Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Spatiotemporal Dynamics and Exposure Analysis of Daily PM2.5 Using a Remote Sensing-based Machine Learning Model and Multi-time Meteorological Parameters

**Binjie Chen**
Zhejiang University https://orcid.org/0000-0002-6094-0702
**Yi Lin**
University of Hong Kong
**Jinsong Deng** ( jsong_deng@zju.edu.cn )
https://orcid.org/0000-0002-1412-478X
**Zheyu Li**
University of California San Diego
**Li Dong**
Zhejiang University
**Yibo Huang**
Zhejiang University
**Ke Wang**
Zhejiang University
**Wu Yang**
Zhejiang University

**Research**

**Keywords:** Particular Matter, Aerosol Optical Depth, Machine learning, Multi-time meteorological parameters, the Yangtze River Delta, Exposure analysis

**Posted Date:** July 15th, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-41719/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Version of Record:** A version of this preprint was published at Atmospheric Pollution Research on February 1st, 2021. See the published version at https://doi.org/10.1016/j.apr.2020.10.005.

## Abstract

# Background

Identifying spatiotemporal characteristics of daily fine particulate matter ($PM_{2.5}$) concentrations is essential for assessing air quality. Exposure analysis can help understand the environmental health impact on human beings and provide basic information for appropriate decision making. This study aimed to estimate daily $PM_{2.5}$ concentrations and analyze the resident exposure level in the economically developed Yangtze River Delta (YRD) from 2016–2018.

# Methods

An integrated method incorporating satellite-based aerosol optical depth (AOD), machine learning models and multi-time meteorological parameters were developed. Ten-fold cross validation (CV) was implemented to evaluate the model performance.

# Results

Compared to the models with daily means of meteorological fields, the models with multi-time meteorological parameters had higher CV $R^2$ and lower CV root mean square error (RMSE) values. The model with the best performance achieved sample- (site-) based CV $R^2$ values of 0.88 (0.88) and RMSE values of 10.33 (10.35) μg/$m^3$. The YRD region is seriously polluted (exceeding the World Health Organization (WHO) Interim Targets (IT)-1 standard of 35 μg/$m^3$) during our study period, especially in Jiangsu Province, but with an improving trend. The residents in Zhejiang Province suffered the least from exposure, with 39 days (4% of the total days) characterized as over polluted (daily average > 75 μg/$m^3$) in our study period. Air pollution in Shanghai Municipality mitigated the most from 2016 to 2018.

# Conclusions

With the advantages of high-accuracy and high-resolution (daily and 0.01°×0.01° resolutions), the proposed method can help explore the effect of air pollution to human health spatiotemporally and guide for environmental policy planning.


## Introduction

Fine particulate matter with aerodynamic diameters less than 2.5 μm, also known as $PM_{2.5}$, not only threatens human health by increasing the risk of many diseases including acute lower respiratory illness, cerebrovascular disease, ischemic heart disease, chronic obstructive pulmonary disease, lung cancer, and stroke [1–6], but also causes economic loss [7]. According to the Global Burden of Diseases study, a total

of approximately 3 million deaths globally were attributable to $PM_{2.5}$ pollution in 2017, making it the eighth leading risk among 84 risk factors [8]. With rapid urbanization and industrialization in China, human activities such as transportation and fuel consumption have emitted increasing $PM_{2.5}$ into the air [9]. To control nationwide air pollution, the Chinese government enacted the Air Pollution Prevention and Control Action Plan in 2013 [10], and established approximately 1500 air quality monitoring sites across the nation. However, the sparse and uneven distribution of these sites imposes restrictions on the spatial resolution and high accuracy assessment of $PM_{2.5}$ concentrations.

In the last few decades, aerosol optical depth (AOD) retrieved via remote sensing has represented a major, though still underused, source of atmospheric environmental monitoring by providing consistent spatiotemporal coverage over large areas. Previous studies have explored the relationship between satellite-derived AOD and ground-level $PM_{2.5}$ [11–16]. AOD represents the integral of the atmospheric extinction coefficient from the surface to space and can be retrieved by various sensors, such as the Sea-viewing Wide Field-of-view Sensor [17], the Multi-angle Imaging SpectroRadiometer [18], the Advanced Himawari-8 Imager [19], and the Moderate Resolution Imaging Spectroradiometer (MODIS) [20], The newly released MODIS AOD product, MCD19A2, has a better spatial resolution of 1 km compared to the traditional AOD products [21]. In addition, based on the advanced Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm, this newly released product could overcome the shortage of dark-target algorithms and have higher accuracy over both dark and bright surfaces [22, 23].

Much less known than remote sensing, the estimation model has been proven to be a useful tool for effectively quantifying the AOD-$PM_{2.5}$ relationship and then mapping the $PM_{2.5}$ distribution spatially. Three major approaches are commonly applied, including physical models based on vertical correction and humidity correction [24–28], chemical models with air transportation simulations [29] and statistical models [30]. Statistical regression models are the most widely used approaches due to the increasingly available site-monitored $PM_{2.5}$ data and their adequate performance. Early studies modeled the AOD-$PM_{2.5}$ relationship using multiple linear regression [31, 32], the linear mixed effect model [33, 34], and the geographically and temporally weighted regression model [35, 36]. Recently, as a branch of artificial intelligence, machine learning methods such as random forest (RF) [37, 38] and gradient boosting regression (GBR) [39] were introduced to map the relationship between AOD and $PM_{2.5}$. As the AOD-$PM_{2.5}$ relationships are both spatially and temporally heterogeneous, a wide range of auxiliary predictors are incorporated in the model [38]. Compared to conventional statistical models, machine learning models can overcome high-order interactions and capture complex nonlinearity among the predictor variables, which also yields superior results [40].

Central to much of the debate is the difficulty in selecting feasible predictor parameters, such as elevation, land-cover types, and local emissions [40, 41], among which the meteorological parameters seemed to be the most critical variables [42, 43]. Meteorological parameters such as wind gust, precipitation and boundary layer height always have a negative relationship with $PM_{2.5}$ concentration, whereas solar radiation and relative humidity have a positive relationship with $PM_{2.5}$ concentration; all of

these variables have significant effects on the transport and transformation of $PM_{2.5}$ pollutants [44–46]. Previous studies always employed the daily means of meteorological fields [14]. However, the meteorological fields at different times of the day vary significantly and cannot be simply demonstrated by daily mean values. For example, due to solar radiation and heat storage in the atmosphere, the temperature at noon is higher than that at dawn and night, whereas relative humidity has the opposite pattern [47]. The diversification of wind speed and wind direction is complex within a day, making it difficult to predict [48]. Moreover, the boundary layer height often follows a gamma distribution and peaks at noon [49]. Despite the fact that we aim to estimate daily mean $PM_{2.5}$ concentrations, there could be a side effect on the accuracy of $PM_{2.5}$ modeling by ignoring the inner-day variations in weather conditions. By incorporating multi-time meteorological parameters, the inner-day variations of $PM_{2.5}$ concentrations could be considered, which helps the improvement of model accuracy. The combination of remote sensing-based machine learning models and multi-time meteorological parameters could provide more accurate and spatially consistent information about urban air quality and consequently allow for an improved understanding of the impacts of urban development on human health.

As a case study in the Yangtze River Delta (YRD), one of the most rapidly developing and economically developed regions in China, this paper explored the spatiotemporal dynamics of the daily ground-level $PM_{2.5}$ concentrations and then exposure risk by using remote sensing-based AOD and machine learning models. Bearing in mind the importance of feasible and effective parameters for the estimation model, in addition to the normalized different vegetation index (NDVI), elevation, day of the year, and location, this study specifically adopted multi-time meteorological parameters instead of the daily means. We evaluated and compared our proposed method using sample- and site-based ten-fold cross validation (CV). Subsequently, the spatiotemporal distribution of $PM_{2.5}$ concentrations with a 0.01°×0.01° resolution was inferred and the population-weighted $PM_{2.5}$ revealing the resident exposure level was analyzed in the YRD during 2016–2018.

# Materials And Methods

# Study Area

The YRD is located in east China, including Jiangsu Province, Zhejiang Province, and Shanghai Municipality (Fig. 1). Its terrain is elevated in the south and depressed in the north, and it is one of the most developed regions in China, with a population of more than 150 million, while the area is only approximately 20 million $km^2$. The high density of residents living in the YRD has resulted in residents suffering from severe air pollution due to rapid urbanization, as high concentrations of $PM_{2.5}$ and $O_3$ have been monitored increasingly frequently [50]. Several policies for controlling the YRD air pollution have been made at both local and regional levels [51], such as Air pollution control plans for key industries in the YRD region [52], and Action plans for comprehensive treatment of air pollution in the YRD region during autumn and winter [53]. Understanding the spatiotemporal distributions of $PM_{2.5}$ in the YRD is still urgent to help develop more effective control strategies.

# Ground-level $PM_{2.5}$ Data

There were 176 air quality monitoring sites in the study area (Fig. 1). We collected hourly $PM_{2.5}$ monitoring data during 2016–2018 from the China National Environmental Monitoring Center (CNEMC, http://www.cnemc.cn/). Many sites had less than 12 valid hourly data points within a day, and the measurements remained the same for more than three continuous hours due to instrument malfunction. We excluded the data under the above conditions in our experiment and then calculated the daily mean $PM_{2.5}$ concentrations for each station [54]. Finally, we applied log transformation to the data.

# MAIAC AOD Data

In this work, we utilized the daily 1 km MAIAC AOD product at 550 nm retrieved from the MODIS Terra and Aqua satellites. All of the MAIAC AOD data from 2016 to 2018 were downloaded from the NASA Level-1 and Atmosphere Archive & Distribution System Distributed Active Archive Center website (https://ladsweb.modaps.eosdis.nasa.gov/). We only selected the highest quality AOD data through the attribution “QA_FLAG” in the dataset.

To handle the non-random missingness in MAIAC AOD product caused by cloud cover, high surface reflectance, and high aerosol loading [55], we combined the Terra and Aqua MAIAC AOD products. For each grid, if both Terra and Aqua AODs were available, the average value was calculated. If only one product was available, either Terra or Aqua AOD, the grid value equaled to the available one.

# AERONET AOD Data

To ensure the accuracy of MAIAC AOD in the YRD, we evaluated the performance of MAIAC AOD product based on the Aerosol Robotic Network (AERONET) observations (http://aeronet.gsfc.nasa.gov/). The selected AERONET sites were listed in Table S1. AERONET AOD at 550 nm was interpolated from AERONET AOD at 500 nm and 675 nm [54]. Figure S1 shows the scatter plots of the MAIAC AOD and AERONET AOD. Despite several anomalies, MAIAC AOD matched well with AERONET AOD. The RMSE and MAE values were 0.12 and 0.09, indicating the good estimation accuracy of MAIAC AOD products in the YRD. Thus, it is possible to use MAIAC AOD product to estimate the daily $PM_{2.5}$ concentration in the YRD.

# Meteorological Data

Six meteorological parameters at a spatial resolution of 0.125°×0.125° were downloaded from the European Centre for Medium-Range Weather Forecasts (ECMWF, http://www.ecmwf.int/). ERA-Interim reanalysis data, as one of the most popular products by the ECMWF, have provided continuous real-time meteorological parameters every 6 hours since 1979 [56], including 2-meter temperature (T2M), surface pressure (SP), boundary layer height (BLH), relative humidity (RH), 10-meter U wind component (WU10M), and 10-meter V wind component (WV10M). Except for the BLH, which was not available at 00:00 UTC, all of the other meteorological parameters were extracted at 00:00, 06:00, and 12:00 UTC, corresponding to local times of 08:00 (morning), 14:00 (afternoon), and 20:00 (evening).

# Other Datasets

Digital elevation model (DEM) data were extracted from the Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Map version 1, which has a spatial resolution of 30 m (http://www.gscloud.cn). LandScan population data (1 km × 1 km) developed by the Oak Ridge National Laboratory were downloaded from the following website: https://landscan.ornl.gov/landscan-datasets. MODIS monthly NDVI products MOD13A3 and MYD13A3 at a 1 km resolution were obtained from the same source as the MAIAC AOD data.

# Data Preprocessing

Because of the inconsistencies in the meteorological data, NDVI data, DEM data, and AOD data, we resampled all of the data to the same spatial resolution of 0.01°×0.01° by linear interpolation. Ground monitoring $PM_{2.5}$ data were assigned to the nearest 0.01°×0.01° grid, and the average value was calculated if there were two or more monitoring sites within one grid. After data preprocessing, there were 31469 matched records in the study period.

# Model Description

Machine learning models usually have two types, the ensemble one and the individual one. The ensemble machine learning models can be further divided into models based on bagging and boosting. Thus, the three different types of machine learning models were considered in our study. First, the RF algorithm, an ensemble machine learning method based on bagging, can be utilized for classification or regression and consists of many independent weak decision trees [57]. Each tree in the model is generated by a random bootstrap sample, and a subset of features is selected to split the tree. The final output of the model is made by averaging the predictions of each individual tree [58]. The second model is the GBR model, which is also an ensemble machine learning model but based on boosting. It is built by iteratively combining weak models via finding the best gradient descent step and minimizing the error of the loss function [59]. The last model is the K-nearest neighbor (KNN) regression model, which is not an ensemble model. By looking for the K-nearest neighbors in the training data at the feature space, the output of the model is the average of these K neighbors [60].

The following predictors were incorporated in these models: MAIAC AOD, NDVI, DEM, day of the year (DOY), longitude (LON), latitude (LAT) and multi-time meteorological parameters (referred to as the Mete-multi model). It is expressed as follows:

$$PM_{2.5\_st} = f(AOD_{st}, NDVI_{st}, DEM_s, DOY_t, LON_s, LAT_s, METE_{8\_st}, METE_{14\_st}, METE_{20\_st}) \quad (1)$$

where $PM_{2.5\_st}$ represents the daily mean $PM_{2.5}$ concentration at location s on day t; f() represents machine learning models (RF, GBR, or KNN); $AOD_{st}$ and $NDVI_{st}$ represent MAIAC AOD and NDVI at location s on day t, respectively; $DEM_s$is the elevation at location s; $DOY_t$ represents day of the year indicating the

temporal variation; LONs and LATs are the longitude and latitude of location s, indicating the spatial variation; METE8_st, METE14_st, and METE20_st represent meteorological parameters at local times of 08:00 (morning), 14:00 (afternoon), and 20:00 (evening)at location s on day t, including T2M, SP, BLH, RH, WU10M, and WV10M.

Moreover, for comparison purposes, models based on the daily means of the meteorological fields were built (referred to as the Mete-mean model). The expression is as follows:

$$PM_{2.5\_st} = f(AOD_{st}, NDVI_{st}, DEM_s, DOY_t, LON_s, LAT_s, METE_{mean\_st}) \quad (2)$$

where $METE_{mean_st}$represents the daily mean value of meteorological parameters at location s onday t, including T2M, SP, BLH, RH, WU10M, and WV10M.

In summary, we developed six models, named RF (Mete-multi), GBR (Mete-multi), KNN (Mete-multi), RF (Mete-mean), GBR (Mete-mean), and KNN (Mete-mean).

# Model Validation and Evaluation

To evaluate the model performance and validate the robustness of the model, we applied the commonly used ten-fold CV based on the samples and sites. The dataset was first randomly split into ten parts, each with 10% of samples in the sample-based CV and 10% of ground sites in the site-based CV. Then, the model was fitted ten times to ensure that all of the data were tested: for each fit, only one part was treated as the testing data, and the rest were treated as training data. The coefficient of determination ($R^2$), root mean square error (RMSE), and mean absolute error (MAE) were utilized to quantitatively assess the performances of the models. The specific formulas of the above statistical metrics are provided in the Supplementary Material.

## Results And Discussion

# Descriptive Statistics of the Dataset

Figure S2 presents the histogram and descriptive statistics of the predictors in our model. Daily $PM_{2.5}$ concentrations range from 4.33 to 303.59 μg/$m^3$, with an average of 54.34 μg/$m^3$. The MAIAC AOD varied from 0.01 to 2, with an average of 0.47, which had a distribution similar to that of $PM_{2.5}$. There were significant variations in the BLH with time, which had a normal distribution at 14:00 but an extremely skewed distribution at 20:00. We also calculated the Pearson correlation coefficients between the variables (Figure S3). The results showed that the AOD had the highest correlation with $PM_{2.5}$ among the predictors ($R = 0.49$, $p < 0.01$), indicating a relatively close relationship between $PM_{2.5}$ concentrations and the AOD. Meteorological parameters at different times of the day were highly correlated with each other, especially the temperature and surface pressure, with correlation coefficients higher than 0.95 ($p <$

0.01). As mentioned previously, the machine learning algorithm can cope with collinearity problems and is applicable for dealing with these highly correlated predictors [61].

# Model Evaluation

The performances of the three types of machine learning models were evaluated through sample- and site-based 10-fold CV (Table 1). Overall, the RF (Mete-multi) model outperformed all of the other models, with site- (sample-) based CV $R^2$, RMSE, and MAE values of 0.88 (0.88), 10.35 (10.33) μg/$m^3$, and 7.00 (6.95) μg/$m^3$, respectively. The GBR models were the most unsatisfactory, with CV $R^2$ values less than 0.70. On the other hand, all three models with multi-time meteorological parameters outperformed the corresponding models with daily mean meteorological parameters. The RF (Mete-multi) and GBR (Mete-multi) models had a slight increase of 0.01 in the $R^2$ value and decreases of 0.3–0.5 μg/$m^3$ in the RMSE value and 0.2–0.3 μg/$m^3$ in the MAE value. The proposed method of utilizing multi-time meteorological parameters promoted the KNN model most, with the CV $R^2$ value increasing by 0.06 and the RMSE (MAE) value decreasing by approximately 2 (1) μg/$m^3$. These results indicated the effectiveness of incorporating the multi-time meteorological parameters in machine learning models.

Table 1
Comparison of different proposed methods through ten-fold CV

| Model | Site-based CV | | | Sample-based CV | | |
|---|---|---|---|---|---|---|
| | $R^2$ | RMSE | MAE | $R^2$ | RMSE | MAE |
| RF (Mete-multi) | **0.88** | **10.35** | **7.00** | **0.88** | **10.33** | **6.95** |
| GBR (Mete- multi) | 0.66 | 17.46 | 12.21 | 0.67 | 17.23 | 12.05 |
| KNN (Mete- multi) | 0.85 | 11.79 | 8.02 | 0.85 | 11.83 | 8.01 |
| RF (Mete-mean) | 0.87 | 10.68 | 7.23 | 0.88 | 10.63 | 7.15 |
| GBR (Mete-mean) | 0.65 | 17.86 | 12.49 | 0.66 | 17.71 | 12.37 |
| KNN (Mete-mean) | 0.79 | 13.77 | 9.27 | 0.79 | 13.74 | 9.25 |

The bold font indicates the best performances among the models. RF: random forest model. GBR: gradient boosting model. KNN: K-nearest-neighbor model. Mete-mean: daily mean of meteorological fields as predictors in the model. Mete-multi: meteorological parameters at different times of the day as predictors in the model.

Given the difference in weather conditions from morning to night, the daily mean values of the meteorological fields could not well reflect the inner-day variations and may be insufficient in estimating $PM_{2.5}$ concentrations. Low temperature and high relative humidity usually result in high $PM_{2.5}$ concentrations because of the proper conditions for new particle formation [45]. However, the sinusoidal curve patterns of the temperature and relative humidity diurnal variations [47] cannot be fully demonstrated by the daily mean values. Diurnal wind speed and wind direction patterns were more complicated and generally not rigorously sinusoidal [48], and diurnal changes could significantly

influence the transport of pollutants [46]. The surface pressure often peaked in the morning and was lowest in the afternoon [62]. Low surface pressure resulted in high wind speed and helped the dispersion of pollutants. The boundary layer height, which had an anti-correlated relationship with $PM_{2.5}$ concentrations [63], was stable and shallow at night but high during the daytime [49]. By utilizing the multi-time meteorological parameters, the diurnal patterns of these meteorological fields and $PM_{2.5}$ concentrations were considered in the estimation model, which improved the model performance. Our results revealed that by replacing the daily mean values with multi-time meteorological parameters, the model performance could be enhanced.

In the following sections, daily $PM_{2.5}$ concentrations in the YRD were estimated using the RF (Mete-multi) model, as it achieved the best performance. As one of the most economically developed regions in China, many studies have tried to generate ground-level $PM_{2.5}$ loadings in the YRD by satellite-driven models [26, 33, 54]. Q Xiao, Y Wang, HH Chang, X Meng, G Geng, A Lyapustin and Y Liu [54] fit the relationship between $PM_{2.5}$ and the MAIAC AOD with a two-stage model, obtaining a CV $R^2$ value of 0.81. Z Ma, Y Liu, Q Zhao, M Liu, Y Zhou and J Bi [33] implemented a nested linear mixed effect model to retrieve $PM_{2.5}$ concentrations at a 3 km resolution with MYD04_3K AOD data and reported a CV $R^2$ value of 0.671. The proposed RF (Mete-multi) model (CV $R^2$ = 0.88) outperformed these previous studies in estimating daily $PM_{2.5}$ concentrations in the YRD at a finer resolution of 0.01°×0.01°. Figure 2 shows the density scatterplots of the model fitting and validation results for the RF (Mete-multi) model, and the spatial site-based CV results. It is worth noting that despite the estimation results of the RF (Mete-multi) model matching well with ground-level measurements (model fitting $R^2$ = 0.98), the proposed model was likely to underestimate the ground-level $PM_{2.5}$ concentrations when greater than approximately 50 μg/$m^3$ (slope = 0.94, intercept = 2.67). The spatial site-based CV results showed local $R^2$ values ranging from 0.37 to 0.98 and RMSE values varying from 2.75 to 19.63 μg/$m^3$, with about 80% of the monitoring sites reporting $R^2$ values higher than 0.80 and RMSE values lower than 12 μg/$m^3$. High-performance sites were mainly located in the inland regions of the study area. Meanwhile, there were two major regions where our proposed model had a slightly unsatisfactory performance. One was the region near the coastline, and the other was the region in northwest Jiangsu Province. Monitoring sites near the seashore often reported a low CV $R^2$ with a high CV RMSE, this phenomenon was mainly caused by the high uncertainty of the MAIAC AOD data near water [64]. The northwestern Jiangsu Province, with a CV RMSE at the monitoring sites greater than 12 μg/$m^3$, suffered from large estimation errors. Because of the severe pollution in the northwestern Jiangsu Province, with annual mean $PM_{2.5}$ concentrations greater than 65 μg/$m^3$, the proposed model had underestimated outputs in most circumstances. This was mainly attributed to the scarcity of matching data at high $PM_{2.5}$ concentrations [65].

# Spatiotemporal Dynamics

The estimated $PM_{2.5}$ concentrations in the 0.01°×0.01° grid by the RF (Mete-multi) model and ground-level measurements are displayed in Fig. 3. Overall, the estimated average $PM_{2.5}$ concentration in the

study area was 48.04 μg/$m^3$ during 2016–2018, while the average of the ground-level measurements was 45.09 μg/$m^3$, indicating a tight estimation consistency. Almost all regions in the study area exceeded the annual Chinese National Ambient Air Quality Standard (CNAAQS) [66] of 35 μg/$m^3$ in 2016 and 2017, which resulted in a serious health concern for the YRD. When it comes to 2018, part of the south Zhejiang Province was up to the standard, while the other regions failed to reach the standard but had a decreasing trend.

The spatial variations are also clear in Fig. 3. There was a declining gradient from north to south, revealing the differences in local emissions, topography and meteorology. The most heavily polluted area was in northwest Jiangsu Province. As the only region with centralized residential heating in the study area, a large number of pollutants was emitted into the air due to fossil fuel burning in winter, which resulted in unsatisfactory air conditions to some degree. The lighter pollution in the south was attributed to the dense vegetation coverage across the rolling mountains with a significant reduction in human activities. In the middle of the study area (north of the Zhejiang Plain, Shanghai Municipality and south of Jiangsu Province), $PM_{2.5}$ concentrations ranged from 35 to 75 μg/$m^3$. This region was the most developed metropolitan region in the study area, and intensive human activities such as vehicle transportation and fuel consumption were the main emission sources [50]. In addition, the high density of skyscrapers and mansions impeded the diffusion of pollutants [67].

Our proposed model captured not only spatial patterns, but also seasonal patterns. Figure 4 illustrates the seasonal patterns of estimated $PM_{2.5}$ during the study period. It was obvious that winter suffered the heaviest air pollution, with a three-year mean $PM_{2.5}$ concentration of 59.74 μg/$m^3$. Summer experienced the slightest fine particulate matter pollution, with the lowest mean $PM_{2.5}$ concentration of 28.14 μg/$m^3$. Apart from the anthropogenic emissions that varied with season, such as the much higher fuel consumption in winter, meteorological conditions also played an important role. As the BLH was relatively lower in the cold season, the temperature inversion was significant, causing the particles difficulty in dispersing [68]. On the other hand, favorable weather conditions, such as frequent precipitation in summer helped wash out the fine particles suspended in air [69], and the high relative humidity accelerated the dilution of pollutants [38].

The temporal changes during the study period are shown in Figure S4. The $PM_{2.5}$ concentrations in the study area were mitigated by approximately 9.55 μg/$m^3$ from 2016 to 2018. There was a slight increase between 2016 and 2017 in western Jiangsu Province. Compared to the regions with fewer human activities, the decline in the $PM_{2.5}$ concentrations in metropolitan areas was more significant. The area with the most enhanced concentration of $PM_{2.5}$ was the north Zhejiang Plain, the Shanghai Municipality and the southern region of Jiangsu Province, where anthropogenic emissions contributed the most to air pollution [51]. Since 2013, the Chinese government intended to mitigate severe air pollution across the nation, and the Air Pollution Prevention and Control Action Plan (also called the Ten Measures of Air) was released [70]. These improving trends suggested that series of strict air pollution control policies and measures had indeed taken effect, as these actions included energy conservation, the promotion of

cleaner fuels and the optimization of the industrial structure. All of these actions focused on reducing local emissions from human activities [66].

# Exposure Analysis

Aside from the spatial and temporal distributions of $PM_{2.5}$, another important issue was the accurate assessment of the resident exposure level to $PM_{2.5}$ in populated areas. Thus, we calculated the population-weighted $PM_{2.5}$ concentrations of Jiangsu Province, Zhejiang Province, and Shanghai Municipality in the period 2016–2018. The annual residential exposure level to $PM_{2.5}$ concentrations is shown in Table 2. Thanks to the implementation of emission control policies, the residential exposure level to $PM_{2.5}$ decreased significantly from 2016 to 2018. Shanghai Municipality improved most, with approximately 20 $\mu g/m^3$ of the decrease, and the other two provinces improved by approximately 10 $\mu g/m^3$. It was worth noting that the exposure analysis revealed the severe air pollution still needed attention in the YRD. Until 2018, $PM_{2.5}$ concentrations still exceeded the World Health Organization (WHO) Interim Targets (IT)-1 standard of annual $PM_{2.5}$ concentration as well as the CNAAQS (35 $\mu g/m^3$), with concentrations of 50.56 $\mu g/m^3$, 38.62 $\mu g/m^3$, and 41.95 $\mu g/m^3$ in Jiangsu Province, Zhejiang Province, and Shanghai Municipality, respectively.

Table 2
Annual exposure $PM_{2.5}$ concentrations in Jiangsu Province, Zhejiang Province and Shanghai Municipality.

| Region | Population (million) | Population-weighted $PM_{2.5}$ concentrations ($\mu g/m^3$) | | |
|---|---|---|---|---|
| | | 2016 | 2017 | 2018 |
| Jiangsu Province | 56.17 | 61.43 | 54.84 | 50.56 |
| Zhejiang Province | 38.18 | 50.47 | 46.80 | 38.62 |
| Shanghai Municipality | 16.38 | 60.45 | 46.75 | 41.95 |

Subsequently, the dynamics of daily exposure $PM_{2.5}$ concentrations are illustrated in Fig. 5(a). The CNAAQS level-2 standard of daily $PM_{2.5}$ concentrations was 75 $\mu g/m^3$, above which poor air conditions may have adverse impacts on residents. There were 139, 33, and 49 days characterized as over-polluted (> 75 $\mu g/m^3$) in Jiangsu Province, Zhejiang Province and Shanghai Municipality, respectively. The number of days when the health of residents was potentially impacted in Jiangsu Province was approximately four times that in Zhejiang Province and two times that in Shanghai Municipality. The time series of $PM_{2.5}$ concentrations also revealed a strong seasonal pattern. It was mostly in summer that $PM_{2.5}$ loadings were under the CNAAQS daily level-1 standard (35 $\mu g/m^3$), and severe air pollution peaked in winter. Although the pollution peaks were more severe in 2018 than in 2017, the number of days that reached the CNAAQS standard increased, showing an improving trend in the residential exposure level to

$PM_{2.5}$. Before 2018, almost all residents in the YRD were exposed to an annual mean $PM_{2.5}$ greater than the WHO-IT1 standard of 35 μg/$m^3$ (Fig. 5(b) (c)). However, 20% of the population in Zhejiang Province lived in areas meeting the WHO IT-1 standard in 2018, while the other two regions still had less than 1% of the population (Fig. 5(d)). The distribution of the population exposure level in Shanghai Municipality was similar to that in Jiangsu Province in 2016. Part of the population in Zhejiang Province was exposed to a higher level of $PM_{2.5}$ than residents in Shanghai Municipality in 2017, whereas 100% of the population in Zhejiang Province suffered less $PM_{2.5}$ exposure in 2016. In 2018, the $PM_{2.5}$ exposure level in Shanghai Municipality was much better than that in Jiangsu Province but worse than that in Zhejiang Province.

# Conclusion

In this study, daily $PM_{2.5}$ concentration was estimated via an integrated method incorporating satellite-based aerosol optical depth (AOD), a machine learning model and multi-time meteorological parameters. The proposed models with multi-time meteorological parameters was compared with the models with daily mean meteorological parameters. The results indicated that multi-time meteorological parameters could enhance the estimation accuracy by considering the inner-day variations in weather conditions. The spatiotemporal distribution and exposure level of $PM_{2.5}$ in the YRD from 2016 to 2018 were obtained via the proposed method. The YRD was seriously polluted (exceeding the WHO IT-1 standard of 35 μg/$m^3$) during the period, especially in Jiangsu Province, but with an improving trend. The residents in Zhejiang Province suffered the least from exposure, with 39 days (4% of the total days) characterized as over polluted (daily average > 75 μg/$m^3$) in our study period. Thanks to the strict air pollution control policies and measures, air condition in Shanghai Municipality promoted the most from 2016 to 2018.

While the proposed method of incorporating a remote sensing-based machine learning model and multi-time meteorological parameters showed good performance in the YRD, larger geographic regions may introduce unpredictable problems. Estimating $PM_{2.5}$ concentrations in large areas, such as China, by machine learning models with multi-time meteorological parameters will be explored in our future work. Second, we will try to fill the data gap in the MAIAC AOD data due to cloud contamination. Several studies have shown that after imputation, the degree of differences in the estimated $PM_{2.5}$ and measured $PM_{2.5}$ could be significantly improved [54, 71]. Finally, associated with long-term epidemical data, the proposed method will help improve the understanding of the impact of $PM_{2.5}$ exposure to residents and help make better recommendations for air pollution control policies.

# Abbreviations

AERONET
Aerosol Robotic Network; AOD:Aerosol optical depth; BLH:Boundary layer height; CNAAQS:Chinese National Ambient Air Quality Standard CV:Cross validation; DEM:Digital elevation model; DOY:Day of the year; GBR:Gradient boosting regression; IT:Interim Targets; KNN:K-nearest neighbor; LAT:Latitude; LON:Longitude; MAE:Mean absolute error; MAIAC:Multi-Angle Implementation of Atmospheric Correction;

MODIS:Moderate Resolution Imaging Spectroradiometer; NDVI:Normalized different vegetation index; $PM_{2.5}$:Fine particulate matter with aerodynamic diameters less than 2.5 μm; $R^2$:Coefficient of determination; RH:Relative humidity; RF:Random forest; RMSE:Root mean square error; SP:Surface pressure; T2M:2-meter temperature; WHO:World Health Organization; WU10M:10-meter U wind component; WV10M:10-meter V wind component; YRD:the Yangtze River Delta.

## Declarations

# Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

### Competing interests

The authors declare that they have no competing interests.

# Funding

This work was supported by the Natural Science Foundation of Zhejiang Province (LY18G030006) and the Ministry of Science and Technology of the People's Republic of China (2016YFC0503404).

# Authors’ contributions

JD and BC designed the study. BC, YL, ZL and LD collected the data. BC, YL and YH performed the analytical work. BC, JD, KW and WY wrote the paper. All authors read and approved the final manuscript.

# Acknowledgments

We thank the China National Environmental Monitoring Center, NASA, Goddard Space Flight Center, European Centre for Medium-Range Weather Forecasts, and Oak Ridge National Laboratory for providing the data.

# Authors' information

[1]College of Environment and Resource Sciences, Zhejiang University, Hangzhou 310058, China. [2]Department of Geography, University of Hong Kong, Hongkong, 999077, China. [3]Eleanor Roosevelt College, University of California-San Diego, La Jolla, CA, 92093, USA.

# References

1. Lelieveld J, Evans JS, Fnais M, Giannadaki D, Pozzer A. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature. 2015;525(7569):367–71.
2. Guan W-J, Zheng X-Y, Chung KF, Zhong N-S. Impact of air pollution on the burden of chronic respiratory diseases in China: time for urgent action. The Lancet. 2016;388(10054):1939–51.
3. Brunekreef B, Hoffmann B. Air pollution and heart disease. The Lancet. 2016;388(10045):640–2.
4. Wang Q, Wang J, He MZ, Kinney PL, Li T. A county-level estimate of PM2.5 related chronic mortality risk in China based on multi-model exposure data. Environ Int. 2018;110:105–12.
5. Landrigan PJ, Fuller R, Acosta NJR, Adeyi O, Arnold R, Basu N, Baldé AB, Bertollini R, Bose-O'Reilly S, Boufford JI, et al. The Lancet Commission on pollution and health. The Lancet. 2018;391(10119):462–512.
6. Han L, Sun Z, He J, Zhang X, Hao Y, Zhang Y. Does the early haze warning policy in Beijing reflect the associated health risks, even for slight haze? Atmos Environ. 2019;210:110–9.
7. Guan Y, Kang L, Wang Y, Zhang N-N, Ju M-T. Health loss attributed to PM2.5 pollution in China's cities: Economic impact, annual change and reduction potential. J Clean Prod. 2019;217:284–94.
8. Stanaway, JDea. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet. 2018;392(10159):1923–94.
9. Butt EW, Turnock ST, Rigby R, Reddington CL, Yoshioka M, Johnson JS, Regayre LA, Pringle KJ, Mann GW, Spracklen DV. Global and regional trends in particulate air pollution and attributable health burden over the past 50 years. Environmental Research Letters 2017, 12(10).
10. Lu X, Lin C, Li W, Chen Y, Huang Y, Fung JCH, Lau AKH. Analysis of the adverse health effects of PM2.5 from 2001 to 2017 in China and the role of urbanization in aggravating the health burden. Sci Total Environ. 2019;652:683–95.
11. Hu X, Waller LA, Lyapustin A, Wang Y, Liu Y. 10-year spatial and temporal trends of PM2.5 concentrations in the southeastern US estimated using high-resolution satellite data. Atmos Chem Phys. 2014;14(12):6301–14.
12. Di Q, Kloog I, Koutrakis P, Lyapustin A, Wang Y, Schwartz J. Assessing PM2.5 Exposures with High Spatiotemporal Resolution across the Continental United States. Environ Sci Technol. 2016;50(9):4712–21.

13. Di Q, Amini H, Shi L, Kloog I, Silvern R, Kelly J, Sabath MB, Choirat C, Koutrakis P, Lyapustin A, et al. An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environ Int. 2019;130:104909.
14. Guo Y, Tang Q, Gong D-Y, Zhang Z. Estimating ground-level PM2.5 concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. Remote Sens Environ. 2017;198:140–9.
15. Zhang T, Zang L, Wan Y, Wang W, Zhang Y. Ground-level PM2.5 estimation over urban agglomerations in China with high spatiotemporal resolution based on Himawari-8. Sci Total Environ. 2019;676:535–44.
16. Yang Y, Christakos G. Spatiotemporal Characterization of Ambient PM2.5 Concentrations in Shandong Province (China). Environ Sci Technol. 2015;49(22):13431–8.
17. Hsu NC, Gautam R, Sayer AM, Bettenhausen C, Li C, Jeong MJ, Tsay SC, Holben BN. Global and regional trends of aerosol optical depth over land and ocean using SeaWiFS measurements from 1997 to 2010. Atmos Chem Phys. 2012;12(17):8037–53.
18. Kahn RA, Garay MJ, Nelson DL, Yau KK, Bull MA, Gaitley BJ, Martonchik JV, Levy RC. Satellite-derived aerosol optical depth over dark water from MISR and MODIS: Comparisons with AERONET and implications for climatological studies. J Geophys Res. 2007;112:D18.
19. Wang W, Mao F, Zou B, Guo J, Wu L, Pan Z, Zang L. Two-stage model for estimating the spatiotemporal distribution of hourly PM1.0 concentrations over central and east China. Sci Total Environ. 2019;675:658–66.
20. Levy RC, Remer LA, Kleidman RG, Mattoo S, Ichoku C, Kahn R, Eck TF. Global evaluation of the Collection 5 MODIS dark-target aerosol products over land. Atmos Chem Phys. 2010;10(21):10399–420.
21. Lyapustin A, Wang Y, Korkin S, Huang D. MODIS Collection 6 MAIAC algorithm. Atmospheric Measurement Techniques. 2018;11(10):5741–65.
22. Lyapustin A, Wang Y, Laszlo I, Kahn R, Korkin S, Remer L, Levy R, Reid JS. Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. Journal of Geophysical Research 2011, 116(D3).
23. Lyapustin A, Martonchik J, Wang Y, Laszlo I, Korkin S. Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. Journal of Geophysical Research 2011, 116(D3).
24. Geng G, Zhang Q, Martin RV, van Donkelaar A, Huo H, Che H, Lin J, He K. Estimating long-term PM2.5 concentrations in China using satellite-based aerosol optical depth and a chemical transport model. Remote Sens Environ. 2015;166:262–70.
25. Li Z, Zhang Y, Shao J, Li B, Hong J, Liu D, Li D, Wei P, Li W, Li L, et al. Remote sensing of atmospheric particulate mass of dry PM2.5 near the ground: Method validation using ground-based measurements. Remote Sens Environ. 2016;173:59–68.

26. Lin CQ, Liu G, Lau AKH, Li Y, Li CC, Fung JCH, Lao XQ. High-resolution satellite remote sensing of provincial PM2.5 trends in China from 2001 to 2015. Atmos Environ. 2018;180:110–6.
27. Yan X, Shi W, Li Z, Li Z, Luo N, Zhao W, Wang H, Yu X. Satellite-based PM 2.5 estimation using fine-mode aerosol optical thickness over China. Atmos Environ. 2017;170:290–302.
28. Zhang Y, Li Z. Remote sensing of atmospheric fine particulate matter (PM2.5) mass concentration near the ground from satellite observation. Remote Sens Environ. 2015;160:252–62.
29. van Donkelaar A, Martin RV, Levy RC, da Silva AM, Krzyzanowski M, Chubarova NE, Semutnikova E, Cohen AJ. Satellite-based estimates of ground-level fine particulate matter during extreme events: A case study of the Moscow fires in 2010. Atmos Environ. 2011;45(34):6225–32.
30. Chen Z-Y, Zhang T-H, Zhang R, Zhu Z-M, Yang J, Chen P-Y, Ou C-Q, Guo Y. Extreme gradient boosting model to estimate PM2.5 concentrations with missing-filled satellite data in China. Atmos Environ. 2019;202:180–9.
31. Benas N, Beloconi A, Chrysoulakis N. Estimation of urban PM10 concentration, based on MODIS and MERIS/AATSR synergistic observations. Atmos Environ. 2013;79:448–54.
32. Gupta P, Christopher SA. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. J Geophys Res. 2009;114:D14.
33. Ma Z, Liu Y, Zhao Q, Liu M, Zhou Y, Bi J. Satellite-derived high resolution PM2.5 concentrations in Yangtze River Delta Region of China using improved linear mixed effects model. Atmos Environ. 2016;133:156–64.
34. Zhang T, Zhu Z, Gong W, Zhu Z, Sun K, Wang L, Huang Y, Mao F, Shen H, Li Z, et al. Estimation of ultrahigh resolution PM2.5 concentrations in urban areas using 160 m Gaofen-1 AOD retrievals. Remote Sens Environ. 2018;216:91–104.
35. He Q, Huang B. Satellite-based mapping of daily high-resolution ground PM2.5 in China via space-time regression modeling. Remote Sens Environ. 2018;206:72–83.
36. Qin K, Zou J, Guo J, Lu M, Bilal M, Zhang K, Ma F, Zhang Y. Estimating PM1 concentrations from MODIS over Yangtze River Delta of China during 2014–2017. Atmos Environ. 2018;195:149–58.
37. Reid CE, Jerrett M, Petersen ML, Pfister GG, Morefield PE, Tager IB, Raffuse SM, Balmes JR. Spatiotemporal prediction of fine particulate matter during the 2008 northern California wildfires using machine learning. Environ Sci Technol. 2015;49(6):3887–96.
38. Wei J, Huang W, Li Z, Xue W, Peng Y, Sun L, Cribb M. Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach. Remote Sens Environ. 2019;231:111221.
39. Zhan Y, Luo Y, Deng X, Chen H, Grieneisen ML, Shen X, Zhu L, Zhang M. Spatiotemporal prediction of continuous daily PM 2.5 concentrations across China using a spatially explicit machine learning algorithm. Atmos Environ. 2017;155:129–39.
40. Stafoggia M, Bellander T, Bucci S, Davoli M, de Hoogh K, De' Donato F, Gariazzo C, Lyapustin A, Michelozzi P, Renzi M, et al. Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. Environ Int. 2019;124:170–9.

41. Liang F, Xiao Q, Wang Y, Lyapustin A, Li G, Gu D, Pan X, Liu Y. MAIAC-based long-term spatiotemporal trends of PM2.5 in Beijing, China. Sci Total Environ 2018, 616–617:1589–1598.
42. Liu Y, Sarnat JA, Kilaru V, Jacob DJ, Koutrakis P. Estimating Ground-Level PM2.5 in the Eastern United States Using Satellite Remote Sensing. Environmental Science Technology. 2005;39(9):3269–78.
43. Pelletier B, Santer R, Vidot J. Retrieving of particulate matter from optical measurements: A semiparametric approach. Journal of Geophysical Research 2007, 112(D6).
44. Barmpadimos I, Hueglin C, Keller J, Henne S, Prévôt ASH. Influence of meteorology on PM10 trends and variability in Switzerland from 1991 to 2008. Atmos Chem Phys. 2011;11(4):1813–35.
45. Easter RC, Peters LK. Binary Homogeneous Nucleation: Temperature and Relative Humidity Fluctuations, Nonlinearity, and Aspects of New Particle Production in the Atmosphere. J Appl Meteorol. 1994;33(7):775–84.
46. Zhang B, Jiao L, Xu G, Zhao S, Tang X, Zhou Y, Gong C. Influences of wind and precipitation on different-sized particulate matter concentrations (PM2.5, PM10, PM2.5–10). Meteorol Atmos Phys. 2017;130(3):383–92.
47. Ephrath JE, Goudriaan J, Marani A. Modelling diurnal patterns of air temperature, radiation wind speed and relative humidity by equations from daily characteristics. Agric Syst. 1996;51(4):377–93.
48. Guo Z, Chang C, Wang R. A Novel Method to Downscale Daily Wind Statistics to Hourly Wind Data for Wind Erosion Modelling. In: Geo-Informatics in Resource Management and Sustainable Ecosystem: 2016. Berlin: Springer Berlin Heidelberg; 2016. pp. 611–9.
49. Liu S, Liang X-Z. Observed Diurnal Cycle Climatology of Planetary Boundary Layer Height. J Clim. 2010;23(21):5790–809.
50. Fu X, Wang S, Zhao B, Xing J, Cheng Z, Liu H, Hao J. Emission inventory of primary pollutants and chemical speciation in 2010 for the Yangtze River Delta region, China. Atmos Environ. 2013;70:39–50.
51. Wang S, Hao J. Air quality management in China: Issues, challenges, and options. J Environ Sci. 2012;24(1):2–13.
52. Air pollution control plan for key industries in the Yangtze River Delta region. http://www.mee.gov.cn/gkml/hbb/bwj/201411/t20141124_291839.htm. Accessed 2020/3/20.
53. Action plan for comprehensive treatment of air pollution in the Yangtze River Delta region during the autumn and winter of 2018–2019
Accessed 2020/3/20
Action plan for comprehensive treatment of air pollution in the Yangtze. River Delta region during the autumn and winter of 2018–2019.
http://www.mee.gov.cn/xxgk2018/xxgk/xxgk03/201811/t20181112_673371.html. Accessed 2020/3/20.
54. Xiao Q, Wang Y, Chang HH, Meng X, Geng G, Lyapustin A, Liu Y. Full-coverage high-resolution daily PM 2.5 estimation using MAIAC AOD in the Yangtze River Delta of China. Remote Sens Environ.

2017;199:437–46.
55. Song Z, Fu D, Zhang X, Han X, Song J, Zhang J, Wang J, Xia X. MODIS AOD sampling rate and its effect on PM2.5 estimation in North China. Atmos Environ. 2019;209:14–22.
56. Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc. 2011;137(656):553–97.
57. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.
58. Li X, Zhang X. Predicting ground-level PM2.5 concentrations in the Beijing-Tianjin-Hebei region: A hybrid remote sensing and machine learning approach. Environ Pollut. 2019;249:735–49.
59. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Statist. 2001;29(5):1189–232.
60. Kudraszow NL, Vieu P. Uniform consistency of NN regressors for functional variables. Statistics Probability Letters. 2013;83(8):1863–70.
61. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. BMC Bioinformatics. 2008;9:307.
62. Mass CF, Steenburgh WJ, Schultz DM. Diurnal Surface-Pressure Variations over the Continental United States and the Influence of Sea Level Reduction. Mon Weather Rev. 1990;119(12):2814–30.
63. Liu Q, Jia X, Quan J, Li J, Li X, Wu Y, Chen D, Wang Z, Liu Y. New positive feedback mechanism between boundary layer meteorology and secondary aerosol formation during severe haze events. Sci Rep. 2018;8(1):6095.
64. Tao M, Wang J, Li R, Wang L, Wang L, Wang Z, Tao J, Che H, Chen L. Performance of MODIS high-resolution MAIAC aerosol algorithm in China: Characterization and limitation. Atmos Environ. 2019;213:159–69.
65. Wang X, Sun W, Wang Z, Wang Y, Ren H. Meteorological Parameters and Gaseous Pollutant Concentrations as Predictors of Ground-level PM2.5 Concentrations in the Beijing-Tianjin-Hebei Region, China. Aerosol Air Quality Research. 2019;19(8):1844–55.
66. Cheng J, Su J, Cui T, Li X, Dong X, Sun F, Yang Y, Tong D, Zheng Y, Li Y, et al. Dominant role of emission reduction in PM2.5 air quality improvement in Beijing during 2013–2017: a model-based decomposition analysis. Atmos Chem Phys. 2019;19(9):6125–46.
67. Guo L, Luo J, Yuan M, Huang Y, Shen H, Li T. The influence of urban planning factors on PM2.5 pollution exposure and implications: A case study in China based on remote sensing, LBS, and GIS data. Sci Total Environ. 2019;659:1585–96.
68. Xie Y, Wang Y, Zhang K, Dong W, Lv B, Bai Y. Daily Estimation of Ground-Level PM2.5 Concentrations over Beijing Using 3 km Resolution MODIS AOD. Environ Sci Technol. 2015;49(20):12280–8.
69. Kim S-W, Yoon S-C, Kim J, Kim S-Y. Seasonal and monthly variations of columnar aerosol optical properties over east Asia determined from multi-year MODIS, LIDAR, and AERONET Sun/sky radiometer measurements. Atmos Environ. 2007;41(8):1634–51.

70. Ding A, Huang X, Nie W, Chi X, Xu Z, Zheng L, Xu Z, Xie Y, Qi X, Shen Y, et al. Significant reduction of PM2.5 in eastern China due to regional-scale emission control: evidence from SORPES in 2011–2018. Atmos Chem Phys. 2019;19(18):11791–801.
71. Shtein A, Kloog I, Schwartz J, Silibello C, Michelozzi P, Gariazzo C, Viegi G, Forastiere F, Karnieli A, Just AC, et al: Estimating Daily PM2.5 and PM10 over Italy Using an Ensemble Model. Environ Sci Technol 2019.

# Figures

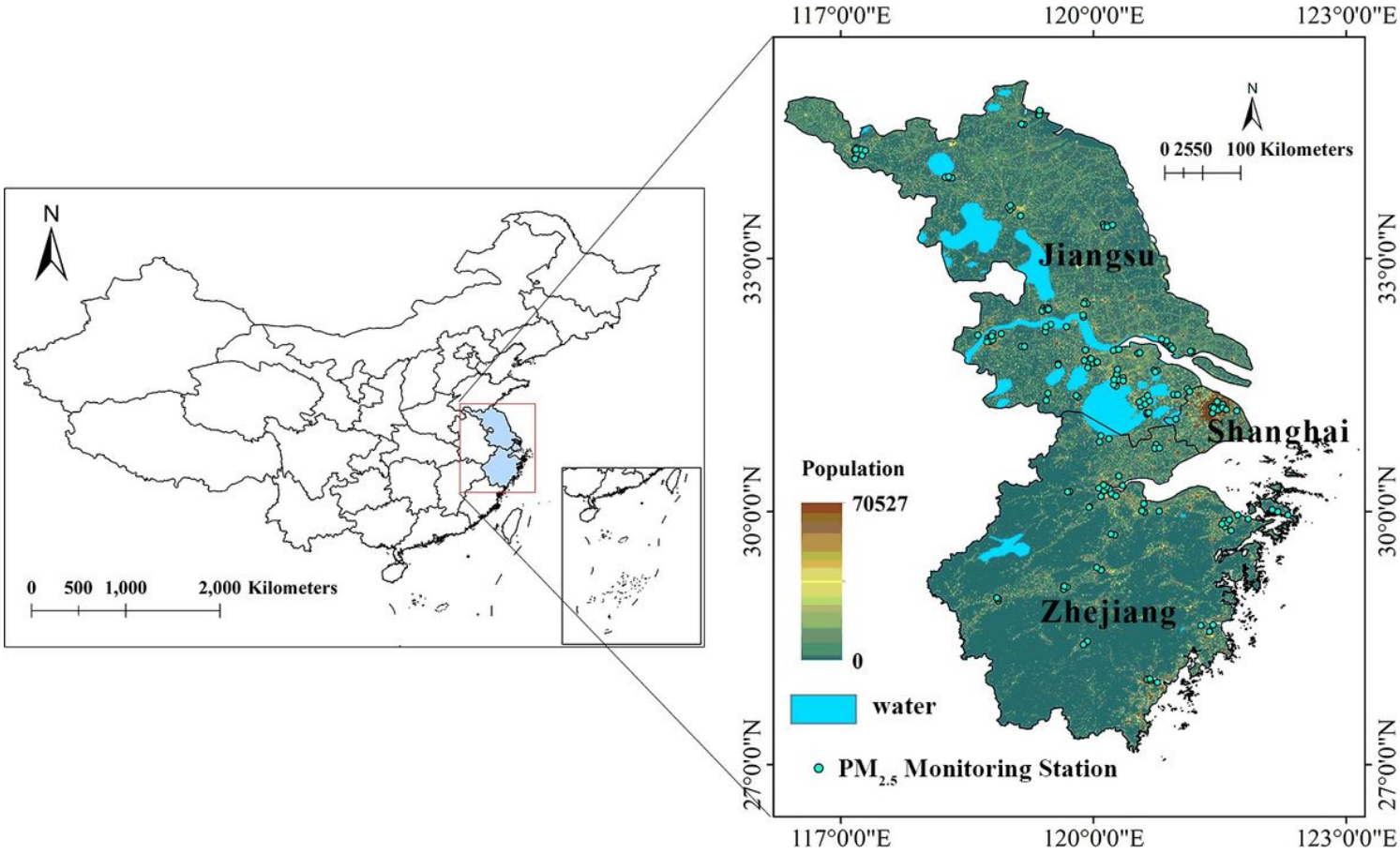
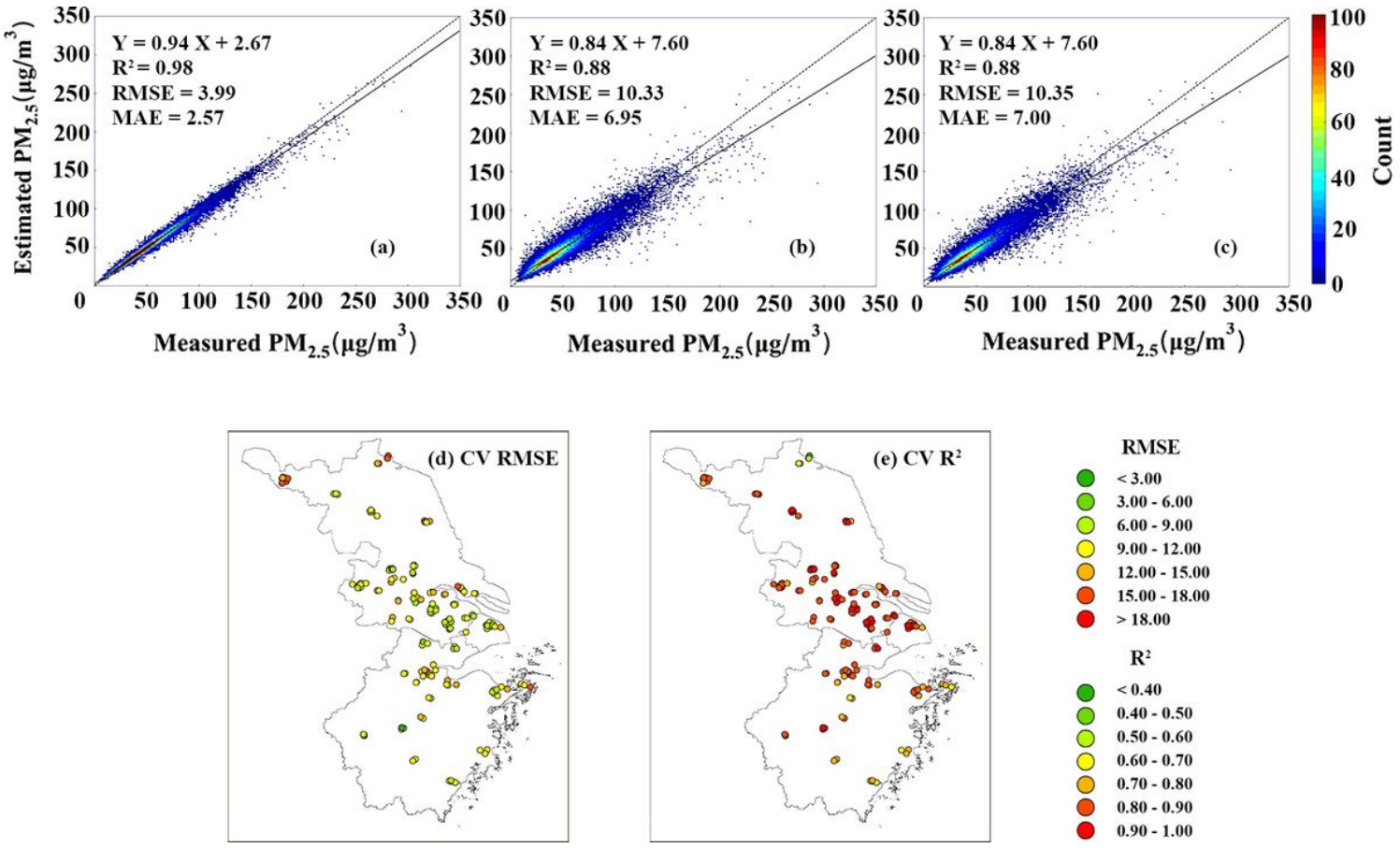


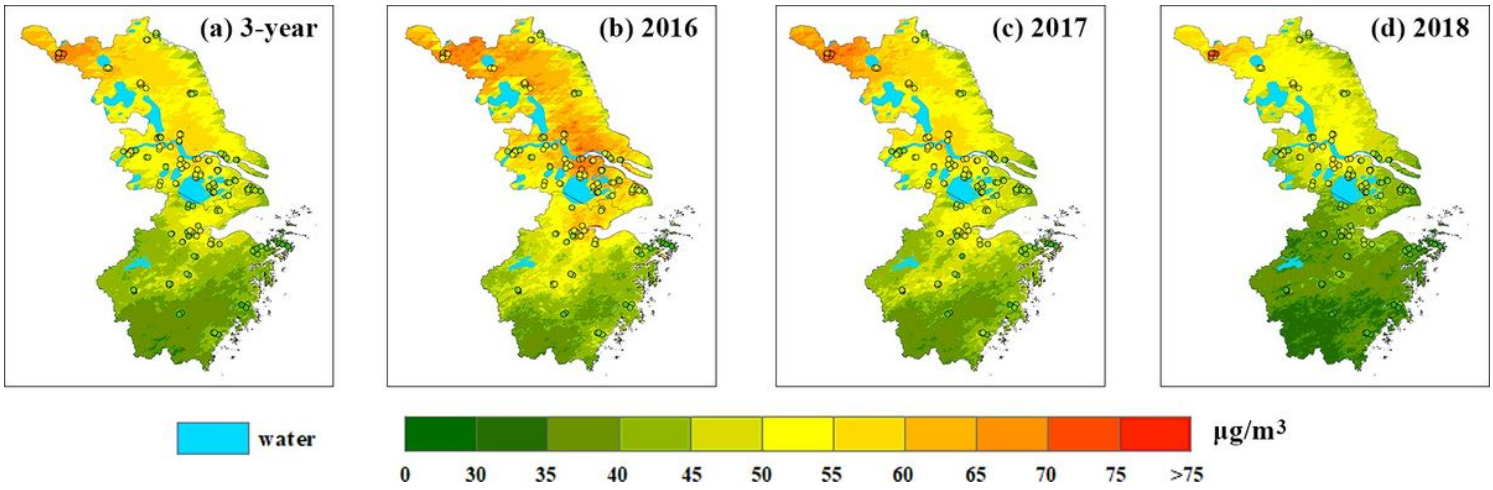

**Figure 1**

Locations of the study area and ground PM2.5 monitoring stations.



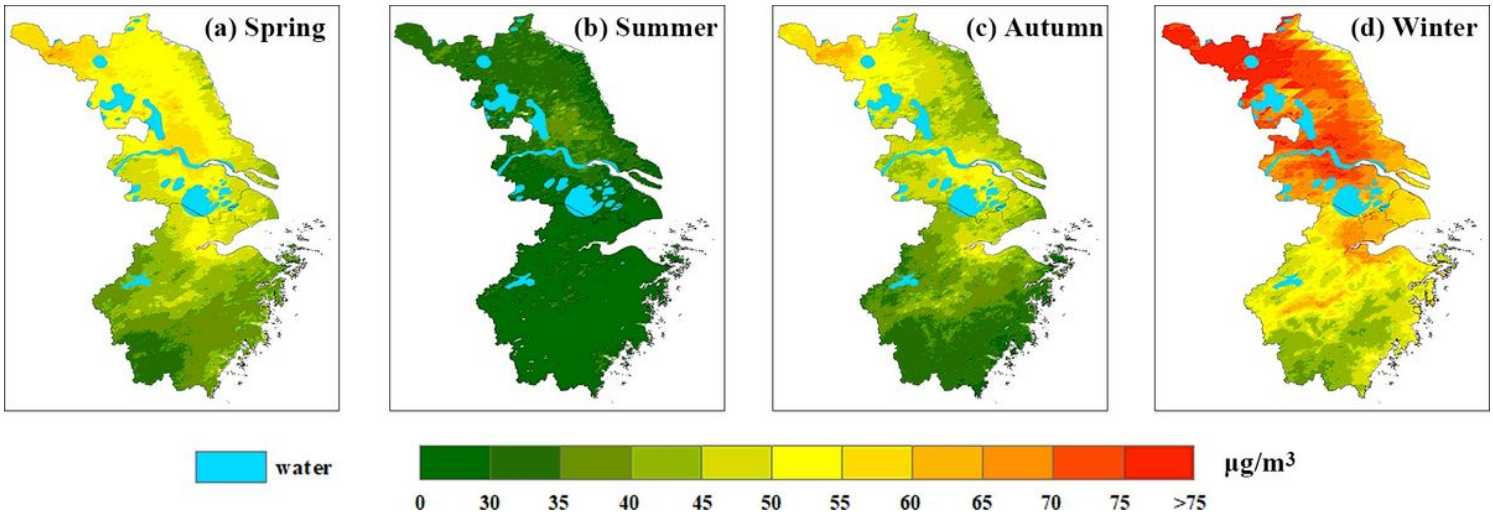

## Figure 2

Density scatterplots of model fitting (a) and the sample-based (b), site-based (c) CV results of the RF (Mete-multi) model. (d) and (e) are the site-based ten-fold CV results at each monitoring station. (d) CV RMSE, where the red color represents high estimation bias and the green colorrepresents low bias. (e) CV R2, wherethe red color represents a high coefficient of determination and the green color represents a low coefficient of determination.




## Figure 3

Yearly mean distribution of estimated PM2.5 concentrations and the ground-level monitoring sites. The red color represents the high PM2.5 concentrations, and the green color represents the low PM2.5 concentrations. (a) 3-year average PM2.5 concentrations. (b), (c), (d) Mean PM2.5 concentration distributions in 2016, 2017 and 2018.




## Figure 4

3-year seasonal mean distribution of estimated PM2.5 concentrations. The red color represents high PM2.5 concentrations, and the green color represents low PM2.5 concentrations. (a) Spring: March, April, and May (b) Summer: June, July, and August (c) Autumn: September, October, and November (d) Winter: December, January, and February.

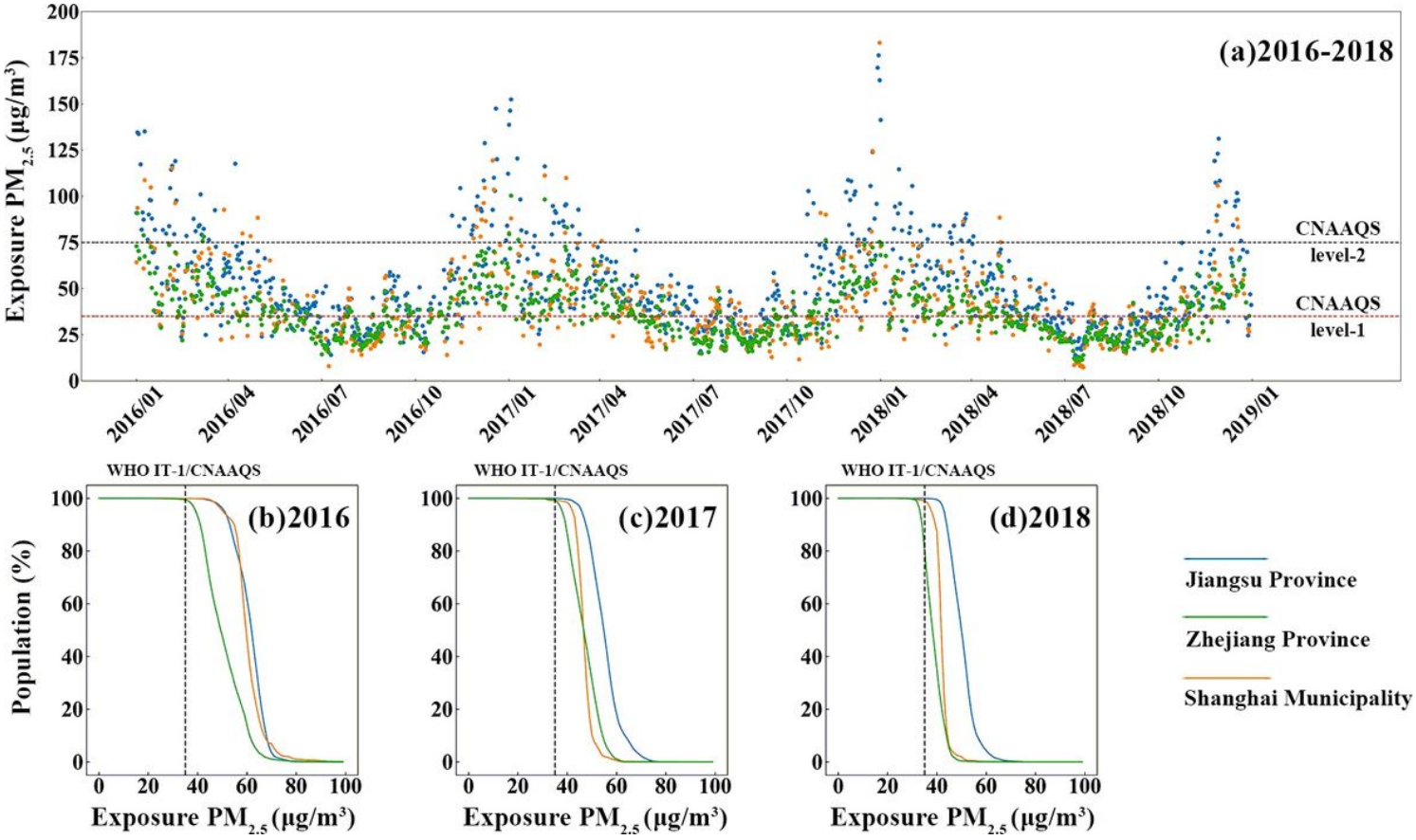


**Figure 5**

Residential exposure level to PM2.5; blue, green and orange represent Jiangsu Province, Zhejiang Province and Shanghai Municipality, respectively. (a) Daily exposure level to PM2.5 during the study period. The black dashed line represents the PM2.5 concentration of 75 μg/m3 (CNAAQS daily level-2 standard), and the red dashed line represents PM2.5 concentration of 35 μg/m3 (CNAAQS daily level-1 standard). (b)(c)(d) show the percentage of the population exposed to PM2.5 higher than the annual mean levels in 2016, 2017, and 2018, respectively. The dashed lines represent the annual PM2.5 concentrations of the WHO IT-1 standard as well as the CNAAQS standard (35 μg/m3).

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.docx