Atmospheric Pollution Research 14 (2023) 101765

Contents lists available at ScienceDirect

Atmospheric Pollution Research

journal homepage: www.elsevier.com/locate/apr

A hybrid prediction model of air quality for sparse station based on spatio-temporal feature extraction

Yue Hu, Xiaoxia Chen *, Hanzhong Xia

Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China

ARTICLE INFO

Keywords:
Air-quality prediction
Hybrid model
Sparse station
Spatio-temporal features
Sudden changes

ABSTRACT

Accurate prediction of air quality is helpful for effective prevention and control of air pollution. However, there are relatively few studies on the air-quality prediction for sparse station. Through the extraction of spatio-temporal features, this research aims to achieve the hybrid prediction of air quality for sparse station. The proposed method is comprised of three parts. The first part is the extraction of multi-scale temporal features. Specifically, the multi-scale time lags of meteorological factors are extracted to cope with sudden changes, and the key features from multi-source data are extracted to avoid the interference of redundant features. The second part is to extract spatial features of related stations based on spatial hierarchy division. The third part is the hybrid model prediction based on spatio-temporal feature groups. The following results are obtained. (1) For sudden-change prediction, the processing method proposed in this paper is more effective. (2) Compared with the baselines and models popular in dense station, the proposed model has superior performance in the air-quality prediction of sparse stations. (3) Compared with the common spatial-related station selection method, the proposed method is more suitable for sparse stations. (4) For multi-step prediction, the proposed model has significant advantages in long-term prediction.

1. Introduction

Air pollution consists of chemicals or particles in the atmosphere, which pose serious threats to health and the environment (Lim et al., 2020). Long-term exposure to air pollution increases the risk of stroke, heart disease, lung cancer, and other chronic lung diseases (Brauer et al., 2021; Li et al., 2019). Therefore, it is of great importance and necessity to prevent air pollution by predicting air-quality.

Air-quality prediction methods are mainly divided into two categories, namely classical physical diffusion models and statistical data-driven models. Classical physical diffusion models have been used for air-quality prediction, such as the Gaussian plume model (Kalhor and Bajoghli, 2017), the community multiscale air quality model (Yang et al., 2019), the geological chemistry model (Lee et al., 2017), and the weather research and forecast chemistry model (Liu et al., 2018). Its core is that mathematical model is constructed based on equations of atmospheric dynamics, atmospheric environmental chemistry, and historical air pollution and meteorological data to calculate the temporal and spatial distribution of pollutants and solve them by computer. However, defects and errors are inevitable in the process of collecting heterogeneous data. Even minor data errors in this model can also lead to relatively large differences in results. Furthermore, the model complexity and computational cost are relatively high.

Statistical data-driven models can be divided into two categories, namely linear statistical models and nonlinear statistical models. The method based on linear statistical model establishes the relationship between related features and air quality through linear model, such as the spatial interpolation method, the land use regression (Ma et al., 2019), the autoregressive integrated moving average model (Zhang et al., 2018), and the multivariable linear regression model (Tai et al., 2010). Since the linear statistical model cannot fit the nonlinear relationship between related features, the nonlinear statistical model is more widely used for air-quality prediction.

In contrast, nonlinear statistical models, such as the decision tree model, the support vector regression model (Liu et al., 2022), the regression tree model, and the artificial neural network model (Perez and Reyes, 2006), are more adaptable for air-quality prediction. However, there is still room for improvement in its prediction accuracy. Many researchers have improved these basic algorithms:

The first is the analysis and extraction of features. For example, empirical mode decomposition (EMD) (Fei, 2016; Guo et al., 2012), ensemble empirical mode decomposition (EEMD) (Xiang et al., 2018), and fast ensemble empirical mode decomposition (FEEMD) (Sun and Wang, 2018) first extract the features of air quality, and then build different predictive sub-model based on different sub-sequence features.

* Corresponding author.
E-mail address: chenxiaoxia@nbu.edu.cn (X. Chen).

https://doi.org/10.1016/j.apr.2023.101765
Received 8 January 2023; Received in revised form 15 April 2023; Accepted 16 April 2023
Available online 20 April 2023

Table 1
Longitude and latitude of each station.

Station	Latitude	Longitude
Gucheng	39.9116	116.1933
Wanshouxigong	39.8802	116.3677
Tiantan	39.8824	116.4647
Guanyuan	39.9327	116.3623
Dongsi	39.9304	116.4246
Nongzhanguan	39.9415	116.4647
Wanliu	39.9598	116.2982
Aotizhongxin	39.9837	116.3995
Shunyi	40.1503	116.6411
Changping	40.2229	116.2202
Dingling	40.2960	116.2234
Huairou	40.3193	116.6338

The second is parameter optimization. The use of different parameters will affect the performance of the model algorithm. For instance, the particle swarm optimization (PSO) (Gu et al., 2020) algorithm is used to optimize the parameters of neural network model.

On the one hand, the algorithm improvement does have a certain effect on the accuracy of the air-quality prediction. On the other hand, researchers are still exploring more methods. According to the finding, deep learning has a powerful ability in extracting big-data features. In the field of air-quality prediction, deep learning methods are used to explore temporal dependence, such as the long short-term memory (LSTM) networks (Seng et al., 2021), the temporal convolution network (TCN) (Samal et al., 2021), the long short-term memory neural network extended (LSTME) model (Li et al., 2017), the gate recurrent unit (GRU) model (Huang et al., 2021) and the bidirectional long short-term memory (Bi-LSTM) networks (Tong et al., 2019). Air-quality data are generally time series. These models have good fitting and generalization ability for time series. Usually, models such as the convolutional neural networks (CNN) (Yan et al., 2021) and the graph convolutional networks (GCN) (Ge et al., 2021) are used to explore spatial characteristics. In the face of more complex situations, only considering temporal dependence or spatial characteristics will lead to lower prediction accuracy. Therefore, more and more researchers use the combination of temporal dependence and spatial characteristics to predict air-quality, such as the combination of the CNN model and the LSTM model (Huang and Kuo, 2018; Wen et al., 2019), the combination of the CNN model and the GRU model (Faraji et al., 2022), and the combination of the GCN model and the LSTM model (Qi et al., 2019). According to the relevant results, the comprehensive utilization of spatio-temporal features can indeed improve the accuracy of air-quality prediction.

In addition, many studies have shown that the hybrid model performs well in air-quality prediction. D. Saravanan and Kumar combined artificial neural networks and auto-regressive moving average methods to evaluate the proportion of air pollution components (Saravanan and Kumar, 2022). Wu and Lin combined the least squares support vector machine and the LSTM neural network to predict the air-quality index (Wu and Lin, 2019). Wang et al. integrated the LSTM neural networks and the random forest for the forecasting and minimization of multiple air pollutants (Wang et al., 2022). Zhang et al. utilized the light gradient boosting machine (LightGBM) model, the gradient boosting decision tree model, and the extreme gradient boosting model to form a scheme for predicting $PM_{2.5}$ concentration (Zhang et al., 2020). Luo et al. facilitated the performance of air-quality prediction by combining the LightGBM, the gated-DNN, and the seq2seq (Luo et al., 2019). Appropriate sub-models constructed based on feature groups can give full play to the advantages of each sub-model, thereby achieving more accurate predictions.

For the above studies, air-quality prediction methods are explored from multiple perspectives, and the advantages of the models are utilized for the respective research objects. However, there are still following deficiencies. (1) The target stations usually choose those stations with dense surrounding stations (Chen and Lin, 2022), and lack air-quality prediction for sparse stations. (2) The selection of related stations of target stations is usually through spatial transformation (Yi et al., 2018), analysis of correlation between station sequences, and distance-based methods (Qin et al., 2019). However, these methods are not suitable for sparse stations to select related stations. (3) Many studies lack the analysis and treatment of sudden changes in sequences. The improvement of the prediction accuracy of sudden changes can enhance the overall prediction accuracy.

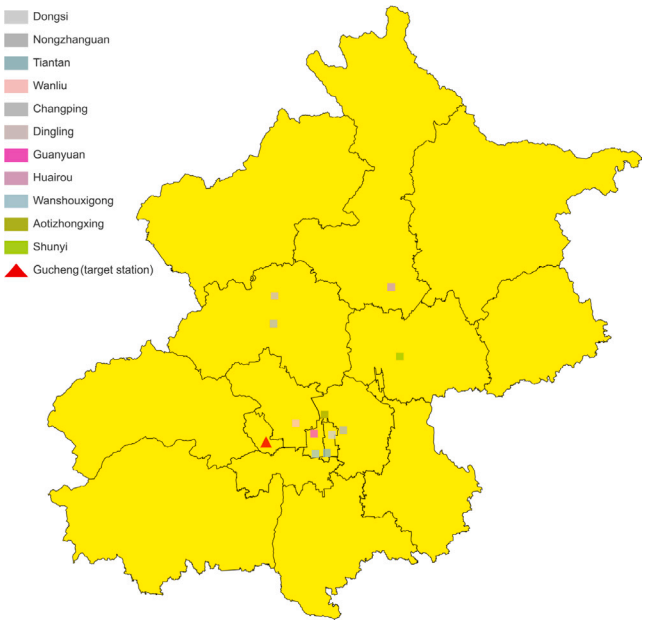


Fig. 1. 12 air-quality monitoring stations in Beijing.

In this paper, the air-quality prediction is carried out for the sparse station, including the analysis and processing of the sudden-change parts, aiming at improving the prediction accuracy of the air-quality prediction of sparse station. The main contributions of this paper are summarized as follows: (1) The method of extracting the multi-scale time lags of the meteorological factors to cope with sudden changes is proposed. (2) The selection method of the related stations of sparse station based on the spatial hierarchy division is obtained. (3) The hybrid model that can give full play to the advantages of each sub-model in dealing with different feature groups is created.

In this paper, a hybrid model based on spatio-temporal features extraction is introduced to solve the problem of air-quality prediction for the sparse station. The rest of the paper is structured as follows. Section 2 briefly describes the study area and materials. Section 3 elaborates the proposed methods. Section 4 analyzes and discusses the simulation results. Section 5 concludes this study. According to the simulation results, the model can effectively predict the air quality of sparse station.

2. Study area and materials

Beijing, the capital of China, has encountered some problems related to pollution control in the process of rapid economic development. The improvement of air quality is a gradual process that requires multiple efforts. In this paper, Beijing is chosen as the study area. Fig. 1 shows the selected 12 air-quality monitoring stations in Beijing, which are: Gucheng, Wanshouxigong, Tiantan, Guanyuan, Dongsi, Nongzhanguan, Wanliu, Aotizhongxin, Shunyi, Changping, Dingling, and Huairou. Table 1 shows their latitudes and longitudes. Compared with the dense station, the stations around the Gucheng station are sparse. Thus, the Gucheng station can be chosen as the target station.

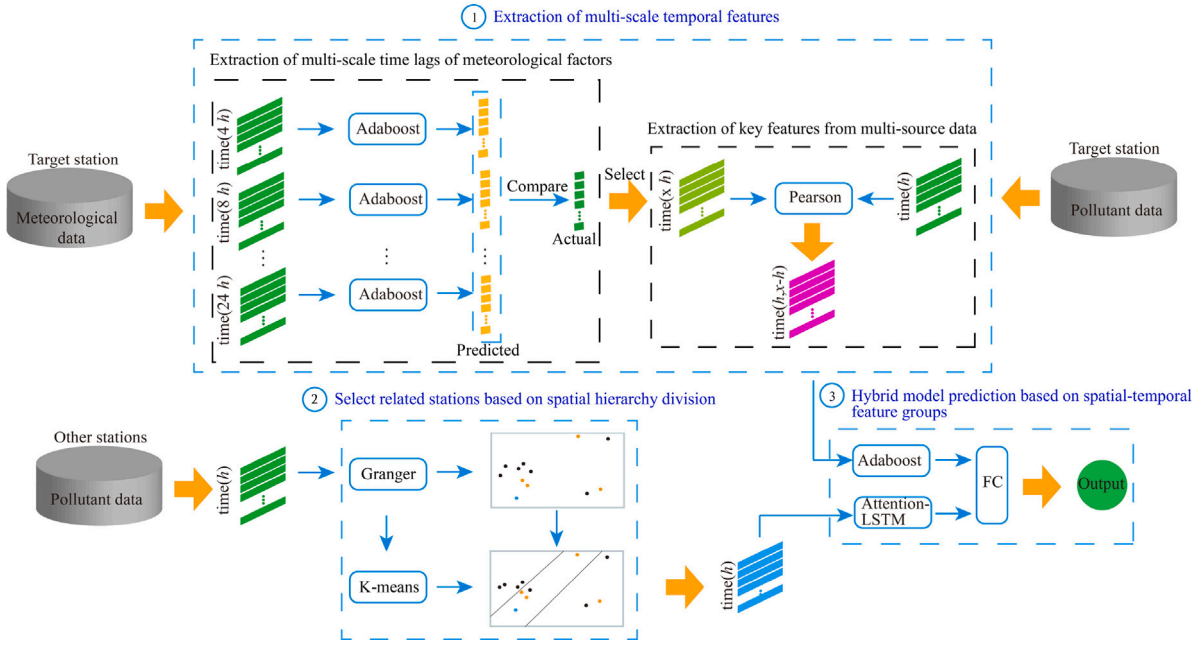


Fig. 2. Overall framework.

The datasets used in this study mainly include the following two categories: (1) Air-quality pollutants of 12 stations, including solid particulate matter ($PM_{2.5}$, PM_{10}) and gas pollutants (SO_2, NO_2, CO, O_3). (2) The meteorological conditions of the sparse station, including wind speed (Wspd), rainfall (Rain), temperature (Temp), dew point (Dewp), and pressure (Pre). These data are collected hourly at each monitoring station, containing 12,000 instances from March 1, 2015 to July 12, 2016 (https://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data).

3. Research methods

3.1. Model configurations

This paper constructed the overall framework shown in Fig. 2, which mainly consists of the following three parts.

The first part is the extraction of multi-scale temporal features, including the extraction of multi-scale time lags of meteorological factors and the extraction of key features from multi-source data. The influencing factors of pollutant concentration to be predicted include historical pollution data and meteorological condition. Among them, historical pollution data affect the pollutant concentration to be predicted in a relatively short time scale, while meteorological conditions affect the pollutant concentration to be predicted in a relatively long time scale. By observing the original sequence, it can be found that there is a sudden change in the sequence. As shown in Fig. 3, the value of $PM_{2.5}$ at the Gucheng station changed significantly between 2:00 am on May 30 and 2:00 am on May 31, surging from 75 to 226. Then, the value of $PM_{2.5}$ fell back to 76 within 12 h. According to practical experience, such sudden changes are usually caused by extreme meteorological conditions, such as rainstorms or typhoons. If it can find the time when the extreme meteorological condition has the most significant effect on the target station, that is, the optimal time lag, the prediction accuracy of the sudden changes can be improved. The historical data of the sparse station contain various pollutant data, as well as meteorological data with optimal time lag. Since too many data features are not conducive to exploring temporal dependence, it is necessary to avoid the interference of redundant features when exploring temporal dependence. In order to select key features, Pearson correlation analysis is used to analyze the correlation of multiple features. Key features are selected as temporal feature group.

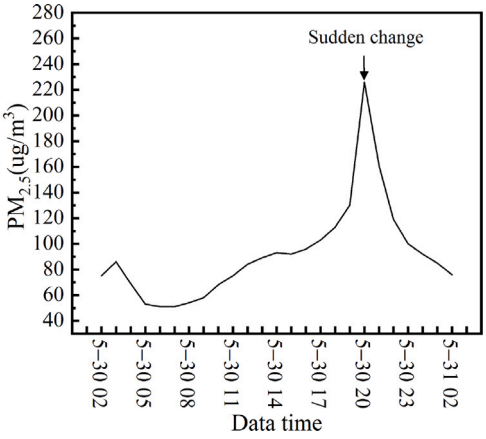


Fig. 3. $PM_{2.5}$ changes over a day at the Gucheng station.

The second part is to select related stations of sparse station based on spatial hierarchy division. Air pollutants are dispersed among different stations. The correlation between stations is complicated due to factors such as geographical location and wind direction. Since other stations are only located on one side of the target station and the target station is relatively far away, the spatial data available for the target station are sparse. To take full advantage of spatial-sparse data, the spatial data are divided into two hierarchies, station and area (Wang and Song, 2018). From the station perspective, this paper can find stations that have an influence on the target station. From the area perspective, clusters of stations with similar influence characteristics on target stations can represent similar areas. By combining these two hierarchies, this paper can find the related stations of the target station more accurately. The features of related stations are selected as the spatial feature group.

The third part is the hybrid model prediction based on the spatio-temporal feature groups. The spatio-temporal feature groups obtained in the first and second parts are used to design the hybrid prediction model. The model consists of two sub-models, Adaboost and

Attention-LSTM. These two sub-models take advantage of their respective strengths to process the corresponding feature groups, and realize hybrid prediction by concatenating the fully connected layer.

3.2. Extraction of multi-scale temporal features

This part includes two modules, which are the extraction of multi-scale time lags of meteorological factors and the extraction of key features from multi-source data.

3.2.1. Extraction of multi-scale time lags of meteorological factors

The air-quality prediction of sparse station also involves sudden changes caused by extreme meteorological conditions. By extracting the multi-scale time lags, the optimal time lag of the extreme meteorological factors is determined. First, the time-lag interval is set to 4 h, with a maximum lag of 24 h. A total of six time-lags are divided. When the time-lag interval is too long, the model may not be able to capture the subtle differences. When the time-lag interval is too short, the lag cannot be reflected. In a relatively short period of time, there may be no difference in sudden-change prediction. Second, the Adaboost model uses six groups of meteorological factors with different time lags to predict pollutant concentration. The group with the lowest error between the actual and predicted values is the meteorological factors with the optimal time lag.

3.2.2. Extraction of key features from multi-source data

The meteorological features with the optimal time lag extracted from 3.2.1 and the historical air pollution data of the sparse station show too many features. There have been many studies proving that too many features may negatively affect prediction accuracy. Therefore, it is necessary to analyze the key features for the prediction of the concentration of pollutants. Pearson correlation analysis is used to measure the degree of linear correlation between sequences, with values ranging from −1 to +1. The larger the absolute value, the more significant the correlation. A value of 0 indicates no correlation. This paper defines two time series S_t and Z_t (t=1, 2, …, n). The Pearson correlation coefficient of the two sequences is calculated as follows: (Eq. (1)):

$$r = \frac{\sum_{t=1}^{n}(S_t - \overline{S})(Z_t - \overline{Z})}{\sqrt{\sum_{t=1}^{n}(S_t - \overline{S})^2}\sqrt{\sum_{t=1}^{n}(Z_t - \overline{Z})^2}} \tag{1}$$

where n is the number of samples per sequence, S_t and Z_t correspond to the tth sample; $\overline{S}$ is the average value of the S sequence, and $\overline{Z}$ is the average value of the Z sequence. Those features with high correlation are selected to be the temporal feature group.

3.3. Selection method of related stations of sparse station based on spatial hierarchy division

This paper explores the spatial correlation between the target station and the other stations from the station hierarchy and the area hierarchy, respectively. In the station hierarchy, the stations that affect the sparse station are selected. There are two main methods for the selection of spatial influence stations. One method is the nearest neighbor method, which selects stations based on distance. However, this method is subject to the interference of wind direction and weather. Furthermore, it does not guarantee a positive correlation between distance and influence relationship. Another method is the correlation-analysis method. This method can only determine whether the two sequences are similar at each time point, but cannot ensure the influence of the sequences of other stations on the sequences of the target station. By analyzing the causal relationship between the two sequences, the sequence of stations that have a direct influence on the sequence of the sparse station can be found. Granger causality test can be used to determine whether one sequence is a delayed expression of another. In this study, the Granger causality test is used to analyze the causal relationship between the two station sequences.

For two time series S_t and Z_t ($t = 1, 2, \ldots, n$), the causality between them is tested by Eqs. (2) and (3). In Eq. (2), Z_t, Z_{t-1}, and S_{t-1} are used for regression, with only one lag. In fact, a longer lag period could be set. For example, Z_t, S_{t-1}, and S_{t-2} are used for regression, with two lags. After the regression, this study needs to check whether the coefficient a_1 in front of S_{t-1} is zero. If a_1 is not equal to zero, it means that the change of S will cause the change of Z, that is, there is a causal relationship between S and Z. Similarly, Eq. (3) tests whether the change of Z could affect the change of S.

$$Z_t = a_0 + a_1 S_{t-1} + a_2 Z_{t-1} + u_{1t} \tag{2}$$

$$S_t = d_0 + d_1 Z_{t-1} + d_2 S_{t-1} + u_{2t} \tag{3}$$

where a_i and d_i ($i = 0, 1, \ldots, n$) are the coefficients in front of the corresponding sequences, and u_{1t} and u_{2t} are assumed to be uncorrelated white noise.

The Granger result of other station sequence to the target station sequence is denoted by $prob(S)$, which refers to the probability of rejecting the null hypothesis. Its value can be directly obtained through the Granger causality test performed by Stata. Similarly, the Granger result of the target station sequence to other station sequence is denoted by $prob(Z)$. If the value of $prob(S)$ is less than 0.05, it indicates that early changes of the station sequence can lead to changes in the sequence at the sparse station. Similarly, if $prob(Z)$ is less than 0.05, it indicates that early changes of the sparse station can lead to changes in the sequence at that station. Accordingly, this study chooses stations with $prob(S)$ less than 0.05 as the stations with influence on sparse station.

In the area hierarchy, this study uses the Granger results $prob(S)$ and $prob(Z)$ as coordinate points of the k-means clustering algorithm to classify stations with similar characteristics. Based on the similar characteristics of stations and the non-crossing principle of areas, stations with similar characteristics are classified into the same area to the greatest extent. Other stations in the same area as the sparse station are considered the related stations of the sparse station. The features of these related stations are selected as the spatial feature group.

3.4. Model components

To improve the prediction accuracy, two different models are adopted for the two feature groups obtained in Sections 3.2 and 3.3. The two models are connected with fully connected layer to achieve hybrid prediction. Two feature groups are used to pretrain multiple prediction models, respectively. By comparing the prediction errors, this study selects Adaboost and Attention-LSTM for the temporal and spatial feature groups, respectively. These model components are described in detail below.

3.4.1. LSTM

In order to solve the problem of gradient explosion or gradient disappearance caused by successive multiplications in the process of back-propagation of recurrent neural networks (RNNs), the LSTM is produced. LSTM has a memory cell with selective memory function, which can select necessary information and filter out noise. Fig. 4 shows the model structure of LSTM.

It can be seen that in addition to the hidden state h_t, the cell state C_t, propagates forward at each sequence index time t. In addition to cell states, LSTM also has a gate structure. The gate of LSTM at the index position t of each sequence generally includes the input gate, forget gate, and output gate. The forget gate controls forgetting. In more detail, it controls whether to forget the hidden unit state of the previous layer with a certain probability. The output f_t of the forget gate is obtained from the hidden state h_{t-1} of the previous sequence

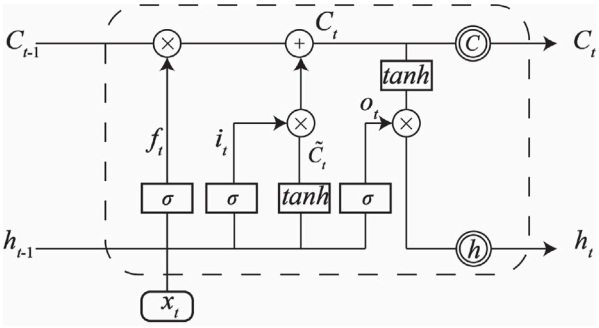


Fig. 4. LSTM structure.

and the data x_t of the current sequence through an activation function (usually sigmoid). Its mathematical expression is shown in Eq. (4):

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f) \tag{4}$$

where W_f, U_f, and b_f are the coefficients and biases of the linear relationship, and σ is the sigmoid activation function.

The input gate is responsible for processing the input of the current sequence position in two parts. The first part uses the sigmoid activation function, and the output is i_t. The second part uses the tanh activation function, and the output is $\widetilde{C}_t$. The results of the two are multiplied to update the cell state. The mathematical expressions are Eqs. (5) and (6):

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \tag{5}$$

$$\widetilde{C}_t = \tanh(W_{\widetilde{C}} h_{t-1} + U_{\widetilde{C}} x_t + b_{\widetilde{C}}) \tag{6}$$

where W_i, U_i, b_i, $W_{\widetilde{C}}$, $U_{\widetilde{C}}$, and $b_{\widetilde{C}}$ are the coefficients and biases of the linear relationship.

The results of both the input gate and the forget gate act on the cell state C_t, which consists of two parts. The first part is the product of the forget gate output f_t and C_{t-1}. The second part is the product of input gate i_t and $\widetilde{C}_t$, as shown in Eq. (7):

$$C_t = C_{t-1} \odot f_t + i_t \odot \widetilde{C}_t \tag{7}$$

where $\odot$ is the Hadamard product.

The hidden state h_t update consists of two parts. The first part is the output o_t of the output gate, which is obtained by the hidden state h_{t-1} of the previous sequence, the current sequence data x_t, and the activation function σ. The second part comprises the hidden state C_t and the tanh activation function. As shown in Eqs. (8) and (9):

$$o_t = \sigma(W_o h_{t-1}) + U_o x_t + b_o \tag{8}$$

$$h_t = o_t \odot \tanh(C_t) \tag{9}$$

where W_o, U_o, and b_o are the coefficients and biases of the linear relationship.

Although the LSTM neural network has a strong ability to extract temporal correlations, it cannot judge the importance of each feature when faced with multiple features. To solve this problem, the Attention mechanism is added to the LSTM model.

3.4.2. Attention-LSTM

Attention mechanism is a technique that enables the model to focus on important information and then learn and absorb it thoroughly. It is a technique that can be used for any sequence model, which is not a complete model. In the encoder–decoder structure, the encoder encodes all the input sequences into a unified feature vector v, which is then decoded. Regardless of the length of the input sequence, it is encoded into a fixed-length vector representation. When the sequence

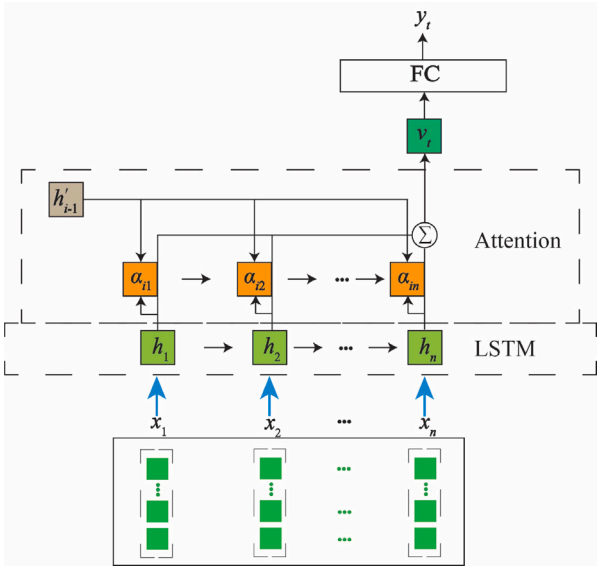


Fig. 5. Attention-LSTM structure.

is too long, v may not contain all the information, which will reduce the prediction accuracy. The addition of the Attention mechanism can solve this problem well. Fig. 5 shows the LSTM model with the Attention mechanism added. Among them, $x_1, x_2, \ldots, x_n$ are the n input vectors before the time t to be predicted. The hidden layer of the LSTM processes the input vector to obtain the hidden state h at each position. The predicted result y at the next moment is calculated by the state h'_{i-1} of the decoder at the previous moment and the result of the decoder (i.e. v_t). The Attention coefficient α_{ij} is used to measure the correlation of the hidden state h_j of position j in the encoder and the state h'_i of the current position i in the decoder. α_{ij} is related to the hidden state h'_{i-1} of the decoder at the previous moment and the hidden state h_j of the encoder at the jth moment, the calculation equations are as Eqs. (10) and (11).

$$e_{ij} = \psi(h'_{i-1}, h_j) \tag{10}$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{m=1}^{M} \exp(e_{im})} \tag{11}$$

where M is the number of hidden vectors in the encoder, and e_{ij} is the correlation between h_j and h'_i. The larger the value of e_{ij}, the stronger the correlation. α_{ij} is the weight distribution obtained by normalizing e_{ij}, and the sum of the probability values is 1.

The Attention coefficients obtained at different moments are assigned to the hidden vector h_j at different moments. v at each moment in the decoder can be obtained by summation (Eq. (12)).

$$v_t = \sum_{m=1}^{M} \alpha_{im} h_m \tag{12}$$

Next, update the current hidden state with the previous hidden state h'_{i-1}, the previous output y_{t-1}, and the vector v_t of the current position (Eq. (13)):

$$h'_i = f(h'_{i-1}, y_{t-1}, v_t) \tag{13}$$

Finally, calculate the current output y_t according to the current hidden state h'_i, the previous output y_{t-1}, and the vector v_t of the current position (Eq. (14)):

$$y_t = g(h'_i, y_{t-1}, v_t) \tag{14}$$

where $f(\cdot)$ and $g(\cdot)$ are corresponding calculation functions. Attention calculates a vector v for each output position so that the output at each

position focuses on the most relevant part. In contrast, it does a better job than the conventional encoder–decoder structure.

The combination of the Attention mechanism and the LSTM model is used for the air-quality prediction of sparse station. The Attention-LSTM model includes input vector, LSTM hidden layer, Attention layer, fully connected layer, and output value. The vectors processed by the input layer and the LSTM hidden layer are used as the input of the Attention layer. To obtain the output value, the Attention layer is responsible for learning a set of Attention coefficients and connecting the fully connected layer.

The influence degree of different features on the prediction moment changes dynamically over time. In other words, the stations and the features that have the most significant influence on the sparse station will change over time. The LSTM model only encodes the input sequence into a fixed-length vector. Since it is impossible to judge which feature significantly affects the current prediction moment, the information utilization is reduced. Attention can judge the importance of each input moment and highlight key factors. Therefore, the model has the effect of improving the prediction accuracy.

3.4.3. Adaboost

Adaboost is the most typical representative of reinforcement ensemble learning. The main logic of this algorithm is to improve the latter on the basis of the former. The Adaboost model first trains a weak learning model and then evaluates the model. In the model, the problem of doing right will decrease the attention to it, while the problem of doing wrong will increase the attention to it. Subsequent new models focus more on overcoming difficulties that previous models are unable to overcome. Finally, all models are integrated to form an extensive framework. Its overall performance improves due to the extended framework having models for both simple and complex problems.

The following are the steps of model training.

Step 1: Train each weak learning model and calculate the sample difficulty.

- Initialize the difficulty of each sample as (Eq. (15)):

$$w_i = \frac{1}{n} \tag{15}$$

where n represents the number of samples.

- Use the training results of the current weak learning model to update the difficulties of all samples.

$$w_i^{new} = \begin{cases} \frac{1}{2(1-\varepsilon)} w_i^{old} \\ \frac{1}{2\varepsilon} w_i^{old} \end{cases} \tag{16}$$

where ε represents the error rate produced by the current learning model. If the current sample is correctly predicted, the first equation of Eq. (16) is adopted to reduce the difficulty of the sample. If the current sample is mispredicted, the second equation of Eq. (16) is used to increase the difficulty of the sample.

- Train the next weak learning model based on the difficulty of the current sample.

Step 2: Learn the weight of each weak learning model.

- Calculate the weight of the current weak learning model according to the current error rate (Eq. (17)).

$$q_k = \frac{1}{2} \log\left(\frac{1-\varepsilon_k}{\varepsilon_k}\right) \tag{17}$$

where ε_k ($k = 1, 2, \ldots, n$) represents the error rate of the kth learning model. These weights are used to integrate the outputs of all weak learning models.

- The final output $\overline{Y}$ can be calculated by Eq. (18):

$$\overline{Y} = q_1 y_1 + q_2 y_2 + \cdots + q_k y_k \tag{18}$$

where q_k and y_k are the weight and output of each weak learning model.

Different from the random forest where each decision tree is independent, the Adaboost model forms a cascade structure that is trained sequentially. The training of the latter is always based on the reanalysis of the previous-model output. Therefore, the Adaboost model has a good fitting ability for the feature group in this paper.

3.5. Algorithms

The specific process of the method is shown in Algorithm 1. Firstly, the multi-scale time lags of meteorological factors are extracted to cope with sudden changes. Secondly, the key features from the multi-source data are extracted. Thirdly, the related stations of the sparse station are selected based on the spatial hierarchy division. Finally, the two models corresponding to the spatio-temporal feature groups are connected with fully connected layers to achieve hybrid prediction.

Algorithm 1 Feature extraction and model prediction

Input: Meteorological conditions of the target station, $Tar.M$; Air pollutants of the target station, $Tar.P$; Pollutants to be predicted from other stations, $Oth.P$;
Output: Hybrid model on the current features;

```
Divide Tar.M into groups with multi-scale time lags, Tar.M_i, i=1, 2, ..., n;
for each Tar.M_i do
    Train the Adaboost;
    Abtain predicted value y_j, j=1, 2, ..., n;
    Compare y_j with the actual value y_t;
    Abtain Tar.M_b with the optional time lag;
end for
Extract key features by Pearson correlation analysis on Tar.P and Tar.M_b;
Abtain the features of related stations Rel.P by spatial hierarchy division;
Train hybrid model with Rel.P and Tar.M_b to get final prediction ŷ;
```

4. Results and discussion

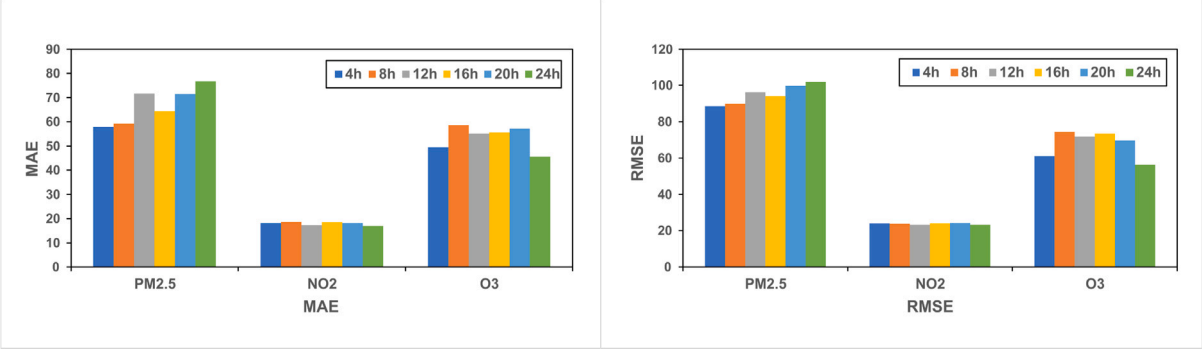
This section presents the simulation settings and gives the analysis and discussion of the results.

4.1. Simulation settings

The LSTM was trained by Adam optimizer with a learning rate lr = 0.005, the batch size is 16 and the epoch is 150. The needed settings of the PSO-LSTM and the LSTME were the same as the LSTM. For the CNN-LSTM, the batch size is 24, the epoch is 100 and the learning rate is 0.005. For the proposed model, the learning rate is 0.1 and the epoch is 100. The above models were implemented using Matlab. For the TCN, the batch size is 64, the epoch is 200, the filter number is 64 and the kernel size is 4. The TCN was implemented using PyCharm. All simulations were conducted on Windows 10 and RTX 3080 GPU.

4.1.1. Data pre-processing

Considering that the raw data from 12 stations inevitably have missing values, it is necessary to perform data preprocessing before simulation. In the following two cases, the similar value imputation method and the mean value imputation method are used to fill the missing values. (1) When some variables of a single station are missing at a certain time, the average value of the same variable of other stations at the same time is used to fill the missing values. (2) When a variable is missing at a certain time for all stations, the average of the data before and after the blank part of the time for each station is used.



(a) (b)

Fig. 6. The combined bar charts of MAE and RMSE values for $PM_{2.5}$, NO_2, and O_3 at different time lags.

Table 2
Results of multiple simulations with NO_2.

	4h		8h		12h		16h		20h		24h	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	18.2090	24.1961	18.9976	24.1406	17.1815	23.1807	18.5307	24.1156	18.2985	24.3298	17.0221	23.4157
2	18.1459	23.9401	18.6334	23.9134	17.3444	23.3638	18.5920	24.1475	18.2339	24.3242	17.1354	23.4393
3	18.1191	24.0829	18.6809	23.9392	17.2709	23.2534	18.3625	24.0441	18.1411	24.3046	16.9328	23.3179
4	18.0131	23.8758	18.6797	23.8549	17.3345	23.3275	18.3266	24.0011	18.2158	24.3484	16.8767	23.1356
5	18.1239	23.9932	18.7186	23.9824	17.4158	23.4095	18.2469	23.9585	18.1463	24.3384	16.8867	23.1974
6	18.0781	23.8880	18.7966	24.0260	17.2716	23.2493	18.4234	24.1479	18.1387	24.3249	16.8690	23.1478
7	18.1332	23.9982	18.6755	23.9581	17.3904	23.2357	18.5726	24.1611	18.1214	24.2900	16.8313	23.1196
8	18.0679	24.0276	18.7759	23.9949	17.2054	23.1800	18.3256	24.0267	18.1388	24.2725	16.9373	23.1805
9	18.2465	24.1196	18.6511	23.8474	17.2929	23.3192	18.5035	24.1918	18.2317	24.3524	17.0218	23.3525
10	18.0699	23.8751	18.7251	23.9435	17.1863	23.1997	18.8252	24.3467	18.1566	24.3902	16.9392	23.2111
Median	18.1215	23.9957	18.6998	23.9508	17.2823	23.2514	18.4635	24.1316	18.1515	24.3274	16.9351	23.2043
Mean	18.1207	23.9997	18.7334	23.9600	17.2894	23.2719	18.4709	24.1141	18.1823	24.3275	16.9452	23.2517

4.1.2. Evaluation metrics

To evaluate the validity of the model, the mean absolute error (MAE), the root mean square error (RMSE) (Abirami and Chitra, 2021), and the coefficient of determination (R^2) are used, as shown in Eq. (19), Eq. (20), and Eq. (21), respectively.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|p_i - y_i| \tag{19}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2} \tag{20}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(p_i - y_i)^2}{\sum_{i=1}^{n}(p_i - \overline{p})^2} \tag{21}$$

where p_i, y_i and $\overline{p}$ are the actual value, the predicted value, and the actual average value, respectively. The smaller the values of MAE and RMSE, the more accurate the results. The larger the value of R^2, the better the result.

4.2. Performance of temporal feature extraction

4.2.1. Extraction of the optimal time lag of meteorological factors

Multi-scale time lags of meteorological factors are extracted. Fig. 6(a) and Fig. 6(b) show the combined bar charts of MAE and RMSE values for $PM_{2.5}$, NO_2, and O_3 at different time lags. It can be seen that the values of MAE and RMSE do not gradually increase with increasing time lags. That is, it is not the case that the closer the time lag is set to the current time, the better the results. Therefore, it is necessary to divide the meteorological factors into different time lags to choose the optimal time lag. Above, the MAE and RMSE values of $PM_{2.5}$ and O_3 are the mean values of multiple simulation results. Since MAE and RMSE values of NO_2 are similar, their median and mean values were compared (Table 2). When the time lag is 24 h, the values of MAE and RMSE of O_3 are the minimum. At the same time, by comparing several metrics of the simulation results of NO_2, the MAE and RMSE values of NO_2 are the smallest at 24 h. Thus, their optimal time lags are 24 h. Similarly, the optimal time lag for $PM_{2.5}$ is 4 h.

4.2.2. Extraction of the key features from multi-source data

First, Pearson correlation analysis is performed on meteorological factors with the optimal time lag obtained in Section 4.2.1 as well as air pollutants of the target station. Second, the features that are highly correlated with the pollutants to be predicted are selected. With $PM_{2.5}$ as an example, the Pearson correlation analysis with other pollutants is performed to obtain the Pearson correlation heat map (Fig. 7). This figure reflects the correlation magnitude of each feature, which can be used to select the key features. In this study, the features with correlation greater than 0.4 with $PM_{2.5}$ are selected as the key features, namely PM_{10}, SO_2, NO_2, CO, and Dewp. Moreover, PM_{10}, CO, O_3, and Wspd are selected as the key features of NO_2, while PM_{10}, NO_2, Temp, and Wspd are selected as the key features of O_3.

4.3. Performance of spatial feature extraction

In the station hierarchy, the Granger results (Table 3) between the target station and the rest stations are obtained through Granger causality test. The target station has absolute Granger causality with itself. Therefore, its Granger result is set to 0. The $prob(S)$ value between each station and the target station is less than 0.05, which indicates that the early change of the station sequence will lead to the change of

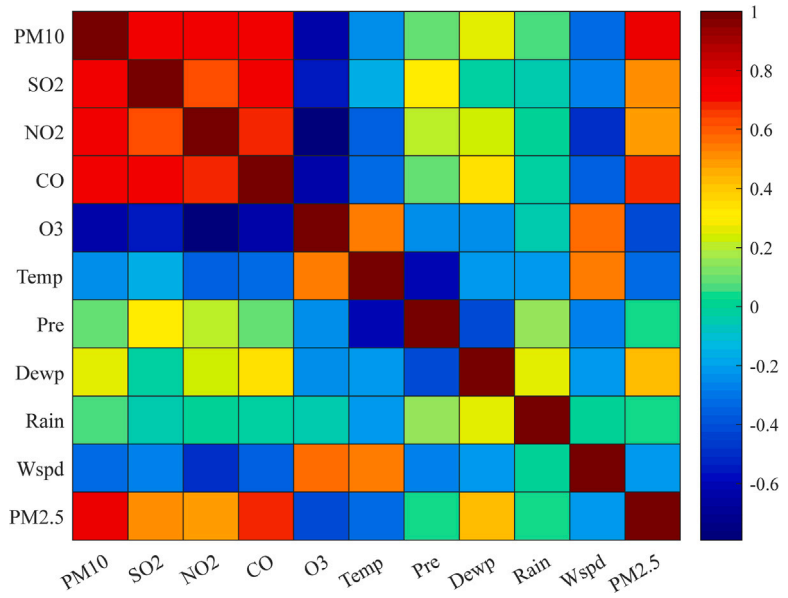
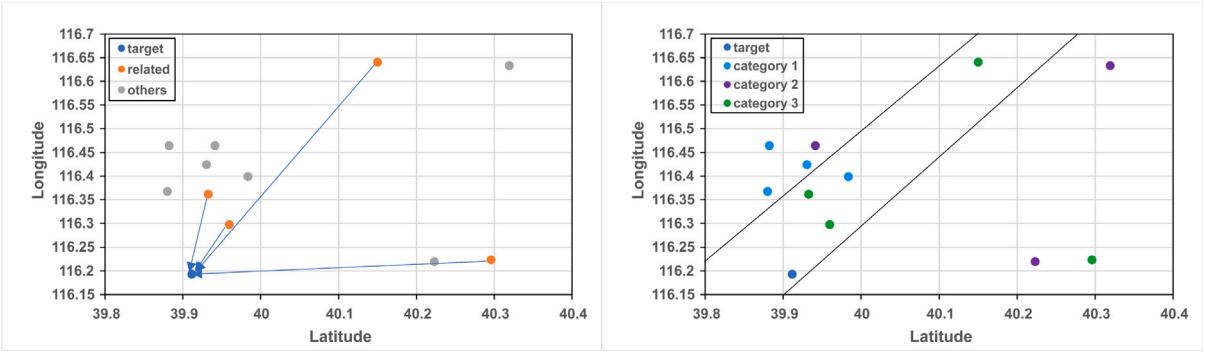


Fig. 7. Heat map of Pearson correlation between features.



(a) Station hierarchy (b) Area hierarchy.

Fig. 8. Spatial feature extraction based on spatial hierarchy division.

Table 3
Granger results.

	DongS	NongZG	TianT	WanL	ChangP	DingL	GuanY	HuaiR	ShunY	WanSXG	AoTZX	GuC
$prob(S)$	0.113	0.857	0.224	0.028	0.792	0.040	0.020	0.884	0.005	0.170	0.137	0.000
$prob(Z)$	0.541	0.006	0.000	0.014	0.014	0.003	0.135	0.000	0.001	0.035	0.161	0.000

the target station sequence. According to the research results, it can be determined that the Wanliu, Dingling, Guanyuan, and Shunyi stations have an influence on the target station. The results of station hierarchy division are shown in Fig. 8(a), where the blue dot represents the target station, the orange dots represent the stations that affect the target station.

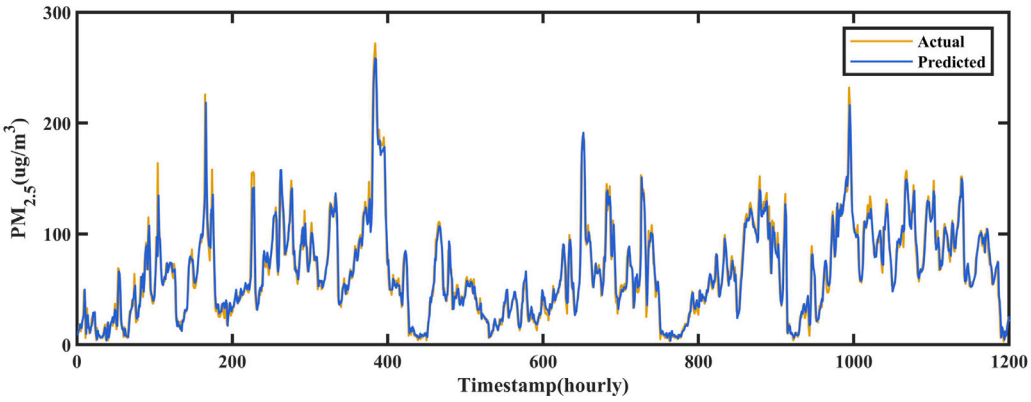
In the area hierarchy, the Granger results $prob(S)$ and $prob(Z)$ of each station are used as the inputs of the k-means clustering algorithm. The stations with similar characteristics are classified into the same category. As shown in Fig. 8(b), the same color indicates that these stations have similar characteristics. Based on the similar characteristics of stations and the non-crossing principle of areas, stations with similar characteristics are classified into the same area to the greatest extent. A total of three similar areas are divided. The stations in the same area as the target station are selected as the related stations of the target station. In this paper, the related stations of the target station are Wanliu, Guanyuan, Shunyi, and Aotizhongxin stations.

4.4. Prediction performance of the proposed model

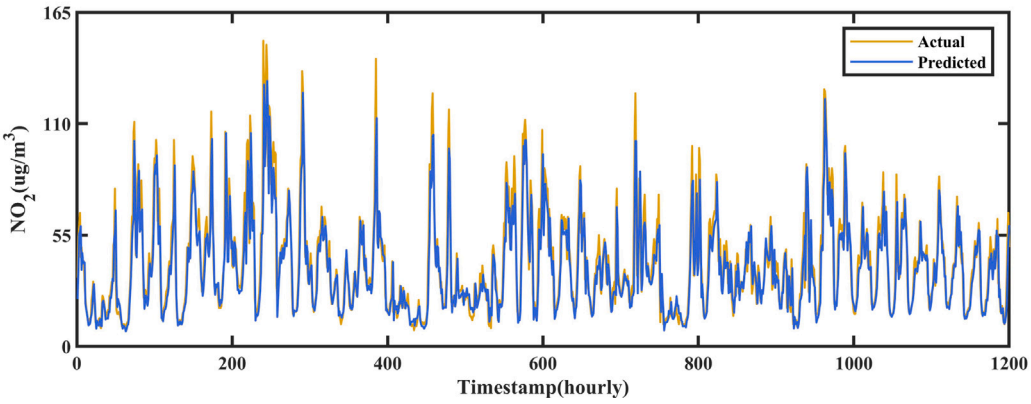
The prediction results of the proposed model for the concentrations of three air pollutants, $PM_{2.5}$, NO_2, and O_3 are shown in Fig. 9, which shows the comparison of the actual and predicted values of $PM_{2.5}$, NO_2, and O_3 concentrations for 1200 consecutive 1-h timestamps, respectively. It can be found that the predicted curves of the three air pollutant concentrations fit well with the actual curves, and most of the points approximately coincide.

4.5. Comparison of sudden changes with other models

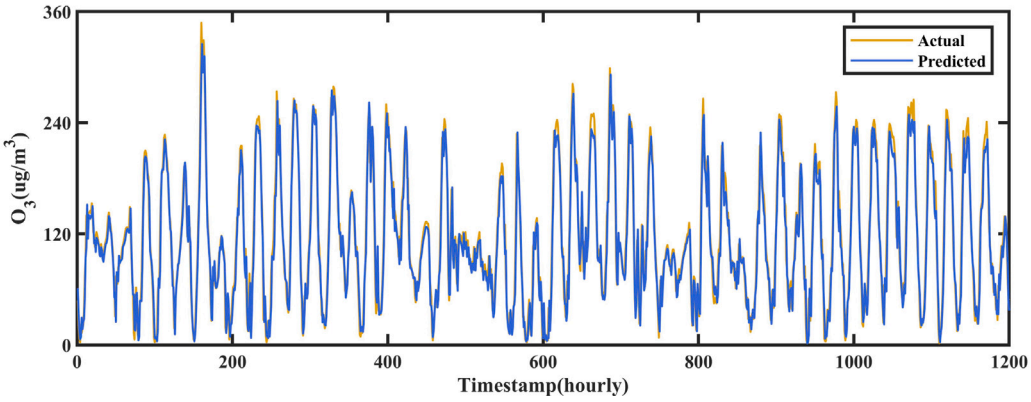
In the case of $PM_{2.5}$, sudden-change predictions of the proposed model (values changing by more than 50 in one hour or by more than 100 in two hours) are compared with those of other models, and the results are shown in the Table 4. The MAE and RMSE values of sudden-change predictions of the proposed model are 38.1934 and 44.9664,



(a) $PM_{2.5}$



(b) NO_2



(c) O_3

Fig. 9. The comparisons of the actual and predicted values of $PM_{2.5}$, NO_2, and O_3 concentrations for 1200 consecutive 1-h timestamps.

respectively, with the smallest error. Other models do not analyze and handle sudden changes accordingly. In this study, the multi-scale time lags of meteorological factors are extracted to cope with sudden changes. By comparing the results, the effectiveness of the proposed method for sudden-change prediction is demonstrated.

Table 4
Comparison of sudden changes with other models.

	MAE	RMSE
Adaboost	41.3754	46.9118
Attention-LSTM	40.0136	45.1101
CNN-LSTM	40.6328	48.1195
PSO-LSTM	47.0017	52.1156
Ours	38.1934	44.9664

4.6. Comparison with the common selection methods of spatial-related stations

In order to verify that the proposed method for selecting related stations can improve the prediction accuracy of the air-quality of sparse station, this paper compares the proposed method for selecting related stations with two other methods, namely the distance method based on nearest-neighbor selection and the Pearson method. Except for the method of selecting the related stations, the other settings for the simulation remain the same. According to the 1-h prediction results (Table 5), it can be found that the distance method has the worst results, while the proposed method has the best results.

Table 5
Comparison with the common selection methods of spatial-related stations.

	$PM_{2.5}$			NO_2			O_3		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
Distance	7.97 ± 0.01	12.64 ± 0.01	0.91 ± 0.00	7.36 ± 0.00	11.28 ± 0.02	0.79 ± 0.01	13.72 ± 0.12	17.76 ± 0.12	0.95 ± 0.00
Pearson	7.99 ± 0.10	12.60 ± 0.22	0.91 ± 0.00	7.32 ± 0.02	11.26 ± 0.03	0.79 ± 0.01	13.14 ± 0.08	17.12 ± 0.13	0.95 ± 0.00
Ours	6.63 ± 0.26	10.13 ± 0.56	0.94 ± 0.01	6.88 ± 0.04	10.26 ± 0.09	0.84 ± 0.01	12.71 ± 0.25	16.65 ± 0.37	0.95 ± 0.00

Table 6
Comparison with baselines.

	$PM_{2.5}$			NO_2			O_3		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
LSTM	16.30 ± 0.56	19.13 ± 0.63	0.92 ± 0.01	7.56 ± 0.14	10.86 ± 0.61	0.83 ± 0.05	15.27 ± 0.86	18.81 ± 0.32	0.93 ± 0.01
PSO-LSTM	7.85 ± 0.04	12.09 ± 0.30	0.92 ± 0.01	7.08 ± 0.06	10.86 ± 0.02	0.81 ± 0.00	13.32 ± 0.03	17.39 ± 0.01	0.95 ± 0.01
Attention-LSTM	8.00 ± 0.03	12.48 ± 0.17	0.92 ± 0.01	7.12 ± 0.11	11.06 ± 0.20	0.80 ± 0.01	13.09 ± 0.06	17.15 ± 0.05	0.95 ± 0.00
Adaboost	8.11 ± 0.30	12.56 ± 0.77	0.92 ± 0.01	7.38 ± 0.17	11.28 ± 0.23	0.80 ± 0.00	13.45 ± 0.02	18.21 ± 0.06	0.94 ± 0.00
Ours	6.63 ± 0.26	10.13 ± 0.56	0.94 ± 0.01	6.88 ± 0.04	10.26 ± 0.09	0.84 ± 0.01	12.71 ± 0.25	16.65 ± 0.37	0.95 ± 0.00

Table 7
Comparison with models popular in dense station.

	$PM_{2.5}$			NO_2			O_3		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
LSTME	12.08 ± 1.83	15.90 ± 1.30	0.92 ± 0.01	8.06 ± 0.91	11.60 ± 0.76	0.79 ± 0.02	13.18 ± 0.14	16.63 ± 0.47	0.94 ± 0.02
TCN	12.13 ± 1.46	16.16 ± 1.59	0.87 ± 0.06	7.74 ± 0.20	11.62 ± 0.11	0.79 ± 0.01	13.94 ± 0.47	18.38 ± 0.11	0.94 ± 0.00
CNN-LSTM	8.54 ± 0.13	12.84 ± 0.50	0.92 ± 0.02	7.22 ± 0.01	11.16 ± 0.10	0.80 ± 0.00	13.19 ± 0.02	17.47 ± 0.01	0.94 ± 0.00
Ours	6.63 ± 0.26	10.13 ± 0.56	0.94 ± 0.01	6.88 ± 0.04	10.26 ± 0.09	0.84 ± 0.01	12.71 ± 0.25	16.65 ± 0.37	0.95 ± 0.00

Meanwhile, the multi-step prediction is also carried out (Fig. 10(a) and (b)). In the 6-h prediction, the errors of the three methods are similar. As the prediction step increases, the error of the proposed model is significantly smallest. Distance method cannot accurately reflect the correlation between stations. Compared to Pearson method, the proposed method finds the stations that are causally related to the target station. Furthermore, the proposed method finds the related stations through a more detailed spatial hierarchy division. The results show that the selection method of spatial stations proposed in this paper is more suitable for the sparse station.

4.7. Comparison with baselines

This paper compared the proposed model with baselines. The baselines chosen for comparison in this paper are LSTM, PSO-LSTM, Attention-LSTM, and Adaboost models. These baselines can be used individually or as components of the hybrid model. LSTM is the most common model in prediction. PSO-LSTM, as an optimization of LSTM, often outperforms LSTM in prediction. The other two baselines are the components of the hybrid model proposed in this paper. Table 6 shows the comparisons of 1-h prediction results for $PM_{2.5}$, NO_2, and O_3 concentrations by each model. In the case of $PM_{2.5}$, the proposed model has the smallest MAE and RMSE values, which are 6.63±0.26 and 10.13±0.56, respectively, followed by the component models Adaboost and Attention-LSTM, whose MAE values are both within 10. The MAE and RMSE values of the LSTM and PSO-LSTM models are larger. It can be found that the proposed model is superior in predicting the 1-h concentration of these three air pollutants.

This study also predicts the $PM_{2.5}$ concentration of the sparse station for the next 6, 12, 18, 24, and 48 h. As can be seen from Fig. 10(c) and (d), the prediction errors of the proposed model are relatively small, especially in long-term prediction. The LSTM model is a conventional encoder–decoder structure. When the input sequence is too long, its prediction performance may be limited. The PSO-LSTM model is still unable to dynamically assign weight to each feature. The Attention mechanism is used in the component of the proposed model, which dynamically assigns weights to the input features and focuses on the important features. The prediction performance of a single model is not as good as that of the hybrid model. The proposed model selects different prediction models for the spatio-temporal feature groups, and fully exploits the advantages of both models in processing the corresponding feature groups. The results demonstrate the superiority of the proposed model over previous models.

4.8. Comparison with models popular in dense station

To demonstrate the validity of the proposed model, it is compared with models popular in dense station. The popular models chosen for comparison in this paper are LSTME, TCN, and CNN-LSTM models, which are commonly used for air-quality prediction in dense station. The 1-h prediction results are shown in Table 7. In the case of $PM_{2.5}$, the proposed model has the smallest MAE and RMSE values, which are 6.63±0.26 and 10.13±0.56, followed by the CNN-LSTM model, whose MAE and RMSE values are 8.54±0.13 and 12.84±0.50, and the TCN model is the largest. The R^2 of the proposed model is the largest, indicating the best fit.

In terms of multi-step prediction, LSTME, CNN-LSTM, and the proposed models show a slight increase in error with increasing prediction step (Fig. 10(e) and (f)). In contrast, the prediction error of the proposed model is small in all prediction steps. In the long-term prediction, the proposed model has the smallest prediction error. With the increase of prediction step, the prediction advantage becomes more pronounced. The TCN model is highly dependent on the temporal correlation explored. As a result, the prediction accuracy decreases as the prediction step increases. LSTME and CNN-LSTM models have limited learning ability with sparse data. The proposed model can comprehensively analyze the relationship of the target station and its surrounding stations. By selecting related stations and key influencing factors, the influence of the redundant feature is eliminated and the prediction accuracy of the sparse station is improved.

5. Conclusions

This paper introduces a hybrid prediction model of air quality for the sparse station based on the extraction of spatio-temporal features. The air-quality prediction for the sparse station relies on the features of the air-quality data of the station and its surrounding stations. First, the spatio-temporal features of the sparse station are obtained through

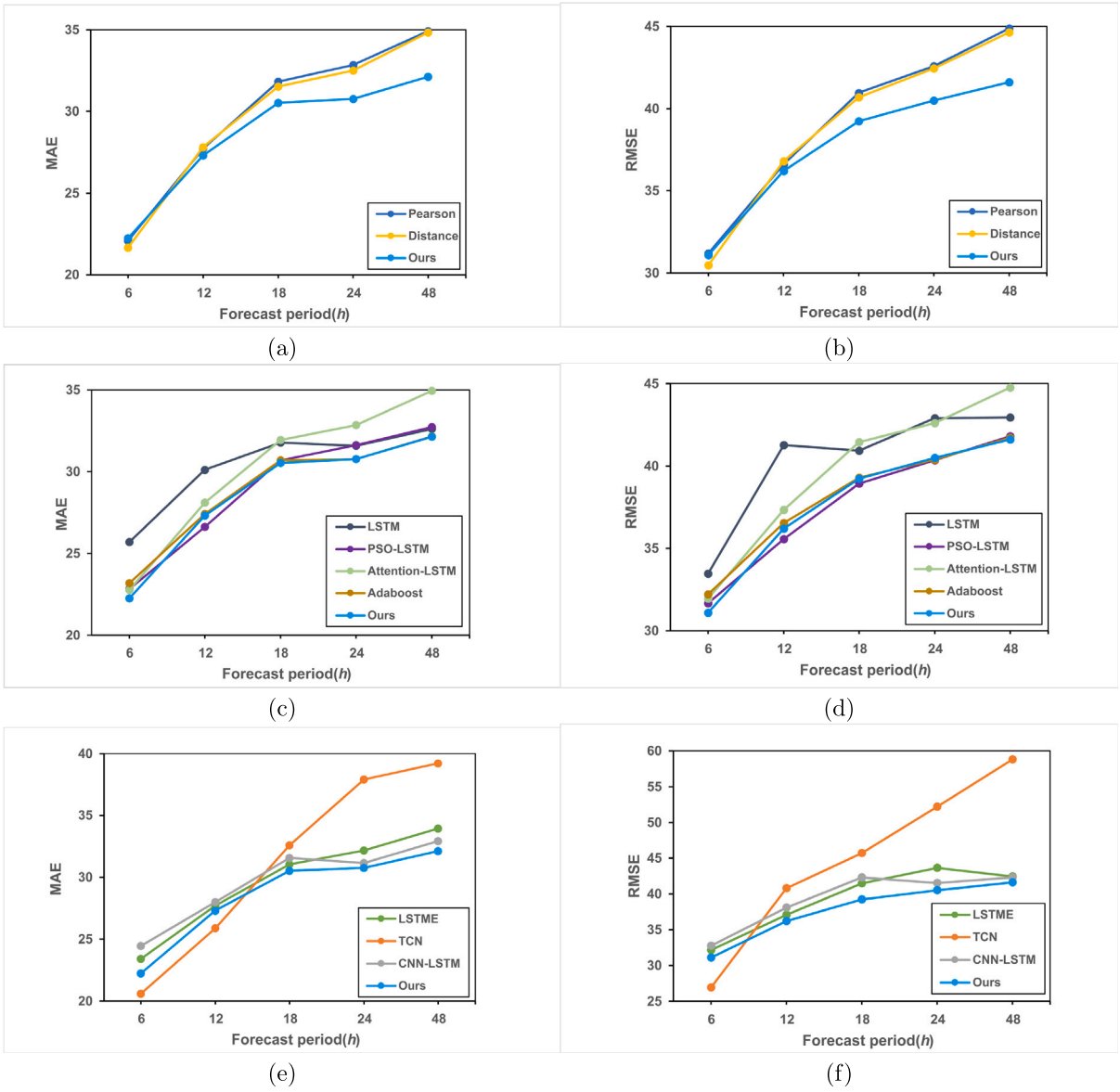


Fig. 10. Multi-step prediction for $PM_{2.5}$.

the extraction of multi-scale temporal features and the selection of related stations based on spatial hierarchy division. Then, the hybrid model prediction based on the spatio-temporal feature groups is performed. The results show that the extraction of multi-scale time lags of meteorological factors in this paper improves the accuracy of sudden-change prediction. In the meanwhile, the proposed model has better performance in the air-quality prediction of the sparse station compared to the models popular in dense station. The following conclusions can be drawn from the simulation results.

First of all, the method of extracting the multi-scale time lags of the meteorological factors can indeed improve the accuracy of sudden-change prediction. In this method, meteorological factors with different time lags are predicted and compared to find the time lag with the greatest influence on the concentration of pollutants to be predicted. This enhances the usefulness of meteorological features for sudden-change prediction.

Secondly, by dividing the study area into the station and area, and combining the stations found that have an influence on the target station in the station hierarchy and the similar areas found in the area hierarchy, the related stations of the target station are found. This method of selecting related stations improves the performance of air-quality prediction of the sparse station.

Finally, the hybrid prediction model is a method to improve the prediction accuracy. The sub-models selected according to different feature groups can give full play to the advantages of their respective models, thus improving the prediction accuracy of the entire hybrid model.

CRediT authorship contribution statement

Yue Hu: Data curation, Writing – original draft, Methodology, Software, Visualization. **Xiaoxia Chen:** Supervision, Reviewing. **Hanzhong Xia:** Visualization, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Zhejiang Provincial Natural Science Foundation of China under Grant LY21F030004; the National Natural Science Foundation of China under Grant 61803214; the Natural Science Foundation of Ningbo, China under Grant 2019A610451; the K.C.Wong Magna Fund in Ningbo University, China.

References

Abirami, S., Chitra, P., 2021. Regional air quality forecasting using spatiotemporal deep learning. J. Clean. Prod. 283, 125341. http://dx.doi.org/10.1016/j.jclepro.2020.125341.

Brauer, M., Casadei, B., Harrington, R.A., Kovacs, R., Sliwa, K., W. H. F. Air Pollution Expert Group, 2021. Taking a stand against air pollution-the impact on cardiovascular disease: A joint opinion from the World Heart Federation, American College of Cardiology, American Heart Association, and the European Society of Cardiology. Circulation 143 (14), e800–e804. http://dx.doi.org/10.1161/CIRCULATIONAHA.120.052666.

Chen, P.-C., Lin, Y.T., 2022. Exposure assessment of PM2. 5 using smart spatial interpolation on regulatory air quality stations with clustering of densely-deployed microsensors. Environ. Pollut. 292, 118401. http://dx.doi.org/10.1016/j.envpol.2021.118401.

Faraji, M., Nadi, S., Ghaffarpasand, O., Homayoni, S., Downey, K., 2022. An integrated 3D CNN-GRU deep learning method for short-term prediction of PM2.5 concentration in urban environment. Sci. Total Environ. 834, 155324. http://dx.doi.org/10.1016/j.scitotenv.2022.155324.

Fei, S.w., 2016. A hybrid model of EMD and multiple-kernel RVR algorithm for wind speed prediction. Int. J. Electr. Power Energy Syst. 78, 910–915. http://dx.doi.org/10.1016/j.ijepes.2015.11.116.

Ge, L., Wu, K., Zeng, Y., Chang, F., Wang, Y., Li, S., 2021. Multi-scale spatiotemporal graph convolution network for air quality prediction. Appl. Intell. 51 (6), 3491–3505. http://dx.doi.org/10.1007/s10489-020-02054-y.

Gu, K., Zhou, Y., Sun, H., Zhao, L., Liu, S., 2020. Prediction of air quality in Shenzhen based on neural network algorithm. Neural Comput. Appl. 32 (7), 1879–1892. http://dx.doi.org/10.1007/s00521-019-04492-3.

Guo, Z., Zhao, W., Lu, H., Wang, J., 2012. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renew. Energy 37 (1), 241–249. http://dx.doi.org/10.1016/j.renene.2011.06.023.

Huang, C.J., Kuo, P.H., 2018. A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities. Sensors 18 (7), 2220. http://dx.doi.org/10.3390/s18072220.

Huang, G., Li, X., Zhang, B., Ren, J., 2021. PM2.5 concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition. Sci. Total Environ. 768, 144516. http://dx.doi.org/10.1016/j.scitotenv.2020.144516.

Kalhor, M., Bajoghli, M., 2017. Comparison of AERMOD, ADMS and ISC3 for incomplete upper air meteorological data (case study: Steel plant). Atmos. Pollut. Res. 8 (6), 1203–1208. http://dx.doi.org/10.1016/j.apr.2017.06.001.

Lee, H.M., Park, R.J., Henze, D.K., Lee, S., Shim, C., Shin, H.J., Moon, K.J., Woo, J.H., 2017. PM2.5 source attribution for Seoul in May from 2009 to 2013 using GEOS-Chem and its adjoint model. Environ. Pollut. 221, 377–384. http://dx.doi.org/10.1016/j.envpol.2016.11.088.

Li, T., Guo, Y., Liu, Y., Wang, J., Wang, Q., Sun, Z., He, M.Z., Shi, X., 2019. Estimating mortality burden attributable to short-term PM2.5 exposure: A national observational study in China. Environ. Int. 125, 245–251. http://dx.doi.org/10.1016/j.envint.2019.01.073.

Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., Chi, T., 2017. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. Environ. Pollut. 231, 997–1004. http://dx.doi.org/10.1016/j.envpol.2017.08.114.

Lim, C.H., Ryu, J., Choi, Y., Jeon, S.W., Lee, W.K., 2020. Understanding global PM2.5 concentrations and their drivers in recent decades (1998–2016). Environ. Int. 144, 106011. http://dx.doi.org/10.1016/j.envint.2020.106011.

Liu, S., Hua, S., Wang, K., Qiu, P., Liu, H., Wu, B., Shao, P., Liu, X., Wu, Y., Xue, Y., Hao, Y., Tian, H., 2018. Spatial-temporal variation characteristics of air pollution in Henan of China: Localized emission inventory, WRF/Chem simulations and potential source contribution analysis. Sci. Total Environ. 624, 396–406. http://dx.doi.org/10.1016/j.scitotenv.2017.12.102.

Liu, C.C., Lin, T.C., Yuan, K.Y., Chiueh, P.T., 2022. Spatio-temporal prediction and factor identification of urban air quality using support vector machine. Urban Clim. 41, 101055. http://dx.doi.org/10.1016/j.uclim.2021.101055.

Luo, Z., Huang, J., Hu, K., Li, X., Zhang, P., 2019. AccuAir: Winning solution to air quality prediction for KDD cup 2018. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1842–1850. http://dx.doi.org/10.1145/3292500.3330787.

Ma, X., Longley, I., Gao, J., Kachhara, A., Salmond, J., 2019. A site-optimised multi-scale GIS based land use regression model for simulating local scale patterns in air pollution. Sci. Total Environ. 685, 134–149. http://dx.doi.org/10.1016/j.scitotenv.2019.05.408.

Perez, P., Reyes, J., 2006. An integrated neural network model for PM10 forecasting. Atmos. Environ. 40 (16), 2845–2851. http://dx.doi.org/10.1016/j.atmosenv.2006.01.010.

Qi, Y., Li, Q., Karimian, H., Liu, D., 2019. A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory. Sci. Total Environ. 664, 1–10. http://dx.doi.org/10.1016/j.scitotenv.2019.01.333.

Qin, Z., Cen, C., Guo, X., 2019. Prediction of air quality based on KNN-LSTM. J. Phys. Conf. Ser. 1237 (4), 042030. http://dx.doi.org/10.1088/1742-6596/1237/4/042030.

Samal, K.K.R., Babu, K.S., Das, S.K., 2021. Multi-directional temporal convolutional artificial neural network for PM2.5 forecasting with missing values: A deep learning approach. Urban Clim. 36, 100800. http://dx.doi.org/10.1016/j.uclim.2021.100800.

Saravanan, D., Kumar, K.S., 2022. IoT based improved air quality index prediction using hybrid FA-ANN-ARMA model. Mater. Today: Proc. 56, 1809–1819. http://dx.doi.org/10.1016/j.matpr.2021.10.474.

Seng, D., Zhang, Q., Zhang, X., Chen, G., Chen, X., 2021. Spatiotemporal prediction of air quality based on LSTM neural network. Alex. Eng. J. 60 (2), 2021–2032. http://dx.doi.org/10.1016/j.aej.2020.12.009.

Sun, W., Wang, Y., 2018. Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. Energy Convers. Manage. 157, 1–12. http://dx.doi.org/10.1016/j.enconman.2017.11.067.

Tai, A.P.K., Mickley, L.J., Jacob, D.J., 2010. Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change. Atmos. Environ. 44 (32), 3976–3984. http://dx.doi.org/10.1016/j.atmosenv.2010.06.060.

Tong, W., Li, L., Zhou, X., Hamilton, A., Zhang, K., 2019. Deep learning PM2.5 concentrations with bidirectional LSTM RNN. Air Qual., Atmos. Health 12 (4), 411–423. http://dx.doi.org/10.1007/s11869-018-0647-4.

Wang, J., Song, G., 2018. A deep spatial-temporal ensemble model for air quality prediction. Neurocomputing 314, 198–206. http://dx.doi.org/10.1016/j.neucom.2018.06.049.

Wang, C., Zheng, J., Du, J., Wang, G., Klemeš, J., Wang, B., Liao, Q., Liang, Y., 2022. Weather condition-based hybrid models for multiple air pollutants forecasting and minimisation. J. Clean. Prod. 352, 131610. http://dx.doi.org/10.1016/j.jclepro.2022.131610.

Wen, C., Liu, S., Yao, X., Peng, L., Li, X., Hu, Y., Chi, T., 2019. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. Sci. Total Environ. 654, 1091–1099. http://dx.doi.org/10.1016/j.scitotenv.2018.11.086.

Wu, Q., Lin, H., 2019. A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors. Sci. Total Environ. 683, 808–821. http://dx.doi.org/10.1016/j.scitotenv.2019.05.288.

Xiang, Y., Gou, L., He, L., Xia, S., Wang, W., 2018. A SVR–ANN combined model based on ensemble EMD for rainfall prediction. Appl. Soft Comput. 73, 874–883. http://dx.doi.org/10.1016/j.asoc.2018.09.018.

Yan, R., Liao, J., Yang, J., Sun, W., Nong, M., Li, F., 2021. Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. Expert Syst. Appl. 169, 114513. http://dx.doi.org/10.1016/j.eswa.2020.114513.

Yang, X., Wu, Q., Zhao, R., Cheng, H., He, H., Ma, Q., Wang, L., Luo, H., 2019. New method for evaluating winter air quality: PM2.5 assessment using community multi-scale air quality modeling (CMAQ) in Xi'an. Atmos. Environ. 211, 18–28. http://dx.doi.org/10.1016/j.atmosenv.2019.04.019.

Yi, X., Zhang, J., Wang, Z., Li, T., Zheng, Y., 2018. Deep distributed fusion network for air quality prediction. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 965–973. http://dx.doi.org/10.1145/3219819.3219822.

Zhang, L., Lin, J., Qiu, R., Hu, X., Zhang, H., Chen, Q., Tan, H., Lin, D., Wang, J., 2018. Trend analysis and forecast of PM2.5 in Fuzhou, China using the ARIMA model. Ecol. Indic. 95, 702–710. http://dx.doi.org/10.1016/j.ecolind.2018.08.032.

Zhang, Y., Zhang, R., Ma, Q., Wang, Y., Wang, Q., Huang, Z., Huang, L., 2020. A feature selection and multi-model fusion-based approach of predicting air quality. Isa Trans. 100, 210–220. http://dx.doi.org/10.1016/j.isatra.2019.11.023.