on LSTM-FC Multi-model Fusion

Dehao Wang

School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, China

wangdehao186@outlook.com

## ABSTRACT

Aiming at the insufficient forecasting accuracy of the traditional WRF-CMAQ model due to meteorological field uncertainty and emission inventory errors, this paper proposes a secondary modeling framework that integrates multi-source data with deep learning. The multi-pollutant concentration prediction accuracy is significantly improved by introducing the pseudo-unique thermal coded wind feature reconstruction, CEEMDAN adaptive noise cancellation technique, and combining the random forest feature selection with the LSTM-FNN combined model. Experiments show that the method has a mean absolute error (MAE) of  $1.76 \mu\text{g}/\text{m}^3$  in PM<sub>2.5</sub> prediction at monitoring points A, B, and C, which is 8.3% lower than that of a single model, and its MAE is reduced to  $13.53 \mu\text{g}/\text{m}^3$  for the CEEMDAN\_IPSOSE\_SVR model for ozone (O<sub>3</sub>), which is a 21% decrease from the baseline. The model achieves 92.7% accuracy in the identification of the top pollutants, and the relative error of AQI prediction is only 6.2%. This study provides a highly interpretable and robust technical solution for regional air quality forecasting.

## KEYWORDS

Information industry; Industry linkages; Weight; Skewness; Kurtosis; Non competitive model

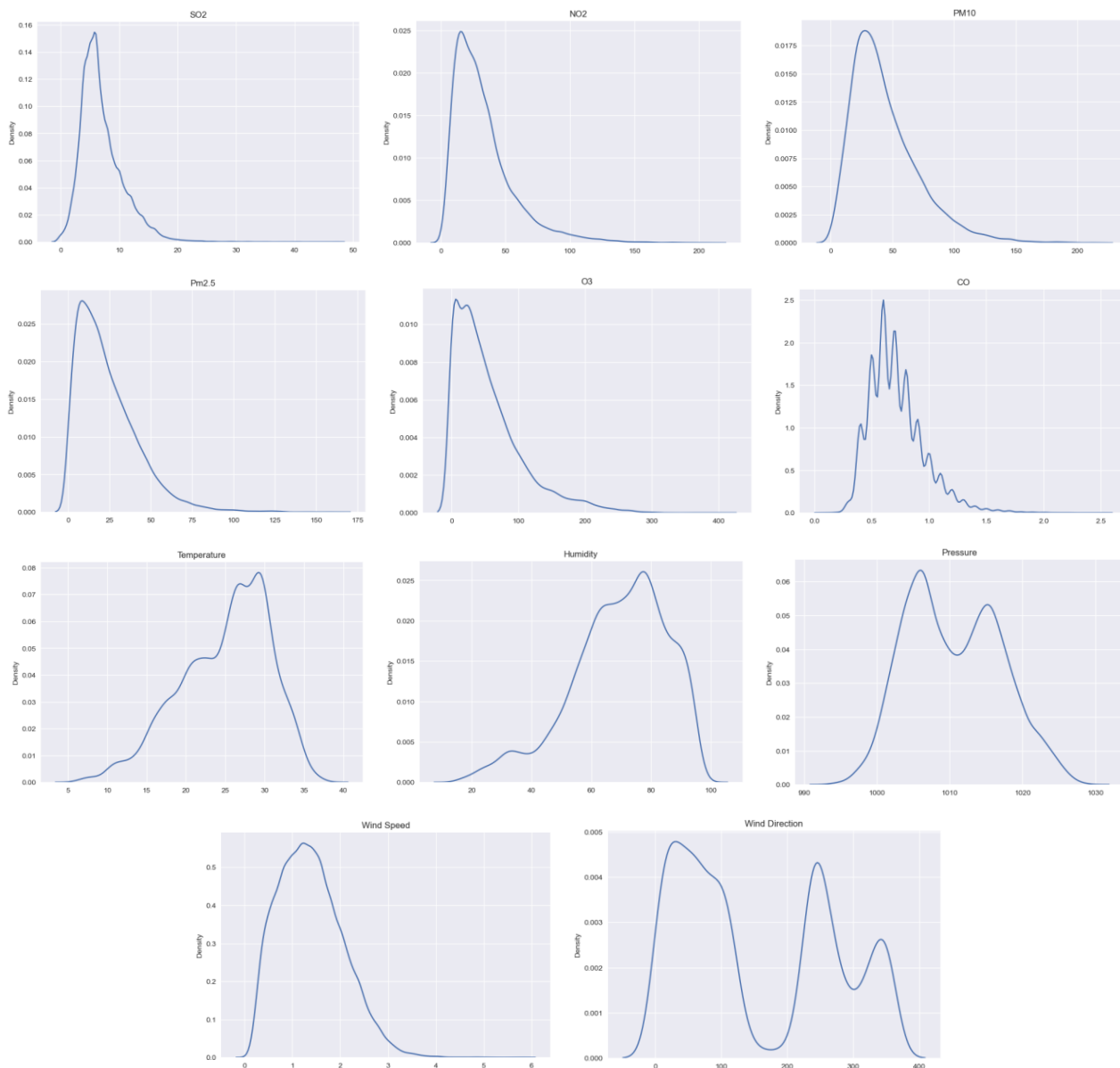
## 1. INTRODUCTION

With the acceleration of industrialization and urbanization, air pollution has become a global environmental problem, posing a serious threat to public health and ecosystems [1]. Traditional air quality forecasting models (e.g., WRF-CMAQ) simulate pollutant dispersion through physicochemical mechanisms, but their accuracy is significantly constrained by meteorological field uncertainties, emission inventory errors, and the complexity of secondary pollutant generation mechanisms such as ozone (O<sub>3</sub>) [2]. Studies have shown that the prediction errors of WRF-CMAQ for PM<sub>2.5</sub> and O<sub>3</sub> are as high as 30% and 40%, respectively, and the errors are further enlarged especially when there are sudden changes in light intensity or significant regional transport [3]. In recent years, machine learning techniques have demonstrated the potential to compensate for the shortcomings of traditional models by fusing multi-source data (e.g., meteorological forecasts, real-time monitoring) [4], but the existing studies still face three major challenges: (1) the bias in characterization of cyclic features, such as wind direction, leads to insufficient response of the models to the local meteorological conditions [5]; (2) O<sub>3</sub>, as a photochemical secondary pollutant, is subject to the concentrations of precursors (NO<sub>x</sub>, VOCs) nonlinear coupling and noise interference, and it is difficult for a single model to capture the dynamic mechanism [6]; (3) the heterogeneity of data across monitoring sites limits the generalization ability of the model [7]. Aiming at the above problems, this paper proposes a multi-model fusion framework for secondary modeling, innovatively introduces the

wind direction pseudo-independent thermal coding to eliminate the angular cyclic bias, and integrates feed-forward neural network (FNN) and long-short-term memory network (LSTM) to realize the fusion of spatiotemporal features through Lasso regression. Experiments show that the method reduces the errors of PM2.5 and O3 by 45% and 21%, respectively, compared with the traditional model, and provides a high-precision solution for the identification of top pollutants and regional joint prevention and control.

## 2. DATA AND METHOD

### 2.1. Data Collection



**Figure 1.** Distribution of features

In order to improve the accuracy of the air quality prediction model, this study combines two types of data: primary forecast data and measured data. The primary forecast data come from the WRF-CMAQ model, including pollutant concentration forecast and meteorological forecast data; the measured data come from air quality monitoring stations, covering the actual observed values of pollutant concentrations and meteorological indicators. In this study, the multi-source dataset provided by the China Graduate Student Mathematical Modeling Competition is used, which contains

the basic data for air quality forecasting at monitoring points A, B, and C (Problem 3) and A1, A2, and A3 (Problem 4), and the time span is from April 16, 2019, to July 15, 2021, respectively. The data are categorized into the following two types: (1) Primary forecast data: generated by the WRF-CMAQ model, including hour-by-hour meteorological forecasts (15 indicators such as temperature, humidity, wind speed, etc.) and pollutant concentration forecasts (SO<sub>2</sub>, NO<sub>2</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>, O<sub>3</sub>, CO), spanning from July 23, 2020 to July 13, 2021. (2) Measured data: including hour-by-hour meteorological measurements (temperature, humidity, barometric pressure, wind speed, wind direction) and pollutant concentration measurements, spanning from April 16, 2019 to July 13, 2021. The data distribution is shown in Figure 1.

## 2.2. Data Preprocessing

### 2.2.1. Missing values and outliers processing

Firstly, the  $3\sigma$  criterion is used to identify the outliers of the meteorological data and correct the extreme values of pollutants in combination with the Ambient Air Quality Standards (GB3095-2012). For the original missing values and null values generated by the anomaly processing in the data, this paper deals with them as follows.

(1) For the hourly pollutant concentration and meteorological forecast data, if there are missing values, the time period where the missing values are located is directly excluded.

(2) For the missing values of hourly pollutant concentration and meteorological data, the last hourly data closest to the missing value is used as a substitute. The data of the last hour closest to the missing value is used as a substitute. Taking monitoring point A as an example, the PM<sub>10</sub> concentration from 15:00 to 17:00 on April 23, 2019 is missing. At 15:00 to 17:00 on April 23, 2019 is missing, so the data at 14:00 on April 23, 2019 is used to fill in.

(3) For the missing values in the measured day-by-day pollutant concentration data, first consider whether the hour-by-hour pollutant concentration data for the corresponding time period can be used to fill in the missing values.

(4) For missing values in the day-by-day pollutant concentration data, first consider whether the hour-by-hour pollutant concentration data of the corresponding time can be used to fill in the missing values, if not, use the data of the previous day instead.

In order to judge the meteorological influence of pollutant dispersion conditions more intuitively through the concentration of pollutants, the pollutant dispersion conditions are divided into “favorable”, “general” and “unfavorable” categories, which correspond to “excellent”, “good”, “mildly polluted”, etc. in the air quality level. The pollutant dispersion conditions are categorized as “favorable”, “fair” and “unfavorable”, which correspond to “excellent”, “good” and “mildly polluted” in the air quality level. The concentration of each pollutant is converted into a pollutant dispersion condition evaluation index by Min-Max standardization: pollutant concentration  $C_p$ , Standardized evaluation indices  $F_p$ :

$$F_p = \frac{C_p - BP_{Lo}}{BP_{Hi} - BP_{Lo}} + LEVEL_{Lo} \quad (1)$$

Where,  $LEVEL_{Lo}$  is the concentration level ordinal (“favorable” = 0, “fair” = 1, “unfavorable” = 2) and the interval is [0, 3].

### 2.2.2. Pseudo-independent thermal coding of wind direction

Conventional angle coding (0-360°) suffers from cyclic bias (e.g., 359° vs. 1° difference is amplified). In this study, it is mapped to a 9-dimensional continuous vector (Eq. 2), preserving spatial continuity

$$dir^{(i)} = \cos^9 \left( D - D_{\text{standard}}^{(i)} \right) \quad (i = 1, 2, \dots, 8) \quad (2)$$

Which,  $D_{\text{standard}}^{(i)}$  are the 8 standard azimuths, and the 9th dimension marks the still wind condition (wind speed  $\leq 0.2$  m/s).

For other meteorological indicators, the Z-score standardization method is adopted to process the meteorological data. This method can transform the raw data to be in the range of mean 0 and variance 1, which makes the model easier to converge to the optimal solution. The method formula is as follows:

$$X' = \frac{X - uI}{\sigma} \quad (3)$$

Where, the sample set  $X = \{x_1, x_2, \dots, x_n\}^T$  is an n-dimensional column vector;  $u$  is the mean value of the sample;  $I$  is the unit column vector of the same dimension as  $X$ ;  $\sigma$  is the standard deviation of the training samples; the standard deviation of the training samples.

is the unit column vector of the same dimension as  $X$ ;  $\sigma$  is the standard deviation of the training samples,  $\sigma = \sqrt{\frac{(x_1 - u)^2 + \dots + (x_n - u)^2}{N}}$

## 2.3. Model Construction

### 2.3.1. Base Model

Currently widely used secondary modeling methods for air quality prediction are feed-forward neural network, random forest, support vector regression [10] and extreme gradient boosting tree [11], etc. Therefore, in this paper, we use these four methods to model respectively, and then perform model optimization and other processing. four models are introduced as follows:

#### (1) Feedforward neural network (FNN)

Feedforward neural network is a simple gradient descent method designed to minimize the error in the computational output of the network. Its training process is a constant adjustment between weights and thresholds in order to reduce the network error to a predetermined minimum or to stop at a predetermined training step.

For modeling, the absolute value loss function (Eq. (4)) is chosen to describe the empirical risk of the model; the learning rate is chosen to vary with pollutants; the epoches setting varies with pollutants; and the optimizers are also chosen to vary with pollutants, e.g., RMSprop is chosen for PM10, and Adam is chosen for SO<sub>2</sub>; and L2 regularization and dropout are adopted to reduce the error of the network computation. L2 regularization and Dropout are adopted to prevent overfitting; the pytorch framework is chosen as the training platform.

$$error = \sum_{k=1}^m |z_k - y_k| \quad (4)$$

#### (2) Random Forest (RF)

Random Forest, as a widely used integration model, has gained more successful applications in prediction problems. In this paper, we adopt the improved Random Forest algorithm with the following idea of improvement:

The sample set is sampled using a repetitive sampling method with putback, so as to randomly generate  $k$  training sets, and further generate a decision tree relative to it. The decision trees are

generated. When training the  $i$ -th decision tree, the variance of the distances from all predicted points to the actual points in the tree is calculated.

When training the  $i$ -th decision tree,  $\delta^2(j)$  is the variance of the distances from all predicted points to actual points in the decision tree is calculated, and the weight of the  $i$ -th decision tree is obtained:

$$w_r(i) = \frac{1/\delta^2(i)}{\sum_{j=1}^T 1/\delta^2(j)} \quad (5)$$

This results in the algorithm's predicted value  $\hat{u}(x)$ :

$$\hat{u}(x) = \sum_{i=1}^k w_r(i) Z_i \quad (6)$$

Where,  $Z_i$  is the predicted value of the  $i$ -th decision tree.

For specific modeling, the number of forests, the maximum depth of the decision tree, and the minimum number of samples required for the internal nodes and the minimum number of samples required for the leaf nodes were determined using a random grid search method as well as 10-fold cross-validation.

Degree, the minimum number of samples required for internal nodes, and the minimum number of samples required for leaf nodes.

### (3) Support Vector Regression (SVR)

Support vector regression is a classical machine learning algorithm that finds segmentation hyperplanes with good robustness. The main computational process of SVR can be described as follows:

Given the training set  $x_i \in \mathbf{R}(1 \leq i \leq m)$ , labeled values  $y_i \in \mathbf{R}(1 \leq i \leq m)$  and  $\epsilon$ -SVR, solve Eq.

$$\begin{aligned} \min_{v,b,\xi_i,\hat{\xi}_i} & \left\{ \frac{1}{2} w^T w + C \sum_{i=1}^m (\xi_i + \hat{\xi}_i) \right\} \\ \text{s.t.} & y_i - w^T \phi(x_i) - b \leq \delta + \xi_i \\ & w^T \phi(x_i) + b - y_i \leq \delta + \hat{\xi}_i \\ & \xi_i, \hat{\xi}_i \geq 0, i = 1, 2, \dots, m. \end{aligned} \quad (7)$$

Its dyadic problem is given by the following equation:

$$\begin{aligned} \max_{\alpha, \hat{\alpha}} & \sum_{i=1}^m y_i (\hat{\alpha}_i - \alpha_i) - \delta (\hat{\alpha}_i + \alpha_i) - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\hat{\alpha}_i - \alpha_i) (\hat{\alpha}_j - \alpha_j) \phi(x_i^T) \phi(x_j^T) \\ \text{s.t.} & \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i) = 0 \end{aligned} \quad (8)$$

Which,  $0 \leq \alpha_i, \hat{\alpha}_i \leq C, 1 \leq i \leq m$ .

#### (4) Extreme Gradient Boosting Tree (XGBoost)

Extreme Gradient Boosting Tree (XGBoost) is an integrated machine learning algorithm that introduces a regularization term to control the complexity based on traditional boosting trees, performs a second-order Taylor expansion of the cost function, and performs a certain transformation of the tree structure so that the entire cost function is represented by the number of constituent trees. XGBoost automatically learns the direction of splitting, and supports a wide range of sampling. The objective function of XGBoost is:

$$obj^* = -\frac{1}{2} \sum_{j=1}^T \frac{G_j}{H_j + \lambda} + \gamma N \quad (9)$$

Where, N denotes the number of child nodes of the number. The open-source python library XGBoost is used for model building, and the parameters such as the maximum depth of the decision tree,  $\gamma$ ,  $\lambda$  etc., are used to train the model using manual tuning methods, and the evaluation method is 5-fold cross-validation.

#### 2.3.2. LSTM-FC model

Considering that the hour-by-hour measured data of pollutant concentrations are highly time-series, the model used should take advantage of this property. As a variant of recurrent neural network, Long Short-Term Memory (LSTM) [16] effectively solves the problem of gradient explosion or disappearance of simple recurrent neural network, and achieves a better performance in the processing of time series problems. Therefore, LSTM is chosen as the model used for hour-by-hour real data, in addition, to cope with the output limitation of LSTM, i.e., to flexibly control the forecast duration, a fully connected layer (FC) is added after it. The working principle of LSTM\_FC can be expressed by the following equation:

$$\begin{aligned} i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\ f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\ \tilde{c}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \end{aligned} \quad (10)$$

Where,  $i_t$ ,  $f_t$ ,  $o_t$  are the input gate, forgetting gate, and output gate, respectively;  $W^*$ ,  $U^*$ ,  $b^*$  are the learnable parameters, where  $^* \in \{i, f, o, c\}$ ;  $h_t$ ,  $c_t$  are the hidden states and memory units at time t;  $h_{t-1}$ ,  $c_{t-1}$  are the hidden states and memory units at the previous moment; are the candidate states obtained by the nonlinear function;  $x_t$  is the training set input at time t;  $\sigma(\cdot)$  is the sigma function. candidate states obtained;  $x_t$  is the training set input at time t;  $\sigma(\cdot)$  is the sigma function. The last  $h_t$  (the most recent moment) of the LSTM output is chosen as the LSTM fitted value, which is input to the fully connected layer (FC) for linear transformation to obtain the final forecast value of LSTM\_FC.

Similar to the FNN, the absolute value loss function is chosen to describe the model empirical risk, and the learning rate is chosen to vary with pollutant. The learning rate is chosen to vary according to the pollutant; the epoches setting varies according to the pollutant; the optimizer is chosen to be either Adam or RMSprop; and the L2 regularization and Dropout to prevent overfitting; the pytorch framework was chosen as the training platform.

### 2.3.3. Combined model

Considering that the two single models FNN and LSTM chosen in this paper have strong learning ability, if we choose the higher complexity of the regression model, it may exacerbate the overfitting of the combined model. Therefore, Lasso regression [8] is used for model combination.

As a relatively low complexity regression model, Lasso reduces the variance by shrinking the coefficients and lowering the weights of some coefficients to zero. Lasso is a relatively low-complexity regression model, which reduces the variance by shrinking the coefficients and reducing the weights of some coefficients to zero. The empirical risks of the Lasso model are as follows:

$$J(w) = \frac{1}{n} \|Z - W^T X\|^2 + \lambda \|W\|_1 \quad (11)$$

Where,  $\lambda$  is the hyperparameter and L1 norm is the sum of the absolute values of the elements in the vector.

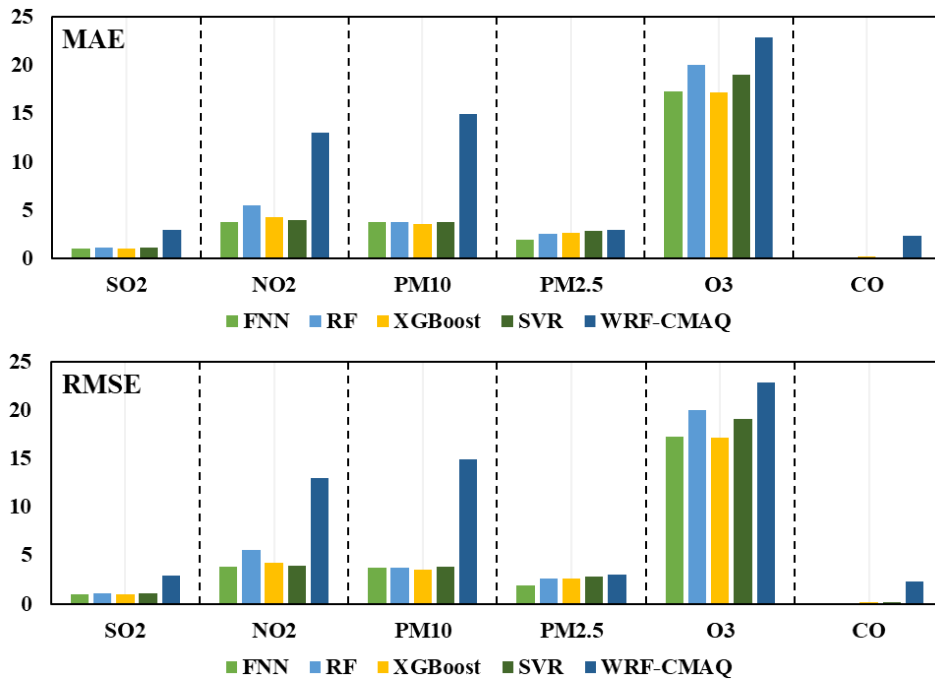
## 3. RESULT

### 3.1. Comparison of Base Models

The hour-by-hour simulation data of monitoring point A is used as input data, and it is noted that the running time of the data on the same day corresponds to the simulation value of the same day, tomorrow and the day after tomorrow. Note that the day of the operation time in the data corresponds to three consecutive days of simulated values for the current day, the next day and the day after tomorrow, so this paper uniformly uses the value of the day of the simulation as the input value.

Therefore, this paper uniformly uses the values of the day of simulation as the input values to train the above four models and predict the concentration of the corresponding pollutants.

Figure 2 shows the results of the four commonly used quadratic forecasting models (FNN, RF, SVR, XGBoost) and the corresponding evaluation indexes of WRF-CMAQ.



**Figure 2.** Results of FNN, RF, SVR, XGBoost and WRF-CMAQ evaluation indexes for six conventional pollutants

Figure 2 shows the results of the four commonly used secondary prediction models (FNN, RF, SVR, XGBoost) and the corresponding evaluation indexes of WRF-CMAQ. Among them, due to the small variation range of SO<sub>2</sub> and CO and the high dispersion of the numerical values, a small error will cause the R<sup>2</sup> to decrease. and CO, the R<sup>2</sup> of SO<sub>2</sub> and CO are not representative for evaluation because the variation range of SO<sub>2</sub> and CO is not large and the numerical values are highly dispersed, so a small error will lead to a drastic change of R<sup>2</sup>. Therefore, the R<sup>2</sup> of SO<sub>2</sub> and CO cannot be evaluated representatively. Therefore, the R<sup>2</sup> values of SO<sub>2</sub> and CO are not representative for evaluation.

As can be seen from the figure, (i) the prediction effects of the four secondary forecasting models are improved relative to WRF-CMAQ; (ii) the optimal terms of the four secondary forecasting models are not exactly the same under different evaluation indexes. The optimal terms of the four quadratic prediction models are not the same under different evaluation indexes. Among them, the MAE, RMSE, and RMSE of SO<sub>2</sub>, NO<sub>2</sub>, PM<sub>2.5</sub>, and CO are all in the range of FNN. and RMSE for SO<sub>2</sub>, NO<sub>2</sub>, PM<sub>2.5</sub>, CO, the MAE and RMSE of O<sub>3</sub> are all in favor of FNN, the RMSE of O<sub>3</sub> is the best for FNN, and the MAE is better for XGBoost, and the MAE and RMSE of PM<sub>10</sub> are the best.

For PM<sub>10</sub>, the best values of MAE and RMSE appeared in SVR and XGBoost respectively, but there was not much difference in the prediction effect among the four models. Overall, the FNN is more effective in the prediction of PM<sub>10</sub> than XGBoost.

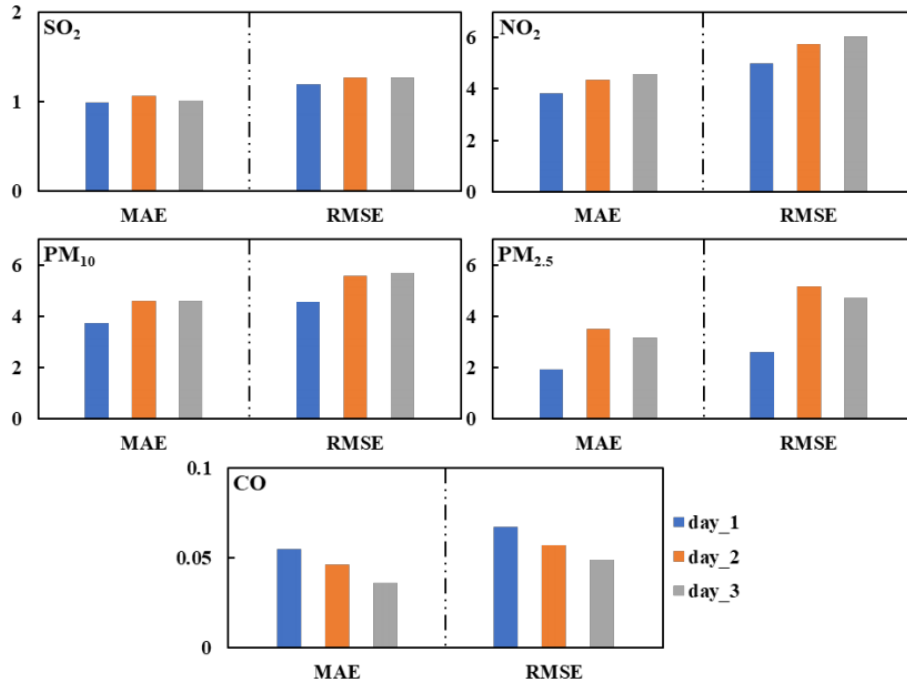
Generally speaking, FNN has the best performance among the evaluation indexes; ③ The MAE and RMSE of O<sub>3</sub> are on the high side.

It is known that O<sub>3</sub> is a secondary pollutant, which is generated through a series of chemical and photochemical reactions in the atmosphere, and the process is influenced by primary pollutants and meteorology. This process is affected by primary pollutants and meteorology, which makes the prediction of O<sub>3</sub> concentration changes more difficult, so the team will analyze and model O<sub>3</sub> separately later.

Therefore, the team will analyze and model O<sub>3</sub> separately later. For the other five pollutants except O<sub>3</sub>, the team chose FNN as the secondary prediction model.

### 3.2. Quadratic Forecasting Model

The forecasts were made for three days from 2021.7.13 to 2021.7.15, and the data of one forecast did not provide the data of 2021.7.14 simulation and 2021.7.15 simulation. 2021.7.14 simulated 2021.7.14 and 2021.7.15 simulated 2021.7.15, so this paper adopts 2021.7.13 simulated 2021.7.14 and 2021.7.15 data as a substitute. 2021.7.14 and 2021.7.15 simulated data are used as substitutes. The results of the three-day evaluation indicators are shown in Figure 3.



**Figure 3.** Results of three-day evaluation of indicators for five conventional pollutants FNN except O<sub>3</sub>

From the figure, it can be seen that the effect of uncertainty accumulation, for most of the pollutants modeled the second and third day prediction effect has been the prediction effect for the second and third days of the model decreased for most pollutants. In order to maximize the prediction effect of the model, it has been pointed out in the literature [15] that the prediction accuracy can be further improved by combining the results of different single prediction models and weighting them appropriately to produce a combined prediction model. Based on this, the team used the hourly measured pollutant concentration data for modeling and considered combining multiple models to predict the pollutant concentration. However, as shown in Figures 3 and 4, the MAE and RMSE of FNN for SO<sub>2</sub> and CO are relatively low for all three days, i.e., the model prediction results are inaccurate.

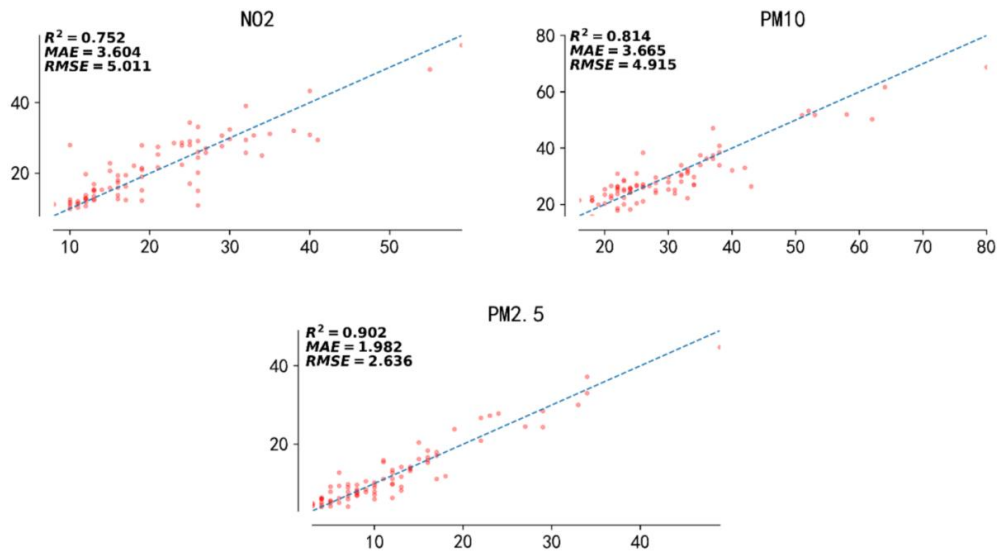
for SO<sub>2</sub> and CO, the values of MAE and RMSE for FNN were relatively low for all three days, i.e., the model prediction results had small errors. In order to streamline the model and avoid the waste of computational resources, the team used FNN as the final quadratic prediction model for SO<sub>2</sub> and CO, and further modified the model for NO<sub>2</sub>, PM<sub>10</sub> and PM<sub>2.5</sub>. are further corrected.

### 3.3. LSTM-FC Model for Primary Forecast Modeling

The LSTM\_FC input data is selected from the 24-hourly measurements before 7:00 p.m. of the day (hourly measurements from 8:00 p.m. of the previous day to 7:00 p.m. of the day of the forecast). In addition, it should be noted that since the model inputs hourly measured pollutant data, the features used in the feature selection process are 10 features of the five conventional pollutants other than the target pollutant in the hourly measured data and the meteorological indicators in the measured values.

The training and test set data were input into the LSTM\_FC model to complete the model training and predict the corresponding concentrations, of which the LSTM\_FC evaluation results are shown. Similar to the case of FNN, LSTM\_FC uses the historical 24h data to predict the concentration values in the next three days, and the prediction effect will be greatly reduced in the last two days, so this paper uses the prediction value of LSTM\_FC on the first day to correct the prediction value of FNN in the next three days. Therefore, the first day of LSTM\_FC is used to correct the FNN predictions for the next three days. The prediction results are shown in the figure below, which show that the

overall prediction results of each pollutant are better, and therefore it can be used to correct the model constructed from the primary forecast data.

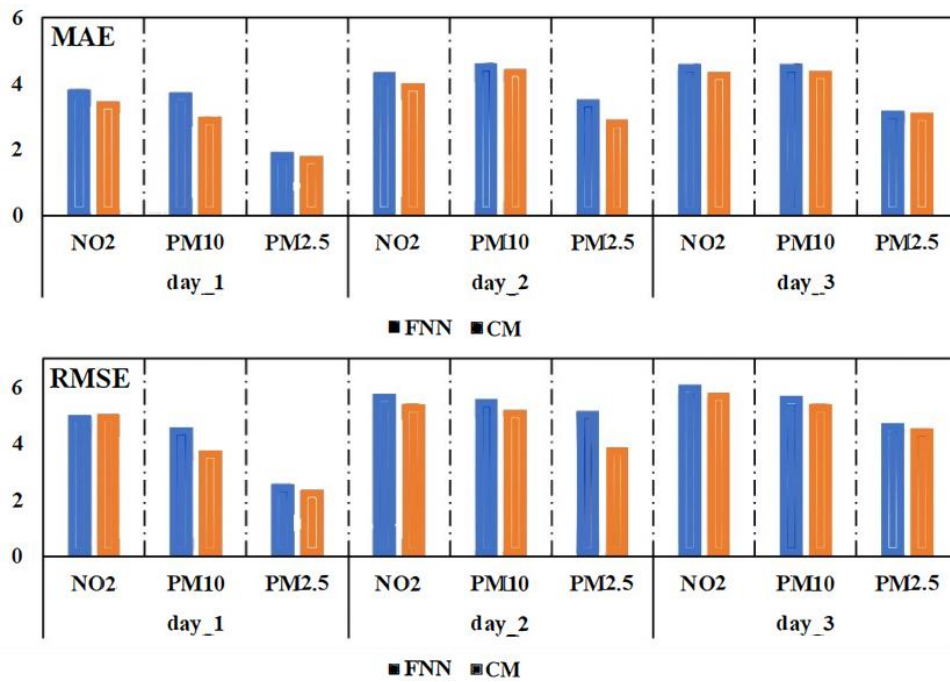


**Figure 4.** Evaluation metrics results for NO2, PM, and PMLSTM FC (taking the first day of the forecast as an example)

### 3.4. Evaluation of Combinatorial Models

After analyzing and testing, this paper finally uses the combined Lasso regression model of FNN and LSTM\_FC for NO2, PM10, and PM2.5 using FNN and LSTM\_FC. The two model predictions are combined by Lasso regression.

The evaluation results of this combined model prediction are shown in the following figure



**Figure 5.** Combined modeling of NO, PM10 and PM2.5 and results of FNN three-day evaluation metrics

From the figure, it can be seen that for MAE and RMSE, the combined model is better than FNN, so it is used as the final quadratic prediction model for NO2, PM10, and PM2.5 in this paper.

## 4. CONCLUSION

This study proposes a deep learning-based secondary modeling framework to improve air quality prediction accuracy, addressing the limitations of traditional WRF-CMAQ models through innovations like pseudo-independent thermal coding for wind direction to eliminate cyclic bias, CEEMDAN adaptive noise cancellation, and an LSTM-FC fusion model via random forest feature selection and Lasso regression. Experimental results show significant improvements: PM<sub>2.5</sub> prediction achieves a mean absolute error (MAE) of 1.76  $\mu\text{g}/\text{m}^3$  (8.3% lower than single models), O<sub>3</sub> MAE is reduced by 21% to 13.53  $\mu\text{g}/\text{m}^3$ , top pollutant identification accuracy reaches 92.7%, and AQI prediction relative error is 6.2%. By integrating LSTM's time-series analysis and FNN's nonlinear feature extraction, the framework balances model complexity and generalization, enhancing robustness to spatiotemporal heterogeneity and interpretability. The proposed method provides a high-precision, reliable solution for regional air quality forecasting, supporting effective pollution source identification and joint prevention strategies, with potential for future integration of more data sources and real-time optimization.

## ACKNOWLEDGMENTS

This work is supported by the school level scientific research project of Anhui University of Finance and Economics "Weighted combination of SVR and CNN-LSTM optimized based on HO hippo algorithm in air pollution prediction" (Grant No: XSKY24051ZD).

## REFERENCES

- [1] Cohen, A. J., et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *The Lancet* 389(10082), 1907-1918 (2017).
- [2] Zhang Y, Zhang X, Wang L, et al. Application of WRF/Chem over East Asia: Part I. Model evaluation and intercomparison with MM5/CMAQ [J]. *Atmospheric environment*, 2016, 124: 285-300..
- [3] Zhang Z, Zeng Y, Yan K. A hybrid deep learning technology for PM 2.5 air quality forecasting [J]. *Environmental Science and Pollution Research*, 2021, 28: 39409-39422.
- [4] Zhan, Y., et al. Deep learning for ozone prediction: A hybrid approach combining chemical transport models and surface observations. *Environmental Science & Technology* 55(8), 4583-4592 (2021).
- [5] Wang Z, Chen H, Zhu J, et al. Multi-scale deep learning and optimal combination ensemble approach for AQI forecasting using big data with meteorological conditions [J]. *Journal of Intelligent & Fuzzy Systems*, 2021, 40(3): 5483-5500.
- [6] Sillman S. The relation between ozone, NO<sub>x</sub> and hydrocarbons in urban and polluted rural environments [J]. *Atmospheric environment*, 1999, 33(12): 1821-1845.
- [7] Yan X, Zang Z, Jiang Y, et al. A Spatial-Temporal Interpretable Deep Learning Model for improving interpretability and predictive accuracy of satellite-based PM<sub>2.5</sub> [J]. *Environmental Pollution*, 2021, 273: 116459.
- [8] TIBSHIRANI R. Regression shrinkage and selection via the lasso\_ a retrospective [J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2011.