https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

A deep learning approach to increase the value of satellite data for $PM_{2.5}$ monitoring in China

Bo Li[1], Cheng Liu[2,3,4,5], Qihou Hu[3], Mingzhai Sun[2], Chengxin Zhang[2], Shulin Zhang[2], Yizhi Zhu[2], Ting Liu[1], Yike Guo[6], Gregory R. Carmichael[7], Meng Gao[8,9]

[1] School of Earth and Space Sciences, University of Science and Technology of China, Hefei, 230026, China
[2] Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei 230027, China
[3] Key Lab of Environmental Optics and Technology, Anhui Institute of Optics and Fine Mechanics, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China
[4] Center for Excellence in Regional Atmospheric Environment, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China
[5] Key Laboratory of Precision Scientific Instrumentation of Anhui Higher Education Institutes, University of Science and Technology of China, Hefei 230027, China
[6] Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China
[7] Department of Chemical and Biochemical Engineering, The University of Iowa, Iowa City, IA 52242, USA
[8] Department of Geography, State Key Laboratory of Environmental and Biological Analysis, Hong Kong Baptist University, Hong Kong SAR, China
[9] John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA

Corresponding author. *E-mail address*: chliu81@ustc.edu.cn (Cheng Liu); mmgao2@hkbu.edu.hk (Meng Gao).

Abstract

Limitations in the current capability of monitoring $PM_{2.5}$ adversely impact air quality management and health risk assessment of $PM_{2.5}$ exposure. Commonly, ground-based monitoring networks are established to measure the $PM_{2.5}$ concentrations in highly populated regions and protected areas such as national parks, yet large gaps exist in spatial coverage. Satellite-derived aerosol optical properties serve to complement the missing spatial information of ground-based monitoring networks. However, such attempts are hampered under cloudy/hazy conditions or during nighttime. Here we strive to overcome the long-standing restriction that surface $PM_{2.5}$ cannot be constrained with satellite remote sensing under cloudy/hazy conditions or during nighttime. We introduce a deep spatiotemporal neural

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

network (ST-NN) and demonstrate that it can artfully fill these observational gaps. We use sensitivity analysis and visualization technology to open the neural network black box data model, and quantitatively discuss the potential impact of the input data on the target variables. This technique provides ground-level $PM_{2.5}$ concentrations with high spatial resolution (0.01°) and 24-hour temporal coverage. Better constrained spatiotemporal distributions of $PM_{2.5}$ concentrations will help improve health effects studies, atmospheric emission estimates, and predictions of air quality.

1 Introduction

Ambient particles raise worldwide concerns due to their impediments on human health(Dockery et al., 1993), and important roles in the Earth's weather and climate system via altering radiation and clouds(Stocker, 2014). Particles with diameter less than 2.5 micrometers ($PM_{2.5}$) are small enough to enter deeply into human lungs, posing the greatest short-term and long-term risks to health(Pope Iii and Dockery, 2006). Accordingly, sources of $PM_{2.5}$ and $PM_{2.5}$ precursors are highly regulated in most industrialized countries.

$PM_{2.5}$ can linger in the atmosphere for days and exhibit substantial spatiotemporal variations(Jia and Jia, 2014; Poet et al., 1972). Diurnal variation of $PM_{2.5}$ concentrations can range from several μg m^{-3} to many hundreds μg m^{-3} within several hours, and appreciable differences in $PM_{2.5}$ concentrations can occur within several kilometers spatially(Guo et al., 2017; Gupta and Christopher, 2009). Such strong spatiotemporal heterogeneity is attributed to both local sources (direct emissions and secondary production) and regional transport(Zheng et al., 2015).

An accurate depiction of the dynamic evolution of $PM_{2.5}$ remains a challenge, but is urgently needed for better regulation of air quality and health risk assessment. The spatiotemporal distribution of $PM_{2.5}$ is commonly obtained from ground sampling instruments or inferred from satellite remote sensing. Over the past several years, China has made remarkable progress in monitoring air quality, with the number of surface monitoring sites exceeding 1600 across the country in 2020(Liu et al., 2021). These sites are mainly concentrated in urban regions, while rural and rural-urban fringes, home to half of China's population, still go unmonitored. Although the density of monitoring sites within urban areas is larger than that in rural areas (Table S1), important sources, especially point sources, can be missed by these sites. Satellite aerosol optical properties serve to complement the missing spatial information

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

of monitoring networks. These capabilities improve with observations from geostationary satellites. For example, Himawari-8 launched by the Japan Meteorological Agency provides aerosol optical depth (AOD) at a 5-km spatial resolution every 10 minutes. However, satellite data provide only indirect constraints on ground-level $PM_{2.5}$ concentrations, as they retrieve column densities instead of surface-level concentrations and challenges remain in resolving the size spectrum of atmospheric aerosol. Furthermore, satellite observations are limited to cloud-free and haze-free scenes.

Numerous efforts have been made to derive or constrain ground-level $PM_{2.5}$ concentrations with satellite AOD, including aerosol data assimilation with sophisticated chemical transport models(Gao et al., 2017; Saide et al., 2012). A forward operator and its adjoint are usually used to link the changes in AOD to aerosol chemical compositions(Gao et al., 2017; Saide et al., 2012). This approach is computationally expensive, and the performance can be degraded by the uncertainties in the operator itself(Saide et al., 2020). There have also been attempts to statistically infer ground-level $PM_{2.5}$ concentration from satellite AOD(Bi et al., 2019; Xiao et al., 2017). Although spatiotemporal gaps were filled with predictions from chemical transport models or with AOD observed by multiple satellite sensors(Bi et al., 2019; Xiao et al., 2017), predictions were obtained at relatively low temporal resolution (daily/monthly)(Bi et al., 2019; Fang et al., 2016; He and Huang, 2018; Li et al., 2017; Ma et al., 2016; Park et al., 2020; Shtein et al., 2019; Wei et al., 2019; Wei et al., 2020; Xiao et al., 2017; You et al., 2016; Yu et al., 2017) and errors would inherit still from the uncertainties in chemical transport modeling or cloudy/hazy conditions(Xiao et al., 2017; Bi et al., 2019). Several studies offered hourly predictions of daytime $PM_{2.5}$ with inputs from geostationary satellites(Chen et al., 2019; Liu et al., 2019; Zhang et al., 2019). Improved predictions(Fu et al., 2018; Wang et al., 2016) were achieved using day-night band sensor (DNB), yet hourly variations remained unclear. A few studies addressed this issue by including temporal predictors, which could indicate the diurnal pattern of $PM_{2.5}$(Jiang et al., 2021; Tang et al., 2019). However, horizontal resolution of most of the input variables were seriously lower than the prediction, and these algorithms exhibited biases stemming from the limited data coverage of AOD retrievals under cloud cover, ice-covered surfaces, or during nighttime. Heavy haze can also be misclassified as clouds in AOD retrievals(Zhang et al., 2020). For example, our statistical analyses suggest that the annual spatial coverage of satellite AOD is only 33% in North China, and even less in other concerned regions in China (Table S2).

Better methods are thus needed to overcome these limitations, particularly for regions with thick

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

clouds and severe haze pollution. In this study, we construct a deep spatiotemporal neural network (ST-NN) model to derive ground-level $PM_{2.5}$ concentrations with inputs of satellite monitored AOD, meteorological elements, and geographical information. With this technique, surface $PM_{2.5}$ concentrations in China can be accurately derived at high spatial and temporal resolution (0.01° and hourly), even during nighttime and under cloudy or hazy conditions.

2 Materials and Methods

We built a deep learning ST-NN model to improve estimates of ground-level $PM_{2.5}$ concentrations, particularly for regions without sampling sites, and for conditions (cloudy, hazy, nighttime, etc.) where satellite retrievals are not available.

2.1 Model Configuration Datasets

We used hourly ground-level observations of $PM_{2.5}$ from the Chinese National Environmental Monitoring Center (CNMEC) network, the daily MODerate Resolution Imaging Spectroradiometer (MODIS) 3km aerosol products(Levy et al., 2013), and the hourly 0.05°×0.05° Himawari-8 AOD products(Yoshida et al., 2018). The original MODIS products were mapped onto regional grids of 0.05°×0.05° resolution to keep them consistent with the geographic coordinate system of Himawari-8 data. Considering the diurnal variation of the solar zenith angles, only daytime satellite data (defined as 00:00-09:00 UTC) were used in this study. The reason for using two different types of satellite aerosol optical thickness is that the Himawari-8 satellite aerosol is effective in capturing the daily variation of aerosols, while the MODIS aerosol product has aerosol optical thickness in different bands to capture information on the properties of aerosols and more accurate numerical results. Other inputs to the neural network include land cover types (MODIS land cover product at 0.05°×0.05° resolution), road network (originally meter level, www.openstreetmap.org; Last access: July 10, 2020), point of interest data (POI), elevation data (1km×1km, the Resource and Environment Science Data Center, RESDC, http://www.resdc.cn; Last access: January 1, 2021), population/gross domestic product data from RESDC, and weather fields (0.05° × 0.05°, hourly) simulated by the Weather Research and Forecasting model version 4.0 with three nested domains (https://www.mmm.ucar.edu/weather-research-and-forecasting-model; Last access: May 15, 2020).

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

Geographical inputs (road network, POI, elevation, etc.) were regridded to 0.01° grids. The initial conditions and boundary conditions of the meteorological fields were derived from the National Centers for Environment Prediction's (NCEP) 6-hour final operational global (FNL) data with a spatial resolution of 0.25° × 0.25°. Validation of the WRF predicted meteorological variables is displayed in Figure. S1. Descriptions and features of these considered datasets are listed in Table S3-S6.

2.2 Data Preprocessing and ST-NN Model Configuration

Figure. S2 displays the architecture of the deep learning ST-NN model. It operates on three major data types, namely AOD, geographical factors, and spatiotemporal distributions of meteorological conditions. AOD data were classified based on the dimension of time as near current moment (AOD at the past four hours, $t_{-3} \sim t_0$ for daytime, and daytime AOD for nighttime prediction), past days (daytime AOD in the past two days), and past week to formulate influencing factors across time. MODIS AOD values retrieved at three bands were used for the past week's results. For aerosol data under cloud cover, a null fill value was used. Meteorological data were arranged as time series of the bottom model level and the vertical features at the current moment t_0 to include the influences of temporal and spatial evolution. All input features were subsequently used to examine the potential relationship with $PM_{2.5}$. Pearson's correlation test was applied for variables that contain dimension of time (e.g., temperature, RH, u-wind, v-wind, AOD, etc.). For time-independent variables (e.g. POI, road network, and land cover type), $PM_{2.5}$ data from CNEMC were classified based on the severity of pollution, and then a Chi-squared test was used. Only those parameters that pass the significance test ($\alpha < 0.01$) were selected (Table S7).

For the exploration of the potential relationships between variables and $PM_{2.5}$, time variable elements with significant influence ($\alpha < 0.01$) were selected (Table S7). Since spatially related geographic information variables (POI, road network, GDP, etc.) were not time-dependent (within the scope of the study), the correlation between land variables and the average annual $PM_{2.5}$ concentration was explored. K-means was used to explore the discreteness of these variables, and then the contingency table was used for significance test. Although neural networks can spontaneously extract valid variables and remove the influence of irrelevant variables, a priori data selection is necessary to

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

effectively reduce model complexity and improve model operational efficiencies.

Inception-ResNet block(Szegedy et al., 2017) and pooling layers were adopted to transform the data and mine the major features of data (more details in Figure. S2). The outputs out of this step were fused and connected, and $PM_{2.5}$ concentrations were then obtained through optimizing the Log-Cosh loss function below.

$$Loss = \frac{1}{n}\sum_{i=1}^{n} \log\left(\cosh\left(y_i^{predict} - y_i^{true}\right)\right) \qquad (1)$$

where $y_i^{predict}$ means the model predicted value and the y_i^{true} represents the observations.

For small differences the log-cosh loss performances similar as $\frac{(y_i^{predict}-y_i^{true})^2}{2}$, and for huge differences (at the beginning of model training), it's closer abs($y_i^{predict} - y_i^{true}$)-log(2).

We initialized all the layers with the built-in Keras glorot uniform initializer as 0, and the biases were initialized with 0. Due to the symmetry of the data, tanh was used as the activation function. We trained the networks for 64 epochs with a batch size of 4, and SGD (Stochastic Gradient Descent) optimizer with an exponential decay of the learning rate α as:

$$\alpha = \begin{cases} 0.001 & epoch <= 32 \\ 0.001\times\exp\left(0.1\times(32\text{-epoch})\right) & epoch > 32 \end{cases} \qquad (2)$$

2.3 Training and Testing

Based on number of samples, dimension of time and dimension of space, the entire datasets (entire year of 2017) were randomly sorted into 10 sections, with 9 sections for training and the rest for testing(Schultz et al., 2021). The testing data doesn't participate in the model training process. We also applied a 10-fold cross-validation to demonstrate the capability of the built model. We used sample-, spatial- and temporal-based cross-validation to evaluate the generalization level of the model. For sample-based cross-validation, we randomly grouped all the data; for spatial-based cross-validation, we randomly grouped the data by site location; and for temporal-based cross-validation, we randomly grouped the data by time. The grouped data are then used for model training and validation of the results. The proposed model was implemented in Python 3.7 with a neural network library named Keras and TensorFlow as the backend.

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

2.4 Sensitivity Analysis

Sensitivity analysis was conducted to explore the influences of input variables on the distribution of ground-level $PM_{2.5}$ concentrations. We also use sensitivity analysis to open black box data mining models(Cortez and Embrechts, 2013). For M input variables $\{X_a: a \in (1,...,M)\}$, each input variable X_a was divided into L levels, and X_{a_j} denotes the j[th] level of X_a. For continuous variables, the L level is evenly divided into 10 parts according to the value range of input variables, and for classified variables, it is equal to the number of channels. N samples from the testing dataset were then selected randomly to replace X_a values with X_{a_j}, and the mean responses of $PM_{2.5}$ ($\hat{y}_{a_j}$) were documented. With the spatial feature considered, the sensitivity of $PM_{2.5}$ to a continuous variable factor (AOD, meteorological variables, etc.) was examined by varying the factor X_a through its range with L levels but keeping the spatial pattern fixed. The X_{a_j} was given as:

$$X_{a_j} = X_a - mean(X_a) + L_j \qquad (3)$$

For classified factors, such as land use type and traffic networks, sensitivity analysis was conducted in the manner of unified feature type:

$$X_{a_j}[N,m,n,j] = \sum_{j=0}^{channels} X_a[N,m,n,j] \qquad (4)$$

where m and n denote the location in spaial coordinate, while j represents the location in category dimension.

Four metrics were calculated to evaluate the relative importance of input variables, namely range (S_r), gradient (S_g), variance (S_v) and Average Absolute Deviation (AAD) (S_d)(Cortez and Embrechts, 2013). For model inputs X_a, evaluation metrics were calculated with equations below:

$$S_r = max\,(\hat{y}_{a_j}: j \in \{1, \dots, L\}) - min\,(\hat{y}_{a_j}: j \in \{1, \dots, L\}) \ (5)$$

$$S_g = \sum_{j=2}^{L} |\hat{y}_{a_j} - \hat{y}_{a_{j-1}}|/(L-1) \ \ (6)$$

$$S_v = \sum_{j=1}^{L} (\hat{y}_{a_j} - \bar{y}_a)^2/(L-1) \ \ (7)$$

$$S_d = \sum_{j=1}^{L} |\hat{y}_{a_j} - \tilde{y}_a|/(L-1) \quad (8)$$

where $\bar{y}_a$ and $\hat{y}_a$ denote the mean and median of the responses. The relative importance (r_a) can be

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

described as:

$$r_a = \varsigma_a / \sum_{i=1}^{M} \varsigma_i \qquad (9)$$

where ς_a means the sensitivity measure for X_a(e.g. range). In this study, the relative importance (r_a) was defined as a vector $\vec{r} = (r_1, r_2, ..., r_M)$.

The influence of errors in input data on predictions of $PM_{2.5}$ concentrations was explored with the equation below:

$$\{input\_data_A[l,i,j,m] = input\_data[l,i,j,m] * (1 + relative\_error) \qquad (10)$$

where l, i, j, m represent the dimensions of input data (l: the batch size, i: latitude, j:longitude, m:channels), and relative error means the uniform distribution of upper and lower bounds of error.

3 Results

3.1 ST-NN model reconstructs observed spatiotemporal (both daytime and nighttime) features of $PM_{2.5}$

Our ST-NN model operated on three major data types, namely AOD, geographical factors, and spatiotemporal distributions of meteorological conditions (details are documented in the Methods section). It was built to improve the predictions of ground-level $PM_{2.5}$, particularly for regions without sampling sites, and for conditions (cloudy, hazy, nighttime, etc.) that satellite retrievals are not available. In this study, we focused on most populated and concerned regions, North China. We also demonstrated that the proposed method can be easily applied to other parts of China including East China, South China, Sichuan Basin, and the heavily polluted Shaanxi province (regions marked in Figure. S3 and Table S8). The performance of the ST-NN model was cross-validated with respect to sampling selection, temporal variability and spatial distribution. As displayed in Figure. 1, our ST-NN model accurately captured the observed spatiotemporal variability of daytime $PM_{2.5}$, with regression slopes close to 1 and intercepts close to 0. Spatial variations of AOD at the past four hours,t_{-3}~t_0 , were used as near moment predictors for daytime $PM_{2.5}$. Daytime AOD values in the past two days and in the past week were also used to formulate influencing factors across time. Multiple validation metrics, including R^2, root mean square error (RMSE, μg m^{-3}), and mean absolute error (MAE, μg m^{-3}), were

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

calculated and listed in Figure. 1, Table 1 and Table 2. R^2 values with respect to sampling selection, temporal variability and spatial distribution in North China were generally above 0.85 (Figure. 1), indicating the applicability of the model under various complex conditions. Its applications to other years and over other regions in China achieved similar pleasant performance, and R^2 value reached even 0.90 when it is applied to Shaanxi Province for year 2019 (Table 1).

Given the relatively long lifetime of aerosols in the atmosphere(Williams et al., 2002) (several days), daytime observed variations of AOD were used in the prediction of nighttime $PM_{2.5}$ (details documented in Methods section). The capability in predicting the diurnal features of $PM_{2.5}$ was demonstrated in Figure. 1 (d-i). Similar values of validation metrics were found for different time windows. R^2 values were generally above 0.80, and RMSE values were below 26 μg m^{-3} for North China. Similar delightful performances were found for other regions also, and the performance of the model in predicting nighttime $PM_{2.5}$ did not exhibit a significant degradation from daytime (Table 1). Despite that satellite AOD retrievals are not available during nighttime, our ST-NN model provides a reasonably reliable prediction of $PM_{2.5}$ during nighttime. This is mainly attributed to the advantage of ST-NN in learning the dynamic transport and dissipation of particles under complex influences of meteorology, terrain, etc., which was exemplified with a haze episode occurrend in North China in 2017 (Figure. 2). And from Figure. S4 we can see that the characteristics of the $PM_{2.5}$ distribution in the Beijing, $PM_{2.5}$ concentrations are influenced by topography and southwest transmission. The data are influenced by meteorological and aerosol data at 0.05°. However it can still be differentiated on a scale of 0.01°.

In addition to cross-validation, independent validation of this ST-NN model was conducted with $PM_{2.5}$ concentrations observed at sites that were not included in the model training. The variability of $PM_{2.5}$ concentrations at these independent stations were also accurately captured by our model, with R^2 values greater than 0.8 (Figure. S5). Independent validation was conducted also with respect to the diurnal variation of $PM_{2.5}$. As indicated in Figure. S6, the diurnal pattern of $PM_{2.5}$ over multiple independent stations across China was reproduced by the ST-NN model.

3.2 Temporal and spatial block cross-validation

To better assess the generalization of our model, additional spatial block cross-validation tests were

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

carried out(Schultz et al., 2021). We selected the East China for the mask testing, and used the sites within the designated area as the validation dataset (Figure. S7). As mask area increases, model performances progressively worsen (Table. S13), but it still captures pollution events (Figure. S8), Even in North China, the most polluted region of China, it still gives good results in different weather conditions(Figure. S9). On the temporal level, we trained the model with data from 2017 to 2019 only, and predicted the dynamic evolution of $PM_{2.5}$ concentrations in 2020. The overall validation is shown in Figure. S10 and Table S14. We find that the model validates worse in scenarios with lower surface $PM_{2.5}$ concentrations, mainly due to the large observational uncertainty at low $PM_{2.5}$ concentrations (Figure. S16).

We further assessed the ability of the model to capture pollution events through accuracy and precision. Accuracy rates were greater than 80% and 75% of sites had precision greater than 60% (Table. S15).

3.3 ST-NN model improves prediction of $PM_{2.5}$ below clouds and during severe haze

A prominent advantage of our ST-NN model is its competence in improving prediction of $PM_{2.5}$ below clouds and during severe haze. Figure. 3 displays satellite images during various episodes in different seasons when the North China region was obscured by clouds. In cloudy conditions, satellites fail to monitor ground-level aerosol pollution, while our ST-NN can fill these observation gaps and provide a complete distribution of $PM_{2.5}$ under cloudy conditions. Compared against ground-level observations, satisfactory performance was found (Figure. 3), with R^2 values exceeding 0.82 in most cases. $PM_{2.5}$ hot spots in South Hebei and Shanxi as observed by the ground-level network were also reproduced by ST-NN. And Figure. S11 shows the overall relative error of cross-validation of the model under different cloud coverage.

We further explored how cloudy conditions would influence the prediction of $PM_{2.5}$ concentrations. Figure. 4 illustrates the predicted $PM_{2.5}$ with full coverage and with cloudy conditions removed for four metropolitan clusters in China. The MODIS Collection 6.1 Cloud mask products were used to track the cloudy conditions in this ST-NN model. Over the study period, 60% of the data in North China were affected by clouds. Heavy haze in China can also be misclassified as cloud by the retrieval algorithm(Zhang et al., 2020). As a result, the influences of clouds on the prediction of $PM_{2.5}$ differ greatly across regions and seasons (Figure. 4). In North China and the Sichuan Basin region, mean

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

$PM_{2.5}$ concentrations with cloudy periods removed exhibit lower values than the full coverage annual mean (Figure. 4c, 3o). On the contrary, negative differences were identified for South China (Figure. 4k), suggesting different driving factors for these regions. In cloudy scenes, $PM_{2.5}$ concentrations exhibited lower values when relative humidity (RH) > 60% in South China (Figure. S12c). This could be related to cloud-related wet removal of air pollutants. Conversely, $PM_{2.5}$ concentrations in North China were biased low using only cloud-free scenes in North China, as indicated with lower satellite observed AOD in cloud-free scenes (Figure. 5a-d). Such underestimation mainly occured during wintertime (Figure. 5).

The CNEMC surface measurements further indicated that $PM_{2.5}$ concentrations were biased low in cloud-free scenes in North China, but biased high in South China, consistent with our ST-NN predictions (Table 3). Over the entire study period, 83% of the data in winter in North China were marked as cloudy, higher than those in other seasons (60%) (Table S2). This is mainly related to the occurrences of snow/ice or severe haze(Zhang et al., 2020). Figure. 4 (d, h, l, p) further justified that groud-level $PM_{2.5}$ under cloudy conditions could be well predicted by our ST-NN model in four metropolitian regions, with high correlation coefficients and low errors. Cross-validation also suggested that our ST-NN model can give valid results under clouds (Figure. S11). Different regions are affected differently by cloud cover, with warmer and more humid regions such as Eastern and Southern China, where errors increase as cloud cover increases, while in dryer regions such as Northern China, the effect of cloud cover has little impact on the results, possibly indicating a potential relationship between surface $PM_{2.5}$ concentrations and cloud formation processes.

3.4 ST-NN model offers better regional representation of $PM_{2.5}$

As the CNEMC stations are concentrated in urban areas (Figure. S13), using only CNEMC data to estimate regional $PM_{2.5}$ concentration would result in an overestimation. As indicated in Table 4, mean $PM_{2.5}$ concentrations at CNEMC stations agree better with the mean over densely populated areas, but are higher than the mean over sparsely populated areas and therefore also higher than the mean over the entire region. This further emphasizes that CNEMC observations might not be able to reflect the pollution in the suburbs and accurately show the overall pollution condition in a region.

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

4 Discussion

A number of studies have explored the prediction of ground-level $PM_{2.5}$ concentrations with statistical methods, as indicated in Table S9. Despite that many of these studies achieved similar performance with respect to R^2, RMSE, MAE and slope as our study, most of them provided only predictions at a low temporal resolution (daily)(Fang et al., 2016; Ma et al., 2016; You et al., 2016; Li et al., 2017; Xiao et al., 2017; Yu et al., 2017; He and Huang, 2018; Bi et al., 2019; Shtein et al., 2019; Wei et al., 2019; Park et al., 2020; Wei et al., 2020). Several studies offered hourly predictions of daytime $PM_{2.5}$ with inputs from geostationary satellites(Chen et al., 2019; Liu et al., 2019; Zhang et al., 2019). However, the nighttime hourly variations are lacked, although endeavors were made with Day-Night Band (DNB) sensor(Wang et al., 2016; Fu et al., 2018). Additionally, few studies addressed this issue by including temporal predictors(Tang et al., 2019; Jiang et al., 2021).

We fully used the spatiotemporal features of aerosol and simulated the dynamic evolution of aerosols under complex influences of meteorology, terrain, etc. in this study. Sampling selection, temporal variation, and spatial distribution based cross-validation demonstrated that the method presented here is skilled in providing reliable ground-level $PM_{2.5}$ concentrations with high spatial resolution (0.01°) and 24-hour temporal coverage, which is challenging especially for heavily polluted regions. Independent validations were also conducted for cloudy conditions and nightime, and no degradation of performance was found.

We examined the importance of satellite observed AOD in the prediction of $PM_{2.5}$ during both daytime and nighttime using four sensitivity measures(Cortez and Embrechts, 2013), namely range S_r, gradient S_g, variance Sv, and average absolute deviation from the median S_{AAD}. AOD accounts for more than 30% of the weight throughout the day, and the relative significance exhibits slightly higher values during nighttime (Table 5), emphasizing the importance of AOD observations in nighttime predictions. Land cover type and meteorological variables also play important roles in the dynamic evolution of $PM_{2.5}$ in North China and other regions, as illustrated in Figure. S14. The effects of the key variables on surface $PM_{2.5}$ concentrations are given in Figure. S15. However, the model tends to better capture moderately polluted conditions, as the relative errors exhibit relatively larger values when observed $PM_{2.5}$ concentrations are above 350 $\mu g\ m^{-3}$ or below 20 $\mu g\ m^{-3}$ (Figure. S16). The relatively poor capability of our ST-NN model in capturing these extremely low or high values are mainly attributed

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

to the rarity of these conditions and the small sampling size (North China: 0.34‰, East China: 0.059‰, South China: 0.068‰, Sichuan Basin: 0.048‰, Shaanxi Province: 0.25‰). Similar uncertainties of the model might be raised by the errors in model input data. Random errors were added to the input data to explore how it would influence the errors of predicted $PM_{2.5}$. Similarly, the quality of AOD data was essential within a very broad range of uncertainty (Figure. S17). When errors of other inputs grow (>20%), the accuracy of prediction would also be significantly degraded (Figure. S17). We examined also how input data quality control process would affect the accuracy, and a negligible role was found (Table S10, S11). During the development of ST-NN models for different regions in China, the loss function decreased in a similar manner, while the decreasing speed and convergence values varied among regions due to differences in the size and feature of data (Figure. S18). We noticed also that the performance of the model varied across regions and seasons, which might be also related to the distinct spatiotemporal features of $PM_{2.5}$ (Figure. S19), and the associated meteorological/geographical characteristics in different regions. The uneven distribution of CNEMC sites might also play a role (Figure. S13).

A long-standing restriction for the use of satellite AOD has been that surface $PM_{2.5}$ cannot be constrained under cloudy conditions, during nighttime or during severe haze(Gao et al., 2017). This limitation has been overcome here with an advanced statistical method. The capability of the built ST-NN model in predicting $PM_{2.5}$ below clouds and during nighttime is mainly due to the consideration of spatiotemporal variation of influencing meteorological/ geographical factors and the dynamic evolution of aerosols. The processes considered are close to those in numerical chemical transport models, but with constraints of satellite AOD. Time-varying and time-invariant factors were processed separately in the ST-NN model to fully explore the dynamic feature of aerosol under complex influences, and the factors on different time scales were considered. Our ST-NN model relies on the regional transport features of air pollution, and it could thus be problematic to track very small point sources. This limitation will be further improved in future studies with more in-depth exploration of the connection between aerosol and clouds(Saide et al., 2012). The issue of rarely observed extreme conditions and small sampling size could be solved also to some extent in the near future when the volume of observations grows with time.

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

Acknowledgments

We acknowledge NASA for providing the MODIS measurements and Japan Meteorological Agency for offering Himawari-8 aerosol optical depth (AOD). We acknowledge Martin G. Schultz for his valuable suggestions on this work.

Author Contributions: C.L. and M.G. conceived the research, and M.S. provided technical support. B.L. conducted model simulations and analyzed results; M.G.S, G.R.C., Q. H., C.Z., S.Z., Y.Z., T.L. and Y.G. assisted with the discussion; all authors contributed to the final interpretation and writing of the manuscript with major contributions by C.L., B. L., and M. G.

Competing Interests: The authors declare that they have no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Funding: This study was supported by grants from the National Natural Science Foundation of China (No. 41722501 and No. 42005084), Research Grants Council of the Hong Kong Special Administrative Region, China (project no. HKBU22201820 and HKBU12202021), the National Key Research and Development Program of China (No. 2018YFC0213104), the National High-Resolution Earth Observation Project of China (No. 05-Y30B01-9001-19/20-3) and Key Research Program of Frontier Sciences, CAS (No. ZDBS-LY-DQC008).

References

Bi, J., Belle, J. H., Wang, Y., Lyapustin, A. I., Wildani, A., and Liu, Y.: Impacts of snow and cloud covers on satellite-derived PM2. 5 levels, Remote sensing of environment, 221, 665-674, 2019.

Chen, J., Yin, J., Zang, L., Zhang, T., and Zhao, M.: Stacking machine learning model for estimating hourly PM2. 5 in China based on Himawari 8 aerosol optical depth data, Science of The Total Environment, 697, 134021, 2019.

Cortez, P. and Embrechts, M. J.: Using sensitivity analysis and visualization techniques to open black box data mining models, Information Sciences, 225, 1-17, 10.1016/j.ins.2012.10.039, 2013.

Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris Jr, B. G., and Speizer, F. E.: An association between air pollution and mortality in six US cities, New England journal of medicine, 329, 1753-1759, 1993.

Fang, X., Zou, B., Liu, X., Sternberg, T., and Zhai, L.: Satellite-based ground PM2. 5 estimation using timely structure adaptive modeling, Remote Sensing of Environment, 186, 152-163, 2016.

Fu, D., Xia, X., Duan, M., Zhang, X., Li, X., Wang, J., and Liu, J.: Mapping nighttime PM2. 5 from

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

VIIRS DNB using a linear mixed-effect model, Atmospheric Environment, 178, 214-222, 2018.

Gao, M., Saide, P. E., Xin, J., Wang, Y., Liu, Z., Wang, Y., Wang, Z., Pagowski, M., Guttikunda, S. K., and Carmichael, G. R.: Estimates of health impacts and radiative forcing in winter haze in eastern China through constraints of surface PM2. 5 predictions, Environmental Science & Technology, 51, 2178-2185, 2017.

Guo, J., Xia, F., Zhang, Y., Liu, H., Li, J., Lou, M., He, J., Yan, Y., Wang, F., and Min, M.: Impact of diurnal variability and meteorological factors on the PM2. 5-AOD relationship: Implications for PM2. 5 remote sensing, Environmental Pollution, 221, 94-104, 2017.

Gupta, P. and Christopher, S. A.: Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach, Journal of Geophysical Research: Atmospheres, 114, 2009.

He, Q. and Huang, B.: Satellite-based mapping of daily high-resolution ground PM2. 5 in China via space-time regression modeling, Remote Sensing of Environment, 206, 72-83, 2018.

Jia, G. and Jia, J.: Atmospheric residence times of the fine-aerosol in the region of south Italy estimated from the activity concentration ratios of 210Po/210Pb in air particulates, J Anal Bioanal Tech, 5, 1-5, 2014.

Jiang, T., Chen, B., Nie, Z., Ren, Z., Xu, B., and Tang, S.: Estimation of hourly full-coverage PM2. 5 concentrations at 1-km resolution in China using a two-stage random forest model, Atmospheric Research, 248, 105146, 2021.

Levy, R., Mattoo, S., Munchak, L., Remer, L., Sayer, A., Patadia, F., and Hsu, N.: The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques, 6, 2989-3034, 2013.

Li, T., Shen, H., Yuan, Q., Zhang, X., and Zhang, L.: Estimating ground - level PM2. 5 by fusing satellite and station observations: a geo - intelligent deep learning approach, Geophysical Research Letters, 44, 11,985-911,993, 2017.

Liu, C., Gao, M., Hu, Q., Brasseur, G. P., and Carmichael, G. R.: Stereoscopic Monitoring: A Promising Strategy to Advance Diagnostic and Prediction of Air Pollution, Bulletin of the American Meteorological Society, 102, E730-E737, 10.1175/bams-d-20-0217.1, 2021.

Liu, J., Weng, F., and Li, Z.: Satellite-based PM2. 5 estimation directly from reflectance at the top of the atmosphere using a machine learning algorithm, Atmospheric Environment, 208, 113-122, 2019.

Ma, Z., Hu, X., Sayer, A. M., Levy, R., Zhang, Q., Xue, Y., Tong, S., Bi, J., Huang, L., and Liu, Y.: Satellite-based spatiotemporal trends in PM2. 5 concentrations: China, 2004–2013, Environmental health perspectives, 124, 184-192, 2016.

Park, Y., Kwon, B., Heo, J., Hu, X., Liu, Y., and Moon, T.: Estimating PM2. 5 concentration of the conterminous United States via interpretable convolutional neural networks, Environmental Pollution, 256, 113395, 2020.

Poet, S., Moore, H., and Martell, E.: Lead 210, bismuth 210, and polonium 210 in the atmosphere: Accurate ratio measurement and application to aerosol residence time determination, Journal of Geophysical Research, 77, 6515-6527, 1972.

Pope III, C. A. and Dockery, D. W.: Health effects of fine particulate air pollution: lines that connect, Journal of the air & waste management association, 56, 709-742, 2006.

Saide, P. E., Carmichael, G. R., Spak, S. N., Minnis, P., and Ayers, J. K.: Improving aerosol distributions below clouds by assimilating satellite-retrieved cloud droplet number, Proceedings

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

of the National Academy of Sciences, 109, 11939-11943, 2012.
Saide, P. E., Gao, M., Lu, Z., Goldberg, D. L., Streets, D. G., Woo, J.-H., Beyersdorf, A., Corr, C. A., Thornhill, K. L., and Anderson, B.: Understanding and improving model representation of aerosol optical properties for a Chinese haze event measured during KORUS-AQ, Atmospheric Chemistry and Physics, 20, 6455-6478, 2020.
Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A., and Stadtler, S.: Can deep learning beat numerical weather prediction?, Philosophical Transactions of the Royal Society A, 379, 20200097, 2021.
Shtein, A., Kloog, I., Schwartz, J., Silibello, C., Michelozzi, P., Gariazzo, C., Viegi, G., Forastiere, F., Karnieli, A., and Just, A. C.: Estimating daily PM2. 5 and PM10 over Italy using an ensemble model, Environmental science & technology, 54, 120-128, 2019.
Stocker, T.: Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change, Cambridge university press2014.
Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A.: Inception-v4, inception-resnet and the impact of residual connections on learning, Thirty-first AAAI conference on artificial intelligence,
Tang, D., Liu, D., Tang, Y., Seyler, B. C., Deng, X., and Zhan, Y.: Comparison of GOCI and Himawari-8 aerosol optical depth for deriving full-coverage hourly PM2. 5 across the Yangtze River Delta, Atmospheric Environment, 217, 116973, 2019.
Wang, J., Aegerter, C., Xu, X., and Szykman, J. J.: Potential application of VIIRS Day/Night Band for monitoring nighttime surface PM2. 5 air quality from space, Atmospheric environment, 124, 55-63, 2016.
Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., and Cribb, M.: Estimating 1-km-resolution PM2. 5 concentrations across China using the space-time random forest approach, Remote Sensing of Environment, 231, 111221, 2019.
Wei, J., Li, Z., Cribb, M., Huang, W., Xue, W., Sun, L., Guo, J., Peng, Y., Li, J., and Lyapustin, A.: Improved 1 km resolution PM 2.5 estimates across China using enhanced space–time extremely randomized trees, Atmospheric Chemistry and Physics, 20, 3273-3289, 2020.
Williams, J., De Reus, M., Krejci, R., Fischer, H., and Ström, J.: Application of the variability-size relationship to atmospheric aerosol studies: estimating aerosol lifetimes and ages, Atmospheric Chemistry and Physics, 2, 133-145, 2002.
Xiao, Q., Wang, Y., Chang, H. H., Meng, X., Geng, G., Lyapustin, A., and Liu, Y.: Full-coverage high-resolution daily PM2. 5 estimation using MAIAC AOD in the Yangtze River Delta of China, Remote Sensing of Environment, 199, 437-446, 2017.
Yoshida, M., Kikuchi, M., Nagao, T. M., Murakami, H., Nomaki, T., and Higurashi, A.: Common retrieval of aerosol properties for imaging satellite sensors, Journal of the Meteorological Society of Japan. Ser. II, 2018.
You, W., Zang, Z., Zhang, L., Li, Y., Pan, X., and Wang, W.: National-scale estimates of ground-level PM2. 5 concentration in China using geographically weighted regression based on 3 km resolution MODIS AOD, Remote Sensing, 8, 184, 2016.
Yu, W., Liu, Y., Ma, Z., and Bi, J.: Improving satellite-based PM 2.5 estimates in China using Gaussian processes modeling in a Bayesian hierarchical setting, Scientific reports, 7, 1-9, 2017.
Zhang, T., Zang, L., Wan, Y., Wang, W., and Zhang, Y.: Ground-level PM2. 5 estimation over urban agglomerations in China with high spatiotemporal resolution based on Himawari-8,

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

Science of the total environment, 676, 535-544, 2019.

Zhang, X., Wang, H., Che, H.-Z., Tan, S.-C., Shi, G.-Y., Yao, X.-P., and Zhao, H.-J.: Improvement of snow/haze confusion data gaps in MODIS Dark Target aerosol retrievals in East China, Atmospheric Research, 245, 10.1016/j.atmosres.2020.105063, 2020.

Zheng, G., Duan, F., Su, H., Ma, Y., Cheng, Y., Zheng, B., Zhang, Q., Huang, T., Kimoto, T., and Chang, D.: Exploring the severe winter haze in Beijing: the impact of synoptic weather, regional transport and heterogeneous reactions, Atmospheric Chemistry and Physics, 15, 2969-2983, 2015.

Figures and Tables

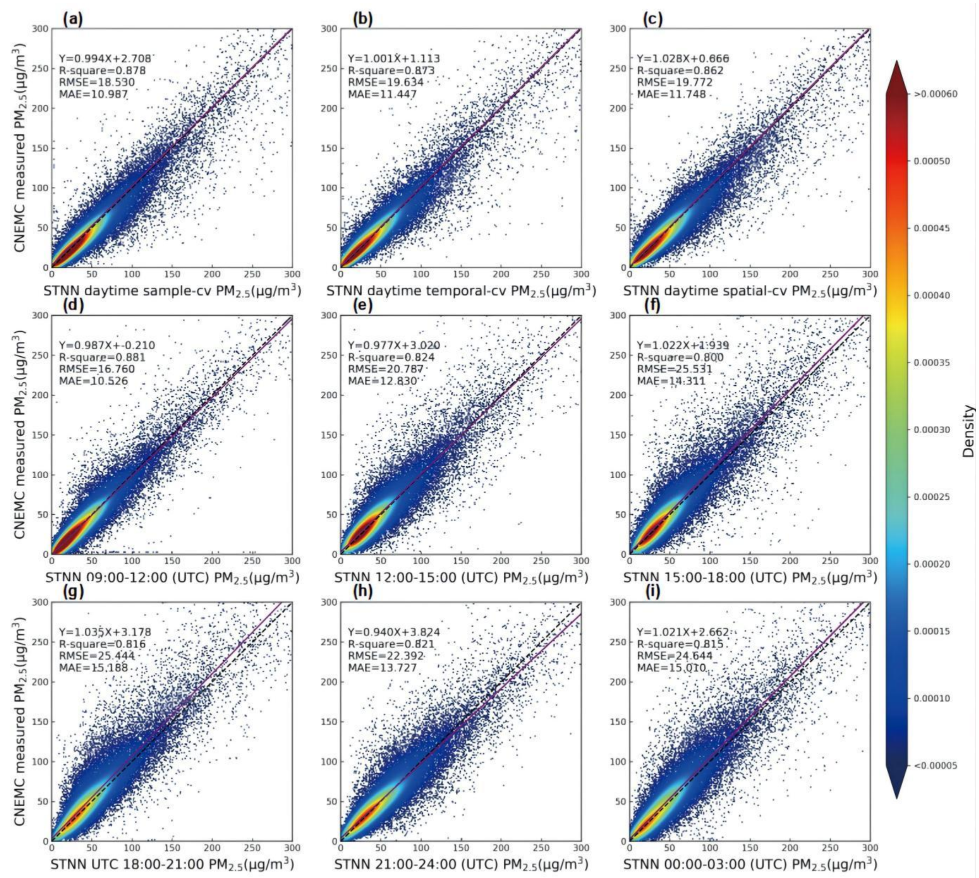

Figure. 1. Density scatterplots of cross-validation with respect to sampling selection, temporal variability, and spatial distribution. (a) daytime, sampling selection; (b) daytime, temporal variability; and (c) daytime, spatial distribution; (d-i): cross-validation across space at different diurnal time slots (both daytime and nighttime). The fitting line is in purple, and the 1:1 standard line is the black dotted line.

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

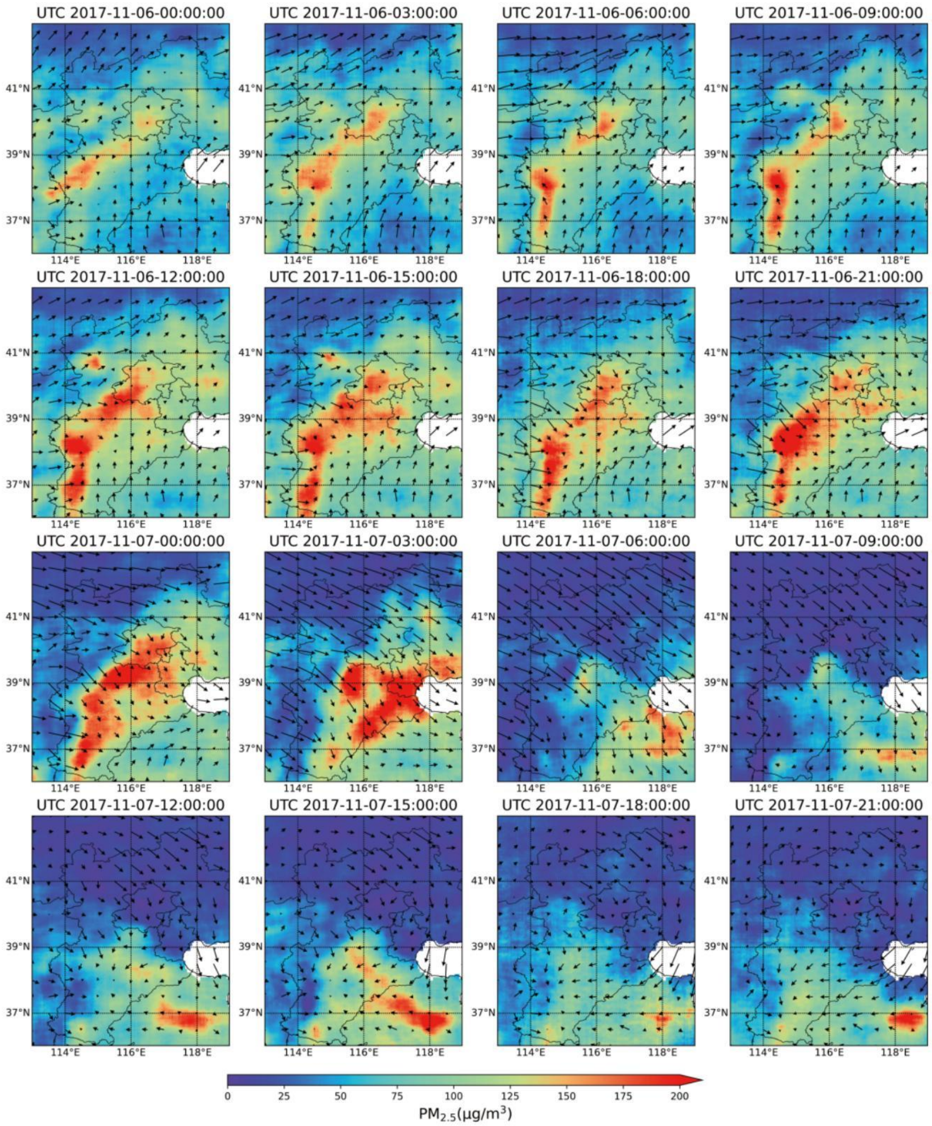

Figure. 2. ST-NN model simulated haze episode on November 16-17, 2017. The spatial distribution of simulated near surface $PM_{2.5}$ concentrations and wind fields.

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

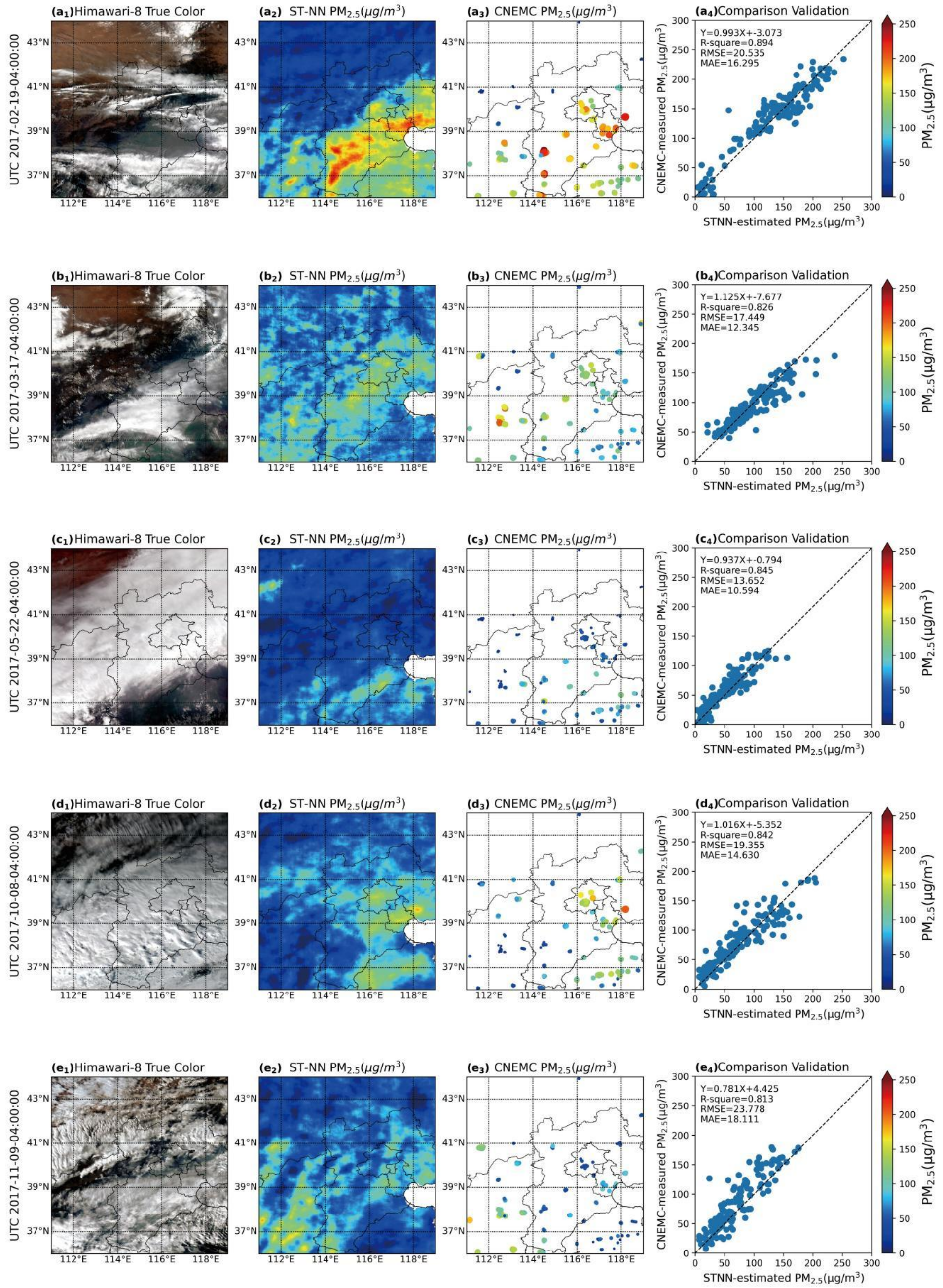

Figure. 3. Performance of the ST-NN model at several cloudy moments (clouded data randomly selected from the results of time-based cross-validation). Left to right columns display true color images from Himawari-8 (five moments, a_1:2017/2/19 04:00 UTC; b_1: 2017/3/17 04:00 UTC; c_1: 2017/5/22 04:00 UTC; d_1: 2017/10/8 06:00 UTC; e_1: 2017/11/9 07:00 UTC), ST-NN model predicted

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

$PM_{2.5}$, at corresponding moments (a_2-e_2), CNEMC observed $PM_{2.5}$ at corresponding moments (a_3-e_3), and the scattered validation (a_4-e_4).

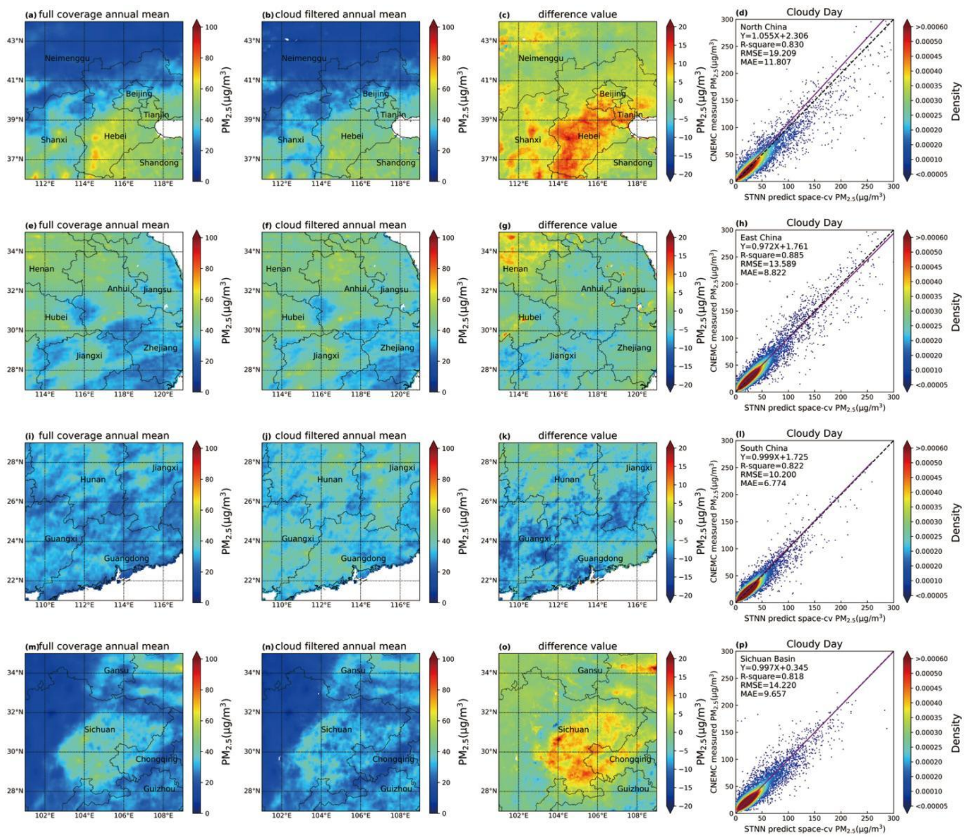

Figure. 4. ST-NN model predicted annual mean $PM_{2.5}$ concentrations in 2017 and validation under cloudy conditions in 2017. ST-NN model predicted full coverage annual mean (a, e, i, m for North China, East China, South China, and Sichuan Basin, respectively); predicted annual mean with MODIS marked cloudy conditions removed (b, f, j, n); the differences between predictions with full coverage and those with MODIS marked cloudy conditions removed (c, g, k, o); Cross-validation with respect to spatial distribution under conditions at stations that were not considered in training.

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

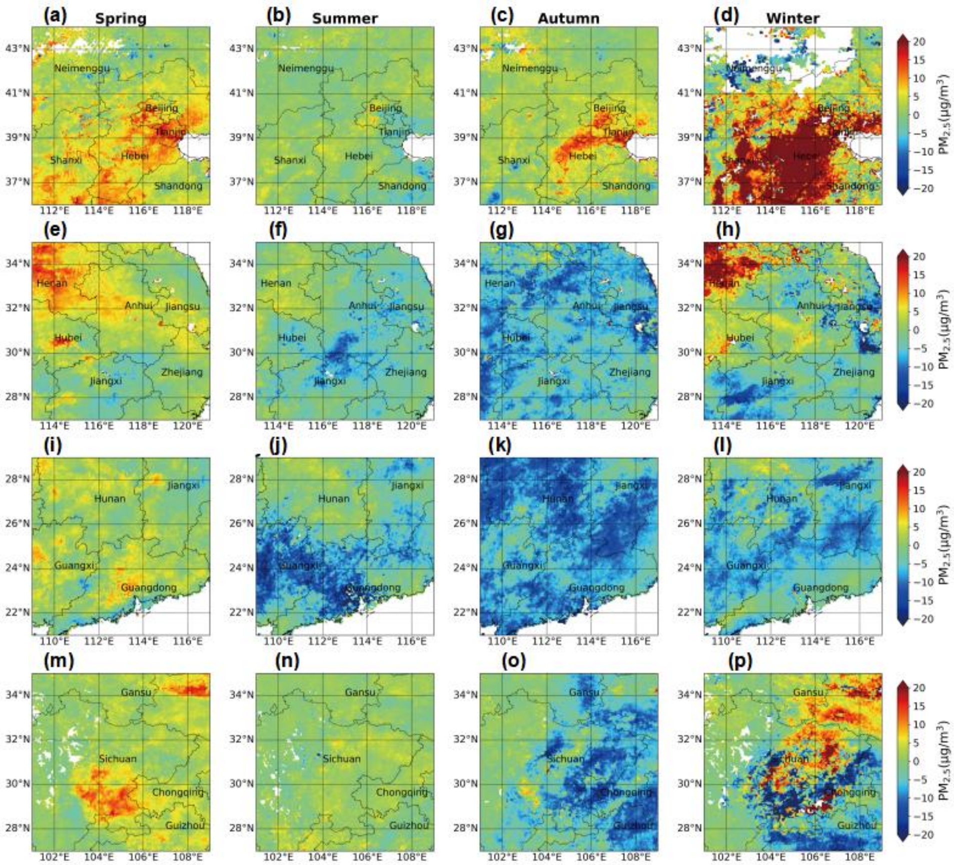

Figure. 5. Seasonal distribution of the difference in $PM_{2.5}$ concentrations between full coverage data and cloud mask filtered data for four metropolitan regions. (a-d) North China. (e-h) East China. (i-l) South China. (m-p) Sichuan Basin.

Table 1. R-square values of cross validation of the model with respect to spatial distribution.

	2017		2018		2019		2020	
	day	night	day	night	day	night	day	night
North China	0.86	0.83	0.82	0.84	0.87	0.85	0.84	0.88
East China	0.81	0.82	0.86	0.85	0.83	0.86	0.86	0.85
South China	0.83	0.84	0.82	0.83	0.83	0.85	0.82	0.80

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

Sichuan Basin	0.84	0.85	0.82	0.80	0.89	0.89	0.87	0.83
Shaanxi Province	0.85	0.84	0.89	0.81	0.90	0.87	0.88	0.88

Table 2. RMSE of cross validation with respect to spatial distribution.

	2017		2018		2019		2020	
	day	night	day	night	day	night	day	night
North China	19.77	22.59	19.92	19.86	16.53	18.44	16.46	13.99
East China	16.15	16.51	13.09	14.04	13.19	12.13	9.88	9.47
South China	11.11	12.81	10.38	11.38	9.52	11.41	6.00	8.96
Sichuan Basin	14.80	17.52	13.90	18.51	10.28	11.86	8.03	10.74
Shaanxi Province	20.15	22.79	15.47	18.88	15.14	17.13	12.01	12.33

Table 3. CNMEC measured and ST-NN predicted $PM_{2.5}$ (μg m^{-3}) concentrations in 2017.

	CNEMC annual mean	CNEMC cloud filtered mean	ST-NN annual mean	ST-NN cloud filtered mean
North China	58.30	43.57	33.84	29.58
East China	48.57	44.33	38.49	40.75
South China	38.27	46.14	29.77	35.94
Sichuan Basin	46.88	36.48	25.80	24.63
Shaanxi Province	51.15	40.47	33.54	30.23

Table 4. ST-NN model estimated $PM_{2.5}$ (μg m^{-3}) concentrations under different population densities.

	North China	East China	South China	Sichuan Basin	Shaanxi Province
CNEMC	58.65	48.65	38.71	46.11	54.08
Populated Regions (>500people/km2)	53.40	43.20	31.31	38.55	46.38
Moderately populated (<500people/km2)	29.43	36.36	27.72	24.36	32.04

https://doi.org/10.5194/egusphere-2022-578
Preprint. Discussion started: 12 July 2022

All areas	34.12	38.53	28.11	25.80	33.52

Table 5. The importance of AOD in the prediction of $PM_{2.5}$, as indicated with sensitivity measures (R_r, R_g, R_v, and R_{AAD}).

	00:00-06:00 (UTC)	06:00-12:00 (UTC)	12:00-18:00 (UTC)	18:00-24:00 (UTC)	day	Night
R_r	0.34	0.31	0.33	0.32	0.34	0.36
R_g	0.42	0.39	0.41	0.40	0.42	0.44
R_v	0.36	0.27	0.32	0.30	0.35	0.39
R_{AAD}	0.36	0.31	0.33	0.32	0.35	0.37