

## ARTICLE OPEN



# Regional aerosol forecasts based on deep learning and numerical weather prediction

Yulu Qiu<sup>1,2,3</sup>, Jin Feng<sup>1,4</sup>✉, Ziyin Zhang<sup>1,4</sup>, Xiujuan Zhao<sup>1,4</sup>, Ziming Li<sup>3</sup>, Zhiqiang Ma<sup>1,4</sup>, Ruijin Liu<sup>1,4</sup> and Jia Zhu<sup>2</sup>

Atmospheric chemistry transport models have been extensively applied in aerosol forecasts over recent decades, whereas they are facing challenges from uncertainties in emission rates, meteorological data, and over-simplified chemical parameterizations. Here, we developed a spatial-temporal deep learning framework, named PPN (Pollution-Predicting Net for PM<sub>2.5</sub>), to accurately and efficiently predict regional PM<sub>2.5</sub> concentrations. It has an encoder-decoder architecture and combines the preceding PM<sub>2.5</sub> observations and numerical weather prediction. Besides, the model proposes a weighted loss function to promote the forecasting performance in extreme events. We applied the proposed model to forecast 3-day PM<sub>2.5</sub> concentrations over the Beijing-Tianjin-Hebei region in China on a three-hour-by-three-hour basis. Overall, the model showed good performance with  $R^2$  and RMSE values of 0.7 and 17.7  $\mu\text{g m}^{-3}$ , respectively. It could capture the high PM<sub>2.5</sub> concentration in the south and relatively low concentration in the north and exhibit better performance within the next 24 h. The use of the weighted loss function decreased the level of “high values underestimation, low values overestimation”, while incorporating the preceding PM<sub>2.5</sub> observations into the encoder phase improved the predictive accuracy within 24 h. We also compared the model result with that from a state-of-the-art numerical model (WRF-Chem with pollutant data assimilation). The temporal  $R^2$  and RMSE from the WRF-Chem were 0.30–0.77 and 19–45  $\mu\text{g m}^{-3}$  while those from the PPN model were 0.42–0.84 and 15–42  $\mu\text{g m}^{-3}$ . The proposed model shows powerful capacity in aerosol forecasts and provides an efficient and accurate tool for early warning and management of regional pollution events.

npj Climate and Atmospheric Science (2023)6:71 | <https://doi.org/10.1038/s41612-023-00397-0>

## INTRODUCTION

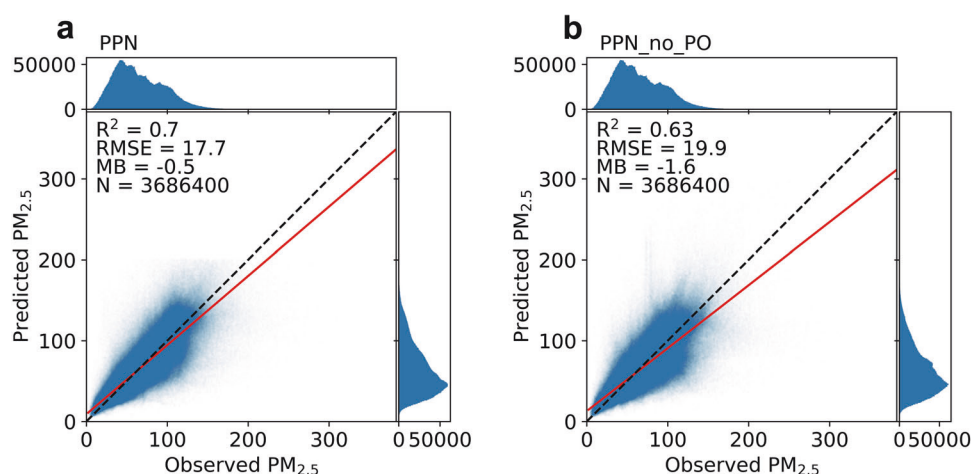
Fine particulate matter (PM<sub>2.5</sub>, with an aerodynamic diameter of <2.5  $\mu\text{m}$ ) has been identified as a key component of atmospheric pollution in China due to large amounts of precursor emissions and unfavorable meteorological conditions<sup>1</sup>, especially during cold seasons. Inhaling PM<sub>2.5</sub> of high concentration poses a serious threat to public health, leading to increased risks of cardiovascular and respiratory diseases<sup>2</sup>. Therefore, accurate prediction of PM<sub>2.5</sub> concentrations is indispensable both for public health warnings and emission controls.

The main approaches of PM<sub>2.5</sub> prediction can be divided in two categories: i.e., numerical models and statistical models. Numerical models, such as the WRF-Chem<sup>3</sup>, CMAQ<sup>4</sup>, and GEOS-Chem<sup>5</sup>, describe the essential atmospheric physical and chemical processes through mathematical formulas, which establish the explicable relationship among atmospheric components, meteorological parameters, and emissions in spatial and temporal dimensions. These numerical models, termed Chemistry Transport Models (CTMs), have been extensively used in PM<sub>2.5</sub> prediction<sup>6,7</sup>, source apportionment<sup>8–10</sup> and mechanism analysis<sup>11–13</sup> worldwide. However, large uncertainties arise from emission rates<sup>14</sup>, meteorological data<sup>15</sup>, as well as the over-simplified chemical parameterizations<sup>16</sup>, may result in certain deviations from the observed PM<sub>2.5</sub> concentrations. To reduce model uncertainties, data assimilation has been extensively utilized to provide more precise initial chemical conditions by incorporating in-situ pollutant observations and satellite retrievals into CTMs<sup>17–21</sup>. It showed beneficial effects on PM<sub>2.5</sub> forecasts for up to 24 h while the benefit from data assimilation diminished rapidly with forecast

range<sup>17,21</sup>. In addition, low computational efficiency is another disadvantage of the CTMs, which usually take several hours for one simulation day with fine resolution<sup>22</sup>.

Statistical methods essentially establish the relationship between multiple predictors and historically observed PM<sub>2.5</sub> concentrations through regression models or machine learning (ML) algorithms, which do not involve the complex physical and chemical processes and usually have less demand for computing resources compared to the CTMs. Initially, linear regression models were applied to PM<sub>2.5</sub> forecast and showed reasonable accuracy<sup>23</sup>. However, the linear assumption in these models sometimes might not precisely capture the connection between predictors and air pollutant concentrations. In recent years, ML algorithms have gained popularity in air quality forecast due to their strong power in effectively handling nonlinear relationship between predictors and targets<sup>24</sup>. For example, Ma et al.<sup>25</sup> utilized the XGBoost algorithm to predict PM<sub>2.5</sub> concentrations in Shanghai and reported a correlation coefficient ( $R$ ) of 0.77 and Root Mean Square Error (RMSE) of about 12–18  $\mu\text{g m}^{-3}$  for 24 h ahead prediction. Bi et al.<sup>26</sup> developed a PM<sub>2.5</sub> forecast system by combining the random forest algorithm with CTM model results in central China, which had  $R^2$  of 0.76 and 0.64 for the next 2 days. These ML methods show good capacities in capturing non-linear relationships between features and aerosol concentrations at single sites. However, air pollutant forecasting is usually a regional issue related to both spatial relations and time sequences. The traditional ML methods are still difficult to resolve the complex spatiotemporal correlations<sup>27</sup>.

<sup>1</sup>Institute of Urban Meteorology, China Meteorological Administration, Beijing 100089, China. <sup>2</sup>Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control (AEMPC), School of Environmental Science and Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China. <sup>3</sup>Beijing Weather Forecast Center, Beijing 100089, China. <sup>4</sup>Key Laboratory of Urban Meteorology, China Meteorological Administration, Beijing 100089, China. ✉email: jfeng@ium.cn



**Fig. 1 Three-hourly scatterplots and probability distributions of observed and predicted  $PM_{2.5}$  concentrations over the BTH region during January 2022.** The initial forecasting time is 00 UTC every day. The red lines denote the linear fitting lines, and the dash lines are  $y = x$  reference lines. **a** Comparison by PPN results. **b** Comparison by PPN\_no\_PO results (The PPN\_no\_PO will be specifically explained in “Impacts of preceding observation restraint”).

Nowadays, deep learning (DL) networks show remarkable capabilities in forecasting air pollutant variations<sup>27,28</sup> due to their successful applications in dealing with nonlinear spatiotemporal correlations and advances in computing resources. DL is based on multiple-layer neural networks. A kind of the representative models is Convolutional Neural Network (CNN) - Recurrent Neural Network (RNN) architectures in which CNNs extract features and spatial relations<sup>29,30</sup> while RNNs handle the temporal dependencies<sup>31,32</sup>. Long Short-Term Memory (LSTM) is a type of RNN specially designed for overcoming gradient vanishing and exploding when managing long-term dependencies<sup>33</sup>, which has been extensively employed in spatial-temporal forecasting of air pollutant<sup>30,34–37</sup>. For example, Yan et al.<sup>30</sup> built a hybrid DL network (CNN-LSTM) to predict  $PM_{2.5}$  concentrations in next 6 h in Beijing based on historical  $PM_{2.5}$  data, meteorological indicators, and spatiotemporal data. Pak et al.<sup>35</sup> combined the CNN-LSTM model with a spatiotemporal feature vector to reflect the relations among parameters to forecast the next day's daily  $PM_{2.5}$  concentrations in Beijing and outperformed the traditional CNN-LSTM method. Yeo et al.<sup>36</sup> integrated the gated recurrent unit (GRU) algorithm, a similar architecture as LSTM but has fewer parameters, with a CNN accounting for the geographical correlation of nearby stations. Using the new method, they improved the 24-h  $PM_{2.5}$  prediction in Seoul by about 10%. Overall, the DL-based model is a promising and efficient tool for  $PM_{2.5}$  forecasts. However, the proposed DL models in above studies are usually for forecasting air pollution concentrations in next day on urban scale, which may not fully meet the requirements in practical applications.

In contrast to previous DL architectures, we develop a more advanced spatial-temporal DL model for short-range (0–72 h)  $PM_{2.5}$  forecast on regional scale, named air Pollution-Predicting Net for  $PM_{2.5}$  (PPN). Our model injects the feature variables in different convolutional layers in terms of their impacts on  $PM_{2.5}$ , to imitate the behavior of CTMs. In addition, the observed  $PM_{2.5}$  concentrations over multiple preceding time-steps are also included to provide accurate initial field of  $PM_{2.5}$  forecast, like the Four-Dimensional Data Assimilation (FDDA) in CTMs. Thus, the PPN model integrates the strengths of DL network, CTMs and assimilation and is expected to achieve better performance in forecasting regional  $PM_{2.5}$  concentration with wide forecast time range. We apply the PPN model for forecasting hourly  $PM_{2.5}$  concentrations over a highly populated and industrialized region in China. Firstly, model performance for  $PM_{2.5}$  forecasts in both temporal and spatial distributions is evaluated, which is followed

by impacts of weighted loss function and preceding observed  $PM_{2.5}$ . Besides, we also compare the PPN model performance with the WRF-Chem results. Finally, model structure and data information are displayed.

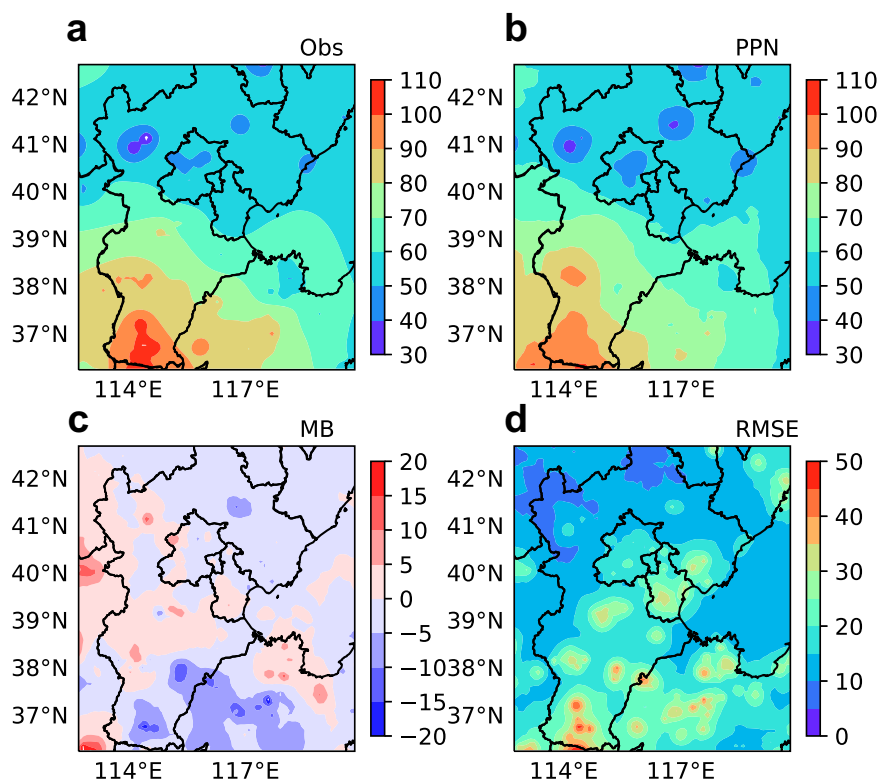
## RESULTS

### Overall performance

Using the trained PPN model and predictors, we forecasted the  $PM_{2.5}$  concentrations in January 2022 at every 9 km × 9 km grid cell over the Beijing-Tianjin-Hebei (BTH) region. Figure 1a displays the scatter plot of predicted and observed  $PM_{2.5}$  concentrations for January 2022 with the initial forecast time of UTC 00 every day. As we see, the  $R^2$  and RMSE values are 0.70 and 17.7  $\mu g m^{-3}$ , respectively. The result for June 2022 when  $PM_{2.5}$  pollution was light is also shown in Supplementary Fig. 1, with  $R^2$  and RMSE values of 0.49 and 6.9  $\mu g m^{-3}$ . The lower  $R^2$  and RMSE in comparison to that in winter may be a result of low  $PM_{2.5}$  concentration in June with small fluctuation. There is little difference among forecast results from different initial forecast time (Supplementary Table 1). The fairly good performance achieved by the PPN model suggests that it is suitable for forecasting  $PM_{2.5}$  concentrations. Aerosol pollution is usually severe in winter due to stagnation and intensive emissions<sup>38</sup>, so we mainly focus on model evaluation for the results in January 2022 in the following.

Figure 2 shows the spatial distributions of observed and forecasted  $PM_{2.5}$  concentrations. It can be found that the heavily polluted region is in the south with average concentration exceeding 100  $\mu g m^{-3}$ . The spatial pattern of  $PM_{2.5}$  over the BTH region is predominantly related to emission, meteorological and orographic factors<sup>39</sup>. The southern region is more industrialized than the northern region, resulting in higher emission rates of  $PM_{2.5}$  and its precursors. North winds can blow away haze, leading to more clean air in the north region, while south winds facilitate pollutant accumulation along mountains in the south. These determinant factors are all considered in the model training process. Therefore, the trained model can accurately capture the spatial pattern of  $PM_{2.5}$  with higher values in the south and lower values in the north over the BTH region.

It should be noted that RMSEs exhibit higher values in the southern region (up to 40–50  $\mu g m^{-3}$ ) and lower values in the north (about 10–30  $\mu g m^{-3}$ ) (Fig. 2c). Although higher base  $PM_{2.5}$  concentration in the south could explain part of it, the under-predicted  $PM_{2.5}$  with MB values of −15–−5  $\mu g m^{-3}$  in the south



**Fig. 2** Spatial performance of the PPN model. **a, b** Comparisons of the spatial distribution of observed and predicted  $\text{PM}_{2.5}$  concentrations by the PPN model over the BTH region during January 2022. **c, d** Corresponding MB (mean bias) and RMSE values. The initial forecasting time is 00 UTC every day.

may be the primary cause of high RMSEs. We further analyze the temporal variations of biases over three representative cities (Beijing, Shijiazhuang and Handan) from north to south (Supplementary Fig. 2). It is found that large biases in Handan, a southern city, mainly occurred around January 12 and 25 when cold air was moving from the north to the south (Supplementary Fig. 3). It dispelled the haze in the north whereas blew pollutants to the southern region. With regards to the southern cities, the “polluted” north winds are beneficial to increasing  $\text{PM}_{2.5}$  concentrations. However, the PPN model still has difficulty in capturing  $\text{PM}_{2.5}$  that is rapidly transported from a long distance. This deficiency may be caused by the consideration of a small range of transport process ( $45 \times 45$  km) in a timestep with limited computing resources.

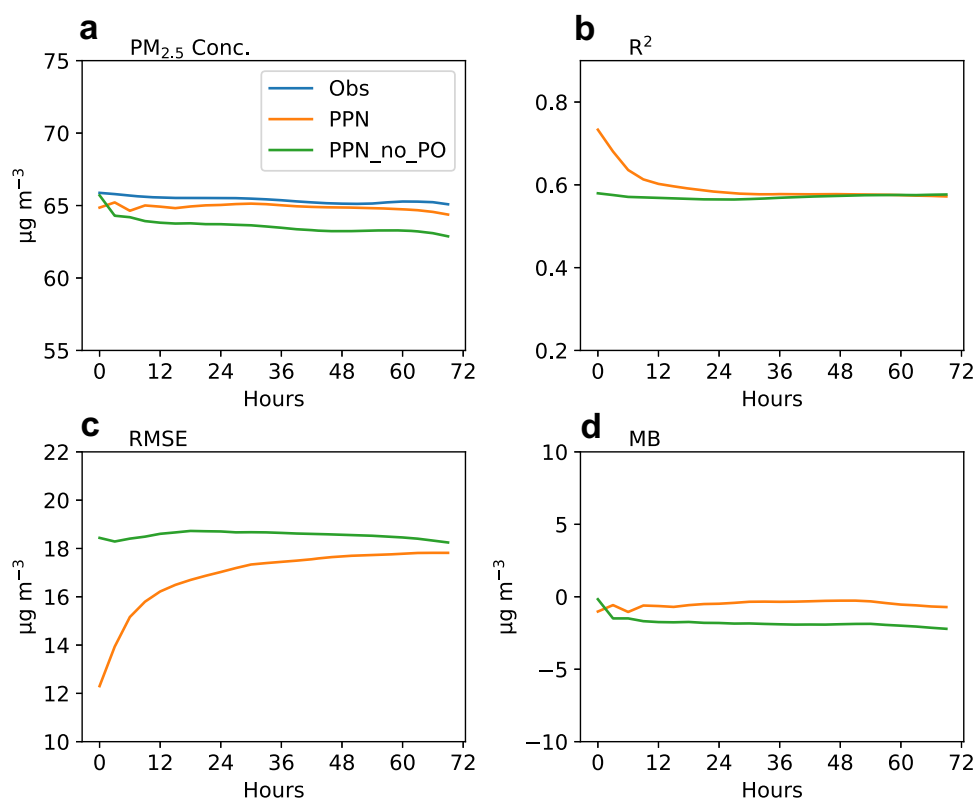
We also compare observations with predicted  $\text{PM}_{2.5}$  concentrations for 0–72 h in advance, as shown in Fig. 3 (orange lines). The predicted mean  $\text{PM}_{2.5}$  concentrations are about  $65 \mu\text{g m}^{-3}$  over the whole region, which are slightly lower than observations (Figs. 3a and d). The PPN results present decreased  $R^2$  and increased RMSE with forecast lead time, and the space-average values of 0.57–0.74 and 12–18  $\mu\text{g m}^{-3}$ , respectively. It is worth noting that the model performance gets worse with forecast lead time. This downward trend is particularly obvious during the 0–24 h and becomes stable after 24 h. As shown in Fig. 3b, the spatial  $R^2$  is about 0.74 at the initial forecast time, then rapidly reduced to 0.58 at the forecast lead time of 24 h. The RMSE value is about 12  $\mu\text{g m}^{-3}$  in the beginning, then decreases to 17  $\mu\text{g m}^{-3}$  at 24 h (Fig. 3c). The model performance levels off during the forecast lead time of 24–72 h with  $R^2$  of about 0.57 and RMSEs of 17–18  $\mu\text{g m}^{-3}$ , which is still an acceptable result. In a word, the model exhibits better performance at 0–24 h and achieves stable and reasonable prediction capacity when the forecast time extends to 3 days.

As mentioned in “Introduction”, ML algorithms have been extensively applied in air quality forecast due to strong capacity in

effectively dealing with nonlinear relationship. Here, we compare the PPN model results with the other three models based on Random Forest, XGBoost and Multilayer Perceptron algorithms, called RF, XGB and MLP for short, respectively. Random Forest and XGBoost are both ensemble learning algorithms based on decision trees for classification and regression, and have been employed in pollutant dataset construction<sup>40</sup> and air quality forecast<sup>25,26</sup>. Multilayer Perceptron is a basic algorithm for neural networks. In order to emphasize the advantage of the deep network of PPN, we employed a shallow network of the MLP model for comparison. The MLP model has 2 hidden layers of 32 and 8 neurons. We trained the RF, XGB and MLP models using the same features and time periods as those in the PPN model. In addition, each grid over the BTH region was utilized in the model training process, which do not fully consider the spatiotemporal correlations between the predictors and the target. The comparison results show that our PPN model has a prediction performance superior to the other three models in forecasting  $\text{PM}_{2.5}$  variations over 13 cities during January 2022 (Supplementary Fig. 4 and Fig. 5). This is highly related to the utilization of multi-convolution layers to capture the spatial relationship among grid cells and LSTM layers for temporal variations, as well as the introduction of preceding observation restraint and weighted loss function that will be elaborated in the following.

### Impacts of weighted loss function

From the above, we can conclude that the proposed PPN model can well capture the spatiotemporal variability of  $\text{PM}_{2.5}$  concentration over the BTH region for the next 3 days. Interpolated  $\text{PM}_{2.5}$  data based on site observations were applied as fitting targets to acquire the gridded results. However, the interpolation may introduce biases into the target dataset, especially in regions with sparsely or unevenly distributed sites. In this study, we introduced



**Fig. 3** Temporal performance of the PPN and PPN\_no\_PO model. Temporal variations of (a) observed and predicted  $\text{PM}_{2.5}$  concentrations by PPN and PPN\_no\_PO during 0–72 h forecast period for the test dataset (January 2022); (b)  $R^2$ ; (c) RMSE; (d) MB.

a weighted loss function in the model training process to eliminate the adverse effects of  $\text{PM}_{2.5}$  interpolation. Detailed information about the weighted loss function can be found in “Methods” section at the end. The influence of the weighted loss function is evaluated in this section to illustrate the innovation of the PPN model. To achieve this, we conducted a comparative experiment that uses Mean Square Error (MSE) as loss function and do not consider the Inverse Distance Weighted (IDW)-based loss function adopted in the PPN model, called PPN\_no\_IDW for short. The comparison between PPN and PPN\_no\_IDW results under different  $\text{PM}_{2.5}$  levels is shown in Fig. 4. The  $\text{PM}_{2.5}$  classification is based on the Technical Regulation on Ambient Air Quality Index (HJ 633–2012). Here, MB is utilized as the evaluation metric instead of RMSE, to better differentiate between overprediction and underprediction. The comparison is based on results from the 114 monitoring sites over the BTH where direct observations are available.

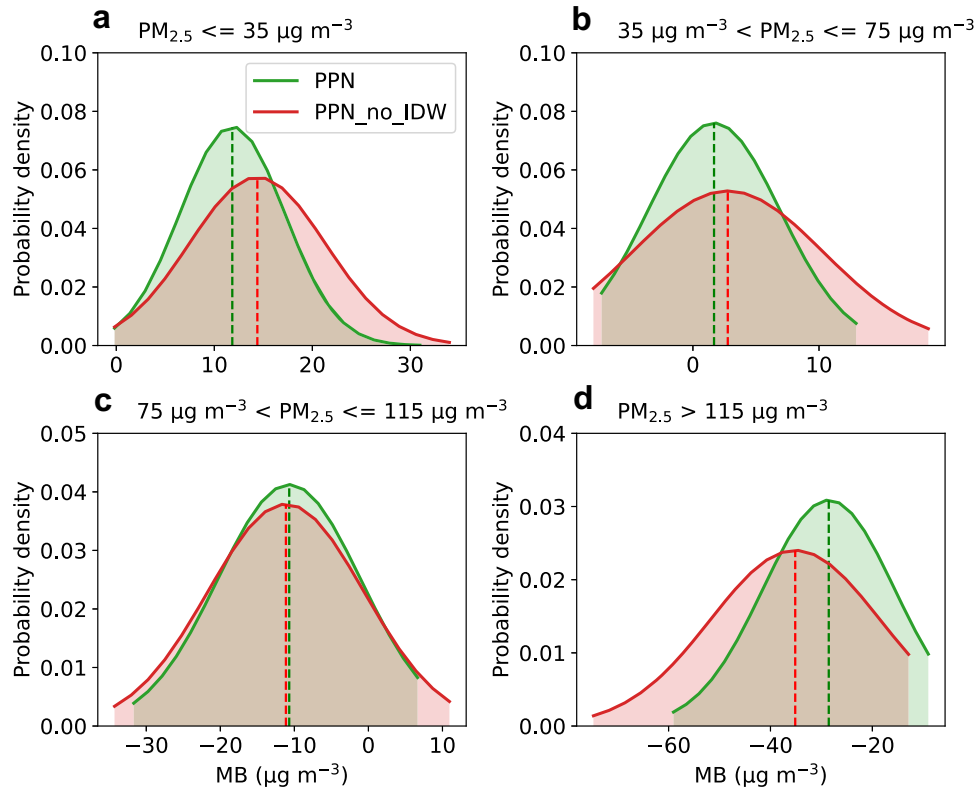
From Fig. 4, we find that the PPN model tends to overpredict  $\text{PM}_{2.5}$  concentrations under clean condition ( $\text{PM}_{2.5}$  concentration  $\leq 35 \mu\text{g m}^{-3}$ ) while underprediction occurs when the pollution level aggravates. The “high values underestimation, low values overestimation” phenomenon was also previously reported by Mao et al.<sup>41</sup>, suggesting the difficulty of capturing extreme values by deep learning algorithms at present. By comparison, the PPN model with weighted loss function outperforms the PPN\_no\_IDW and decreases the level of “high values underestimation, low values overestimation”. The improvement by the PPN is significant when  $\text{PM}_{2.5}$  concentration is  $\leq 35 \mu\text{g m}^{-3}$  (clean condition) and  $> 115 \mu\text{g m}^{-3}$  (moderate or heavy polluted condition). The probability of MB values  $> 15 \mu\text{g m}^{-3}$  from the PPN model under clean condition is much less than that from the PPN\_no\_IDW (Fig. 4a). Under moderate or heavy polluted condition, the probability of large bias is also

reduced by the PPN model. The average MB value by PPN is  $-28 \mu\text{g m}^{-3}$ , whereas that from the PPN\_no\_IDW is  $-35 \mu\text{g m}^{-3}$  (Fig. 4d). The improvement by the PPN model is less significant when  $\text{PM}_{2.5}$  concentration is  $\leq 115 \mu\text{g m}^{-3}$  and  $> 35 \mu\text{g m}^{-3}$  (Fig. 4b, c). Overall, the PPN model, which considers more information of in-situ observations in the loss function, can achieve higher prediction accuracy at the monitoring sites.

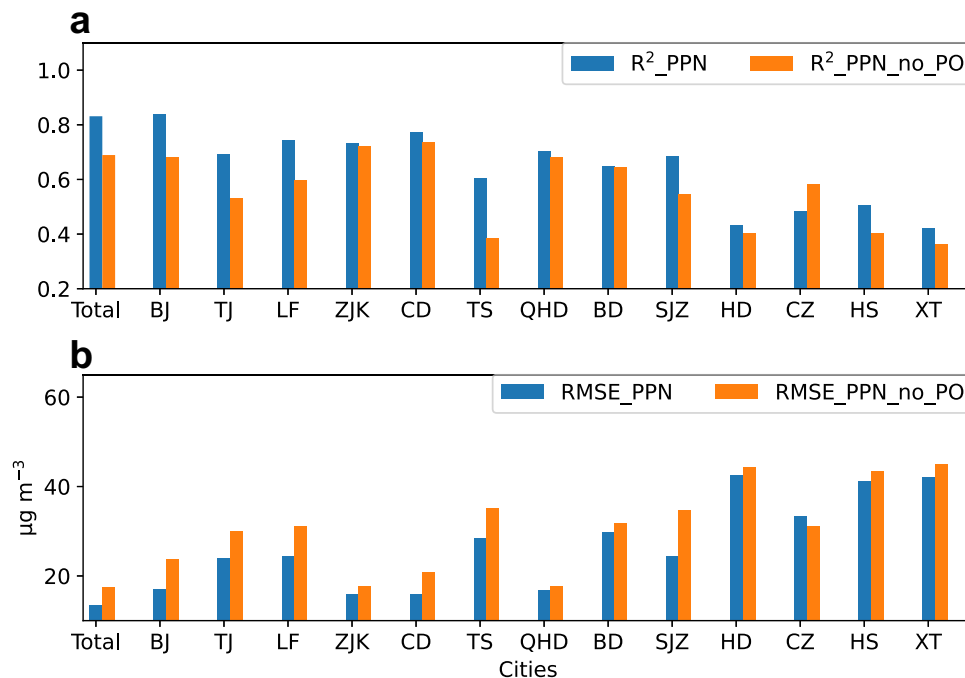
### Impacts of preceding observation restraint

To imitate the FDDA behavior for CTMs, another key feature of the PPN model is to consider the preceding  $\text{PM}_{2.5}$  observations within the encoder phase. To verify the significance of the observation restraint, our study also established a model similar to the PPN framework but without the Preceding Observation restraint in the encoder phase (called PPN\_no\_PO). Figure 1b shows the scatterplot of observed and predicted  $\text{PM}_{2.5}$  concentrations by PPN\_no\_PO for the test dataset. Compared to the PPN results, the PPN\_no\_PO model shows poorer performance with a lower  $R^2$ -value of 0.63 and larger RMSE value of  $19.9 \mu\text{g m}^{-3}$ . Figure 2 also displays the predicted  $\text{PM}_{2.5}$  concentration by the PPN\_no\_PO for multi-step ahead prediction (green lines) and the comparison with the standard PPN result. The spatial  $R^2$  and RMSE values from the PPN\_no\_PO result remain at about 0.57 and  $18 \mu\text{g m}^{-3}$  while those from the PPN model are 0.57–0.74 and  $12$ – $18 \mu\text{g m}^{-3}$ , respectively. The better performance of the PPN model is readily evident during the 0–24 h forecast lead time. It highlights the importance of considering the preceding  $\text{PM}_{2.5}$  concentration in the encoder phase and indicates that the proposed PPN model is more advanced in  $\text{PM}_{2.5}$  forecast than traditional DL methods.

We also compare the PPN and PPN\_no\_PO results over the monitoring sites since most of these sites are close to human settlements and susceptible to attracting public attention. The predicted temporal characteristics of  $\text{PM}_{2.5}$  by PPN and



**Fig. 4** Probability density of MB values from PPN and PPN\_no\_IDW results. The results are for the testing dataset (January 2022) under different  $PM_{2.5}$  concentration levels. The dash lines denote the corresponding average values. The MB values are derived from results at the 114 monitoring sites over the BTH region. **a**  $PM_{2.5} \leq 35 \mu g m^{-3}$ ; **b**  $35 \mu g m^{-3} < PM_{2.5} \leq 75 \mu g m^{-3}$ ; **c**  $75 \mu g m^{-3} < PM_{2.5} \leq 115 \mu g m^{-3}$ ; **d**  $PM_{2.5} > 115 \mu g m^{-3}$ .



**Fig. 5** Evaluation metrics for predicted  $PM_{2.5}$  concentrations by PPN and PPN\_no\_PO. The results are derived from 13 cities of the BTH region during January 2022. The comparisons are based on results in 0–24 h forecast lead time when the difference between PPN and PPN\_no\_PO is most prominent. **a**  $R^2$ , **b** RMSE.



PPN\_no\_PO over 13 cities of the BTH region are shown in Supplementary Fig. 6. The corresponding evaluation metrics are shown in Fig. 5. The comparisons in Supplementary Fig. 6 and Fig. 5 are based on results in 0–24 h forecast lead time when the difference between PPN and PPN\_no\_PO is most prominent. In general, the PPN-predicted  $\text{PM}_{2.5}$  concentrations are in good agreement with the observations at the monitoring sites. It well captured two regional pollution events during Jan 6–12 and Jan 22–26. The values of  $R^2$  and RMSE are 0.42–0.84 and 15–42  $\mu\text{g m}^{-3}$ , respectively. Similar to the previous results, the PPN predicted  $\text{PM}_{2.5}$  in the northern cities (BJ, TJ, LF, ZJK, CD, TS, QHD) performs better than those over the southern region (BD, SJZ, HD, CZ, HS). By comparison, the PPN architecture exhibits better performance than PPN\_no\_PO in forecasting  $\text{PM}_{2.5}$  concentration over most monitoring sites with larger  $R^2$  and smaller RMSE (Fig. 5). The performance improvement is more significant in the northern cities, such as BJ, TJ, LF and TS, with RMSE values decreased by 20–28% compared with those from PPN\_no\_PO. However, the reduction ratios are just 3–6% over the southern cities (HD, XT, HS).

### Comparison with the WRF-Chem results

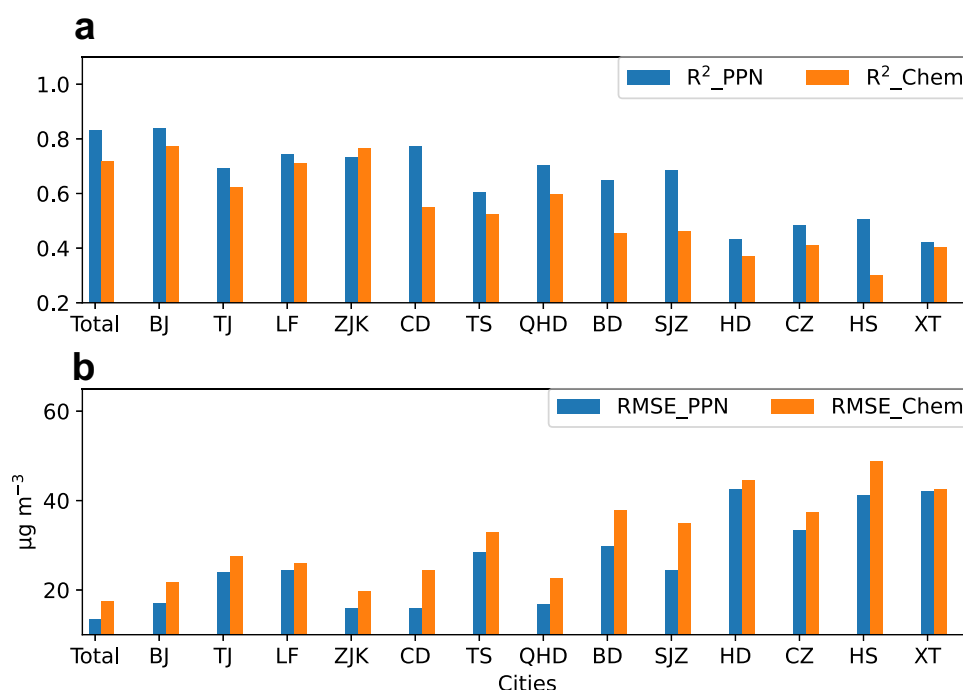
To further evaluate the performance of the PPN model in forecasting  $\text{PM}_{2.5}$  concentrations, we also compare it to the state-of-the-art CTM WRF-Chem model. In contrast to the PPN model, the WRF-Chem model integrated with data assimilation is built on the numerical methods that fully takes into account the atmospheric physical and chemical processes. It is a more interpretable tool for air quality forecast. Supplementary Fig. 7 shows the time series of predicted  $\text{PM}_{2.5}$  by PPN and WRF-Chem over monitoring sites during January 2022, and the corresponding evaluation indices are displayed in Fig. 6. The temporal  $R^2$  and RMSE values from the WRF-Chem are 0.30–0.77 and 19–45  $\mu\text{g m}^{-3}$  while those from the PPN model are 0.42–0.84 and 15–42  $\mu\text{g m}^{-3}$ . Among 13 cities over the BTH, the RMSEs from the PPN model are 1–35% lower than those from the WRF-Chem. The reduction ratios in 10 cities are >10%. The PPN forecast result in June 2022 is also superior to the WRF-Chem result with larger  $R^2$

and lower RMSE (Supplementary Figs. 8 and 9). Therefore, we conclude that the PPN model outperforms the WRF-Chem. Moreover, we also evaluate the model capacity in forecasting  $\text{PM}_{2.5}$  concentration at different forecast lead times over the monitoring sites, as illustrated in Fig. 7. It is noted that the average RMSE from the PPN model is 25  $\mu\text{g m}^{-3}$  at 0–24 h which is 17% lower than that from the WRF-Chem (30  $\mu\text{g m}^{-3}$ ) (Fig. 7a). The reduction ratios slightly decrease to 14–16% with increased forecast lead time (24–72 h). The average  $R^2$ -values of the PPN model are 0.61–0.68, which are higher than those of the WRF-Chem model (0.53–0.57) (Fig. 7d). No matter what the forecast lead time is, the PPN model outperforms the WRF-Chem in forecasting  $\text{PM}_{2.5}$  variations. We also find that the PPN model exhibits better predictive capacity particularly under clean and good air quality conditions ( $\text{PM}_{2.5} < 75 \mu\text{g m}^{-3}$ ), up to 25% lower than the RMSE from WRF-Chem. However, the maximum reduction ratio on polluted days is just 13%.

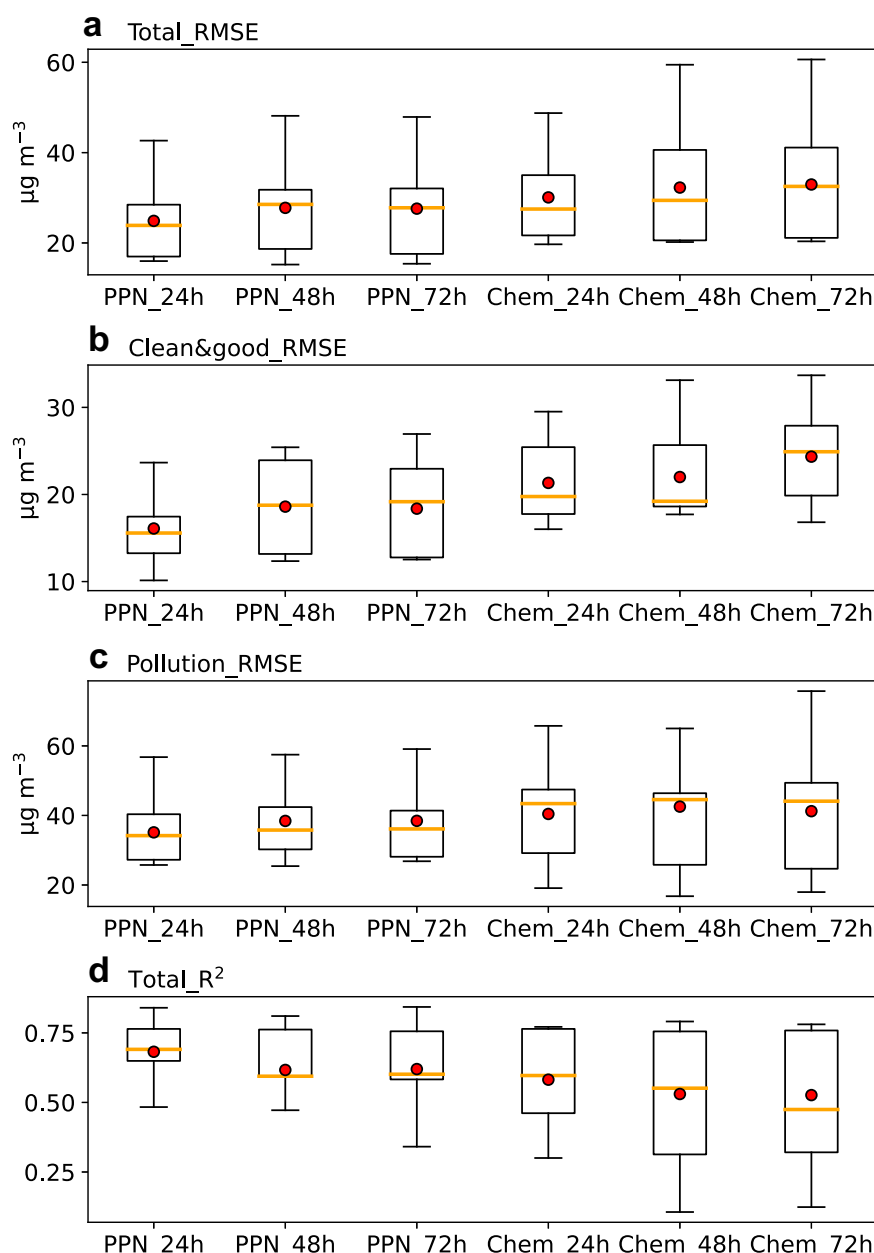
### SUMMARY AND DISCUSSION

In this study, we developed a new spatial-temporal PPN model for short-range  $\text{PM}_{2.5}$  forecasting and applied it to the BTH region. The model separated the input features into “local” or “non-local” layers according to their direct impacts on  $\text{PM}_{2.5}$  to imitate the behavior of CTMs. We trained and validated the PPN model with meteorological, emission and derived data during 2020 and 2021, and utilized the trained model to predict  $\text{PM}_{2.5}$  variations in January 2022, which is served as the testing dataset. The predictive ability of the model was evaluated on a test dataset with observations.

Overall, the proposed model exhibits good accuracy in predicting spatial and temporal variations of  $\text{PM}_{2.5}$  concentrations over the BTH, with  $R^2$  and RMSE values of 0.70 and 17.7  $\mu\text{g m}^{-3}$ , respectively. Regarding spatial characteristics, the model successfully captures the high  $\text{PM}_{2.5}$  concentrations in the south and relatively low concentrations in the north. The model shows better performance in the next 24 h with spatial  $R^2$  and RMSE of 0.58–0.74 and 12–17  $\mu\text{g m}^{-3}$ , respectively. Then the performance



**Fig. 6** Evaluation metrics for predicted  $\text{PM}_{2.5}$  concentration by PPN and WRF-Chem. The results are derived from 13 cities of the BTH region during January 2022 with the forecast lead time of 0–24 h. **a**  $R^2$ , **b** RMSE.



**Fig. 7** Box plots of RMSEs and  $R^2$  from PPN and WRF-Chem. The results are at 0–24 h, 24–48 h and 48–72 h forecast lead time over 13 cities on the test dataset (January 2022). The red dots, orange lines denote mean and medium values, and boxes and whiskers represent 25%–75 and 10–90% percentiles. **a** RMSEs for total test dataset; **b** RMSEs for clean and good conditions ( $PM_{2.5} \leq 75 \mu g m^{-3}$ ); **c** RMSEs for pollution conditions ( $PM_{2.5} > 75 \mu g m^{-3}$ ); **d**  $R^2$  for total test dataset.

gets stable as the forecast lead time increases. In a word, the model has good capacity in forecasting spatial and temporal variability of  $PM_{2.5}$  over the BTH.

The consideration of the weighted loss function and preceding  $PM_{2.5}$  observation in the encoder phase are two innovative points in the model structure. To further explore the advantage of the PPN model, we conducted two sensitivity tests, in which the aforesaid two points are not taken into account. Results show that the PPN model with weighted loss function outperforms the PPN\_no\_IDW and decreases the level of “high values underestimation, low values overestimation”. Considering the preceding  $PM_{2.5}$  concentration into encoder phase greatly improves the results in the next 24 h with  $R^2$  increased from 0.57 to 0.58–0.74 and RMSE decreased from  $18 \mu g m^{-3}$  to  $12$ – $17 \mu g m^{-3}$ . This improvement is also evident over the monitoring stations,

especially over the northern cities. We also compare the model performance with the state-of-the-art WRF-Chem results. By comparison, the PPN model shows better predictive accuracy than WRF-Chem. The temporal  $R^2$  and RMSE values from the WRF-Chem are 0.30–0.77 and  $19$ – $45 \mu g m^{-3}$  while those from the DL model are 0.42–0.84 and  $15$ – $42 \mu g m^{-3}$ . This better performance exists within all the forecasting lead times.

Spatial-temporal DL algorithms exhibit powerful capacity in dealing with non-linear correlations, making them suitable for air quality forecast. Several hybrid DL networks (CNN-LSTM and Graph Convolutional-LSTM) have already been successfully applied in  $PM_{2.5}$  forecast in winter Beijing with RMSE values of about  $22$ – $53 \mu g m^{-3}$  in the first 6 h<sup>30</sup> and  $24 \mu g m^{-3}$  for next 24 h<sup>42</sup>, respectively. Unlike the PPN model, these two results are based on a single or several surrounding sites in Beijing. The

corresponding RMSEs from PPN are  $11\text{--}17\text{ }\mu\text{g m}^{-3}$ . Although the comparison is based on different forecast period, the better performance of PPN can still demonstrate the advanced network structure that imitates the behavior of CTMs with inclusion of the preceding  $\text{PM}_{2.5}$  observations and weighted loss function. In summary, the PPN model demonstrates strong potential for the application of spatiotemporal  $\text{PM}_{2.5}$  forecast over the BTH region, and it is an efficient and accurate tool for regional  $\text{PM}_{2.5}$  forecast.

The application of the PPN model could be extended to other components of air pollution, e.g., ozone and nitrogen oxides, with their related meteorological data and emission data as the input parameters. Moreover, it should be noted that we designed the PPN model for short-range air quality forecasts. For medium-range (over 5 days) forecasts on a day-to-day basis, spatial information would be more important than short-range one due to long-distance transport. Therefore, there would be a different DL model for the medium-range air quality forecasts.

## METHODS

### Model structure

PPN has an encoder-decoder architecture and uses the PredRNN<sup>43</sup> as the backbone network for capturing the spatiotemporal variation for  $\text{PM}_{2.5}$  concentrations. The PredRNN uses multi convolution layers to capture the spatial relationship among grid cells and LSTM layers for temporal variations. It is an updated version of ConvLSTM<sup>44</sup>, which adds a shortcut connecting the last convolutional layer and the first convolutional layer between adjacent timesteps. The model structure is shown in Fig. 8.

Spatially, the model is designed following the order of aerosol-related processes in CTMs. There is one convolutional layer with the convolutional kernel sizes of  $1 \times 1$  ( $9 \times 9$  km), representing the local processes (i.e., chemical and turbulence diffusion processes). Two non-local layers follow, representing the short-distance transport and long-distance transport, with  $3 \times 3$  ( $27 \times 27$  km) and  $5 \times 5$  ( $45 \times 45$  km) respectively. Rather than considering all input features in the same layer as the PredRNN, the PPN framework firstly separate the feature variables into two parts, i.e., local variables and non-local variables (Table 1), according to their main physical-chemical effects on aerosol. The local variables including emissions, temperature, humidity, precipitation, sea level pressure (SLP) and planetary boundary layer height (PBLH),

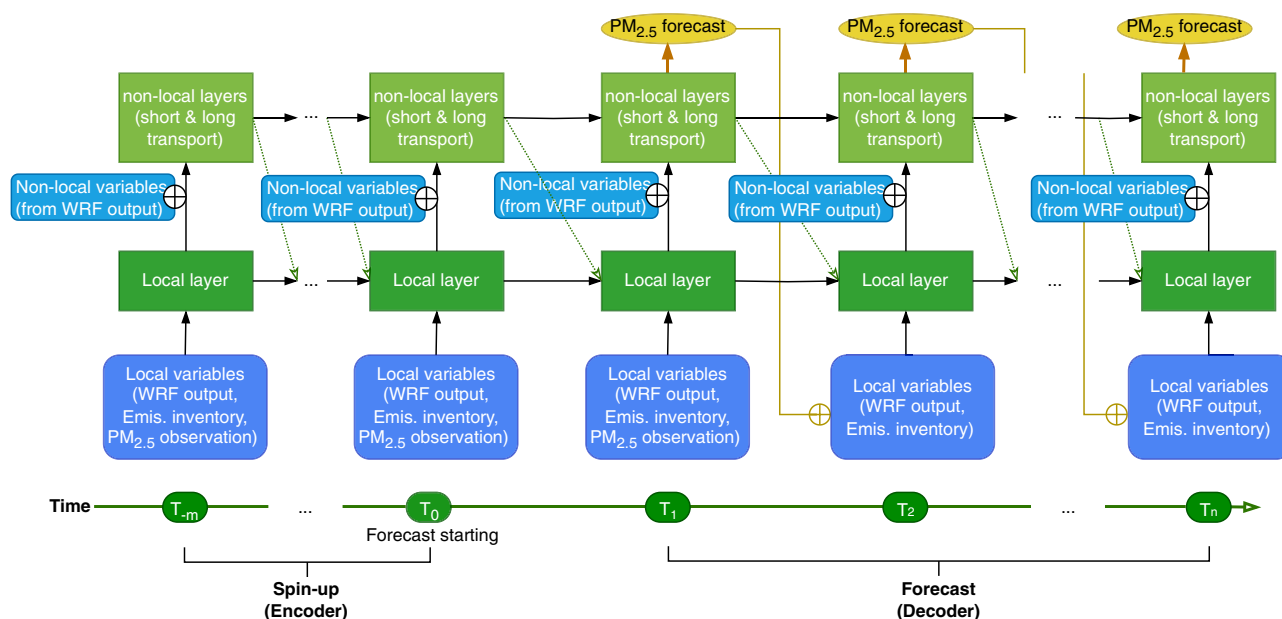
etc., which are more directly to affect local  $\text{PM}_{2.5}$  formation and depletion, thus are gone into the first “local layer”. These variables concerning synoptic patterns and transport process, such as wind, geopotential height at 700 hPa and terrain height, introduce to the PPN model at the beginning of “short transport layer”. It should be noted that now the PPN model only considers a slightly small range of transport process (within  $5 \times 5$  grid cells) in a timestep due to the limitation of our computing resources. Therefore, large-scale inter-regional transport within one timestep may not be well represented by the model. More computing resources that can afford a large size of convolutional kernel could help to solve the problem in the future.

Temporally, the model uses an asymmetric encoder-decoder structure, all the timesteps before the initial time are encoding phase, like the spin-up phase in CTMs and Numerical Weather Prediction (NWP) models. The forecasting timesteps are decoding phase. The outputs of decoder in every timestep are the  $\text{PM}_{2.5}$  forecasts. During the decoding phase, the model inputs the meteorological variables, emissions and  $\text{PM}_{2.5}$  forecast from the last timestep as the feature variables. And in every timestep of the encoding phase, in addition to the above variables, the PPN model added the  $\text{PM}_{2.5}$  observation in the first convolutional layer. Since the LSTM is able to learn and remember the input information for the past period, we expect the PPN to provide a better initial field by adding multi observations across a preceding time period. This idea is similar to a sequential FDDA such as grid nudging<sup>45,46</sup> or Ensemble Kalman Filter<sup>47</sup>.

### Experimental region and feature selection

We took the Beijing-Tianjin-Hebei (BTH) region ( $36.3\text{--}42.0^\circ\text{N}$ ,  $111.8\text{--}121.6^\circ\text{E}$ ) as our experimental region, which is surrounded by mountains in the north and west and consists of megacities like Beijing (BJ) and Tianjin (TJ) (Fig. 9b). The BTH region is a highly industrialized and densely populated region confronting with severe aerosol pollution<sup>48</sup>, thus has strong demands for air quality prediction to formulate emission reduction strategy and alert the public health.

$\text{PM}_{2.5}$  concentration is influenced by both meteorological conditions and precursor emissions. Therefore, we classified the input parameters into three categories: i.e., meteorological variables, emission data and derived parameters, as listed in



**Fig. 8** Structure of PPN model. The PPN model structure with an encoder-decoder architecture. The spin-up timesteps are encoding phase ( $T_{-m}, \dots, T_0$ ) and the forecasting timesteps are decoding phase ( $T_1, \dots, T_n$ ).



**Table 1.** Meteorological data were from the Weather Research and Forecasting (WRF) model simulation that provides a downscaled meteorological field based on the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis data (<https://cds.climate.copernicus.eu/>, last access: 2023/02/27). The WRF model was configured to cover the whole China region (Fig. 9a) following Feng et al.<sup>49</sup>, with a horizontal resolution of 9 km and 38 vertical layers up to 50 hPa. The initial and boundary conditions were obtained from the ERA5 reanalysis data with horizontal resolution of 0.25° × 0.25°. Supplementary Table 2 shows the detailed parameterization schemes used in the WRF simulation.

Parameters	Sources	Layers
<b>Meteorological variables</b>		
Temperature at 2 m (T2)	WRF	Local
Relative humidity at 2 m (RH2)	WRF	Local
Sea level pressure (SLP)	WRF	Local
Precipitation rate (Precp)	WRF	Local
Planetary boundary layer height (PBLH)	WRF	Local
U and V wind at 10 m (U10, V10)	WRF	Non-local
U and V wind at 850hPa (U850, V850)	WRF	Non-local
Wind direction at 10 m (wd10)	WRF	Non-local
Wind speed at 10 m (ws10)	WRF	Non-local
Terrain height (HGT)	WRF	Non-local
Geopotential Height at 700 hPa	WRF	Non-local
<b>Emission variables</b>		
Emissions rates of SO <sub>2</sub> , NO <sub>x</sub> , VOCs and primary PM <sub>2.5</sub>	MEIC inventory	Local
<b>Derived variables</b>		
Changes of SLP compared to 24 h ahead (dSLP)	WRF	Local
Changes of temperature at 850 hPa compared to 24 h ahead (dT850)	WRF	Local
Annual mean PM <sub>2.5</sub> concentration	In-situ observation	Local
Preceding PM <sub>2.5</sub> concentration	In-situ observation	Local

The meteorological variables included factors that relate to synoptic pattern (SLP, geopotential height at 700 hPa), aerosol chemistry (temperature, relative humidity), transport (wind at 10 m and 850hPa, terrain height), vertical diffusion (PBLH), and wet scavenging (precipitation).

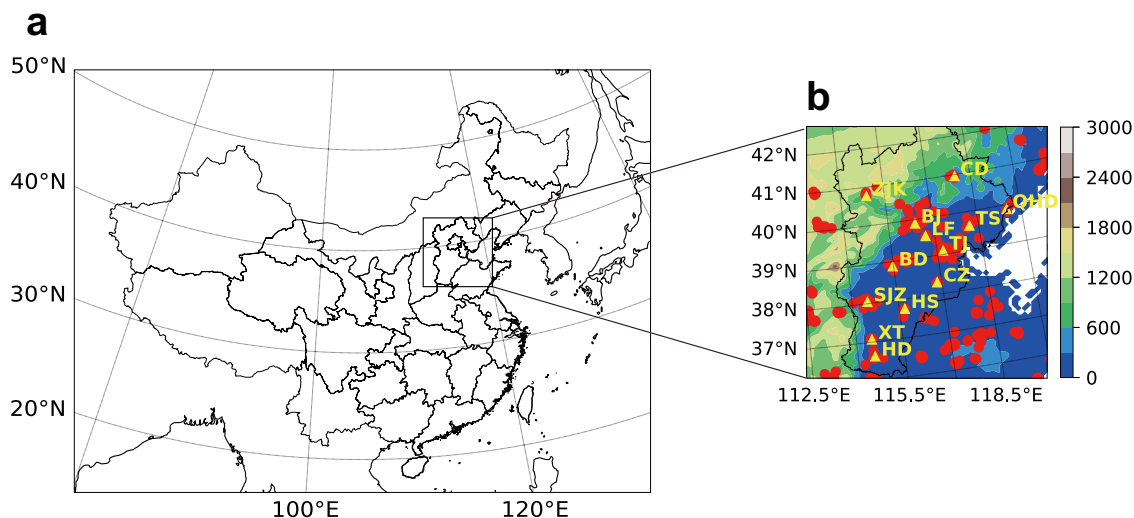
We also considered the emission data as input features to distinguish spatial and monthly variations of PM<sub>2.5</sub> precursors. The emission data were based on the Multi-resolution Emission Inventory for China (MEIC) developed by Tsinghua University (<http://meicmodel.org/>, last access: 2022/6/30). It is known that SO<sub>2</sub>, NO<sub>x</sub>, VOCs and primary PM<sub>2.5</sub> directly contribute to the formation of sulfate, nitrate, black carbon, organic carbon and other inorganics that constitute PM<sub>2.5</sub> in the atmosphere. Thus, emission rates of these species were utilized as input features in this study. As the original emission data are in the resolution of 0.25°, we firstly interpolated them to the WRF grid with a horizontal resolution of 9 km.

In addition to the meteorological and emission variables, several derived parameters were also used as input features, as suggested by aerosol-meteorology interactions. Tai et al.<sup>50</sup> concluded that high PM<sub>2.5</sub> concentrations in the continents were correlated with reduced SLP due to simultaneous presence of cyclones and convergence. Therefore, the change of SLP compared to 24 h ahead was selected as an input feature in our study. Additionally, cold fronts bring a decrease in temperature while stagnation is usually associated with increasing or steady temperature. Thus, changes of temperature indicate synoptic variations thus affecting aerosol concentrations. Rather than utilizing temperature changes at surface, we employed those at 850 hPa to minimize topographic and urban effects. Besides, annual mean PM<sub>2.5</sub> concentrations (averages during 2020–2021) in each grid were also adopted as a static feature for learning spatial correlations. Another important feature is the preceding PM<sub>2.5</sub> observation in the encoder phase and forecast start time to reduce the forecast error that has already been previously introduced.

All of the input variables were firstly standardized to eliminate impacts of unit inconsistency on forecasting results. We assumed that all data obey the standard normal distribution with a mean value of 0 and a standard deviation of 1, thereby they could be standardized one by one following:

$$z = \frac{x - \bar{x}}{\sigma} \quad (1)$$

where  $z$  and  $x$  represent the standardized and original data,  $\bar{x}$  and  $\sigma$  are mean and standard deviation values.



**Fig. 9** WRF domain configuration and the experimental region. The colors, red dots and yellow characters in the right panel denote the terrain height (m), PM<sub>2.5</sub> observation sites and city locations over the BTH, respectively. **a** WRF domain; **b** PPN experimental region.

## Observation data

Surface PM<sub>2.5</sub> concentrations were obtained from monitoring stations over the BTH region and its surrounding area, accessed from the China National Environmental Monitoring Centre (CNEMC, <http://www.cnemc.cn/>). The distribution of these stations is shown in Fig. 9b. To keep consistent with the input feature map, we interpolated the station data of PM<sub>2.5</sub> concentration to grid data with a horizontal resolution of 9 km using the Inverse Distance Weighted (IDW) method. The stations with high density over the BTH and its surrounding areas could reduce the bias from interpolation to some extent. The gridded PM<sub>2.5</sub> data were applied both in the fitting target and input features. These input features are the annual mean PM<sub>2.5</sub> and the preceding PM<sub>2.5</sub> which have been introduced in “Experimental region and feature selection”.

## WRF-Chem results

The purpose of our study is to develop a more efficient and accurate PM<sub>2.5</sub> forecasting system based on the PPN framework. For comparison purposes, we also adopted the WRF-Chem model to predict hourly PM<sub>2.5</sub> concentrations with a resolution of 9 km, covering the North China Plain and its surrounding areas. The WRF-Chem model has the same meteorological configuration as the WRF described in “Experimental region and feature selection”. For chemistry, the Carbon Bond Mechanism version Z (CBMZ)<sup>51</sup> and the Model for Simulating Aerosol Interactions and Chemistry (MOSAIC)<sup>52</sup> were used as the gas-phase chemistry and aerosol mechanisms to conduct PM<sub>2.5</sub> forecasting. Emission rates of anthropogenic pollutants were also from the MEIC inventory, the same as those in the PPN framework. To improve the forecast skill, the three-dimensional variational (3DVAR) algorithm was applied in the WRF-Chem model. Surface pollutant observations (PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, and CO) were assimilated into the model following the method in Sun et al.<sup>20</sup>. The predicted PM<sub>2.5</sub> concentrations generated from the WRF-Chem model were compared with the PPN results to further demonstrate the performance of the proposed DL model.

## Experiment

In this study, the input features together with the PM<sub>2.5</sub> observations in 2 years (2020–2021) were adopted as the training and validating datasets, 90% of which were randomly chosen as the training set (for model training) and 10% of which were the validating set (for model hyper-parameter optimization or tuning). Data from January and June 2022 were employed as the test dataset to evaluate the model performance. Aerosol pollution is usually severe in winter due to stagnation and intensive emissions<sup>38</sup>, thus we mainly focus on model evaluation for the January 2022 results. There are 5800 sequences with 80 × 80 grids in the training and validating dataset. Each sequence in the encoder-decoder architecture contained 2 days for inputs and 3 days for prediction with a time resolution of 3 h. In the PPN model training, Adam Optimizer and one-cycle learning rate schedule with a weight decay parameter of 10<sup>−4</sup> were utilized. This learning rate schedule can substantially speed up training process with high test accuracy<sup>53</sup>. The gradient clipping strategy was also adopted in our model to avoid “gradient explosion” phenomenon.

As described in “Observation data”, the interpolated PM<sub>2.5</sub> data were applied as fitting targets to acquire a gridded result. However, the interpolation may introduce biases into the target dataset, especially in regions with sparsely or unevenly distributed sites. To eliminate the influence of interpolation on model training, we proposed a weighted loss function based on the IDW method and mean square error (MSE) metric (WMSE). The

WMSE is defined as:

$$WMSE = \frac{\sum_{i=1}^n w_i \times (\hat{y}_i - y_i)^2}{\sum_{i=1}^n n \times w_i} \quad (2)$$

$$w_i = \begin{cases} \frac{\sum_{j=1}^{1 \leq d_j \leq 3} \frac{1}{d_j^2}}{m_i} & 1 \leq d_j \leq 3 \\ 1 & d_j = 0 \\ 0.1 & d_j > 3 \end{cases} \quad (3)$$

where  $w_i$  denotes the weight factor for MSE in grid  $i$ ,  $y_i$  and  $\hat{y}_i$  represent the observed and predicted PM<sub>2.5</sub> concentrations in grid  $i$ .  $w_i$  is considered as 1 if there is one or more PM<sub>2.5</sub> monitoring stations in grid  $i$ . If there is no station in grid  $i$ ,  $w_i$  is the average of square sum of the inverse distance between grid  $i$  and the stations in 3 × 3 grids. If there is no station in the surrounding 3 × 3 grids,  $w_i$  is set to a minimum value 0.1. The application of WMSE increases the weights of MSE in areas near the PM<sub>2.5</sub> monitoring sites, thus reducing biases introduced by interpolation over regions with sparse sites during the model learning process.

## DATA AVAILABILITY

The ERA5 reanalysis data are obtained from <https://cds.climate.copernicus.eu/>. The observed PM<sub>2.5</sub> concentrations are obtained from the China National Environmental Monitoring Centre (CNEMC, <http://www.cnemc.cn/>). All WRF data and PPN forecast data are available upon request from the corresponding author (Dr. Jin Feng).

## CODE AVAILABILITY

The pre-trained model and inference code can be found at [https://gitee.com/jfengcode/pub\\_ppn](https://gitee.com/jfengcode/pub_ppn). They are also available upon request from the corresponding author (Dr. Jin Feng).

Received: 9 December 2022; Accepted: 13 June 2023;

Published online: 21 June 2023

## REFERENCES

- Zhai, S. et al. Fine particulate matter (PM<sub>2.5</sub>) trends in China, 2013–2018: separating contributions from anthropogenic emissions and meteorology. *Atmos. Chem. Phys.* **19**, 11031–11041 (2019).
- Xiao, Q. et al. Tracking PM<sub>2.5</sub> and O<sub>3</sub> pollution and the related health burden in China 2013–2020. *Environ. Sci. Technol.* **56**, 6922–6932 (2022).
- Grell, G. A. et al. Fully coupled “online” chemistry within the WRF model. *Atmos. Environ.* **39**, 6957–6975 (2005).
- Appel, K. W. et al. The Community Multiscale Air Quality (CMAQ) model versions 5.3 and 5.3.1: system updates and evaluation. *Geosci. Model Dev.* **14**, 2867–2897 (2021).
- Bey, I. et al. Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *J. Geophys. Res. Atmos.* **106**, 23073–23095 (2001).
- Goldberg, D. L. et al. Using gap-filled MAIAC AOD and WRF-Chem to estimate daily PM<sub>2.5</sub> concentrations at 1 km resolution in the Eastern United States. *Atmos. Environ.* **199**, 443–452 (2019).
- Kong, Y. et al. Improving PM<sub>2.5</sub> forecast during haze episodes over China based on a coupled 4D-LETKF and WRF-Chem system. *Atmos. Res.* **249**, 105366 (2021).
- Marmur, A., Park, S. K., Mulholland, J. A., Tolbert, P. E. & Russell, A. G. Source apportionment of PM<sub>2.5</sub> in the southeastern United States using receptor and emissions-based models: Conceptual differences and implications for time-series health studies. *Atmos. Environ.* **40**, 2533–2551 (2006).
- Wang, L. et al. Source apportionment of PM<sub>2.5</sub> in top polluted cities in Hebei, China using the CMAQ model. *Atmos. Environ.* **122**, 723–736 (2015).
- Guo, H. et al. Source apportionment of PM<sub>2.5</sub> in North China using source-oriented air quality models. *Environ. Pollut.* **231**, 426–436 (2017).
- Qiu, Y., Liao, H., Zhang, R. & Hu, J. Simulated impacts of direct radiative effects of scattering and absorbing aerosols on surface-layer aerosol concentrations in China during a heavily polluted event in February 2014. *J. Geophys. Res. Atmos.* **122**, 5955–5975 (2017).

12. Li, K. et al. Ozone pollution in the North China Plain spreading into the late-winter haze season. *Proc. Natl Acad. Sci. USA* **118**, e2015797118 (2021).
13. Zhu, J., Chen, L. & Liao, H. Multi-pollutant air pollution and associated health risks in China from 2014 to 2020. *Atmos. Environ.* **268**, 118829 (2022).
14. Aleksankina, K., Reis, S., Vieno, M. & Heal, M. R. Advanced methods for uncertainty assessment and global sensitivity analysis of an Eulerian atmospheric chemistry transport model. *Atmos. Chem. Phys.* **19**, 2881–2898 (2019).
15. Vautard, R. et al. Evaluation of the meteorological forcing used for the Air Quality Model Evaluation International Initiative (AQMEII) air quality simulations. *Atmos. Environ.* **53**, 15–37 (2012).
16. Foley, K. M. et al. Incremental testing of the Community Multiscale Air Quality (CMAQ) modeling system version 4.7. *Geosci. Model Dev.* **3**, 205–226 (2010).
17. Jiang, Z. et al. Probing into the impact of 3DVAR assimilation of surface PM<sub>10</sub> observations over China using process analysis. *J. Geophys. Res. Atmos.* **118**, 6738–6749 (2013).
18. Dai, T., Schutgens, N. A., Goto, D., Shi, G. & Nakajima, T. Improvement of aerosol optical properties modeling over Eastern Asia with MODIS AOD assimilation in a global non-hydrostatic icosahedral aerosol transport model. *Environ. Pollut.* **195**, 319–329 (2014).
19. Jung, J. et al. The impact of the direct effect of aerosols on meteorology and air quality using aerosol optical depth assimilation during the KORUS-AQ campaign. *J. Geophys. Res. Atmos.* **124**, 8303–8319 (2019).
20. Sun, W., Liu, Z., Chen, D., Zhao, P. & Chen, M. Development and application of the WRFDA-Chem three-dimensional variational (3DVAR) system: aiming to improve air quality forecasting and diagnose model deficiencies. *Atmos. Chem. Phys.* **20**, 9311–9329 (2020).
21. Lee, S. et al. Seasonal dependence of aerosol data assimilation and forecasting using satellite and ground-based observations. *Remote Sens.* **14**, 2123 (2022).
22. Misenis, C. & Zhang, Y. An examination of sensitivity of WRF/Chem predictions to physical parameterizations, horizontal grid spacing, and nesting options. *Atmos. Res.* **97**, 315–334 (2010).
23. Thomas, S. & Jacko, R. B. Model for forecasting expressway fine particulate matter and carbon monoxide concentration: application of regression and neural network models. *J. Air Waste Manage. Assoc.* **57**, 480–488 (2007).
24. Karimian, H. et al. Evaluation of different machine learning approaches to forecasting PM<sub>2.5</sub> mass concentrations. *Aerosol Air Qual. Res.* **19**, 1400–1410 (2019).
25. Ma, J., Yu, Z., Qu, Y., Xu, J. & Cao, Y. Application of the XGBoost machine learning method in PM<sub>2.5</sub> prediction: a case study of Shanghai. *Aerosol Air Qual. Res.* **20**, 128–138 (2020).
26. Bi, J., Knowland, K. E., Keller, C. A. & Liu, Y. Combining machine learning and numerical simulation for high-resolution PM<sub>2.5</sub> concentration forecast. *Environ. Sci. Technol.* **56**, 1544–1556 (2022).
27. Zhang, B. et al. Deep learning for air pollutant concentration prediction: a review. *Atmos. Environ.* **290**, 119347 (2022).
28. Feng, J., Li, Y., Qiu, Y. & Zhu, F. Capturing synoptic-scale variations in surface aerosol pollution using deep learning with meteorological data. *Atmos. Chem. Phys.* <https://doi.org/10.5194/acp-23-375-2023> (2022).
29. Sayeed, A. et al. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 h in advance. *Neural Netw.* **121**, 396–408 (2020).
30. Yan, R. et al. Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Syst. Appl.* **169**, 114513 (2021).
31. Athira, V., Geetha, P., Vinayakumar, R. & Soman, K. P. DeepAirNet: applying recurrent networks for air quality prediction. *Procedia Comput. Sci.* **132**, 1394–1403 (2018).
32. Ong, B. T., Sugiyara, K. & Zetsu, K. Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM<sub>2.5</sub>. *Neural Comput. Appl.* **27**, 1553–1566 (2016).
33. Yu, Y., Si, X., Hu, C. & Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **31**, 1235–1270 (2019).
34. Du, S., Li, T., Yang, Y. & Horng, S. J. Deep air quality forecasting using hybrid deep learning framework. *IEEE Trans. Knowl. Data Eng.* **33**, 2412–2424 (2021).
35. Pak, U. et al. Deep learning-based PM<sub>2.5</sub> prediction considering the spatio-temporal correlations: A case study of Beijing, China. *Sci. Total Environ.* **699**, 133561 (2020).
36. Yeo, I., Choi, Y., Lops, Y. & Sayeed, A. Efficient PM<sub>2.5</sub> forecasting using geographical correlation based on integrated deep learning algorithms. *Neural Comput. Appl.* **33**, 15073–15089 (2021).
37. Zhu, J., Deng, F., Zhao, J. & Zheng, H. Attention-based parallel networks (APNet) for PM<sub>2.5</sub> spatiotemporal prediction. *Sci. Total Environ.* **769**, 145082 (2021).
38. Wang, S. et al. Spatial distribution, seasonal variation and regionalization of PM<sub>2.5</sub> concentrations in China. *Sci. China Chem.* **58**, 1435–1443 (2015).
39. Yan, D. et al. Evolution of the spatiotemporal pattern of PM<sub>2.5</sub> concentrations in China—A case study from the Beijing-Tianjin-Hebei region. *Atmos. Environ.* **183**, 225–233 (2018).
40. Li, H. et al. Constructing a spatiotemporally coherent long-term PM<sub>2.5</sub> concentration dataset over China during 1980–2019 using a machine learning approach. *Sci. Total Environ.* **765**, 144263 (2021).
41. Mao, W., Wang, W., Jiao, L., Zhao, S. & Liu, A. Modeling air quality prediction using a deep learning approach: Method optimization and evaluation. *Sustain. Cities and Soc.* **65**, 102567 (2021).
42. Sun, Q., Zhu, Y., Chen, X., Xu, A. & Peng, X. A hybrid deep learning model with multi-source data for PM<sub>2.5</sub> concentration forecast. *Air Qual. Atmos. Health.* **14**, 503–513 (2021).
43. Wang, Y. et al. PredRNN: A Recurrent Neural Network for Spatiotemporal Predictive Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* <https://doi.org/10.1109/TPAMI.2022.3165153> (2022).
44. Shi, X. et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* **28**, 802–810 (2015).
45. Bowden, J. H., Otte, T. L., Nolte, C. G. & Otte, M. J. Examining interior grid nudging techniques using two-way nesting in the WRF model for regional climate modeling. *J. Clim.* **25**, 2805–2823 (2012).
46. Jeon, W. et al. A quantitative analysis of grid nudging effect on each process of PM<sub>2.5</sub> production in the Korean Peninsula. *Atmos. Environ.* **122**, 763–774 (2015).
47. Houtekamer, P. L. & Zhang, F. Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **144**, 4489–4532 (2016).
48. Feng, J., Quan, J., Liao, H., Li, Y. & Zhao, X. An air stagnation index to qualify extreme haze events in northern China. *J. Atmos. Sci.* **75**, 3489–3505 (2018).
49. Feng, J., Huang, X. & Li, Y. Improving surface wind speed forecasts using an offline surface multilayer model with optimal ground forcing. *J. Adv. Model. Earth Syst.* **14**, 1–16 (2022).
50. Tai, A. P., Mickley, L. J. & Jacob, D. J. Correlations between fine particulate matter (PM<sub>2.5</sub>) and meteorological variables in the United States: Implications for the sensitivity of PM<sub>2.5</sub> to climate change. *Atmos. Environ.* **44**, 3976–3984 (2010).
51. Zaveri, R. A. & Peters, L. K. A new lumped structure photochemical mechanism for large-scale applications. *J. Geophys. Res. Atmos.* **104**, D23 (1999).
52. Zaveri, R. A., Easter, R. C., Fast, J. D. & Peters, L. K. Model for Simulating Aerosol Interactions and Chemistry (MOSAIC). *J. Geophys. Res. Atmos.* **113**, D13 (2008).
53. Smith, L. N. & Nicholay, T. Super-convergence: very fast training of neural networks using large learning rates. *arXiv* <https://doi.org/10.48550/arXiv.1708.07120> (2019).

## ACKNOWLEDGEMENTS

This research was supported by National Natural Science Foundation of China (Grant No. 42275009), the Open fund by Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control (Grant No. KHK2001), National Natural Science Foundation of China (Grant No. 41975168), the Beijing Natural Science Foundation (Grant No. 8194078).

## AUTHOR CONTRIBUTIONS

J.F. conceived the study and design the model. Y.Q. performed the data gathering, model training and validation, analyses, visualizations and wrote the paper. Y.Q., J.F., Z.Z., X.Z., and R.L. contributed to the WRF and WRF-Chem simulation. Z.Z., and Z.L. helped data gathering. J.F., Z.M. and J.Z. helped in editing and revising the paper.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41612-023-00397-0>.

**Correspondence** and requests for materials should be addressed to Jin Feng.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023