Joint Retrieval of $PM_{2.5}$ Concentration and Aerosol Optical Depth over China Using Multi-Task Learning on FY-4A AGRI

Bo LI[1,2,3], Disong FU[2, *], Ling YANG[1], Xuehua FAN[2, **], Dazhi YANG[4], Hongrong SHI[5], Xiangao XIA[2]

1 Electronic Engineering College, Chengdu University of Information Technology, Chengdu 610225, China

2 Key Laboratory of Middle Atmosphere and Global Environment Observation, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

3 Research Centre on Meteorological Observation Engineering Technology, China Meteorological Administration, Beijing 100081, China

4 School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin 150001, China

5 Key Laboratory of Cloud-Precipitation Physics and Severe Storms, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

*, ** Corresponding author: Disong Fu, Xuehua Fan
Email: fudisong@mail.iap.ac.cn; fxh@mail.iap.ac.cn

ABSTRACT

Aerosol optical depth (AOD) and fine particulate matter with a diameter of less than 2.5μm ($PM_{2.5}$) play crucial roles in air quality, human health, and climate change. However, the complex correlation of AOD–$PM_{2.5}$ and the limitations of existing algorithms pose a significant challenge in realizing the accurate joint retrieval of these two parameters at the same location. On this point, a multi-task learning (MTL) model, which enables the joint retrieval of $PM_{2.5}$ concentration and AOD, is proposed and applied on the top-of-the-atmosphere reflectance (TOAR) data gathered by the Fengyun-4A Advanced Geosynchronous Radiation Imager (FY-4A AGRI), and compared to that of two single-task learning (STL) models, namely, random forest (RF) and deep neural network (DNN). Specifically, The MTL model achieved a coefficient of determination (R^2) of 0.88 and a root mean square error (RMSE) of 0.10 in AOD retrieval. In comparison to the RF model, the R^2 increased by 0.04, the RMSE decreased by 0.02, and the percentage of retrieval results falling within the expected error range (Within-EE) rose by 5.55%. The R^2 and RMSE of $PM_{2.5}$ retrieval by MTL model are 0.84 and 13.76 $\mu g \cdot m^{-3}$, respectively. Compared with the RF model, the R^2 increased by 0.06, the RMSE decreased by 4.55 $\mu g \cdot m^{-3}$, and the Within-EE increased by 7.28%. Additionally, compared to the DNN model, the MTL model showed an increase of 0.01 in R^2 and a decrease of 0.02 in RMSE in AOD retrieval, with a corresponding increase of 2.89% in Within-EE. For $PM_{2.5}$ retrieval, the MTL model exhibited an increase of 0.05 in R^2, a decrease of 1.76 $\mu g \cdot m^{-3}$ in

RMSE, and an increase of 6.83% in Within-EE. The evaluation suggests that the MTL model is able to provide simultaneously improved AOD and $PM_{2.5}$ retrievals, offering a critical advantage in capturing the actual distribution of fine particulate matter with high time efficiency.

Key words: AOD; $PM_{2.5}$; FY-4A; multi-task learning; joint retrieval

https://doi.org/10.1007/s00376-024-3222-y

Article Highlights:

- The simultaneous retrieval of AOD and $PM_{2.5}$ concentration at a large spatial scale is achieved by employing a multi-task learning algorithm.
- Multi-task learning can retrieve $PM_{2.5}$ and AOD with higher accuracy, compared to Single-task learning.
- Multi-task learning provides new insights on high-efficiency monitoring of aerosol pollution.

1. Introduction

Fine particulate matter ($PM_{2.5}$) refers to suspended particles with aerodynamic diameter less than or equal to 2.5 μm, which is an important indicator of air pollution (Hill et al., 2023). On the other hand, aerosol optical depth (AOD) is a critical parameter describing the attenuation of solar radiation as caused by aerosols in the atmosphere. Both $PM_{2.5}$ and AOD play pivotal roles in air quality, human health (Ho et al., 2018), and climate change research (Zhang et al., 2017). Therefore, realizing high-precision detection of aerosols over large areas is important for the environment and human health. Whereas ground-based measurement constitutes a means of acquiring reliable and high-precision information, the measurement sites are, more often than not, distributed unevenly and sparsely. For instance, approximately 90% of $PM_{2.5}$ monitoring sites in China are concentrated in the eastern and coastal regions, but in the western regions, such as the Tibetan Plateau or other high-altitude areas, measurement sites are grossly sparse (Xiao et al., 2016). No better than that, the existing number of Aerosol Robotic Network (AERONET) sites for AOD measurements within the Asian-Pacific region is fewer than 90. Therefore, AERONET AOD data are usually used only for regional studies of aerosols and validation of remotely sensed data, and effective monitoring of aerosols on a large scale cannot be realized (Zhang et al., 2018; Hu et al., 2014). The unevenness and scarcity of ground-based measurement sites, which implies limited observation records, pose significant challenges to conducting comprehensive aerosol studies at a sub-continental scale, such as over China, which is the country concerning this work.

The estimation of surface $PM_{2.5}$ using satellite AOD data (hereafter, AOD–PM method) has emerged as the predominant approach over the past decade (Xue et al., 2020). The mapping from predictors (satellite AOD and auxiliary variables) to the predictand ($PM_{2.5}$) are mostly data-driven, as exemplified by multiple linear regression (Gupta & Christopher, 2009), linear mixed-effect model (Lee et al., 2011), geographically weighted regression (Hu et al., 2013), random forest (RF) (Chen et al., 2018), convolutional neural network (CNN), long short-term memory (LSTM) network (Xu et al., 2021; Pak et al., 2018), among other models. An important premise and theoretical foundation for retrieving surface $PM_{2.5}$ concentration with satellite AOD is the strong correlation and connection between $PM_{2.5}$ and AOD (Li et al., 2015). Therefore, the availability and reliability of such satellite-derived $PM_{2.5}$ necessarily depend upon those of satellite AOD. In this regard, to alleviate the errors introduced by the intermediate AOD estimation, researchers have explored the possibility of directly obtaining $PM_{2.5}$ from the top-of-the-atmosphere reflectance (TOAR) captured by imagers onboard satellites; this approach is known as the end-to-end method (Yang et al., 2020). For example, Shen et al. (2018) introduced a TOAR-PM method that established the relationship between TOAR, observation angles, meteorological factors and $PM_{2.5}$; taking the Wuhan urban agglomeration as the study area, they achieved an in-sample cross-validated coefficient of determination (R^2) and a root mean square error (RMSE) of 0.87 and 9.89 μg·m^{-3}, respectively. Based on an RF model and TOAR data from Himawari-8, Bai et al. (2021) estimated the $PM_{2.5}$ over the Yangtze River delta region; their finding revealed an R^2 value of 0.75 and a RMSE of 18.71 μg·m^{-3}. Similarly, Mao et al. (2021) proposed a RF-based method to directly estimate hourly ground-level $PM_{2.5}$ in China from the Fengyun-4A Advanced Geosynchronous Radiation Imager (FY-4A AGRI) TOAR, and evaluated all training samples by cross-validation method and keep an acceptable accuracy (R^2 = 0.90, RMSE = 15 μg·m^{-3}). Yin et al. (2021) also utilized the Himawari-8 TOAR to estimate the concentration of $PM_{2.5}$ over China, but by employing a light gradient boosting machine (LightGBM); the R^2 and RMSE of $PM_{2.5}$ estimated are 0.83 and 23.7 μg·m^{-3}.

As for AOD, several physical retrieval algorithms have long been well known, including the dark target (DT) (Kaufman et al., 1997), deep blue (DB) and Multi-Angle Atmospheric Correction (MAIAC) algorithms (Lyapustin et al., 2018). Similar to the case of $PM_{2.5}$, end-to-end AOD estimation algorithms have also been used as promising alternatives to the physical algorithms. For instance, Ding et al. (2022) employed a so-called "neural network aerosol retrieval for geostationary satellite (NNAeroG)" method to estimate AOD over the full disk area of FY-4A. The validation result demonstrated an RMSE of 0.24, an R^2 of 0.73, and 58.7% AOD values falling within an expected error envelope of ± (0.05 + 15%×$AOD_{AERONET}$), which is abbreviated as EE15 hereafter. She et al. (2022) proposed a Landsat-8 AOD retrieval algorithm based on DNN, which successfully retrieved AOD in a longitude range of 30°W–160°E and a latitude range of 60°S–60°N. The estimated AOD was found to exhibit excellent agreement with AERONET AOD, with an R^2 of 0.71, an RMSE of 0.15, and 61% of the retrieved AOD values falling within an expected error envelope of $\pm$ (0.05 + 20%×$AOD_{AERONET}$).

Regardless of the retrieval subject (i.e., $PM_{2.5}$ or AOD), most current TOAR-based retrieval methods are single-task learning (STL) models. In STL, individual tasks employ independently trained models without explicit mechanisms to support information sharing across different tasks. However, applying STLs on TOAR data, though simple in conception, are confined by many factors. For instance, there is a conceptual difference between $PM_{2.5}$, which reflects the near-surface turbidity of the atmosphere, and TOAR, which covers the atmospheric information from surface to several hundred kilometers in altitude. This vertical distribution depth mismatch makes $PM_{2.5}$ retrieval more challenging. On the other hand, retrieving AOD from TOAR faces two general difficulties. First, ground-based observatories based on ground-based instruments (e.g., sun photometers) are mainly located in the eastern and coastal regions of China, such as the North China Plain region, the Yangtze River Delta region, and the Pearl River Delta region (Xun et al., 2021). Due to the sparse distribution of these stations, data-driven models may not be able to capture more complex regional features, thus risking overfitting. Second, missing values often occur due to cloud cover (Kokhanovsky et al., 2007), thereby reducing the number of available samples. During the cloud detection process, heavy aerosols can be misclassified as clouds (Song et al., 2019), which enhances the mutual interference between clouds and aerosols. Therefore, there are significant limitations in improving the retrieval accuracy of $PM_{2.5}$ and AOD. It is encouraging that multiple studies have shown a correlation between $PM_{2.5}$ and AOD (Zheng et al., 2017; Yang et al., 2019). Given the existence of this correlation, is it possible to employ a method that effectively utilizes this relationship to jointly retrieve $PM_{2.5}$ and AOD using TOAR data, even with a limited number of samples?

In what follows, any method that allows for the joint retrieval of multiple correlated parameters is referred to as multi-task learning (MTL). One distinct advantage of MTL over STL is that it allows parameters sharing to a certain degree between several related learning tasks, thereby improving the performance of all tasks (Ruder, 2017; Zhang & Yang, 2017). In fact, MTL is a general concept that goes far beyond the retrieval of atmospheric parameters. But in recent years, several MTL approaches have been adopted by atmospheric scientists. For instance, Zhang et al. (2020) proposed an MTL model that combines CNN and gated recurrent unit (GRU) for multi-step-ahead multi-station $PM_{2.5}$ prediction. It is tested on three monitoring stations located in three different districts of Lanzhou, China, with an RMSE of 7.96 $\mu g \cdot m^{-3}$, indicating better performance in intensive air quality prediction than previous models based on simple hybridization. Further, to fully utilize the meteorological information from the monitoring stations, Xu and Yoneda (2021) proposed a long short-term memory (LSTM) autoencoder MTL model to predict $PM_{2.5}$ time series in multiple locations city wide, which greatly improved the prediction accuracy and calculation cost compared with the traditional LSTM model. Meanwhile, in order to verify the performance difference between MTL model and STL model, Song et al. (2022) employed attentive MTL model to predict air quality in urban stations. In the comparison of results in Seoul, the proposed attentive model with MTL outperformed the STL attentive model in terms of accuracy performance, with RMSE values for $PM_{2.5}$ being 8.36 and 8.66 $\mu g \cdot m^{-3}$, respectively. On the other hand, in order to enhance the input, the spectral and spatial information is jointly used to retrieve fine mode fraction (FMF), Chen et al., (2020) proposed an artificial neural network for aerosol retrieval (NNAero) to jointly retrieve AOD and (FMF). The input data are the Moderate Resolution Imaging Spectroradiometer (MODIS) TOAR together with MODIS-derived surface reflectance in 5 spectral bands, and the labels are AERONET FMF and AOD. The results show that 68% of the NNAero AOD values are within EE15, which is better than the DT algorithm (31% within EE15).

This contribution aims to simultaneously retrieve AOD and $PM_{2.5}$ concentrations over China using FY-4A AGRI data based on an MTL algorithm. The performance of the model is thoroughly evaluated using ground-based observations and compared with that of the classical RF and DNN models. The rest of the paper is organized as follows: Section 2 introduces the various datasets used in this study, among which the FY-4A AGRI data and European Center for Medium-Range Weather Forecasts'

(ECMWF's) fifth-generation reanalysis (ERA5) are used as model inputs, whereas the Multi-Angle Atmospheric Correction (MAIAC) AOD data and the ground-based $PM_{2.5}$ observations from two sources act as the training targets and out-of-sample verifications. In Section 3, a detailed description of data processing and the setups of the machine-learning models are provided. Section 4 summarizes the statistical results and evaluates the performance of the models. Finally, conclusion follows in Section 5.

2. Data

This study utilized FY-4A AGRI and ECMWF ERA5 data, spanning from March 12, 2018, to March 11, 2019, covering the geographical expanse of China, as inputs for the MTL model. For the same region and time frame, the MAIAC AOD and the $PM_{2.5}$ measurements, collected by the China Environmental Monitoring Center (CEMC), are used as the targets during learning. Furthermore, $PM_{2.5}$ measurements obtained from the Environmental Protection Department (EPD) in Hong Kong (HK) and AERONET AOD data are utilized for out-of-sample verification. The Long-term Gap-free High-resolution Air Pollutant (LGHAP) concentration dataset was employed to assess and validate the performance of the MTL model in Northwest China.

2.1FY-4A /AGRI data

FY-4A, which was launched on December 11, 2016, has capabilities that are greatly enhanced, as compared to its predecessors, in terms of environmental and weather monitoring, warning, and forecasting. The AGRI sensor onboard FY-4A has 14 channels and a wavelength range of 0.45–13.8 μm, covering the visible (VIS), near-infrared (NIR), medium infrared, and long infrared spectra, with a spatial resolution of 1km (Fu et al., 2024). The level-1 FY-4A AGRI TOAR data are available online (http://satellite.nsmc.org.cn/). In this study, three channels that are known to related to aerosol are selected, namely, 0.45–0.49 μm (CH01), 0.75–0.90 μm (CH03), and 2.1–2.35 μm (CH06). Furthermore, geometrical properties including solar zenith angle (SOZ), solar azimuth angle (SOA), satellite zenith angle (SAZ), and satellite azimuth angle (SAA) are also computable parameters that are to be used as inputs.

2.2 Meteorological data

ERA5, which is a global atmospheric reanalysis product developed by ECMWF with a spatial resolution of 0.25°×0.25°, is used as an auxiliary dataset to provide the meteorological factors that can affect the composition and transport of aerosols (Deng et al., 2012). The selected meteorological factors include the u- and v-component wind at a height of 10 m (U10, V10, m/s), temperature at a height of 2 m (T2m, K), surface pressure (Sp, Pa), boundary layer height (Blh, m), total column ozone (Tco3, DU), total column water vapor (Tcwv, kg/m^2); these can be directly downloaded from the C3S climate data store (https://cds.climate.copernicus.eu/) with relative ease, owing to the mature data dissemination system of ECMWF.

2.3 MODIS MAIAC AOD

The Terra and Aqua satellites, with MODIS sensors onboard, were launched in December 1999 and May 2002, respectively. There are four standard land MODIS AOD products, which have been retrieved using DT, DB, DT–DB combined, and MAIAC algorithms (Xie et al., 2019). Among these products, the MAIAC AOD has the highest spatial resolution of 1 km. Validation results with AERONET AOD show that MAIAC AOD retrievals are highly correlated with ground-based AOD measurements. The correlation coefficients (R) are greater than 0.8 at more than 68% of AERONET sites. The accuracy of MAIAC AOD retrievals is high (within expected error (EE) = 87.49% and 83.15%) in the regions of tropical rainforest climate and tropical open forest climate (Qin et al., 2021). For that reason, the daily MAIAC AOD product at the wavelength of 550 nm is used in this work. This product has passed quality assurance, cloud screening, and adjacency testing (Lyapustin et al., 2018).

2.4 Ground-based observation data

Quality-controlled hourly $PM_{2.5}$ concentration measurements are collected from CEMC (http://www.cnemc.cn), with regularly calibrated sensors located at 1590 sites across the China (see Fig. 1); these data are to be used as training targets. For out-of-sample verification, the hourly $PM_{2.5}$ concentrations data at 15 air-quality monitoring stations in HK are acquired from the Environmental Protection Department (EPD) website (http://www.epd.gov.hk/epd/). The hourly $PM_{2.5}$ concentration were measured using the micro-oscillation balance method and β absorption method, and the uncertainty is said to be within 5 μg·m^{-3} (Miao and Liu, 2019). As for AOD data, they are sources from the AERONET, which is the world's largest ground-based aerosol observation network. AERONET provides observations of aerosol properties (https://aeronet.gsfc.nasa.gov/), such as optical, microphysical or radiative properties. AERONET provides AOD in seven wavelength bands (340, 380, 440, 500, 675, 870, and 1020 nm). In this study, AOD at 550 nm is least-squares fitted using the well-known quadratic relationship between AOD and wavelength (Fu et al., 2023).

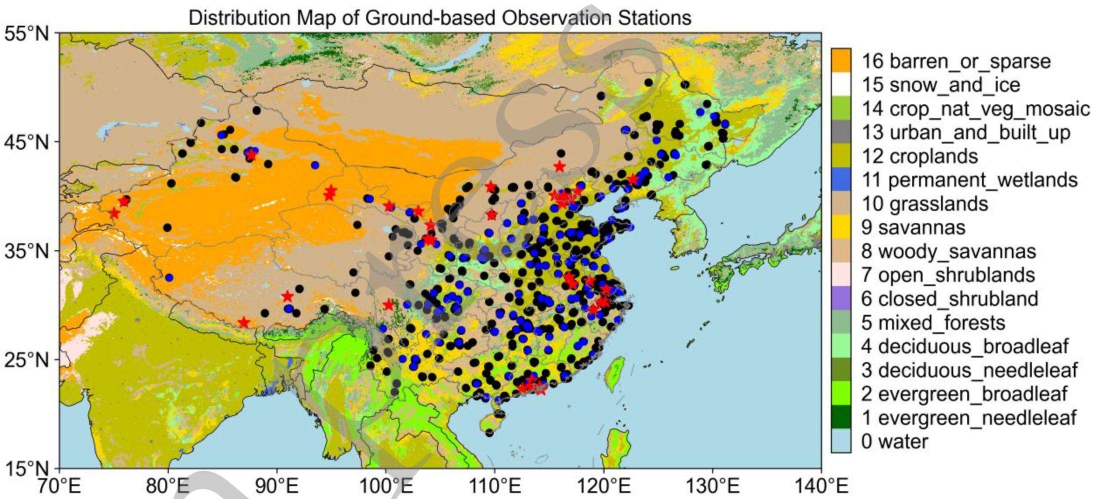


Fig. 1. Spatial distribution of AERONET sites (red stars) and $PM_{2.5}$ ground stations (blue/black dots). Annotated with black and blue dots are $PM_{2.5}$ sites used for model training and testing, respectively. The background map indicates the type of ground cover from MODIS in 2018.

2.5 LGHAP concentration dataset

We obtained the LGHAP dataset from the Earth System Science Data website (https://www.earth-system-science-data.net). The dataset synergistically integrates multimodal aerosol data from different sources using a tensor-flow-based data fusion method to generate daily gapless AOD products for China from 2000 to 2020 with a spatial resolution of 1 km. Subsequently, the $PM_{2.5}$ concentration on continuous space was estimated using an integrated learning approach. The data were stored in NetCDF format, and code was provided to help users read and visualize the data (Bai et al., 2022). The verification results of ground observation data show that LGHAP AOD data is highly correlated with AERONET AOD data, with an R of 0.91 and RMSE equaling 0.21. In addition, $PM_{2.5}$ estimates were highly correlated with ground-based measurements, with an R of 0.95 and RMSE equaling 12.03 μg·m^{-3}. In this study, we used LGHAP AOD and LGHAP $PM_{2.5}$ data to verify the estimation results of MTL model in Northwest China.

3 Methodology

Figure 2 illustrates the workflow of this study, including dataset preprocessing, model training and validation, and retrieval of AOD and $PM_{2.5}$ concentrations. The workflow is applied not only to the MTL model but also two classical STL models (RF and DNN).

3.1 Data preprocessing

Because data from multiple sources are used in the study, in order to ensure spatial scale consistency, satellite and auxiliary reanalysis data are resampled to a grid size of 0.01°×0.01° using bilinear interpolation (Wei et al., 2019). Subsequently, all data are spatially matched based on the nearest-neighbor grid cells to the $PM_{2.5}$ stations. Regarding temporal matching, $PM_{2.5}$ values from two hours adjacent to MODIS overpass time (02:30 UTC for Terra and 05:30 UTC for Aqua) and FY-4A data within 2.5 min of the overpass time are averaged. Reanalysis data meeting this temporal requirement are drawn from the time series. To filter for data with high quality and reliability, samples containing zero and invalid values are rejected. At the same time, because $PM_{2.5}$ and AOD have lognormal distributions, a logarithmic transformation is applied, as to make the data more symmetric and stable. After data filtering, a total of 132,540 samples remains for model training and validation. Of these, 90% of the sites are used for training the model (marked as black dot in Fig.1), while the remaining are used for testing the model (marked as red dot in Fig. 1). It is worth noting that the proportion of MODIS MAIAC AOD data divisions based on $PM_{2.5}$ sites matched is the same as for $PM_{2.5}$. To mitigate the impacts of outliers on model fitting, normalization and standardization must be applied to all variables prior to training. In the feature selection stage, we adopted the method of permutation feature importance to select 17 most representative features as the input of the model, as shown in Table 1. Permutation feature importance is a model inspection technique that measures the contribution of each feature to a fitted model's statistical performance on a given tabular dataset (https://scikit-learn.org/).

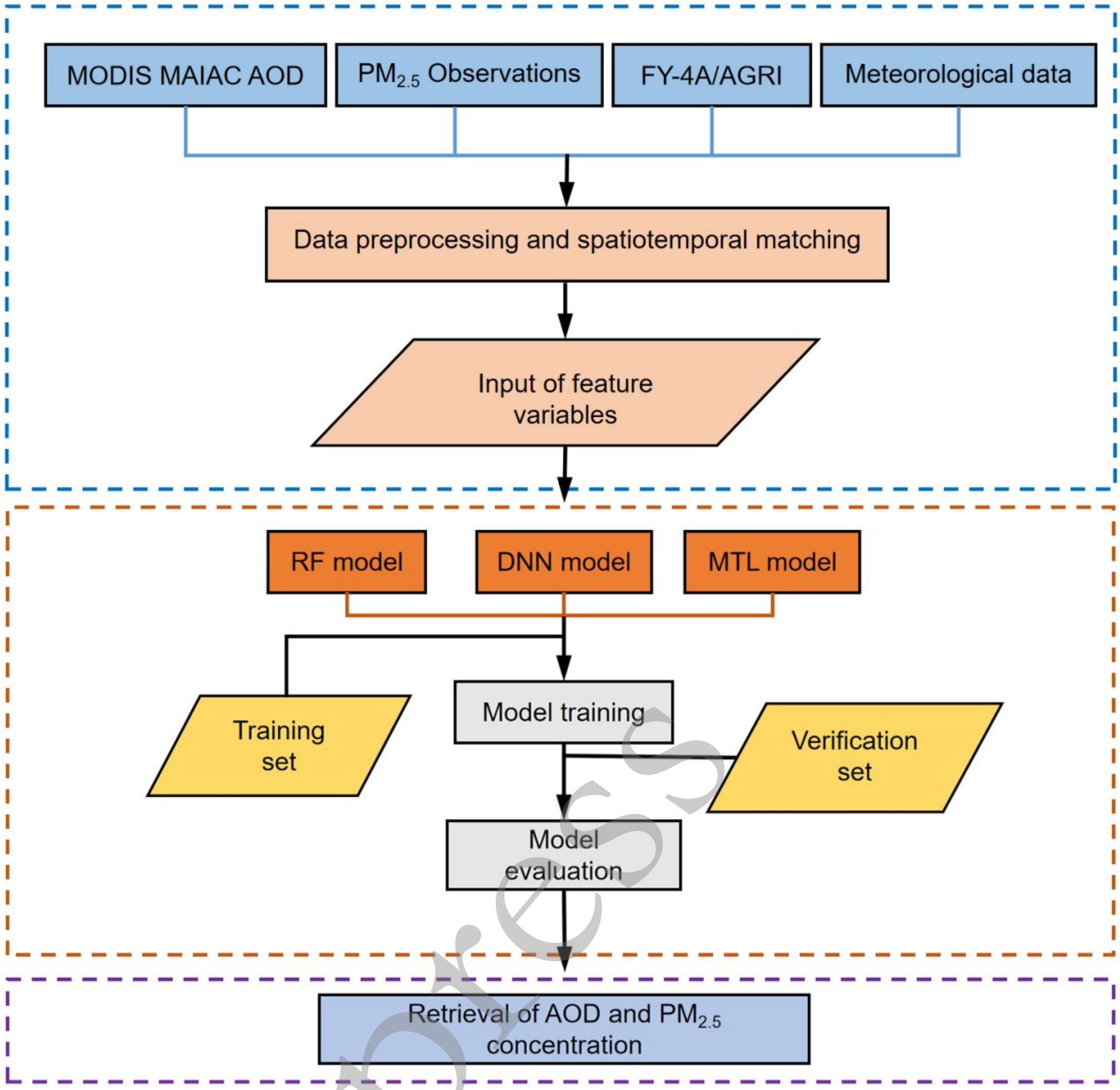


Fig. 2. Machine learning framework to retrieve AOD and $PM_{2.5}$ concentrations from various sites in China.

Table 1. Details of input datasets used in this study.

Field	Variables	Spatial resolution (km)	Main scientific objectives
FY-4A AGRI	0.45–0.49um (CH01)	1	Small particle aerosol, true color
	0.75–0.90um (CH03)	1	Vegetation, aerosols
	2.1–2.35um (CH06)	2-4	Cirrus cloud, aerosol, particle size
FY-4A geo-data	SAA	4	Position information from ground stations and satellite
	SAZ	4	
	SOA	4	
	SOZ	4	
	Lon	-	
	Lat	-	
ECMWF/ERA5	Blh	25	Planetary boundary layer height
	T2m	25	2-m air temperature
	U10	25	10-m eastward wind
	V10	25	10-m northward wind
	Tcwv	25	Total column water vapor
	Tco3	25	Total column ozone
	Sp	25	Surface pressure
Others	Mon	-	The month values were converted to a cosine distribution in the range of -1 to 1 to better represent seasonal effects.

3.2 STL models

In the field of satellite remote sensing, RF and DNN are two widely used machine-learning algorithms. RF is a popular ensemble learning method that has been widely applied for the retrieval of surface solar radiance (Shi et al., 2023), $PM_{2.5}$ and AOD estimation (Wei, et al., 2019; Chen et al., 2018). Similarly, She et al. (2020)

demonstrated that AOD retrieved with a DNN model has good reliability. Given such previous experience, RF and DNN are selected in this study as the STL models for benchmarking purposes.

RF is a classic ensemble learning method based on decision trees, which can quantify nonlinear relationships; its network structure is depicted in Fig. 3a. The algorithm has good robustness and the ability to deal with high-dimensional data and can deal with missing values and outliers effectively. In addition, the output results are easily interpretable and understandable.

The DNN is simply a multiplayer perceptron with more than two hidden layers (Yuan et al., 2020). The DNN architecture used in this paper is presented in Fig. 3b, which consists of one input layer (17 predictors), five hidden layers, and one output layer, which outputs either AOD or $PM_{2.5}$ depending on the learning task. Each hidden layer contains neurons, whose value is computed by taking a linear combination of all the neuron values from the previous layer using multiple weights and a bias term; in that the network is a fully connected one. This is followed by the application of a nonlinear activation function, to make the output more relevant to the predicted value. In this study, the rectified linear unit (ReLU) as the chosen as the activation function (Glorot et al., 2011).

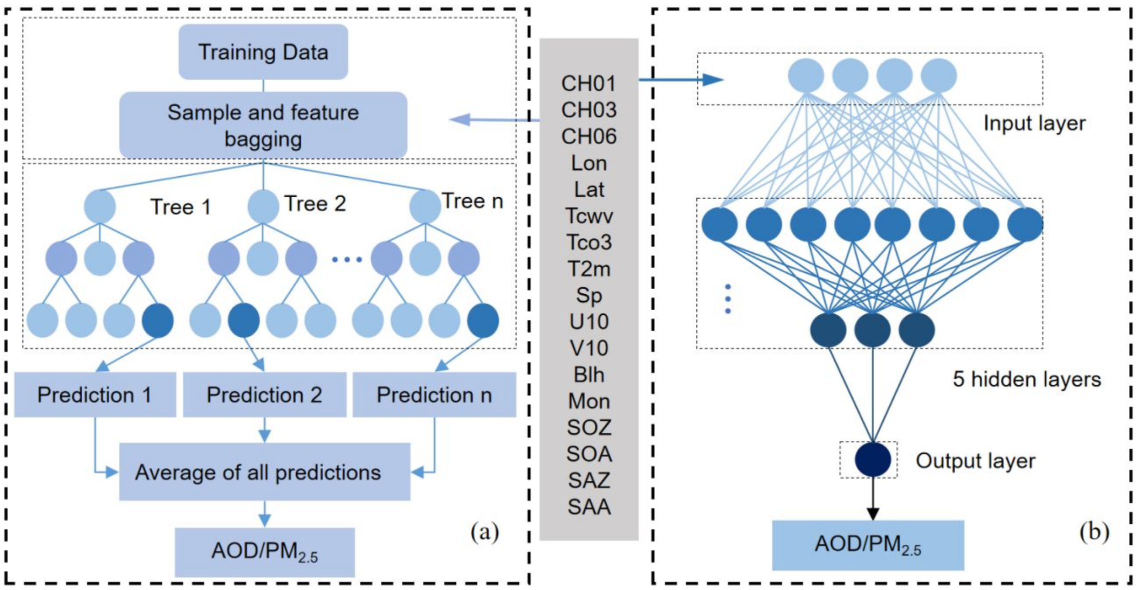


Fig. 3. The architecture of (a) random forest algorithm and (b) deep neural network.

3.3 MTL model

To incorporate the potential nonlinear spatial correlations between AOD and $PM_{2.5}$ retrievals, MTL is thought functional, as it is able to learn multiple related prediction tasks at the same time and to share the feature information of multiple tasks (Ruder, 2017). Compared with STL, MTL has stronger abilities in generalization and feature learning, and thus has become increasingly popular in the field of artificial intelligence (Ranjan et al., 2019). When performing MTL, there lie two main difficulties: handling multiple loss functions and designing an efficient parameter sharing mechanism.

When dealing with the loss function during MTL, a common practice is to add each individual loss function linearly, as illustrated in Eq. (1):

$$L_{\text{total}} = \sum w_i L_i \qquad (1)$$

where w_i and L_i represent the weights and losses of task *i*, respectively, and L_{total}

represents the total loss function. Since the overall performance heavily relies on the weights assigned to each loss function, traditional weighing methods lack a reliable basis and adjusting these weights can be time-consuming and laborious, which subsequently making it difficult to achieve the desired optimization result. In addition, there often exist conflicting objectives among multiple tasks, such as improving the performance of one task may lead to a decline in the performance of another task, which makes finding an effective set of weights to achieve simultaneous optimization of two tasks accompanied by high uncertainty.

Speaking of uncertainty, MTL is usually associated with both the cognitive uncertainty and accidental uncertainty. More specifically, cognitive uncertainty is caused by the model itself, where the outcome to be predicted is beyond the knowing range of the model; however, this uncertainty may be reduced as the training data points get more numerous. Accidental uncertainty includes heteroscedastic and homoscedastic uncertainties. Whereas the former refers to the uncertainty caused by differences in the input data, also known as data dependence, the latter is commonly used to characterize the data noise between different tasks, in that, the optimal weight for each task depends on its noise size (Kendall et al., 2018). In view of that, the main consideration, as to the weight assignment of the two learning tasks, is the homoscedastic uncertainty. In short, the total loss function is assumed to take the form:

$$L_{(w,\sigma_1,\sigma_2,)} = \frac{1}{2\sigma_1^2}L_1(w) + \frac{1}{2\sigma_2^2}L_2(w) + \log\sigma_1\sigma_2 \qquad (2)$$

where σ_1 and σ_2 signify uncertainties arising from data noise, serving as learnable parameters. As σ increases, the associated weight decreases; conversely, with decreasing noise σ, the corresponding weight increases. To mitigate overfitting, regularization terms are appended to the end of the loss function. During training, $\log \sigma_1 \sigma_2$ is introduced as a trainable variable, effectively constraining the loss function's variation range and preventing division by zero anomalies. In this study, the initial value of s is set as a random number greater than 0. As the number of training increases, the s representing the noise characteristics of $PM_{2.5}$ and AOD eventually converges to 3.7 and 0.7.

Parameter sharing in MTL has two main mechanisms: hard parameter sharing and soft-parameter sharing (Sun et al., 2019). Hard parameter sharing is suitable for tasks with strong correlation and can effectively reduce the risk of overfitting. However, when the correlation of outputs between tasks is poor, MTL with only hard parameter sharing may not fully satisfy the requirements of all tasks and deteriorate model performance instead. On the other hand, soft-parameter sharing allows each task to have separate parameters and hidden layers, while still enables information access among tasks. The way to achieve this is to regularize the loss at the output layer, by measuring the distance between the models (Maurer et al., 2012). The distance, as typically gauged with L_1 or L_2 loss, describes the parameter similarity between the same layer of different models, thereby encouraging the similarity of parameters across multiple tasks. This strategy achieves the goal of preventing complete independence in the specific layer and increasing the utilization of pertinent information.

The structure of MTL used in the study is shown in Fig. 4, which contains four shared layers and three specific layers for each task. Each layer is fully connected to the adjacent ones, with other techniques and settings including batch normalization, ReLU activation and dropout mechanism implemented for the network. Batch normalization seeks to accelerate convergence during training by normalizing the inputs in each small training batch of data. The shared layers adopt hard parameter sharing, which ensures that each neuron has identical parameters for both tasks. Consequently, a set of feature weights (w_1, w_2, ..., w_n) that exhibit generalization performance is obtained at the end of shared layers. The next steps involve specific layers. In these specific layers, each task is equipped with an independent model, possessing its own set of parameters. During training, the parameters of each model corresponding to each task are updated independently. Following each update, the distance between model parameters (L2 norm) is introduced as a regularization term to ensure similarity among parameters as much as possible. This notion of soft-parameter sharing is greatly inspired by MTL regularization techniques.

In addition to setting up the loss function and establishing the model structure, the selection of hyperparameters is also very important. For example, if the learning rate is set too high, then the parameter update may be too drastic, causing the loss function to not converge effectively; Larger batch sizes improve the stability of parameter updates. In the training process, we used a 10-fold cross-validation method to select the best hyperparameters of the MTL model. Specifically, we set the number of iterations, learning rate and batch size to 378, 0.001 and 1000, respectively, and selected the Adam optimizer to optimize the weight parameters in the neural network model.

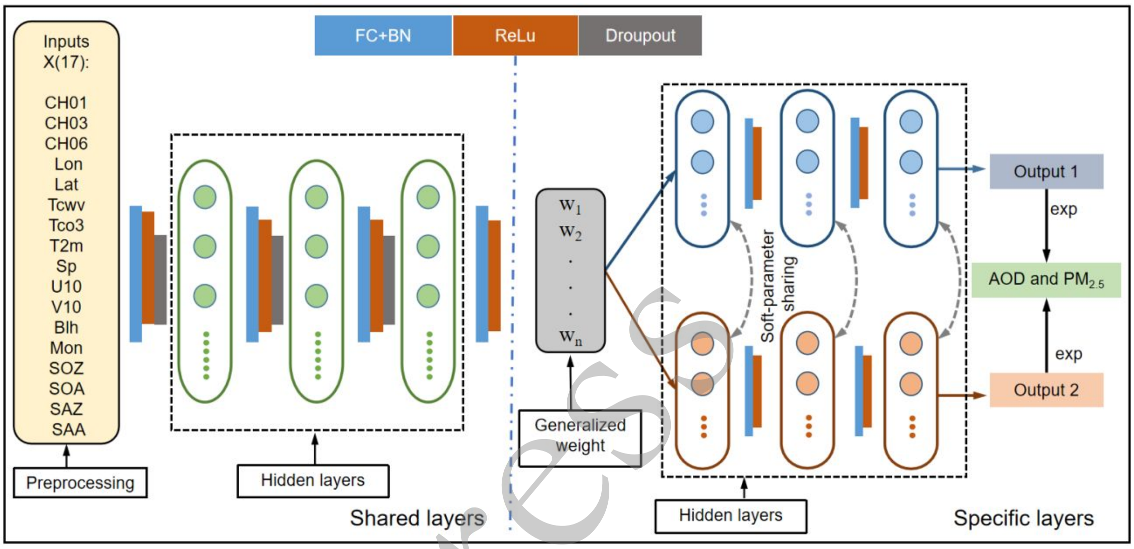


Fig. 4. The architecture of MTL model. Both shared layers (left) and specific layers (right) are full-connected (FC). Hard parameter sharing is used in shared layers while soft-parameter sharing is applied to the specific layers. ReLU is the activation layer, BN is a batch normalization layer.

3.4 Model Validation

A 10-fold cross-validation method is applied in the training dataset (recall Section 3.1) to select the best model parameters (Rodríguez et al., 2010). The testing dataset is used to evaluate the model performance based on statistical metrics R^2, RMSE, and mean absolute error (MAE). The EE is used to evaluate the accuracy of the MTL model, and the calculation formula is:

$$\mathrm{EE} = (1 \pm 0.15)\mathrm{y_i} \pm \mathrm{b_i} \qquad (3)$$

where y_i represents the label value of ground-based AOD or $PM_{2.5}$, and b_i represents the intercept, which takes the value of 0.05 for AOD verification and 5 $\mu g \cdot m^{-3}$ for $PM_{2.5}$ verification.

4 Results and discussion

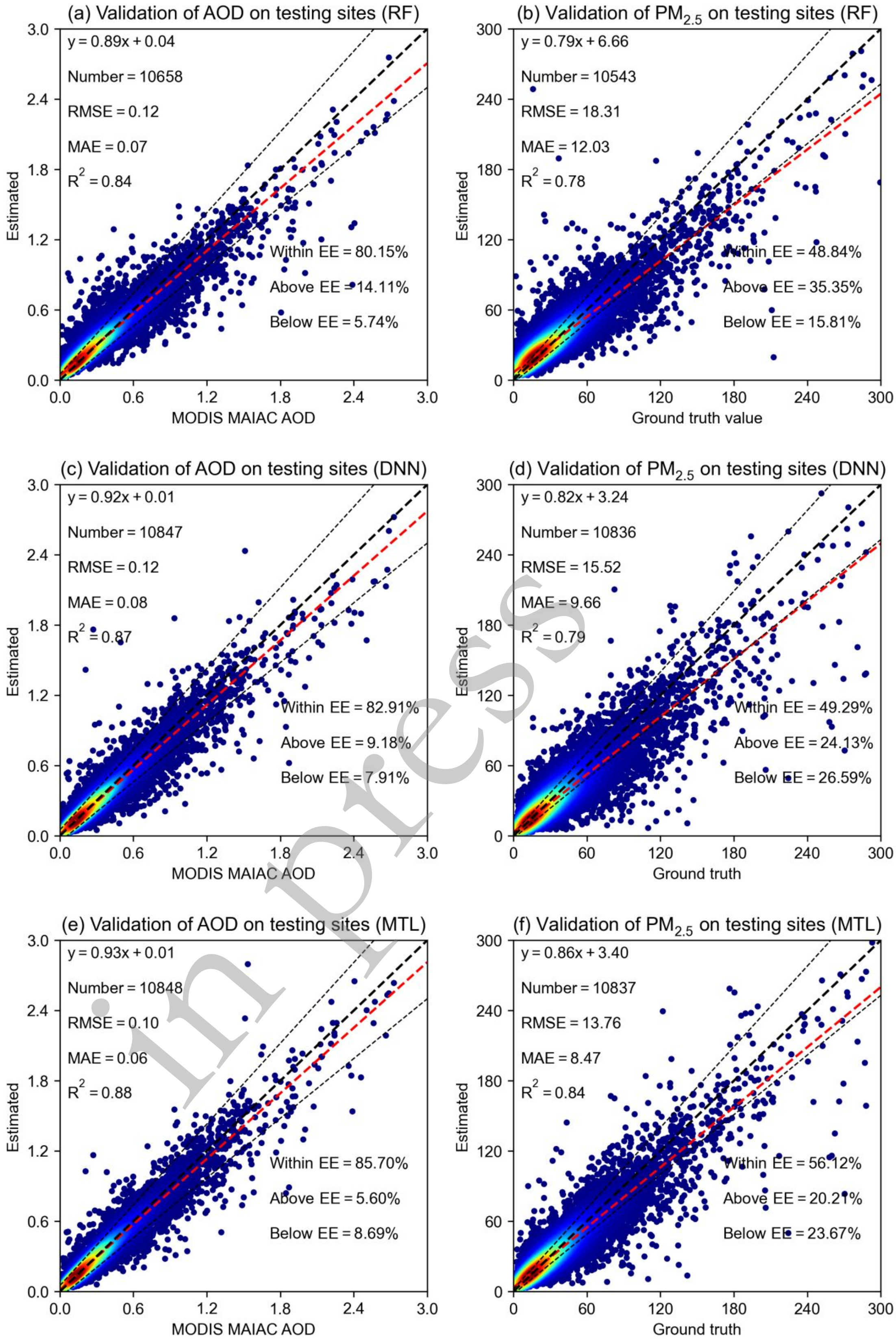
4.1 Comparison among three models

Here, both STL models and MTL models are validated with test data. First, according to the spatial distribution of ground stations, 10% of $PM_{2.5}$ observation stations (159 blue dots in Fig. 2) are uniformly selected as the test data set (out-of-station validation), and through spatio-temporal matching between these stations and MODIS MAIAC AOD, the AOD verification data set at the test station is obtained. The parameters used to evaluate model performance are R^2, RMSE, MAE, and EE.

Figure 5a displays the AOD predictions by the RF model, with respect to MAIAC AOD. The model has an R^2 of 0.84, an RMSE of 0.12, and an MAE of 0.07, with 80.15% of Within-EE predictions. However, the RF model tends to overestimate AOD for low MAIAC AOD values and underestimate it for high MAIAC AOD values. The DNN model (Fig. 5c) sees improvements over the RF model, with an R^2 of 0.87. However, there are no significant changes in RMSE and MAE (0.12 and

0.08). The MTL model (Fig. 5e) performs the best, with the evaluation metrics taking R^2 = 0.88, RMSE = 0.10, MAE = 0.06, and 85.70% Within-EE predictions.

In terms of the $PM_{2.5}$ prediction, results of STL model and MTL model are compared. It can be seen from Fig. 5b and Fig. 5d that the result of DNN-STL (R^2 = 0.79, RMSE = 15.52 $\mu g \cdot m^{-3}$, MAE = 9.66 $\mu g \cdot m^{-3}$) is better than that of RF-STL (R^2 = 0.78, RMSE = 18.31 $\mu g \cdot m^{-3}$, MAE = 12.03 $\mu g \cdot m^{-3}$), but the optimization effect is not obvious, Within-EE only increased by 0.45%. The R^2, RMSE, MAE, and Within-EE for the MTL model are 0.84, 13.76 $\mu g \cdot m^{-3}$, 8.47 $\mu g \cdot m^{-3}$ and 56.12% (Fig. 5f), respectively, with significant improvement in all metrics compared with the STL model. Compared with the traditional STL model, the MTL model developed in this study optimized both tasks (AOD and $PM_{2.5}$) to a certain extent. The results further show that the MTL model is more effective in optimizing $PM_{2.5}$ than AOD.



447

448 **Fig. 5.** Validation of estimated AOD and $PM_{2.5}$ on testing sites of RF (a, b), DNN (c,

d), and MTL model (e, f). The light dotted line represents the EE line, the dark dashed line represents the 1:1 line, and the red dashed line is the linear regression fitting line.

4.2 Site-specific model performance

Figure 6 illustrates the spatial distributions of the four-evaluation metrics for the MTL model on testing sites, revealing significant regional disparities and trends. In most sites across eastern and coastal China, the values of R^2 exceed 0.8 (Fig. 6a and Fig. 6c), indicating its capability to effectively capture variations in AOD and $PM_{2.5}$ with a high accuracy. However, the performance of the MTL model experiences a substantial decline in western China (R^2 values are generally less than 0.5), particularly in areas characterized by complex topography like the Tibetan Plateau. This observation can be primarily attributed to the scarcity of ground-based observation sites in that region, hindering the ability of the model to adapt to the unique aerosol patterns prevalent in this distinctive environment (Fang et al., 2016).

Moreover, AOD and $PM_{2.5}$ exhibit a pronounced north-to-south gradient in terms of their RMSE distribution in China (Fig. 6b and Fig. 6d), which is consistent with the findings of previous studies (Wei, et al., 2019). Specifically, sites in the North China Plain (NCP), which is an area known for its significant anthropogenic emissions, and Northwest China, which is characterized by substantial natural dust emissions (Gui et al., 2020), show higher RMSE values for AOD and $PM_{2.5}$, exceeding 0.18 and 18 $\mu g \cdot m^{-3}$.

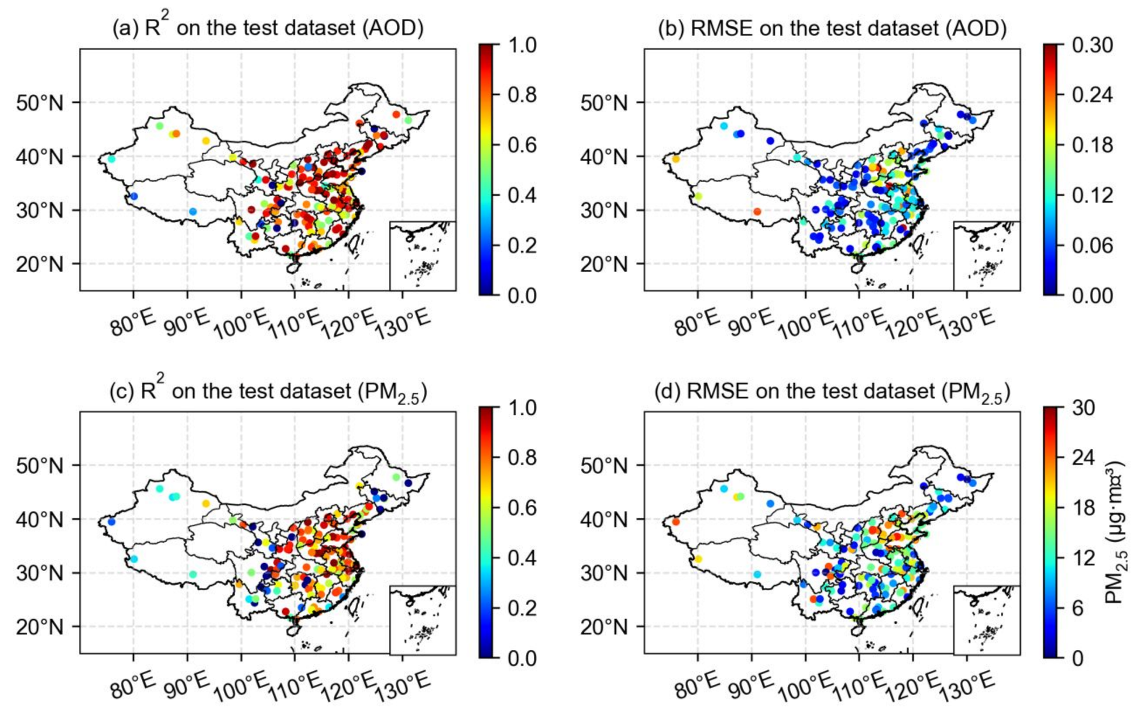


Fig. 6. Spatial distribution of MTL model evaluation metrics (a, c: R^2 for AOD and $PM_{2.5}$, b, d: RMSE for AOD and $PM_{2.5}$.)

In order to comprehensively assess the discrepancy between the estimation results of the MTL model and the ground-based observations,we utilized the AOD and $PM_{2.5}$ ground-based observations for the period of March 12, 2018 to March 11, 2019 to validate the accuracy of the MTL model. These data include AOD observations from 38 AERONET stations in China and $PM_{2.5}$ data from 15 air quality monitoring stations in Hong Kong. Fig 7a demonstrates the comparison results between the AERONET AOD and the MTL AOD, with an R^2 of 0.79, an RMSE of 0.14, an MAE of 0.10, and a Within-EE of 74.09%. In addition, Fig 7b demonstrates the results of the comparison between the MTL $PM_{2.5}$ and the ground-level $PM_{2.5}$ observations in Hong Kong, with an R^2 of 0.76, an RMSE of 8.11 $\mu g \cdot m^{-3}$, an MAE of

5.86 μg·m^{-3}, and a Within-EE of 61.35%. These results show that the predictions of the MTL model are in good agreement with the ground observations, verifying the reliability of the MTL model.

Due to the limited number of ground-based observation sites in Northwest China, it is difficult to provide sufficient data for validating the performance of the MTL model in the region. Therefore, in this study, observation sites distributed in 28 cities in the region were selected and LGHAP AOD and LGHAP $PM_{2.5}$ data from March 12, 2018, to March 11, 2019, were obtained by spatio-temporal matching. In Fig. 7c, the results of the comparison between MTL AOD and LGHAP AOD are demonstrated, with R^2 of 0.76, RMSE of 0.13, MAE of 0.10, and Within-EE of 61.28%. And in Fig. 7d, the results of comparison between MTL $PM_{2.5}$ and LGHAP $PM_{2.5}$ are demonstrated with R^2 of 0.70, RMSE of 15.44 μg·m^{-3}, MAE of 9.90 μg·m^{-3}, and Within-EE of 51.45%. The results show that the validation results of the MTL model with ground-based observation sites are better than the validation results of the MTL model with the LGHAP AOD and LGHAP $PM_{2.5}$ in Northwest China, especially in terms of $PM_{2.5}$. This is mainly due to the limited training data in the region and the fact that most of the region is in plateaus, basins and mountains with cold and dry winters and high summer temperatures. In particular, the Taklamakan Desert surrounded by high mountains and the Gobi Desert in Inner Mongolia are one of the dust source areas in China, which makes the aerosol types in Northwest China more complex and diverse.

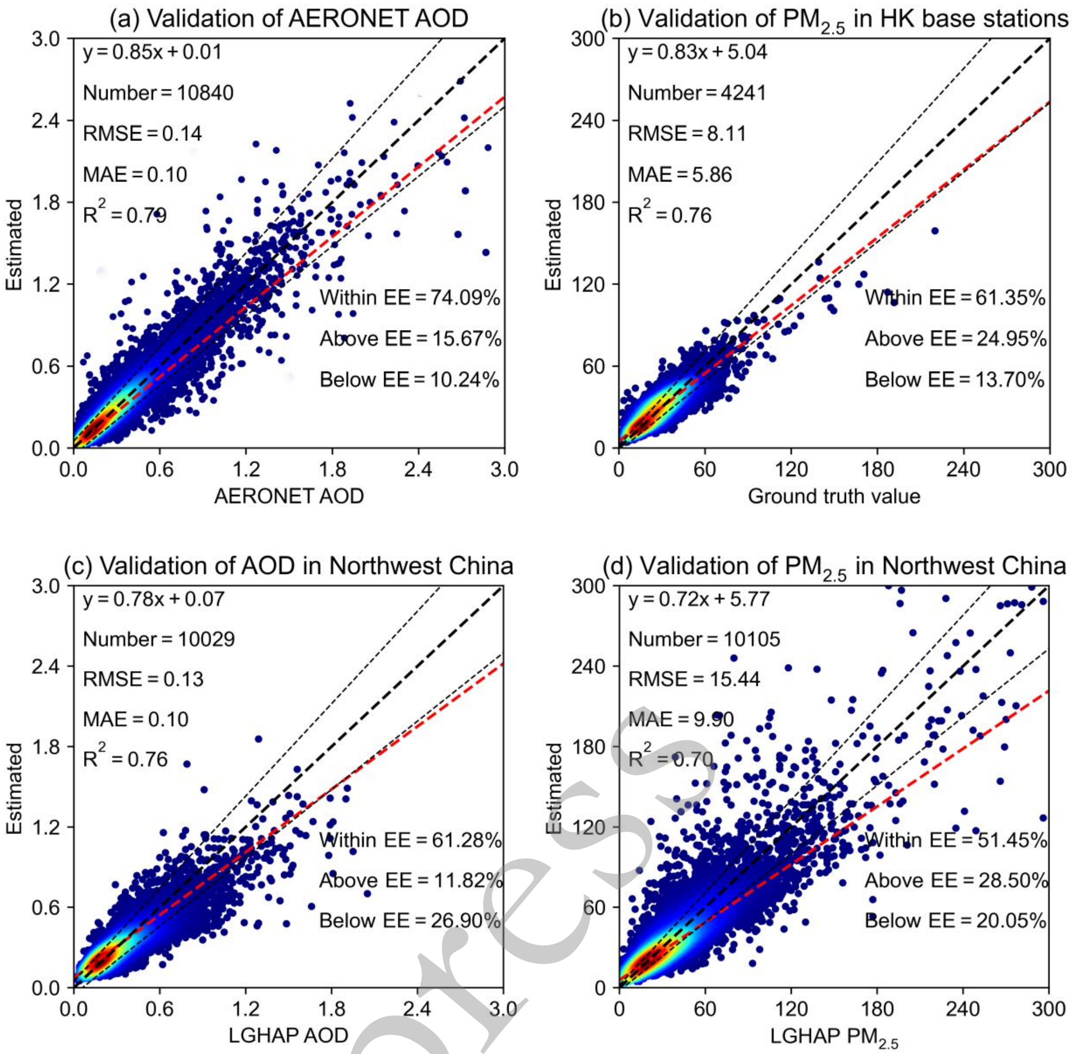


Fig. 7. Scatter plot of validation results of MTL AOD and $PM_{2.5}$ in different regions. (a) Comparison results of AERONET AOD and MTL AOD. (b) Comparison results of HK Ground-based Observations $PM_{2.5}$ and MTL $PM_{2.5}$. (c), (d):The results of MTL AOD ($PM_{2.5}$) and LGHAP AOD ($PM_{2.5}$) were compared in Northwest China.

4.3 Feature importance of MTL model

Owing to the "black box" nature of ML, interpreting the output has hitherto been a demanding task. In this study, this issue is addressed by employing the DeepLift algorithm from the Captum core library. The DeepLift algorithm allows us to compare model predictions to a reference, enabling the quantification of the importance of each feature through a modified backpropagation technique. It is worth noting that we set 0 as the reference baseline for all features when computing imputation values. Comprehensive Captum tutorials are available on the official website (https://captum.ai/).

Figure 8 shows the attribution score exhibiting the significance of the independent variable on the dependent variable (AOD or $PM_{2.5}$). Positive attribution scores indicate that the feature positively contributes to the predicted value, while negative scores suggest the opposite. In general, the feature importance of the two tasks has a certain similarity owing to inner connection between two related tasks and the parameter sharing mechanism. Specifically, the attribution scores of the visible channels (CH01 and CH03) are relatively high and have negative directions, –0.19 and –0.15 respectively. Larger values of these two features result in smaller values of corresponding estimation. When the reflectance of the visible channels is lower, it usually indicates a stronger aerosol extinction capability in the atmosphere. The attribution scores of Tcwv in AOD (0.12) and $PM_{2.5}$ (0.04) are both positive values. In general, when Tcwv increases, it will cause more extinction by water vapor and also affect the aggregation and sedimentation of particulate matter, thereby increasing AOD and $PM_{2.5}$ concentration. The relative lower attribution score of Tcwv in $PM_{2.5}$ may be associated with the fact that $PM_{2.5}$ represents the dry mass concentration of fine particulate, which is hardly affected by water vapor.

Furthermore, the seasonal effects (Mon) and geographical factor (Lon and Lat) are critical in AOD and $PM_{2.5}$ estimation, which represent different temporal and spatial heterogeneity. It is worth noting that due to the independence of the two tasks, there will also be some features with opposite attribution score signs (i.e., CH06, U10 and SAZ), which also indicates that the MTL model will have a certain negative transfer due to different task requirements. Therefore, the MTL model requires high correlation between different tasks.

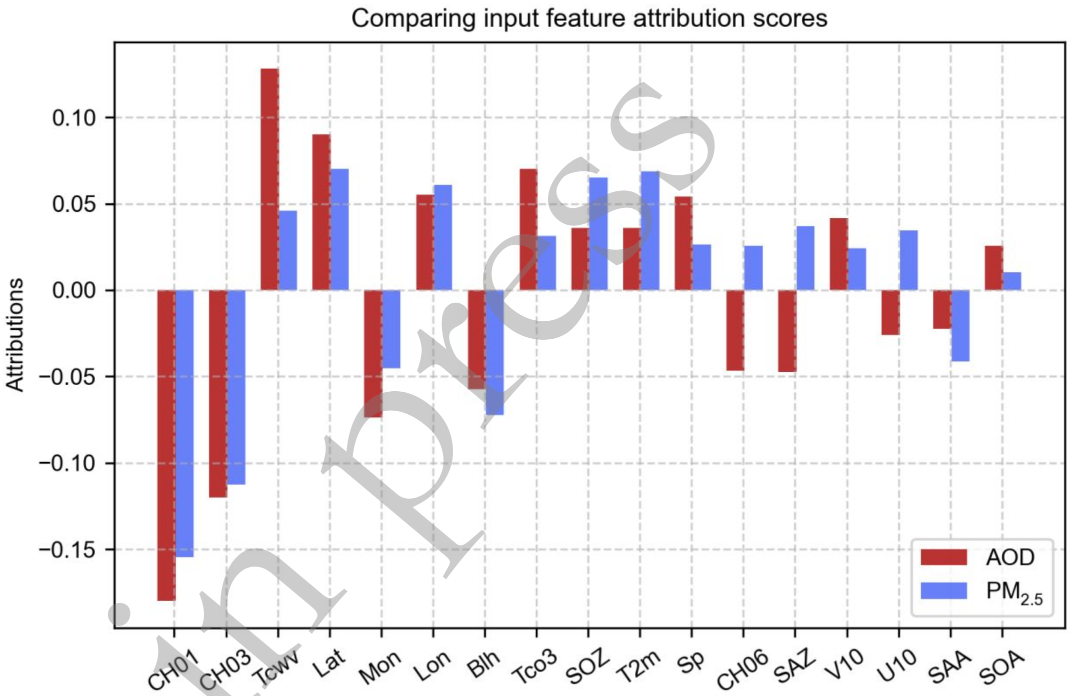


Fig. 8. Attribution scores for MTL model input features. The red and blue bars represent the importance of the features for the AOD and $PM_{2.5}$ estimation, respectively. Positive values indicate a positive contribution to the estimation, while negative values signify the opposite.

4.4 Spatial distribution of seasonal averages

Figure 9 shows the spatial distribution of seasonal means of MTL AOD and MTL $PM_{2.5}$ for the period from March 12, 2018 to March 11, 2019. The seasonal mean values of MTL AOD are higher in spring (0.45) and summer (0.40) and relative lower in autumn (0.33) and winter (0.38), which is consistent with previous findings (Chen et al., 2023). The rise in aerosol loading in spring (Fig. 9a) can be attributed to frequent spring dust events in the north, leading to peaks in natural dust and windborne sand at the surface (He et al., 2016). Higher Blh in summer lead to vertical transport of aerosol particles to higher altitudes, which further enhances complete mixing of aerosols with water within the boundary layer. As a result, smaller aerosol particles within the boundary layer grow to optically active sizes (Qu et al., 2016). Higher temperatures in summer enhance photochemical reactions and also lead to an increase in aerosol loading during this season (Qi et al., 2013). Autumn is usually accompanied by stable atmospheric circulation and favorable diffusion conditions that favor dispersion and dilution of particulate matter, thus reducing aerosol loading. In contrast, in winter, lower Blh may result in particulate matter not being efficiently transported and mixed to higher altitudes, leading to a reduction in AOD (Qu et al.,2016). However, higher spatial distributions of AOD may still occur in areas with high emissions and complex topography, such as Sichuan and Chongqing, where special topography makes aerosol transport difficult.

Previous studies have indicated that $PM_{2.5}$ concentrations are higher in winter and lower in summer (Li et al., 2017; Leung et al., 2020), which is similar to the MTL results. $PM_{2.5}$ concentration are generally low in summer (26.45 μg·m^{-3}) and autumn (30.81 μg·m^{-3}) in China and showed similar spatial distributions (Fig. 9f and Fig. 9g). On the contrary, $PM_{2.5}$ concentrations were significantly higher in spring (37.67 μg·m^{-3}) and winter (41.51 μg·m^{-3}), especially in winter (Fig. 9e and Fig. 9h). The main reasons are the frequent sandstorms and the long-distance transmission of sand and dust in spring and the burning of coal and fossil fuels for heating in winter leading to more pollutant emissions in northern China (Wei et al., 2021).

A comparison of seasonal AOD and $PM_{2.5}$ distribution reveals elevated aerosol levels over the North China Plain (NCP), Sichuan Basin, and Chongqing, attributed to factors such as intense human activities, adverse climatic conditions, and geographical features like basins that intensify anthropogenic aerosol emissions (Wang et al., 2018). Conversely, Northeast China, including Heilongjiang and Jilin, as well as Southwest China, such as Tibet and Yunnan, exhibit comparatively lower AOD and $PM_{2.5}$ levels due to sparse populations and reduced anthropogenic aerosol emissions. Furthermore, the favorable terrain and climatic conditions in these regions facilitate the dispersion of pollutants (Su et al., 2018).

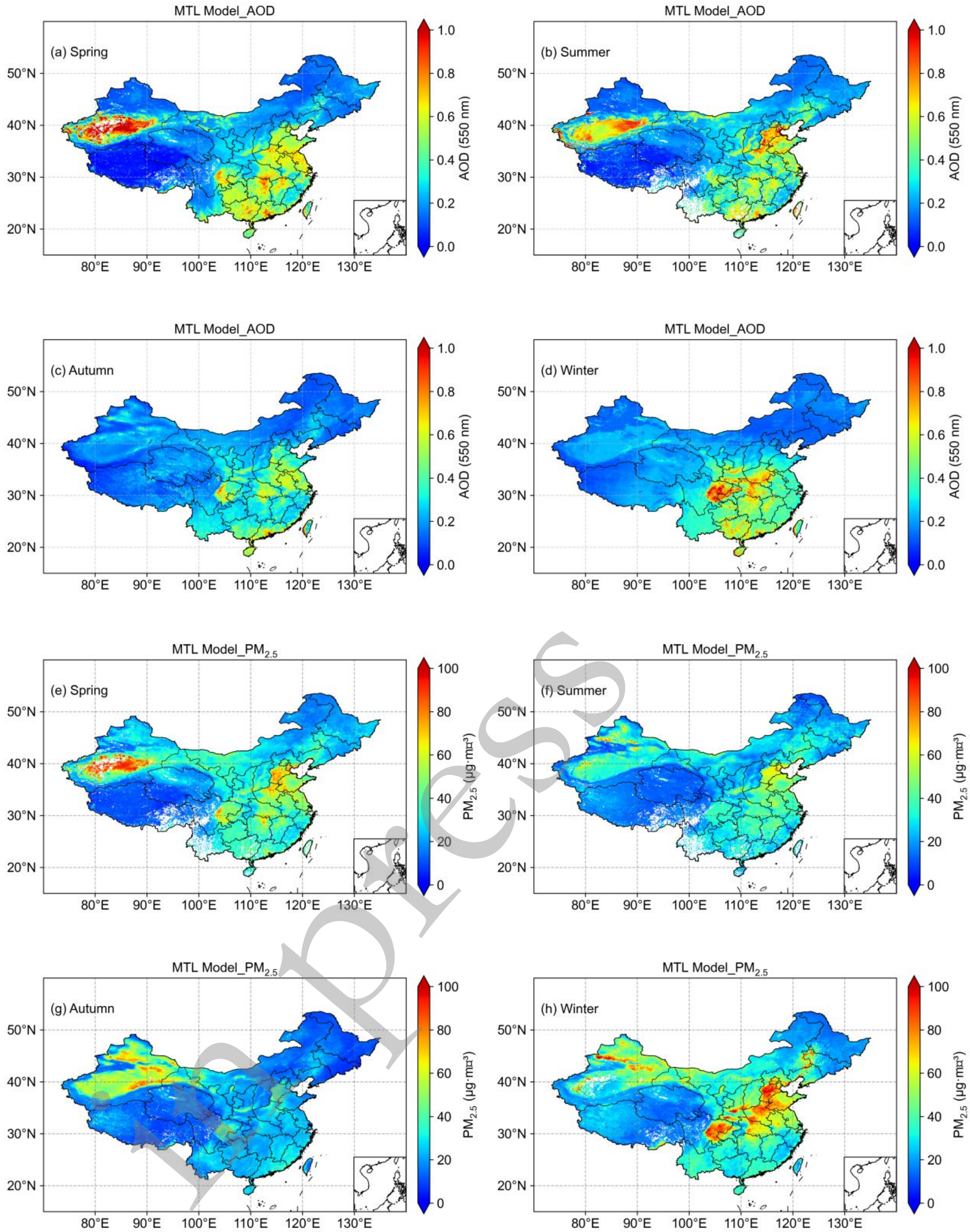


Fig. 9. Spatial distribution of seasonal mean AOD and $PM_{2.5}$ in China in 2018 for (a), (e) spring (March to May); (b), (f) summer (June to August); (c), (g) autumn (September to November); and (d), (h) winter (December to February).

Fig. 10a and Fig. 10b show the spatial distribution of MTL AOD and LGHAP AOD, which have high consistency. The high values of MTL AOD and LGHAP AOD are located in the Taklamakan Desert and Guanzhong region of Shaanxi, and the annual average AOD ranges from 0.35 to 0.70. In Qinghai, northern Xinjiang, and the vast portion of Inner Mongolia, the annual average values of AOD are generally less than 0.3. Fig. 10c and Fig. 10d show spatial distribution of MTL $PM_{2.5}$ and LGHAP $PM_{2.5}$, with annual mean $PM_{2.5}$ values exceeding 50 $\mu g \cdot m^{-3}$ in the Tarim Basin, northern Xinjiang, and Guanzhong region of Shaanxi. The topographic and climatic characteristics of the Tarim Basin restrict air movement and dispersion, leading to the accumulation of pollutants in the basin. The Guanzhong region of Shaanxi has well-developed industrial and transportation activities, resulting in a large amount of near-surface release of particulate matter. Overall, the annual mean value of LGHAP $PM_{2.5}$ is significantly higher than that of MTL $PM_{2.5}$, especially in Qinghai Province, where LGHAP $PM_{2.5}$ is about 30 $\mu g \cdot m^{-3}$ compared to MTL $PM_{2.5}$ of about 20 $\mu g \cdot m^{-3}$. The main reason is that the LGHAP dataset is a gap-free aerosol product generated by reconstructing MODIS daily AOD gaps, while the MTL model needs to retrieve AOD under cloud-free conditions. The presence of clouds reduces the spatial and temporal continuity of the MTL model estimates, which results in the differences in the spatial distributions of AOD and $PM_{2.5}$.

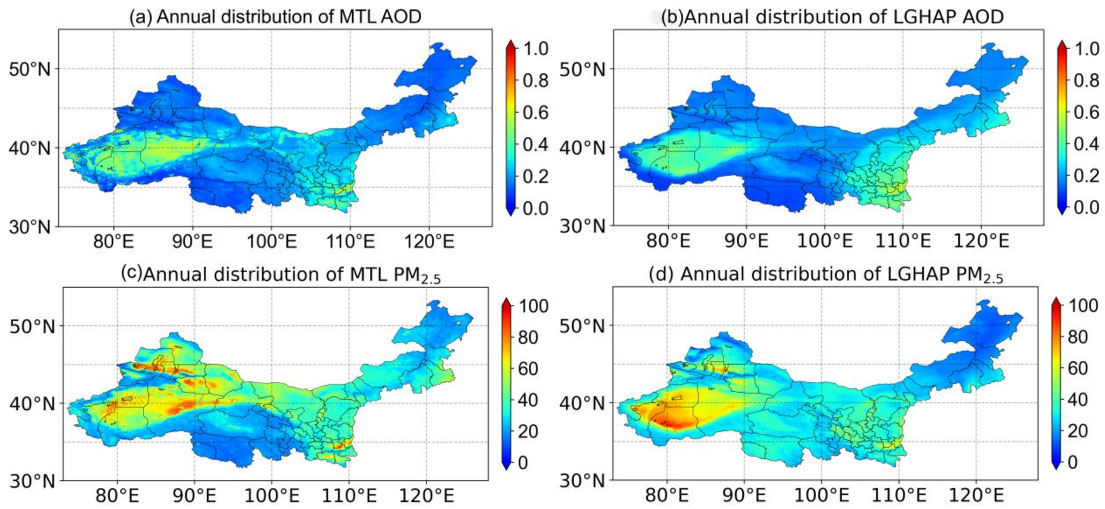


Fig. 10. In the figure, (a) and (b) are the spatial distribution maps of MTL AOD and LGHAP AOD in Northwest China, and (c) and (d) are the spatial distribution maps of MTL $PM_{2.5}$ and LGHAP $PM_{2.5}$ in Northwest China.

5 Conclusion

A high-precision and high-efficiency method for aerosol observation is of great significance for ameliorating air quality, mitigating climate change, protecting ecosystem, and improving human health. In order to improve the retrieval accuracy and efficiency of AOD and $PM_{2.5}$, a MTL algorithm based on parameter sharing mechanism is proposed in this study. The constructed MTL model is able to retrieve AOD and $PM_{2.5}$ simultaneously based on FY-4A AGRI data and ERA5 reanalysis. The main conclusions are as follows.

(1) The estimated AOD and $PM_{2.5}$ from three models were evaluated at the testing sites. The results showed that the MTL model predicted AOD (R^2 = 0.88, RMSE = 0.10, Within-EE = 85.70%) and $PM_{2.5}$ (R^2 = 0.84, RMSE = 13.76 $\mu g \cdot m^{-3}$, Within-EE = 56.12%) the best compared to the conventional STL models (RF and

DNN). Furthermore, independent validation against AERONET sites in China and $PM_{2.5}$ observation sites in HK confirmed the generalization and reliability of MTL model (AOD: R^2 = 0.79, RMSE = 0.14, Within-EE = 74.09% and $PM_{2.5}$: R^2 = 0.76, RMSE = 8.11 $\mu g \cdot m^{-3}$, Within-EE = 61.35%). The performance of MTL model in Northwest China is mainly affected by geographical conditions and the number of training data.

(2) The feature contribution score of MTL model is calculated by using attribution algorithm. The results show that the feature contribution distributions of the two tasks are similar, but a few features have opposite contributions to the two tasks, which is related to the independence of the tasks. Poor correlation between tasks can lead to negative transfer effects in multiple MTL models across different tasks.

(3) The MTL model was used to estimate the AOD and $PM_{2.5}$ concentrations in China in 2018, and the spatial distribution map was drawn. The results showed that the highest and lowest AOD values were found in spring (0.45) and autumn (0.33). In contrast, the seasonal variation of $PM_{2.5}$ is large, with the highest and lowest $PM_{2.5}$ concentrations in winter (41.51 $\mu g \cdot m^{-3}$) and summer (26.45 $\mu g \cdot m^{-3}$). In addition, the difference of spatial distribution in Northwest China mainly depends on the spatio-temporal continuity of MTL model estimation results, and an effective interpolation algorithm can improve the integrity of MTL model estimation results.

The results of this study are based on the sample data collected from 1590 ground-based sites in China. However, the distribution of these sites is uneven, which further affects the overall training effect of the model, therefore, the estimation results in Northwest China have further room for improvement. In future work, using more efficient cloud detection algorithms and interpolation algorithms can further improve the accuracy and completeness of MTL model estimation results.

Acknowledgments

The work was supported by the Opening Foundation of Key Laboratory of Atmospheric Sounding, the China Meteorological Administration and the CMA Research Center on Meteorological Observation Engineering Technology (U2021Z03), the National Natural Science Foundation of China (Grant No. 42030708, 42375138, 42030608, 42105128), the Opening Foundation of Key Laboratory of Atmospheric Chemistry, China Meteorological Administration (2022B02).

REFERENCES

Bai, H., Zheng, Z., Zhang, Y., Huang, H., & Wang, L. (2021). Comparison of satellite-based $PM_{2.5}$ estimation from aerosol optical depth and top-of-atmosphere reflectance. *Aerosol and Air Quality Research*, ***21*****(2)**, 1–17. https://doi.org/10.4209/aaqr.2020.05.0257.

Bai, K., Li, K., Ma, M., Li, K., Li, Z., Guo, J., et al., 2022. LGHAP: the Long-term Gap-free High-resolution Air Pollutant concentration dataset, derived via tensor-flow-based multimodal data fusion. Earth Syst. Sci. Data **14 (2)**, 907–927. https://doi.org/ 10.5194/essd-14-907-2022.

Chen, G., Li, S., Knibbs, L. D., Hamm, N. A. S., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M. J., & Guo, Y. (2018). A machine learning method to estimate $PM_{2.5}$ concentrations across China with remote sensing, meteorological and land use information. Science of the Total Environment, 636, 52–60. https://doi.org/10.1016/j.scitotenv.2018.04.251.

Chen, X., de Leeuw, G., Arola, A., Liu, S., Liu, Y., Li, Z., & Zhang, K. (2020). Joint retrieval of the aerosol fine mode fraction and optical depth using MODIS spectral reflectance over northern and eastern China: Artificial neural network method. Remote Sensing of Environment, **249**. https://doi.org/10.1016/j.rse.2020.112006.

Chen, A., Yang, J., He, Y., Yuan, Q., Li, Z., & Zhu, L. (2023). High spatiotemporal resolution estimation of AOD from Himawari-8 using an ensemble machine learning gap-filling method. Science of the Total Environment, **857**. https://doi.org/10.1016/j.scitotenv.2022.159673.

Deng, X., Shi, C., Wu, B., Chen, Z., Nie, S., He, D., & Zhang, H. (2012). Analysis of aerosol characteristics and their relationships with meteorological parameters over Anhui province in China. Atmospheric Research, **109–110**, 52–63. https://doi.org/10.1016/j.atmosres.2012.02.011.

Ding, H., Zhao, L., Liu, S., Chen, X., de Leeuw, G., Wang, F., Zheng, F., Zhang, Y., Liu, J., Li, J., She, L., Si, Y., & Gu, X. (2022). FY-4A/AGRI Aerosol Optical Depth Retrieval Capability Test and Validation Based on NNAeroG. Remote Sensing, **14(21)**. https://doi.org/10.3390/rs14215591.

Fang, X., Zou, B., Liu, X., Sternberg, T., & Zhai, L. (2016). Satellite-based ground $PM_{2.5}$ estimation using timely structure adaptive modeling. Remote Sensing of Environment, **186**, 152–163. https://doi.org/10.1016/j.rse.2016.08.027.

Fu, D., Gueymard, C. A., & Xia, X. (2023). Validation of the improved GOES-16 aerosol optical depth product over North America. Atmospheric Environment, **298**. https://doi.org/10.1016/j.atmosenv.2023.119642.

Fu, D., Shi, H., Gueymard, C. A., Yang, D., Zheng, Y., Che, H., Fan, X., Han, X., Gao, L., Bian, J., Duan, M., & Xia, X. (2024). A Deep-Learning and Transfer-Learning Hybrid Aerosol Retrieval Algorithm for FY4-AGRI: Development and Verification over Asia. Engineering.

https://doi.org/10.1016/j.eng.2023.09.023.

Glorot X, Bordes A, Bengio Y., 2011: Deep Sparse Rectifier Neural Networks. Journal of Machine Learning Research, **15**,315-323.

Gui, K., Che, H., Zeng, Z., Wang, Y., Zhai, S., Wang, Z., Luo, M., Zhang, L., Liao, T., Zhao, H., Li, L., Zheng, Y., & Zhang, X. (2020). Construction of a virtual $PM_{2.5}$ observation network in China based on high-density surface meteorological observations using the Extreme Gradient Boosting model. Environment International, **141**. https://doi.org/10.1016/j.envint.2020.105801.

Gupta, P., & Christopher, S. A. (2009). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. Journal of Geophysical Research Atmospheres, **114(14)**. https://doi.org/10.1029/2008JD011496.

He, Q., Zhang, M., & Huang, B. (2016). Spatio-temporal variation and impact factors analysis of satellite-based aerosol optical depth over China from 2002 to 2015. Atmospheric Environment, **129**, 79–90. https://doi.org/10.1016/j.atmosenv.2016.01.002.

Hill, W., Lim, E. L., Weeden, C. E., Lee, C., Augustine, M., Chen, K., Kuan, F. C., Marongiu, F., Evans, E. J., Moore, D. A., Rodrigues, F. S., Pich, O., Bakker, B., Cha, H., Myers, R., van Maldegem, F., Boumelha, J., Veeriah, S., Rowan, A., … Swanton, C. (2023). Lung adenocarcinoma promotion by air pollutants. Nature, **616(7955)**, 159–167. https://doi.org/10.1038/s41586-023-05874-3.

Ho, H. C., Wong, M. S., Yang, L., Shi, W., Yang, J., Bilal, M., & Chan, T. C. (2018). Spatiotemporal influence of temperature, air quality, and urban environment on cause-specific mortality during hazy days. Environment International, **112**, 10–22. https://doi.org/10.1016/j.envint.2017.12.001.

Hu, X., Waller, L. A., Al-Hamdan, M. Z., Crosson, W. L., Estes, M. G. Jr., Estes, S. M., et al. (2013). Estimating ground-level $PM_{2.5}$ concentrations in the southeastern U.S. using geographically weighted regression. Environmental Research, **121**, 1–10. https://doi.org/10.1016/j.envres.2012.11.003.

Hu, X., Waller, L. A., Lyapustin, A., Wang, Y., Al-Hamdan, M. Z., Crosson, W. L., Estes, M. G., Estes, S. M., Quattrochi, D. A., Puttaswamy, S. J., & Liu, Y. (2014). Estimating ground-level $PM_{2.5}$ concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. Remote Sensing of Environment, **140**, 220–232. https://doi.org/10.1016/j.rse.2013.08.032.

Kaufman, Y. J., Tanré, D., Remer, L. A., Vermote, E. F., Chu, A., & Holben, B. N. (1997). Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. Journal of Geophysical Research Atmospheres, **102(14)**, 17051–17067. https://doi.org/10.1029/96jd03988.

Kendall, A., Gal, Y., & Cipolla, R. (2017). Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. [Available online from http://arxiv.org/abs/1705.07115]

Kokhanovsky, A. A., Breon, F. M., Cacciari, A., Carboni, E., Diner, D., Di Nicolantonio, W., Grainger, R. G., Grey, W. M. F., Höller, R., Lee, K. H., Li, Z., North, P. R. J., Sayer, A. M., Thomas, G. E., & von Hoyningen-Huene, W. (2007). Aerosol remote sensing over land: A comparison of satellite retrievals using different algorithms and instruments. Atmospheric Research, **85(3–4)**, 372–394. https://doi.org/10.1016/j.atmosres.2007.02.008.

Lee, H. J., Liu, Y., Coull, B. A., Schwartz, J., & Koutrakis, P. (2011). A novel calibration approach of MODIS AOD data to predict $PM_{2.5}$ concentrations. Atmospheric Chemistry and Physics, **11(15)**, 7991–8002. https://doi.org/10.5194/acp-11-7991-2011.

Leung, D. M., Shi, H., Zhao, B., Wang, J., Ding, E. M., Gu, Y., Zheng, H., Chen, G., Liou, K. N., Wang, S., Fast, J. D., Zheng, G., Jiang, J., Li, X., & Jiang, J. H. (2020). Wintertime Particulate Matter Decrease Buffered by Unfavorable Chemical Processes Despite Emissions Reductions in China. Geophysical Research Letters, **47(14)**. https://doi.org/10.1029/2020GL087721.

Li, J., Carlson, B. E., & Lacis, A. A. (2015). How well do satellite AOD observations represent the spatial and temporal variability of $PM_{2.5}$ concentration for the United States? Atmospheric Environment, **102**, 260–273. https://doi.org/10.1016/j.atmosenv.2014.12.010.

Li, T., Shen, H., Zeng, C., Yuan, Q., & Zhang, L. (2017). Point-surface fusion of station measurements and satellite observations for mapping $PM_{2.5}$ distribution in

China: Methods and assessment. Atmospheric Environment, **152**, 477–489. https://doi.org/10.1016/j.atmosenv.2017.01.004.

Lyapustin, A., Wang, Y., Korkin, S., & Huang, D. (2018). MODIS Collection 6 MAIAC algorithm. Atmospheric Measurement Techniques, **11(10)**, 5741–5765. https://doi.org/10.5194/amt-11-5741-2018.

Mao, F., Hong, J., Min, Q., Gong, W., Zang, L., & Yin, J. (2021). Estimating hourly full-coverage $PM_{2.5}$ over China based on TOA reflectance data from the Fengyun-4A satellite. Environmental Pollution, **270**. https://doi.org/10.1016/j.envpol.2020.116119.

Maurer, A., Pontil, M., & Romera-Paredes, B. (2012). Sparse coding for multitask and transfer learning. [Available online from http://arxiv.org/abs/1209.0738]

Miao, Y., & Liu, S. (2019). Linkages between aerosol pollution and planetary boundary layer structure in China. Science of the Total Environment, **650**, 288–296. https://doi.org/10.1016/j.scitotenv.2018.09.032.

Pak, U., Kim, C., Ryu, U., Sok, K., & Pak, S. (2018). A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction. Air Quality, Atmosphere and Health, **11(8)**, 883–895. https://doi.org/10.1007/s11869-018-0585-1.

Qi, Y. L., Ge, J. M., & Huang, J. P. (2013). Spatial and temporal distribution of MODIS and MISR aerosol optical depth over northern China and comparison with AERONET. Chinese Science Bulletin, **58(20)**, 2497–2506.

https://doi.org/10.1007/s11434-013-5678-5.

Qin, W., Fang, H., Wang, L., Wei, J., Zhang, M., Su, X., Bilal, M., & Liang, X. (2021). MODIS high-resolution MAIAC aerosol product: Global validation and analysis. Atmospheric Environment, 264. https://doi.org/10.1016/j.atmosenv.2021.118684.

Qu, W., Wang, J., Zhang, X., Sheng, L., & Wang, W. (2016). Opposite seasonality of the aerosol optical depth and the surface particulate matter concentration over the north China Plain. Atmospheric Environment, **127**, 90–99. https://doi.org/10.1016/j.atmosenv.2015.11.061.

Ranjan, R., Patel, V. M., & Chellappa, R. (2016). HyperFace: A Deep Multi-task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition. [Available online from http://arxiv.org/abs/1603.01249]

Rodríguez, J. D., Pérez, A., & Lozano, J. A. (2010). Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, **32(3)**, 569–575. https://doi.org/10.1109/TPAMI.2009.187.

Ruder, S. (2017). An Overview of Multi-Task Learning in Deep Neural Networks. [Available online from http://arxiv.org/abs/1706.05098]

She, L., Zhang, H. K., Bu, Z., Shi, Y., Yang, L., & Zhao, J. (2022). A Deep-Neural-Network-Based Aerosol Optical Depth (AOD) Retrieval from

Landsat-8 Top of Atmosphere Data. Remote Sensing, **14(6)**. https://doi.org/10.3390/rs14061411.

She, L., Zhang, H. K., Li, Z., de Leeuw, G., & Huang, B. (2020). Himawari-8 aerosol optical depth (Aod) retrieval using a deep neural network trained using aeronet observations. Remote Sensing, **12(24)**, 1–20. https://doi.org/10.3390/rs12244125.

Shen, H., Li, T., Yuan, Q., & Zhang, L. (2018). Estimating Regional Ground-Level PM 2.5 Directly From Satellite Top-Of-Atmosphere Reflectance Using Deep Belief Networks. Journal of Geophysical Research: Atmospheres, 123(24), 13,875-13,886. https://doi.org/10.1029/2018JD028759.

Shi, H., Yang, D., Wang, W., Fu, D., Gao, L., Zhang, J., Hu, B., Shan, Y., Zhang, Y., Bian, Y., Chen, H., & Xia, X. (2023). First estimation of high-resolution solar photovoltaic resource maps over China with Fengyun-4A satellite and machine learning. Renewable and Sustainable Energy Reviews, **184**, 113549. https://doi.org/10.1016/j.rser.2023.113549.

Song, S., Bang, S., Cho, S., Han, H., & Lee, S. (2022). Attentive Multi-Task Prediction of Atmospheric Particulate Matter: Effect of the COVID-19 Pandemic. IEEE Access, **10**, 10176–10190. https://doi.org/10.1109/ACCESS.2022.3144588

Song, Z., Fu, D., Zhang, X., Han, X., Song, J., Zhang, J., Wang, J., & Xia, X. (2019). MODIS AOD sampling rate and its effect on $PM_{2.5}$ estimation in North China.

Atmospheric Environment, **209**, 14–22. https://doi.org/10.1016/j.atmosenv.2019.04.020.

Su, T., Li, Z., & Kahn, R. (2018). Relationships between the planetary boundary layer height and surface pollutants derived from lidar observations over China: Regional pattern and influencing factors. Atmospheric Chemistry and Physics, **18(21)**, 15921–15935. https://doi.org/10.5194/acp-18-15921-2018.

Sun, T., Shao, Y., Li, X., Liu, P., Yan, H., Qiu, X., & Huang, X. (2019). Learning Sparse Sharing Architectures for Multiple Tasks. [Available online from http://arxiv.org/abs/1911.05034]

Wang, X., Dickinson, R. R. E., Su, L., Zhou, C., & Wang, K. (2018). $PM_{2.5}$ pollution in China and how it has been exacerbated by terrain and meteorological conditions. Bulletin of the American Meteorological Society, **99(1)**, 105–120. https://doi.org/10.1175/BAMS-D-16-0301.1.

Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., & Cribb, M. (2019). Estimating 1-km-resolution $PM_{2.5}$ concentrations across China using the space-time random forest approach. Remote Sensing of Environment, **231**. https://doi.org/10.1016/j.rse.2019.111221.

Wei, J., Li, Z., Pinker, R. T., Wang, J., Sun, L., Xue, W., Li, R., & Cribb, M. (2021). Himawari-8-derived diurnal variations in ground-level $PM_{2.5}$ pollution across China using the fast space-time Light Gradient Boosting Machine (LightGBM). Atmospheric Chemistry and Physics, **21(10)**, 7863–7880.

https://doi.org/10.5194/acp-21-7863-2021.

Wei, J., Li, Z., Sun, L., Peng, Y., Zhang, Z., Li, Z., Su, T., Feng, L., Cai, Z., & Wu, H. (2019). Evaluation and uncertainty estimate of next-generation geostationary meteorological Himawari-8/AHI aerosol products. Science of the Total Environment, **692**, 879–891. https://doi.org/10.1016/j.scitotenv.2019.07.326.

Xiao, Q., Zhang, H., Choi, M., Li, S., Kondragunta, S., Kim, J., Holben, B., Levy, R. C., & Liu, Y. (2016). Evaluation of VIIRS, GOCI, and MODIS Collection 6 AOD retrievals against ground sunphotometer observations over East Asia. Atmospheric Chemistry and Physics, **16(3)**, 1255–1269. https://doi.org/10.5194/acp-16-1255-2016.

Xie, G., Wang, M., Pan, J., & Zhu, Y. (2019). Spatio-temporal variations and trends of MODIS C6.1 Dark Target and Deep Blue merged aerosol optical depth over China during 2000–2017. Atmospheric Environment, **214**. https://doi.org/10.1016/j.atmosenv.2019.116846.

Xu, X., & Yoneda, M. (2021). Multitask Air-Quality Prediction Based on LSTM-Autoencoder Model. IEEE Transactions on Cybernetics, **51(5)**, 2577–2586. https://doi.org/10.1109/TCYB.2019.2945999.

Xue, Y., Li, Y., Guang, J., Tugui, A., She, L., Qin, K., Fan, C., Che, Y., Xie, Y., Wen, Y., & Wang, Z. (2020). Hourly $PM_{2.5}$ Estimation over central and eastern China based on Himawari-8 data. Remote Sensing, **12(5)**. https://doi.org/10.3390/rs12050855.

Xun, L., Lu, H., Qian, C., Zhang, Y., Lyu, S., & Li, X. (2021). Analysis of aerosol optical depth from sun photometer at Shouxian, China. Atmosphere, **12(9)**. https://doi.org/10.3390/atmos12091226.

Yang, L., Xu, H., & Yu, S. (2020). Estimating $PM_{2.5}$ concentrations in Yangtze River Delta region of China using random forest model and the Top-of-Atmosphere reflectance. Journal of Environmental Management, **272**. https://doi.org/10.1016/j.jenvman.2020.111061.

Yang, Q., Yuan, Q., Yue, L., Li, T., Shen, H., & Zhang, L. (2019). The relationships between $PM_{2.5}$ and aerosol optical depth (AOD) in mainland China: About and behind the spatio-temporal variations. Environmental Pollution, **248**, 526–535. https://doi.org/10.1016/j.envpol.2019.02.071.

Yin, J., Mao, F., Zang, L., Chen, J., Lu, X., & Hong, J. (2021). Retrieving $PM_{2.5}$ with high spatio-temporal coverage by TOA reflectance of Himawari-8. Atmospheric Pollution Research, **12(4)**, 14–20. https://doi.org/10.1016/j.apr.2021.02.007.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., & Zhang, L. (2020). Deep learning in environmental remote sensing: Achievements and challenges. Remote Sensing of Environment, **241**. https://doi.org/10.1016/j.rse.2020.111716.

Zhang, Q., Jiang, X., Tong, D., Davis, S. J., Zhao, H., Geng, G., Feng, T., Zheng, B., Lu, Z., Streets, D. G., Ni, R., Brauer, M., Van Donkelaar, A., Martin, R. V., Huo, H., Liu, Z., Pan, D., Kan, H., Yan, Y., … Guan, D. (2017). Transboundary health impacts of transported global air pollution and international trade. Nature, **543(7647)**, 705–709. https://doi.org/10.1038/nature21712.

Zhang, Q., Wu, S., Wang, X., Sun, B., & Liu, H. (2020). A $PM_{2.5}$ concentration prediction model based on multi-task deep learning for intensive air quality monitoring stations. Journal of Cleaner Production, **275**. https://doi.org/10.1016/j.jclepro.2020.122722.

Zhang, M., Wang, Y., Ma, Y., Wang, L., Gong, W., & Liu, B. (2018). Spatial distribution and temporal variation of aerosol optical depth and radiative effect in South China and its adjacent area. Atmospheric Environment, **188**, 120–128. https://doi.org/10.1016/j.atmosenv.2018.06.028.

Zhang, Y., & Yang, Q. (2017). A Survey on Multi-Task Learning. [Available online from http://arxiv.org/abs/1707.08114]

Zheng, C., Zhao, C., Zhu, Y., Wang, Y., Shi, X., Wu, X., Chen, T., Wu, F., & Qiu, Y. (2017). Analysis of influential factors for the relationship between $PM_{2.5}$ and AOD in Beijing. Atmospheric Chemistry and Physics, **17(21)**, 13473–13489. https://doi.org/10.5194/acp-17-13473-2017.