Environ Monit Assess (2023) 195:1510
https://doi.org/10.1007/s10661-023-12081-0

RESEARCH

Hourly forecasting on $PM_{2.5}$ concentrations using a deep neural network with meteorology inputs

Yanjie Liang · Jun Ma · Chuanyang Tang · Nan Ke · Dong Wang

Received: 12 July 2023 / Accepted: 31 October 2023 / Published online: 22 November 2023

Abstract The $PM_{2.5}$ (particulate matter with a diameter of fewer than 2.5 μm) has become a global topic in environmental science. The neural network that based on the non-linear regression algorithm, e.g., deep learning, is now believed to be one of the most facile and advanced approaches in $PM_{2.5}$ concentration prediction. In this study, we proposed a $PM_{2.5}$ predictor using deep learning as infrastructure and meteorological data as input, for predicting the next hour $PM_{2.5}$ concentration in Beijing Aotizhongxin monitor point. We efficiently use the parameter's spatiotemporal correlation by concatenating the dataset with time series. The predicted $PM_{2.5}$ concentration was based on meteorology changes over a period. Therefore, the accuracy would increase with the period growing. By extracting the intrinsic features between meteorological and $PM_{2.5}$ concentration, a fast and accurate prediction was carried out. The *R* square score reached maximum of 0.98 and remained an average of 0.9295 in the whole test. The average bias of the model is 9 μg on the validation set and 1 μg on the training set. Moreover, the differences between the predictions and expectations can be further regarded as the estimation for the emission change. Such results can provide scientific advice to supervisory and policy workers.

Keywords Air quality forecasting · Machine learning · Data analytics · $PM_{2.5}$ · Meteorology

Yanjie Liang and Jun Ma contributed equally to this work.

Y. Liang · D. Wang (✉)
School of Energy and Power Engineering, Shandong University, Jinan 250061, China
e-mail: dwang@sdu.edu.cn

J. Ma · C. Tang · N. Ke
College of Engineering, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA

Introduction

The $PM_{2.5}$ (particulate matter with a diameter of fewer than 2.5 μm) has become a global topic in environmental science, mainly linked to lung cancer, heart disease, stroke, respiratory infections, and chronic lung disease (WHO, 2016). In China, air pollutants caused by human activities caused severe problems in the past few decades. Ecological governance has drawn much more attention from both the government and the people for the last 10 years (Yang, 2016). Nevertheless, poor diffusion conditions, direct emission from residential energy consumption, and high-emission factories in and around Beijing are still great challenges for the government (Chen et al., 2020; Yang et al., 2019). It is a tough issue to benefit environmental improvement and economic development (Song et al., 2022).

Accurate theoretical models are expected to be used in forecasting the $PM_{2.5}$ concentration. They are based on the laws of physics and chemistry and usually applied to simulate the diffusion and emission of pollutants (Cui et al., 2015). However, the disadvantages of the theoretical models are also evident as their advantages. In the use of the community multiscale air quality (CMAQ) with meteorological conditions for predicting the $PM_{2.5}$ concentration, one of the well-known theoretical models in the pollutant prediction, we need to process the data through a meteorology-chemistry interface processor (MCIP), which is complicated (Lightstone et al., 2017; Zhang et al., 2021). This process requires a heavy load of computational calculations. This makes the CMAQ a slow and complex process to predict the $PM_{2.5}$ concentration after we input the initial variable. Sometimes, it may need manual handling, which makes the process even more tedious.

The empirical model, which is usually in the form of the production of machine learning, is prevalent in the meteorology prediction. In earlier studies, researchers prefer to use the linear regression algorithm (one algorithm to build the empirical models) to predict the $PM_{2.5}$ concentration (Vlachogianni et al., 2011; Zhou et al., 2014). They tried to use meteorological conditions only and made some progress (Tai et al., 2010). However, most of them are rudimentary with either some sample correlations between the pollutant concentration and meteorology conditions or imprecise predictions of the pollutant concentration. Therefore, the neural network based on the non-linear regression algorithm, e.g., deep learning, convolutional neural networks (CNN), and recurrent neural network (RNN), is now believed to be one of the most facile and advanced approaches in $PM_{2.5}$ concentration prediction. They already showed enhanced performance in meteorological forecasting in most recent studies (Pak et al., 2020; Suriya et al., 2023). However, it is widely acknowledged that the learning of neural networks heavily relies on the choice of data preprocessing methods and algorithmic selection. As for input data preprocessing methods, Yang et al. investigated the predictive performance of $PM_{2.5}$ concentration using CNN, LSTM, and CNN-LSTM models with different input configurations, model architectures, and prediction timeframes, highlighting the importance of optimizing inputs and structures for improved multi-hour $PM_{2.5}$ forecasting (Yang et al., 2021). Lee et al. achieved a substantial enhancement in observational accuracy ($R^2 \approx 0.8$) by employing unconventional satellite-derived aerosol optical depth from GOCI and MODIS as a key predictor (Lee et al., 2021).

As for algorithmic selection, Dai et al., proposed a VAR-XGBoost model for O_3 prediction in China, achieving superior accuracy compared to other models. Their study identifies key factors influencing O_3 concentrations, emphasizes the importance of wind speed and temperature, and highlights the spatial distribution of O_3 with east–west gradients. Additionally, they introduced a PCA-MEE-ISPO-LightGBM model (matter-element extension–principal component analysis–improved particle swarm optimization–light gradient boosting machine algorithm) to assess urban haze risk. Their findings show that the full index improves evaluation accuracy by 4–16% compared to using only causative factor indices (Dai et al., 2022, 2023). Teng et al. proposed a hybrid graph deep neural network (GNN_LSTM) model to dynamically capture the spatiotemporal correlations of $PM_{2.5}$. Such a design substantially improves the model performance in 72-h $PM_{2.5}$ forecasting (overall R^2 increases from 0.6 to 0.79) (Teng et al., 2023). Zhu et al. developed an automated hourly $PM_{2.5}$ forecasting model using a parallel multi-input 1D-CNN-biLSTM model by combining data from both target and nearby monitoring stations, achieving RMSE, MAE, and R^2 values of 3.88, 2.52, and 0.94, respectively (Zhu et al., 2023).

The achievements resulting from these efforts are exciting, but they result in increased computational requirements or additional data processing (Huang et al., 2021; Kumar & Sahu, 2021). What is more, the inaccuracy of the emission inventory always limits the reliability of the forecast results. The optimization of comprehensive cost factors, including computational resources, data sources, and accuracy, presents a formidable challenge in achieving a harmonious equilibrium.

In this study, we proposed a facile and accurate way of predicting the hourly $PM_{2.5}$ concentration using a non-linear regression algorithm based neural network approach. The hourly meteorological conditions are set up as the only input. The parameter's spatiotemporal correlation was fully and efficiently considered by concatenating the dataset with time series. A fast and accurate prediction was carried out. By repeating rigorous tests, the highest R^2 in our test dataset reached 0.98 and generally around 0.95. This means if the emission is unchanged, our model is accurate enough to predict the concentration. Therefore, with higher R^2 and fewer input parameters, we proposed a brand new method of evaluating the emission reduction from the actual $PM_{2.5}$ concentration.

Data sources

In this study, we collected data from two organizations and merged them into one dataset. The air quality data are collected from Beijing Municipal Environmental Monitoring Center (http://www.bjmemc.com.cn/), and the meteorological data are downloaded from the National Oceanic and Atmospheric Administration (NOAA, https://www.ncdc.noaa.gov/). We use air quality data from the National Olympic Sports Center Aotizhongxin air quality monitor point in Beijing and Beijing national airport meteorological data. In the meteorological dataset, there are a lot of missing data presented by − 9999. Some of them are caused by sensor and server downtime, and some of them are caused by the limitation of technology that the method we monitor the meteorological is longer than 1 h. We make up those missing values by some function. The parameters of the merged dataset are presented in Table 1.

The data presented in Table 1 contained 19,633 rows collected from 01/01/2018 to 03/28/2020. The dataset includes missing values presented by the blanks. We tried to find the function that can calculate $PM_{2.5}$ concentration by giving the meteorological data.

Table 1 Merged dataset

Parameters	Timing	Unit
$PM_{2.5}$	–	$\mu g{\cdot}m^{-3}$
Temperature	–	°C
Dew point	–	°C
Wind speed rate	–	$m{\cdot}s^{-1}$
Liquid precipitation	6-h average	mm
Sea level pressure	–	hPa

Methodology

Data analysis

As shown in Fig. 1, the $PM_{2.5}$ concentration has a seasonal variation. Based on various factors like central heating, holidays, and weather, the Ministry of Environmental Protection of China usually divides the Beijing annual climate into two seasons. Accordingly, from May to October is the non-heating season, and from October to next year's April is the heating season. Like most cities in the north part of China, Beijing provides central heating from early November to the end of March (Yue

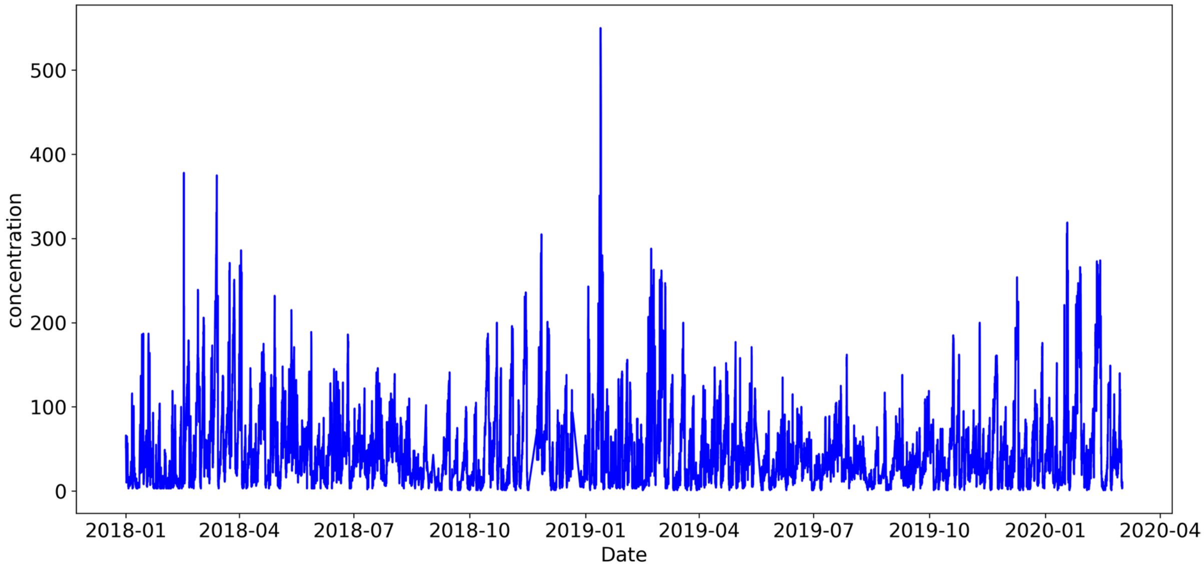


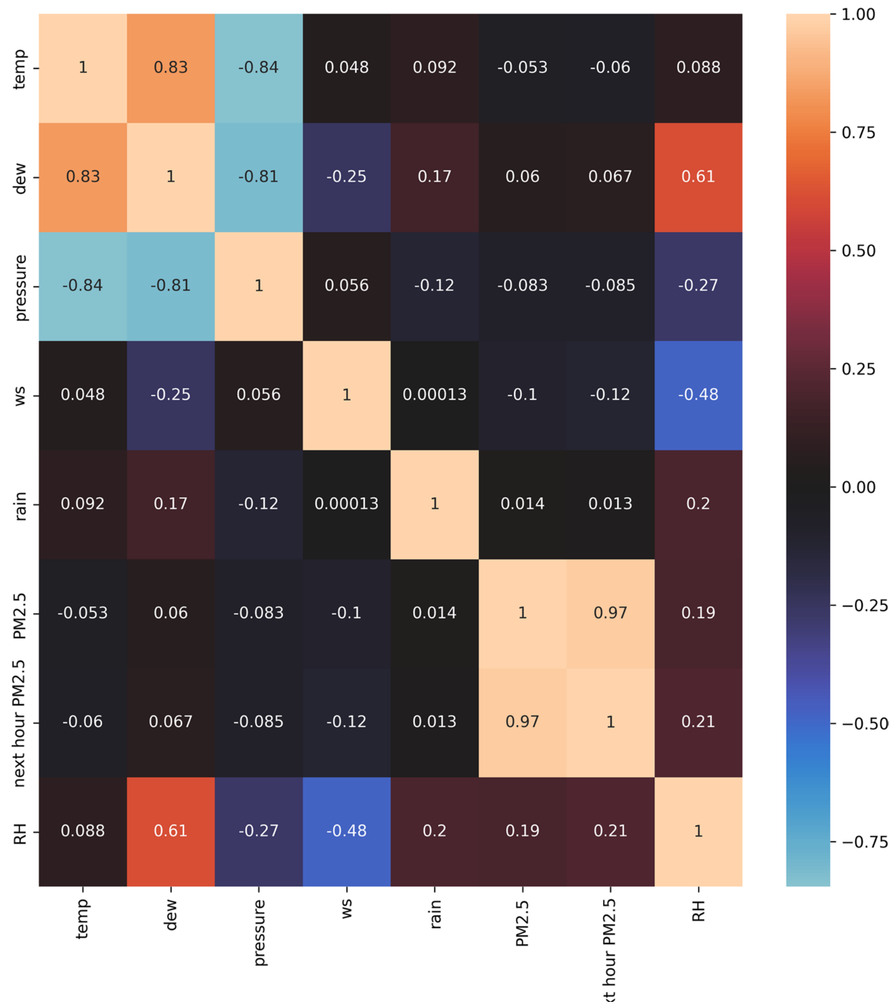
Fig. 1 $PM_{2.5}$ concentration in Aotizhongxin point

et al., 2018). During the heating season, some festival traditions will cause a rapid increase in pollutants. For instance, the grand fireworks display during the lunar new year (Seidel & Birnbaum, 2015). Those also happen in many other festivals in winter. Since the weather conditions in winter usually reduce the diffusion effect, the pollutant may last longer than in summer.

However, the diffusion conditions could also influence this seasonal change. According to the meteorological conditions, we believe that the climate conditions also show seasonal variation. Moreover, in Beijing, the anthropogenic influence will decrease during winter since most of the residents will travel out of the city in the winter vacation. Accordingly, the emission throughout one calendar year can be assumed as a constant in the *n*th layer of the deep neural network. In other words, we might have a way to express emission in terms of meteorological conditions. To minimize the factors that would influence the prediction of $PM_{2.5}$ concentration, we chose to get rid of the approximate constant. In that case, the only thing that will affect the $PM_{2.5}$ concentration in our model is the diffusion conditions.

To reveal the relationship between $PM_{2.5}$ concentration and meteorological parameters, we first tried linear regression and plotted the correlation heatmap, as shown in Fig. 2. As we indicated empirically, the $PM_{2.5}$ concentration will decrease when the meteorological conditions favor diffusion (Wang & Ogawa, 2015). We can also find a similar trend in the correlation map. However, based on the proposed data in Fig. 2, we are not supposed to use simple linear regression to describe the relationship between meteorological conditions and $PM_{2.5}$ concentration. Since the relationship

Fig. 2 Heatmap of all variables



is not close. To get more reliable results, we need more powerful tools to generalize the relationship.

The data in Fig. 2 indicated that there is a strong correlation between the temperature, the atmospheric pressure, and the dew point. A stronger correlation will cause multicollinearity, which will influence the accuracy in the linear regression (Yoo et al., 2014). Besides, it will lead to a non-linear regression inaccuracy. So, we use the relative humidity (*RH*) instead of the temperature (*temp*) and the dew point (*DP*), which shows a stronger correlation with the $PM_{2.5}$ concentration and is independent of other variables. The covert function that is backward from the Goff-Gratch equation and the Magnus equation is shown as Formula 1 (Alduchov & Eskridge, 1996; Buck, 1981).

$$RH = e^{\frac{17.27DP}{237.7+DP} - \frac{17.27temp}{237.7+temp}} \tag{1}$$

In addition, by retrieving the data published in journals and our analyzing for parameters, we found that the current $PM_{2.5}$ concentration has a stronger relationship with the previous meteorological conditions (Huang & Kuo, 2018; Pak et al., 2020). From the heating map, we know that the *R* score for the current $PM_{2.5}$ concentration and the wind speed is -0.1. While the *R* score for the next hour's $PM_{2.5}$ and wind speed is -0.12. Consequently, we should get better results when we use time series prediction.

Deep learning

From the data analysis, we concluded that the $PM_{2.5}$ concentrations are related to climate conditions within a specified period. In other words, we can use machine learning based on the time series to help us increase the accuracy of the predicted result.

There are several machine learning algorithms based on time series, for example, the long-short-term-memory (LSTM) network (Hochreiter & Schmidhuber, 1997). However, algorithms like this would not work on our dataset because the dataset we are using is not large enough and contains missing values. In dealing with the insufficient dataset, we need to drop the missing values and shuffle our dataset for getting an accurate result. This operation will change the time series information in the dataset. To address this issue, we append time series by inserting columns that are shifted by the other columns in our dataset. We used $(T-n)$ to represent the value of the parameter for *n* hours before the predicted time. Accordingly, the input and output of our model is shown in Table 2.

Table 2 Input and output parameters

Input	
	RH $(T-n)$[1]
	Pressure $(T-n)$
	Wind speed $(T-n)$
	Precipitation $(T-n)$
	Temperature $(T-n)$
	RH $(T-(n-1))$
	......
	RH $(T-1)$
	Pressure $(T-1)$
	Wind speed $(T-1)$
	Precipitation $(T-1)$
	Temperature $(T-1)$
Output	
	$PM_{2.5}$ concentration (T)

We used deep neural network to build our model (Bai et al., 2019; Pak et al., 2020). The deep neural network consists of multiple logical layers to progressively extract features from the original data. Generally, a deep neural network has three types of layers, which are the input layer, the output, and the hidden layers. Figure 3 shows a classical deep neural network. There are full connections between each layer.

In this study, we used dense layers to build our network, which is a kind of the optimized LSTM model. The dense layers are similar to the linear layers, but they need activation to complete linear or non-linear regression. Within a limited part of the network, it can be described as a function shown as Formula 2.

$$f(\sum_{i=1}^{n} wx + b) \tag{2}$$

For example, the *Relu* is an activation that represents a linear function, and *Tanh* is a non-linear activation. Both of them show advantages and disadvantages in regression. We combined the three different for building our model. The *Tanh* and the *Sigmoid* are similar. They both have excellent convergence rates; however, although with similar advantages, they also have negative impacts, like gradient vanishing

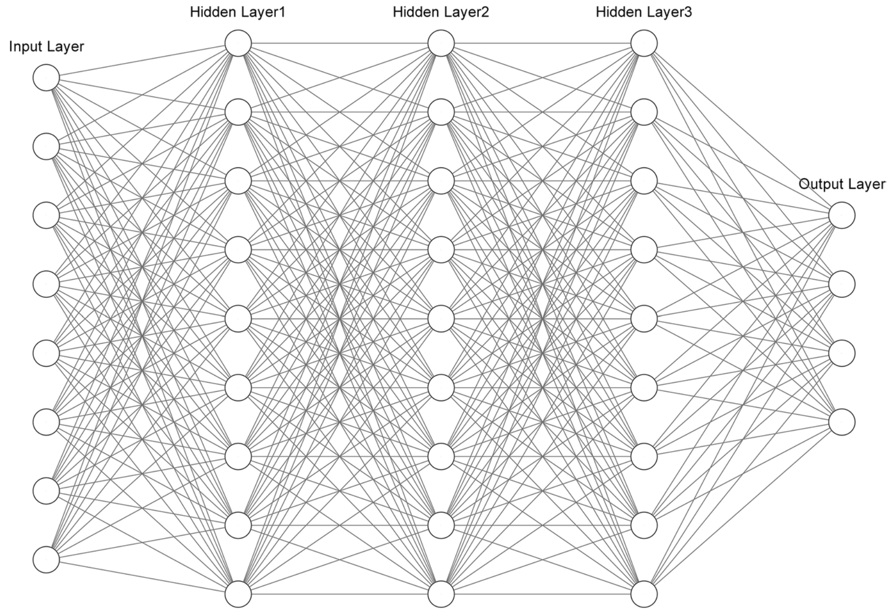


Fig. 3 Classical deep neural network

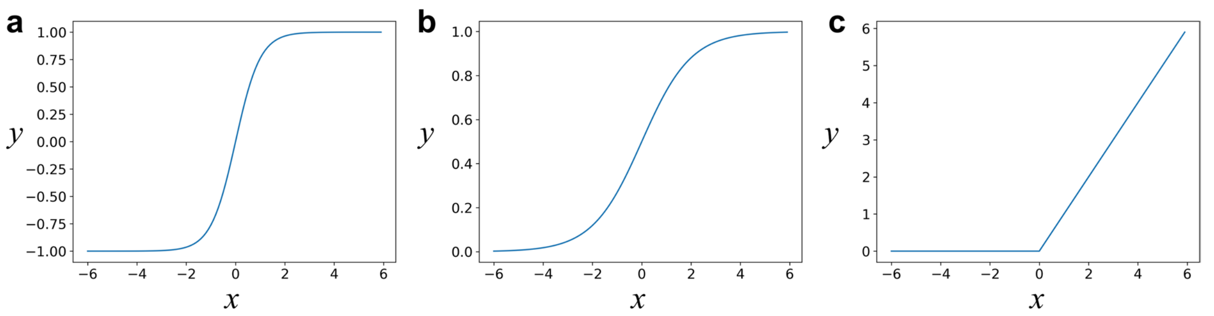


Fig. 4 Activation models of **a** *Tanh*; **b** *Sigmoid*; and **c** *Relu*

problems. The *Relu* can effectively solve the gradient vanishing problem. However, it will lead to a new dying *Relu* problem: the neuron will never activate on any data point again. The three activation models are shown in Fig. 4.

Two of the most important factors that would influence the model training are the loss calculator and the optimizer. The mean absolute error (MAE) is a common measure in statistics and is usually applied in machine learning and neural networks. In this study, the MAE is utilized as the loss calculator since the MAE is unambiguous and more natural when representing the measure of average (Cort & Kenji, 2005). The MAE is calculated following Formula 3:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{3}$$

The other important role in deep learning is the optimizer. The process of deep learning is an optimization issue that minimizes the target function of $\mathbf{J}(\boldsymbol{\theta})$ (Heaton, 2018). The optimizer is the algorithm on which the optimization process depends. There are two types of optimizers: the stochastic gradient descent (SGD) series and the adaptive learning rate (ALR) series (Yan et al., 2018). Adam is an SGD optimizer with momentum. It introduces the ALR to make the optimizing process more quickly and the results more precise. The Adamax is an ALR optimizer. Compared to

Table 3 Extra parameters

Parameters	Remarks
look_back_step	In this study, we set this parameter to 6, 12, and 24. It means our input will contain the meteorological data for the past 6, 12, and 24 h
predict_forward	In this study, we set this parameter to 1. It means the targeted predicted $PM_{2.5}$ concentration is for the next hour
train_set_length	This parameter determines the length of the training data. In this study, it equals 75% of the total dataset
valid_set_length	This parameter determines the length of the validation data. In this study, it equals 15% of the total dataset
test_set_length	This parameter determines the length of the testing data. In this study, it equals 10% of the total dataset

Table 4 Layers in detail

Layers	Units	Activation
Dense	128	*Tanh*
Dense	256	*Sigmoid*
Dense	256	*Tanh*
Dense	512	*Relu*
Dense	512	*Tanh*
Dense	1024	*Tanh*
Dense	1024	*Relu*
Dense	1	None

Adam, the Adamax use infinite norm to converge to a more stable state.

Workflow

Step 1: setting special parameters

In our model, five extra parameters need to be set as special parameters, as shown in Table 3.

Step 2: data processing

First, read the dataset we prepared before, change the temperature and the dew point to the relative humidity, and make it a normalization set. Then, use the parameters that we set in step 1 to prepare the training, the validation, and the dataset testing.

Step 3: building the model and training the data

We used TensorFlow to implement our neural network, the mean absolute error as the loss calculation, and the Adamax as the optimizer.

The layers in the model are shown in Table 4.

Step 4: prediction and display

Use the testing data to predict the next hour's $PM_{2.5}$ concentration. Compare the results to the real $PM_{2.5}$ concentration. After predicting, we denormalized the prediction and expectation values and displayed them in the same figure.

Results and discussion

Prediction performance

We evaluated the performance of the predictor by predicting the test dataset after training our predictor with validation and training dataset. After shuffling

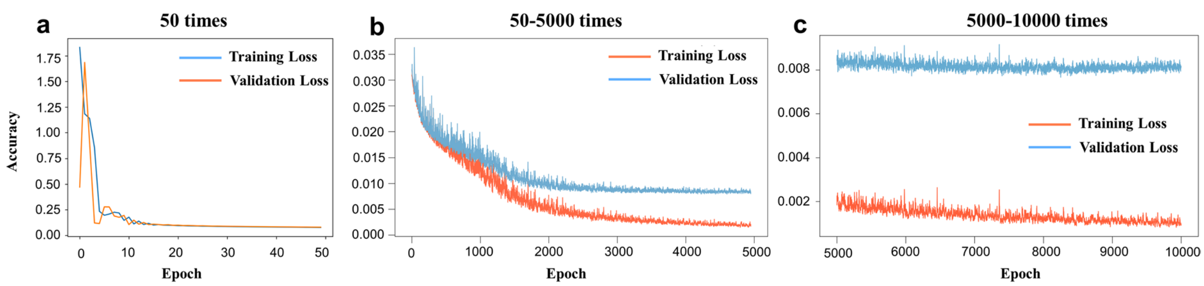


Fig. 5 Visualization of training process

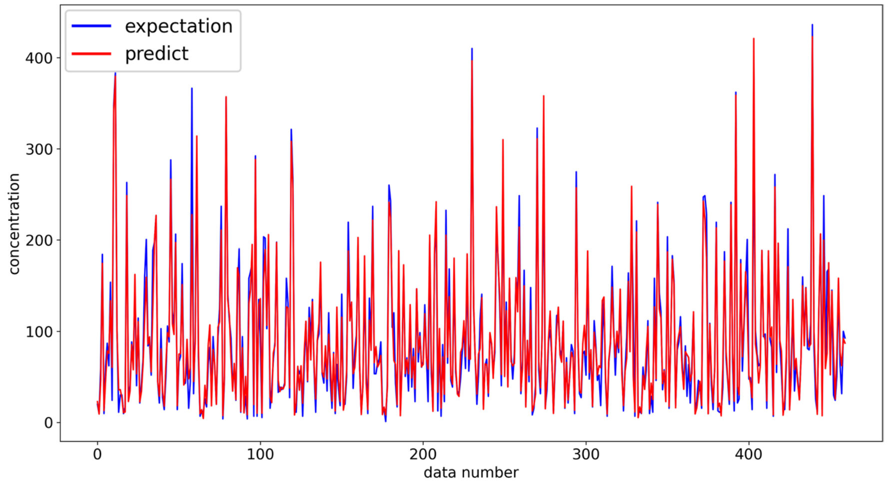


Fig. 6 Prediction of test dataset

Table 5 Key evaluation scores for different models

Look-back hours	Training loss (last value)	Validation loss (last value)	*R* square
6	0.0020	0.0200	0.6812
12	0.0016	0.0111	0.8827
24	0.0009	0.0090	0.9295
36	0.0006	0.0081	0.9456
48	0.0007	0.0079	0.9490

the whole dataset, the training data and test data are 85% and 15% of the total sample data, respectively. The key indicator of the evaluation is the *R* square score between prediction and expectation. Still, we also used mean absolute error in validation loss and training loss as an auxiliary reference to evaluate our model during the training process. We used 24 h as our look-back parameter and trained our model. The training process is shown in Fig. 5. The validation loss is about 0.009, the training loss is about 0.0009, and our scale for $PM_{2.5}$ concentration is 0–1000, which means that the average bias of the model is only 9 µg on the validation set and only 1 µg on the training set.

In Fig. 6, we displayed the prediction based on the test dataset. In this test, our *R* square score is 0.9295. Moreover, the prediction in some singular sample data also showed a precise value. The

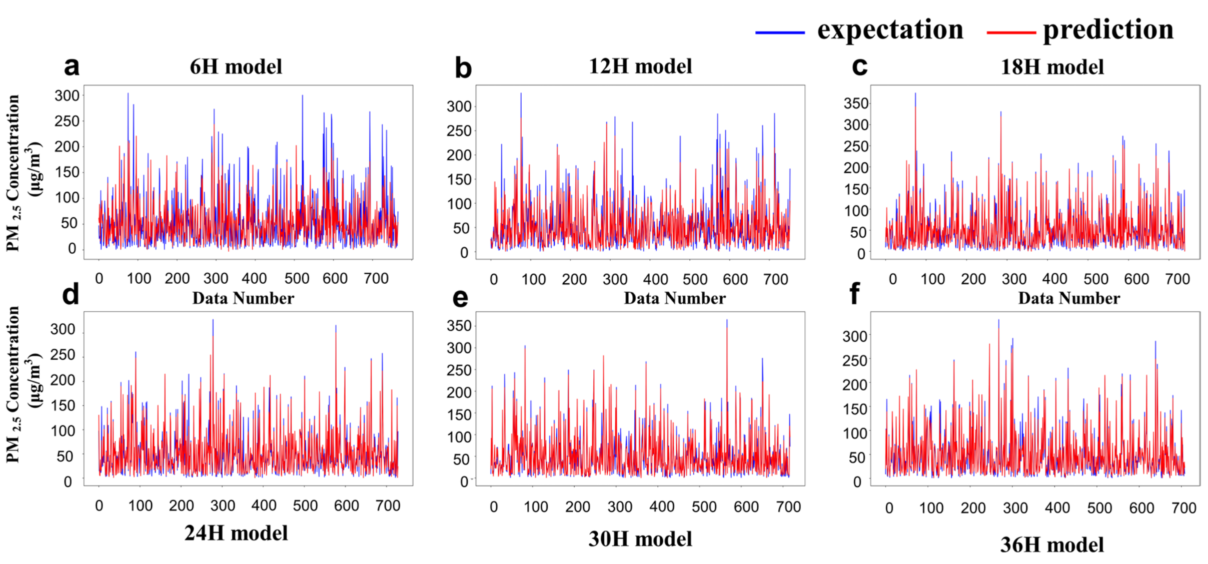


Fig. 7 Six different models' prediction comparisons

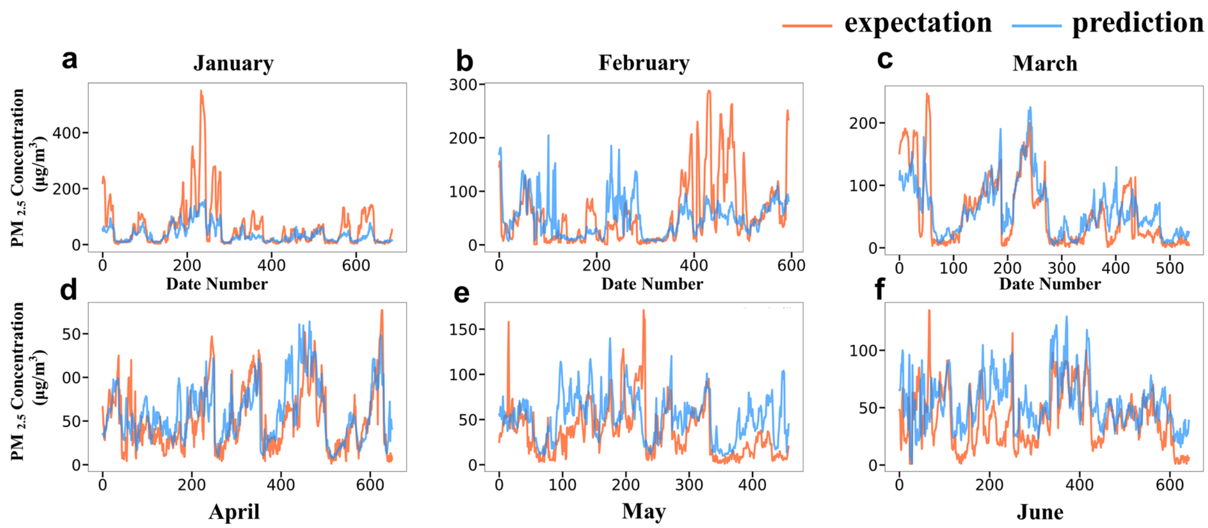


Fig. 8 Prediction of 2019 (January to June)

precise prediction shows a strong correlation between $PM_{2.5}$ concentrations and meteorology conditions. This result also confirms our conjecture that we can assume that the emission over a period can be replaced by some constant or a number covert from the meteorological conditions.

Significant variables

In this study, we had observed the difference between the models by changing the parameter of look-back time. We set our look-back time to 6 h, 12 h, 24 h, 36 h, and 48 h. With increasing the look-back hours, it would take more time to preprocess the dataset and train the model, and also, it takes more space to store the dataset. At the same time, we will have a more accurate prediction. Table 5 shows key evaluation scores for our models. A significant space and time increment showed up when the look-back hour changed from 6 to 12 h. The growth was not evident when it changed from 12 to 24 h.

Figure 7 shows our results for models. The prediction model made with 6 h is precise in the low concentration part but does not perform well at high concentration. The accuracy would be improved by increasing the look-back hours. However, when the look-back hour is more than 36 h, the progress is almost to nothing. This phenomenon proves that the $PM_{2.5}$ staying time in this area is about one and a half days.

Table 6 Average concentration of $PM_{2.5}$

Month	Prediction	Actual value
All	54.90	46.76
1	38.37	65.41
2	51.88	58.28
3	59.20	55.88
4	58.21	52.48
5	49.13	36.91
6	44.47	39.83
7	49.75519	39.54215
8	42.906364	26.591953
9	48.57876	41.165897
10	72.35884	42.368332
11	74.83305	46.587044
12	70.13672	47.830585

Table 7 Key indicators

Value name	Prediction	Expectation
Average	54.74	54.26
R square score	0.9833	
75% data error[a]	1.94	
Max error value	160.61	
Error more than 10	5% data, 399 rows/6890 rows	

[a]75% of the data errors are less than 1.94

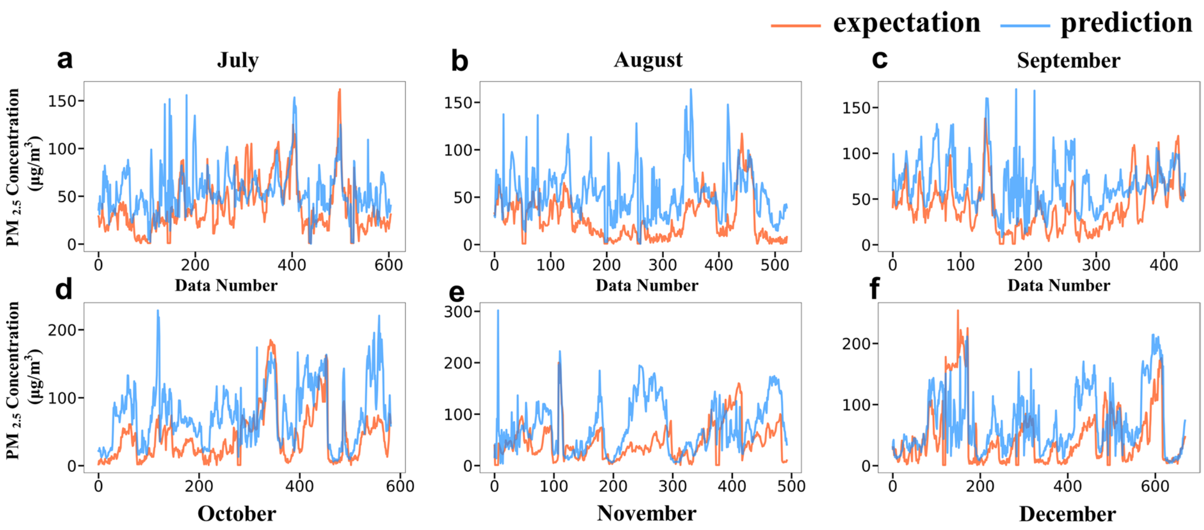


Fig. 9 Prediction of 2019 (July to December)

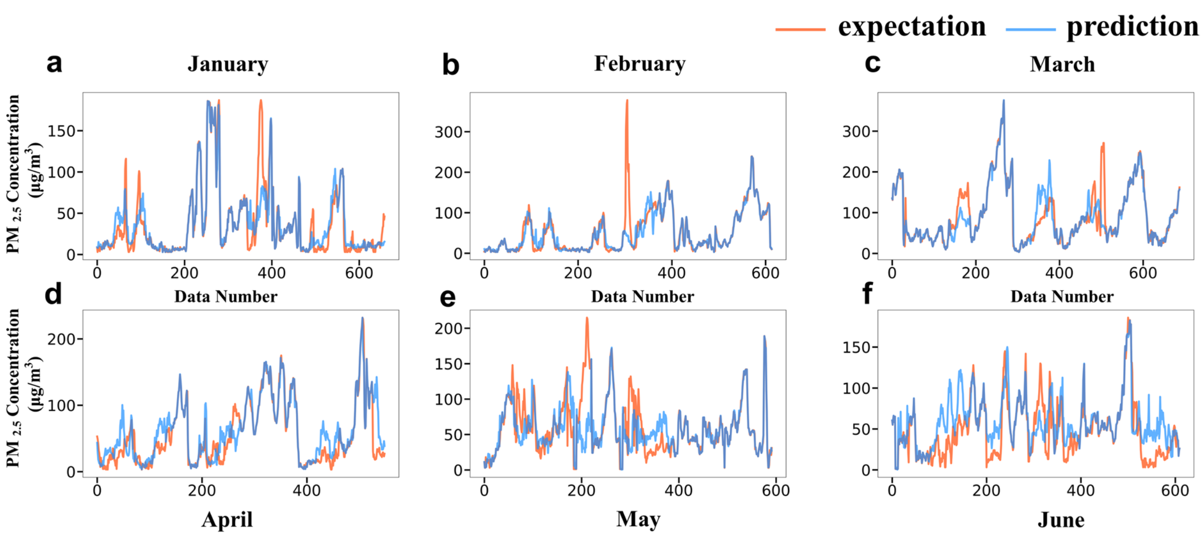


Fig. 10 Prediction of 2018 (January to June)

This result also confirms our previous studies. The meteorological condition's impact is time-sensitive. From our research, we could assume that most condition's effect is mostly within 36 h. And the data in the first 12 h were strongly correlated to the prediction.

Limitations and improvement

Although our model's value uses as less information as possible to predict the concentration of $PM_{2.5}$ and the predictor has pretty good performance in prediction, the model has some limitations like follows. Since our hypothesis has some limitations that, we assume the emission is calculated by meteorological conditions and it is a constant in the *n*th derivative; however, in the real world, Beijing's emissions have been decreasing year by year, and our model cannot predict well in the future. However, from the other direction, we can use this model to evaluate the effect of action on energy conservation and emission

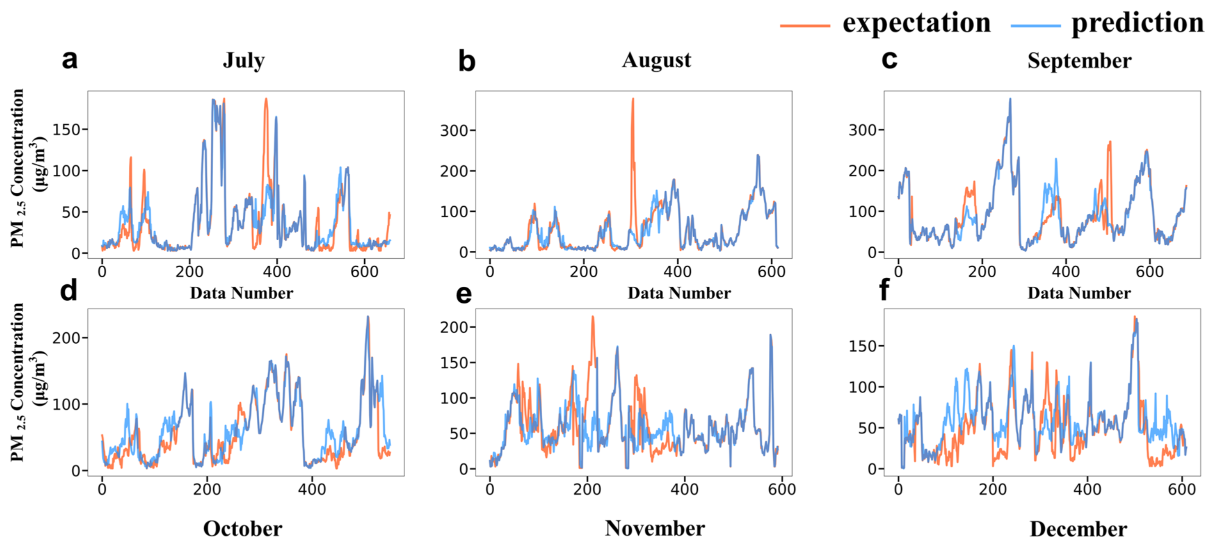


Fig. 11 Prediction of 2018 (July to December)

reduction in Beijing. Take 48-h model as an example, and we change the input from 2018 to 2019 year. The prediction and the real concentration is shown in Fig. 8 and Fig. 9. This difference is caused by the emission change in different periods. In other words, we can calculate the change of emission level by our model with some functions. We did a calculation and displayed it in Table 6. The results of the comparison show that the average concentration decreased by 15%, which is basically in line with the official report from the Beijing government.

To verify the prediction and our assumption, we also used data from 2018 to observe the results. The key value of the judgment is shown in Table 7 and Fig. 10 and Fig. 11. We can see that this model perfectly match the data from 2018, which proves that our method is feasible for projections of overall emissions.

Conclusion

In this paper, we proposed a $PM_{2.5}$ predictor using deep learning as infrastructure and meteorological data as input for predicting the next hour's $PM_{2.5}$ concentration in the Beijing Aotizhongxin monitor point. We fully consider and efficiently use the parameter's spatiotemporal correlation by concatenating the dataset with time series. By extracting the intrinsic features between meteorological and $PM_{2.5}$ concentration, a fast and accurate prediction was carried out. We predicted the $PM_{2.5}$ concentration based on a meteorology change over a period. Therefore, the accuracy improved with the period of growth. The effect of meteorology is time-based: the $PM_{2.5}$ concentration is highly related to the meteorological conditions for the first 12 h and started going down after that. Based on the previous studies, the $PM_{2.5}$ concentration was commonly calculated based on parameters like CO concentration or via likelihood estimation after sampling. In these studies, the inaccuracy of the emission inventory always limits the reliability of the forecast results. From this perspective, it is complicated and unreliable. That is the reason why we believe it is a simpler and more efficient approach to complete the prediction through meteorological conditions. Moreover, the differences between the predictions and expectations can be further regarded as the estimation for the emission change. Such results can provide scientific advice to supervisory and policy workers. We believe it is possible to get a precise future prediction by adding data variables that contain the emission change information.

Acknowledgements The scientific calculations in this paper have been done on the HPC Cloud Platform of Shandong University. The views and ideas expressed herein are solely those of the authors and do not represent the opinions of the funding

agencies in any form. The views and ideas expressed herein are solely those of the authors.

Author contribution All authors contributed to the study's conception and design. YL and JM contributed equally to this work. Material preparation, data collection, and analysis were performed by YL, JM, CT, and NK. YL and JM wrote the first draft of the manuscript, and all authors commented on previous versions. DW: writing—review and editing, supervision. All authors read and approved the final manuscript.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding authors upon reasonable request.

Declarations

Ethics approval and consent to participate All authors have read, understood, and have complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors and are aware that with minor exceptions, no changes can be made to authorship once the paper is submitted.

Competing interests 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.

References

Alduchov, O. A., & Eskridge, R. E. (1996). Improved Magnus form approximation of saturation vapor pressure. *Journal of Applied Meteorology and Climatology, 35*(4), 601–609. https://doi.org/10.1175/1520-0450(1996)035%3c0601:IMFAOS%3e2.0.CO;2

Bai, Y., Li, Y., Zeng, B., Li, C., & Zhang, J. (2019). Hourly PM2.5 concentration forecast using stacked autoencoder model with emphasis on seasonality. *Journal of Cleaner Production, 224*, 739–750. https://doi.org/10.1016/j.jclepro.2019.03.253

Buck, A. L. (1981). New equations for computing vapor pressure and enhancement factor. *Journal of Applied Meteorology and Climatology, 20*(12), 1527–1532. https://doi.org/10.1175/1520-0450(1981)020%3c1527:NEFCVP%3e2.0.CO;2

Chen, X., Yang, T., Wang, Z., Hao, Y., He, L., & Sun, H. (2020). Investigating the impacts of coal-fired power plants on ambient PM2.5 by a combination of a chemical transport model and receptor model. *Science of The Total Environment, 727*, 138407. https://doi.org/10.1016/j.scitotenv.2020.138407

Cort, J. W., & Kenji, M. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research, 30*(1), 79–82. https://doi.org/10.3354/cr030079

Cui, H., Chen, W., Dai, W., Liu, H., Wang, X., & He, K. (2015). Source apportionment of PM2.5 in Guangzhou combining observation data analysis and chemical transport model simulation. *Atmospheric Environment, 116*, 262–271. https://doi.org/10.1016/j.atmosenv.2015.06.054

Dai, H., Huang, G., Zeng, H., & Yu, R. (2022). VAR-tree model based spatio-temporal characterization and prediction of O3 concentration in China. *Systems, 10*(6), 263. https://doi.org/10.3390/systems1006026

Dai, H., Huang, G., Wang, J., & Zeng, H. (2023). VAR-tree model based spatio-temporal characterization and prediction of O3 concentration in China. *Ecotoxicology and Environmental Safety, 257*, 114960. https://doi.org/10.1016/j.ecoenv.2023.114960

Heaton, J. (2018). Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning. *Genetic Programming and Evolvable Machines, 19*(1), 305–307. https://doi.org/10.1007/s10710-017-9314-z

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

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. https://doi.org/10.3390/s18072220

Huang, C., Hu, J., Xue, T., Xu, H., & Wang, M. (2021). High-resolution spatiotemporal modeling for ambient PM2.5 exposure assessment in China from 2013 to 2019. *Environmental Science Technology, 55*(3), 2152–2162. https://doi.org/10.1021/acs.est.0c05815

Kumar, V., & Sahu, M. (2021). Evaluation of nine machine learning regression algorithms for calibration of low-cost PM2.5 sensor. *Journal of Aerosol Science, 157*, 105809. https://doi.org/10.1016/j.jaerosci.2021.105809

Lee, A., Jeong, S., Joo, J., Park, C.-R., Kim, J., & Kim, S. (2021). Potential role of urban forest in removing PM2.5: A case study in Seoul by deep learning with satellite data. *Urban Climate, 36*, 100795. https://doi.org/10.1016/j.uclim.2021.100795

Lightstone, S. D., Moshary, F., & Gross, B. (2017). Comparing CMAQ forecasts with a neural network forecast model for PM2.5 in New York. *Atmosphere, 8*(9), 161. https://doi.org/10.3390/atmos8090161

Pak, U., Ma, J., Ryu, U., Ryom, K., Juhyok, U., Pak, K., & Pak, C. (2020). Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: A case study of Beijing, China. *Science of the Total Environment, 699*, 133561. https://doi.org/10.1016/j.scitotenv.2019.07.367

Seidel, D. J., & Birnbaum, A. N. (2015). Effects of Independence Day fireworks on atmospheric concentrations of fine particulate matter in the United States. *Atmospheric Environment, 115*, 192–198. https://doi.org/10.1016/j.atmosenv.2015.05.065

Song, M., Jeong, R., Kim, D., Qiu, Y., Meng, X., Wu, Z., & Ahn, J. (2022). Comparison of phase states of PM2.5 over megacities, Seoul and Beijing, and their implications on particle size distribution. *Environmental Science Technology, 56*(24), 17581–17590. https://doi.org/10.1021/acs.est.2c06377

Suriya, Natsagdorj, N., Aorigele, Zhou, & H., Sachurila. (2023). Prediction of PM2.5 concentration in Ulaanbaatar with deep learning models. *Urban Climate, 47*, 101357. https://doi.org/10.1016/j.uclim.2022.101357

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. *Atmospheric Environment, 44*(32), 3976–3984. https://doi.org/10.1016/j.atmosenv.2010.06.060

Teng, M., Li, S., Xing, J., Fan, C., Yang, J., Wang, S., Song, G., Ding, Y., Dong, J., & Wang, S. (2023). 72-Hour real-time forecasting of ambient PM(2.5) by hybrid graph deep neural network with aggregated neighborhood spatiotemporal information. *Environment International, 176*, 107971. https://doi.org/10.1016/j.envint.2023.107971

Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S., & Kukkonen, J. (2011). Evaluation of a multiple regression model for the forecasting of the concentrations of NOx and PM10 in Athens and Helsinki. *Science of the Total Environment, 409*(8), 1559–1571. https://doi.org/10.1016/j.scitotenv.2010.12.040

Wang, J., & Ogawa, S. (2015). Effects of meteorological conditions on PM2.5 concentrations in Nagasaki, Japan. *International Journal of Environmental Research and Public Health, 12*(8), 9089–9101. https://doi.org/10.3390/ijerph120809089

World Health Organization (WHO). (2016). *Ambient air pollution: A global assessment of exposure and burden of disease*. World Health Organization.

Yan, L., Zhou, M., Wu, Y., & Yan, L. (2018). *Long short term memory model for analysis and forecast of PM2.5.* Paper presented at the Cloud Computing and Security, Cham.

Yang, F. (2016). Under the dome: "Chinese" smog as a viral media event. *Critical Studies in Media Communication, 33*(3), 232–244. https://doi.org/10.1080/15295036.2016.1170172

Yang, J., Song, D., Fang, D., & Wu, F. (2019). Drivers of consumption-based PM2.5 emission of Beijing: A structural decomposition analysis. *Journal of Cleaner Production, 219*, 734–742. https://doi.org/10.1016/j.jclepro.2019.02.109

Yang, J., Yan, R., Nong, M., Liao, J., Li, F., & Sun, W. (2021). PM2.5 concentrations forecasting in Beijing through deep learning with different inputs, model structures and forecast time. *Atmospheric Pollution Research, 12*(9), 101168. https://doi.org/10.1016/j.apr.2021.101168

Yoo, W., Mayberry, R., Bae, S., Singh, K., Peter He, Q., & Lillard, J. W., Jr. (2014). A study of effects of multicollinearity in the multivariable analysis. *International Journal of Applied Science and Technology, 4*(5), 9–19.

Yue, T., Zhang, X., Wang, C., Zuo, P., Tong, Y., Gao, J., & Gao, X. (2018). Environmental impacts of the revised emission standard for air pollutants for boilers during the heating season in Beijing, China. *Aerosol and Air Quality Research, 18*(11), 2853–2864. https://doi.org/10.4209/aaqr.2018.02.0046

Zhang, Z., Zang, Z., Cheng, X., Lu, C., Huang, S., Hu, Y., & Ye, L. (2021). Development of three-dimensional variational data assimilation method of aerosol for the CMAQ model: An application for PM2.5 and PM10 forecasts in the Sichuan Basin. *Earth and Space Science, 8*(5), e2020EA001614. https://doi.org/10.1029/2020EA001614

Zhou, Q., Jiang, H., Wang, J., & Zhou, J. (2014). A hybrid model for PM2.5 forecasting based on ensemble empirical mode decomposition and a general regression neural network. *Science of the Total Environment, 496*, 264–274. https://doi.org/10.1016/j.scitotenv.2014.07.051

Zhu, M., & Xie, J. (2023). Investigation of nearby monitoring station for hourly PM2.5 forecasting using parallel multiinput 1D-CNN-biLSTM. *Expert Systems with Applications, 211*, 118707. https://doi.org/10.1016/j.eswa.2022.118707