



## Article

# A National-Scale 1-km Resolution PM<sub>2.5</sub> Estimation Model over Japan Using MAIAC AOD and a Two-Stage Random Forest Model

Chau-Ren Jung <sup>1,2</sup>, Wei-Ting Chen <sup>3</sup> and Shoji F. Nakayama <sup>1,\*</sup>

- <sup>1</sup> Japan Environment and Children's Study Programme Office, National Institute for Environmental Studies, Tsukuba 305-8506, Japan; crjung@mail.cmu.edu.tw
- <sup>2</sup> Department of Public Health, College of Public Health, China Medical University, Taichung 406040, Taiwan
- <sup>3</sup> Department of Atmospheric Sciences, National Taiwan University, Taipei 106319, Taiwan; weitingc@ntu.edu.tw
- \* Correspondence: fabre@nies.go.jp

**Abstract:** Satellite-based models for estimating concentrations of particulate matter with an aerodynamic diameter less than 2.5  $\mu\text{m}$  (PM<sub>2.5</sub>) have seldom been developed in islands with complex topography over the monsoon area, where the transport of PM<sub>2.5</sub> is influenced by both the synoptic-scale winds and local-scale circulations compared with the continental regions. We validated Multi-Angle Implementation of Atmospheric Correction (MAIAC) aerosol optical depth (AOD) with ground observations in Japan and developed a 1-km-resolution national-scale model between 2011 and 2016 to estimate daily PM<sub>2.5</sub> concentrations. A two-stage random forest model integrating MAIAC AOD with meteorological variables and land use data was applied to develop the model. The first-stage random forest model was used to impute the missing AOD values. The second-stage random forest model was then utilised to estimate ground PM<sub>2.5</sub> concentrations. Ten-fold cross-validation was performed to evaluate the model performance. There was good consistency between MAIAC AOD and ground truth in Japan (correlation coefficient = 0.82 and 74.62% of data falling within the expected error). For model training, the model showed a training coefficient of determination ( $R^2$ ) of 0.98 and a root mean square error (RMSE) of 1.22  $\mu\text{g}/\text{m}^3$ . For the 10-fold cross-validation, the cross-validation  $R^2$  and RMSE of the model were 0.86 and 3.02  $\mu\text{g}/\text{m}^3$ , respectively. A subsite validation was used to validate the model at the grids overlapping with the AERONET sites, and the model performance was excellent at these sites with a validation  $R^2$  (RMSE) of 0.94 (1.78  $\mu\text{g}/\text{m}^3$ ). Additionally, the model performance increased as increased AOD coverage. The top-ten important predictors for estimating ground PM<sub>2.5</sub> concentrations were day of the year, temperature, AOD, relative humidity, 10-m-height zonal wind, 10-m-height meridional wind, boundary layer height, precipitation, surface pressure, and population density. MAIAC AOD showed high retrieval accuracy in Japan. The performance of the satellite-based model was excellent, which showed that PM<sub>2.5</sub> estimates derived from the model were reliable and accurate. These estimates can be used to assess both the short-term and long-term effects of PM<sub>2.5</sub> on health outcomes in epidemiological studies.

**Keywords:** aerosol optical depth; PM<sub>2.5</sub>; random forest model; satellite-based estimation model



**Citation:** Jung, C.-R.; Chen, W.-T.; Nakayama, S.F. A National-Scale 1-km Resolution PM<sub>2.5</sub> Estimation Model over Japan Using MAIAC AOD and a Two-Stage Random Forest Model. *Remote Sens.* **2021**, *13*, 3657. <https://doi.org/10.3390/rs13183657>

Academic Editors: Muhammad Bilal and Janet E. Nichol

Received: 19 August 2021  
Accepted: 10 September 2021  
Published: 13 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Particulate matter (PM) is a mixture of solid particles and liquid droplets suspended in the air that can be classified into coarse (PM with an aerodynamic diameter less than 10  $\mu\text{m}$ ; PM<sub>10</sub>) and fine (PM with an aerodynamic diameter less than 2.5  $\mu\text{m}$ ; PM<sub>2.5</sub>) fractions based on their size [1,2]. PM<sub>2.5</sub> is considered more toxic than PM<sub>10</sub> because of its smaller diameter, meaning it can deposit into the alveolar region of the lungs and induce a series of immune responses [2]. Many epidemiological studies reported that exposure to PM<sub>2.5</sub> is associated with adverse health outcomes, such as allergic respiratory diseases, asthma, cardiovascular

diseases, neuropsychological diseases, and mortality [3–7]. The major sources of PM<sub>2.5</sub> in Japan include sea salt, biomass combustion, soil dust, and secondary aerosols, which are derived from local emission sources and long-range transportation [8]. The higher PM<sub>2.5</sub> concentrations in western Japan, particularly in spring and winter, are mainly due to long-range transportation from Asia [9]. Owing to the significant adverse health effects of PM<sub>2.5</sub>, there is an urgent need to better monitor its temporal and spatial distributions.

In Japan, the PM<sub>2.5</sub> monitoring network was launched in 2010 with only a few monitoring stations, and the number of monitoring stations has increased over time. However, the lack of extensive PM<sub>2.5</sub> measurements before 2010, and sparse monitoring stations in mountainous areas and Hokkaido, may limit the application of PM<sub>2.5</sub> data in long-term epidemiological studies in Japan and cause exposure misclassification of subjects. Land use regression (LUR) is a cost-effective approach to estimate large-scale PM<sub>2.5</sub> distributions [10], but lacks temporal resolution owing to the unavailability of temporally resolved land use data. Hence, it is difficult to estimate daily PM<sub>2.5</sub> concentrations using LUR. With the merits of large-scale coverage and daily routine records, satellite-based aerosol optical depth (AOD), defined as the integrated extinction coefficient over the vertical atmospheric column above the earth's surface, is a useful and reliable indicator of daily ground PM<sub>2.5</sub> concentrations [11]. However, no study has used AOD to develop a satellite-based PM<sub>2.5</sub> estimation model in Japan.

AOD generated from the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the NASA Terra and Aqua satellites is the most popular product, which has been applied widely with meteorological variables and land use data to estimate PM<sub>2.5</sub> concentrations across the world [12]. However, most previous studies were conducted in continental regions, such as the United States [13–15], Europe [16,17], and mainland China [18–21]. Only a few studies were conducted on islands in moist environment because the missing rates of AOD are generally high in coastal regions and islands, mainly owing to frequent cloud occurrence [22,23]. Early studies relied on MODIS AOD with coarse spatial resolution using the Dark Target (10 or 3 km resolution) and Deep Blue (10 km resolution) algorithms [24]. A new advanced algorithm with 1-km-resolution, the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm, was released in 2018 [25]. The high-resolution MAIAC AOD product is potentially useful for estimating the ground PM<sub>2.5</sub> in islands with complex topography and land use type distribution such as Japan.

Several traditional statistical models, such as multivariate linear regression, a linear mixed effect model, geographically weighted regression, and a generalised additive model, have been utilised with AOD to establish the empirical models for PM<sub>2.5</sub> (as reviewed by [26]). However, these statistical models generally make strong assumptions about the data. Machine learning algorithms that take into account the complex non-linear interactions between variables without needing to make assumptions about distributions were also proposed to develop satellite-based AOD-PM<sub>2.5</sub> models. These machine learning algorithms include random forest [14,27], gradient boosting [28], extreme gradient boosting (XGBoost) [29], support vector machine [21], and deep learning [30,31]. These studies have proved that machine learning algorithms outperform traditional statistical models.

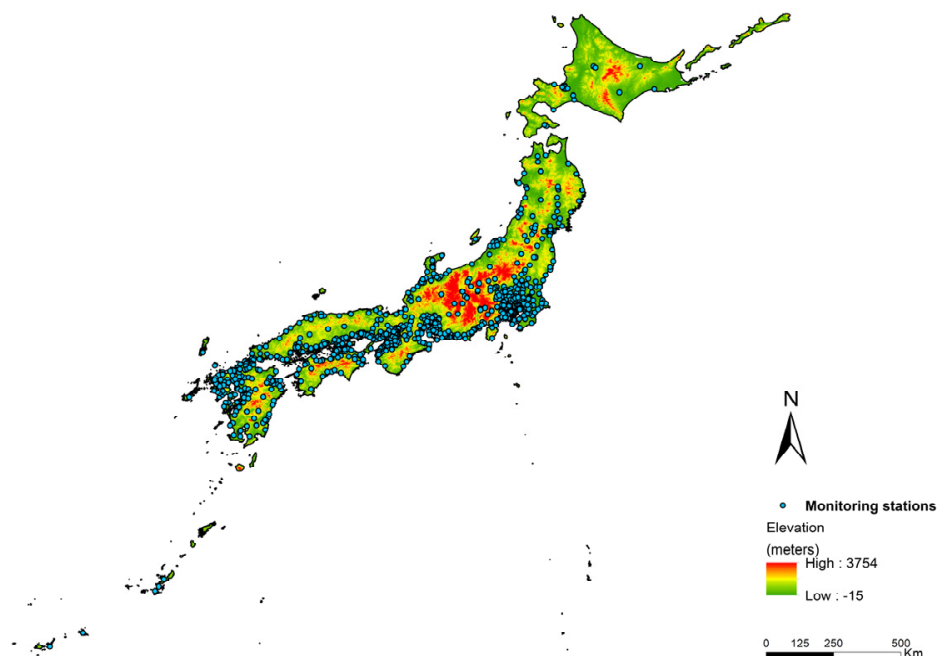
In this study, we validated MAIAC AOD with ground measurements of AOD, and established a 1-km-resolution satellite-based model over Japan between 2011 and 2016 to estimate daily PM<sub>2.5</sub> concentrations. The objectives of this study were (1) to validate performance of satellite-based AOD over Japan; (2) to apply a two-stage random forest model incorporating AOD, meteorological variables, and land use data to estimate daily PM<sub>2.5</sub> concentrations; and (3) to identify important predictors of outdoor PM<sub>2.5</sub> concentrations.

## 2. Materials and Methods

### 2.1. Study Area

The study area included the main islands of Japan and the island group of Okinawa within 24° to 46°N and 123° to 149°E (Figure 1). The remote offshore islands, i.e., the Bonin

Islands, were excluded from the study. To facilitate data integration, the study area was divided into 379,643 grids with a resolution of 45 s in longitude and 30 s in latitude based on the Japan third mesh data (approximately 1-km  $\times$  1-km) [32].



**Figure 1.** Study area and locations of PM<sub>2.5</sub> monitoring stations in 2016.

The study collected satellite-based AOD, meteorological data, and land use data (the data sources, coverage years, and spatial and temporal resolutions are summarised in Table S1). The details of these data are described in the following sections.

## 2.2. MAIAC AOD

The 1-km-resolution MAIAC algorithm-based level-gridded AOD data (MCD19A2 collection 6 product) between 1 January 2011 and 31 December 2016 were downloaded from the Level 1 and Atmospheric Archive and Distribution System Distributed Active Archive Center (LAADS DAAC; <https://ladsweb.modaps.eosdis.nasa.gov/search/>, accessed on 12 September 2021). MAIAC AOD at 550 nm (SDS name: Optical\_Depth\_055) from Terra (overpass at around 10:30 a.m.) and Aqua (overpass at around 1:30 p.m.) with the best quality (QA.CloudMask = Clear (code: 001) and QA.AdjacencyMask = Clear (code: 000)) were retrieved separately. A simple linear regression between MAIAC AOD from Terra and Aqua was developed by using complete pairs of AOD from these two satellites. If the MAIAC AOD from Terra was missing for a specific day, the MAIAC AOD from Aqua was used to impute the missing value by simple linear regression, and vice versa [23]. Then, the arithmetic means of AOD from Terra and Aqua were calculated to represent daily values.

## 2.3. AERONET AOD

The Aerosol RObotic NETwork (AERONET) programme is a global ground-based remote sensing aerosol network established by NASA and PHOTONS (PHOTométrie pour le Traitement Opérationnel de Normalisation Satellitaire; Univ. of Lille 1, CNES and CNRS-INSU) with several collaborators, including national agencies, institutes, and universities across the world [33]. The AOD from the AERONET can be used as ‘ground truth’ to validate satellite-based measurements. Total optical depths based on the AOD levels (Level 2.0: cloud screened and quality assured) of 12 stations in Japan (Figure S1) were downloaded from the Goddard Space Flight Center ([https://aeronet.gsfc.nasa.gov/new\\_](https://aeronet.gsfc.nasa.gov/new_)

[web/index.html](#), accessed on 12 September 2021). The AERONET did not provide the AOD at 550 nm and therefore this was interpolated from the AOD at 500 and 675 nm [34].

#### 2.4. Ground PM Measurements

Hourly PM<sub>2.5</sub> measurements from 2011 to 2016 were downloaded from the environmental numerical database of the National Institute for Environmental Studies [35]. There were 1,068 stations measuring PM<sub>2.5</sub> that were maintained by local government in 2016 (Figure 1). The monitoring stations used a  $\beta$ -ray attenuation method, tapered element oscillating microbalance, and a light-scattering method to measure PM<sub>2.5</sub> concentrations continuously and hourly. The air quality data were collected, summarised, and released by the National Institute for Environmental Studies [35], and are available online (<https://www.nies.go.jp/igreen/index.html>, accessed on 12 September 2021).

#### 2.5. Meteorological Variables

The in situ measurements of daily temperature, relative humidity, precipitation, and surface pressure recorded by the automated meteorological data acquisition system (AMeDAS) during 2011–2016 were downloaded from the Japan Meteorological Agency. There are around 1300 rain gauges across Japan, and approximately 921 and 147 monitoring stations observing daily temperature and relative humidity across Japan, respectively [36]. Missing values of meteorological variables were excluded. The regression-kriging method was applied to estimate daily temperature and surface pressure in a 1-km-resolution grid lacking weather-monitoring stations [37]. Additionally, ordinary kriging was utilised to estimate daily values of relative humidity and precipitation in 1-km grids.

Boundary layer height, 10-m-height zonal wind (u10), and 10-m-height meridional wind (v10) were obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) Re-analysis Interim (ERA-Int). Data at 00.00 and 12.00 UTC, and at four steps (i.e., 3, 6, 9, and 12 h) after the observation time points, with a spatial resolution of 0.125° longitude  $\times$  0.125° latitude, were downloaded from the ECMWF online depository (<https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/>, accessed on 12 September 2021). The inverse distance weighted (IDW) method was used to interpolate the boundary layer height, u10, and v10 data to the centroids of grids. The fixed search radius was set to 15 km for each centroid of fixed grids, and all the observed values of boundary layer height within the 15 km searching radius of a specific grid were averaged. The same IDW method was applied in our previous study [23].

Five-kilometre cloud-fraction data were accessed from the Terra and Aqua MODIS Collection 6 Level cloud product (MOD06\_L2 and MYD06\_L2), which consisted of cloud optical and physical parameters downloaded from the LAADS DAAC. The cloud-fraction data from Terra and Aqua were averaged as daily values. The IDW method with a fixed search radius of 7.5 km was used to interpolate the cloud fraction values into each grid.

#### 2.6. Land Use Data

Urban and built-up area data with 50 m resolution during 2006–2011 (ID number of the category = #2) were retrieved from the High-Resolution Land Use and Land Cover (HRLULC) map published by the Japan Aerospace Exploration Agency (JAXA) Earth Observation Research Center ([https://www.eorc.jaxa.jp/ALOS/en/lulc/lulc\\_index.htm](https://www.eorc.jaxa.jp/ALOS/en/lulc/lulc_index.htm), accessed on 12 September 2021). Industrial area data created in 2009 were accessed from the Ministry of Land, Infrastructure, Transport and Tourism (<https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-L05.html>, accessed on 12 September 2021). Road network data were retrieved from the Global Map Japan v2.2 Vector data released in 2016 by the Geoinformation Authority of Japan ([https://www.gsi.go.jp/kankyochiri/gm\\_japan\\_e.html](https://www.gsi.go.jp/kankyochiri/gm_japan_e.html), accessed on 12 September 2021). The distances from the centroid of the grid to the nearest primary road and highway were calculated, as well as the total length of roads within each grid.

The normalised difference vegetation index (NDVI), a measurement of plant health, was extracted from the Terra MODIS MOD13Q1 product with a 16-day interval and 250 m spatial resolution in the sinusoidal coordinate system.

### 2.7. Population Counts

Annual population count data with 100 m resolution during 2011–2016 were downloaded from the WorldPop website (<https://www.worldpop.org/geodata/listing?id=29>, accessed on 12 September 2021). The annual average population count in each 1-km grid was calculated.

### 2.8. Elevation

Elevation data with 30 m spatial resolution was accessed from the Advance Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model v002 (ASTGTM.002) announced in 2011. The average elevation within each 1-km grid was calculated.

### 2.9. Statistical Methods

The Pearson's correlation coefficient (R), RMSE, mean bias error (MBE), percentage of data within the expected error (EE;  $\pm(0.05 + 0.15\text{AERONET AOD})$ ), and slope of the linear regression were selected to evaluate the consistency between MAIAC AOD and AERONET AOD [38,39].

Random forest is an ensemble learning model, which has outperformed most traditional machine learning algorithms [40]. The random forest model is an extended machine learning algorithm of bagging (bootstrap aggregation) that not only builds a number of decision trees on bootstrapped training samples, but also randomly selects a subset of predictors ( $m_{try}$ ) from the full set of predictors when building these trees [40,41]. For regression, the random forest model aggregates predictions from trees by averaging the predicted values from trees. In the present study, a two-stage random forest model was applied to develop the satellite-based PM<sub>2.5</sub> model. In the first stage, random forest was used to impute the missing values of MAIAC AOD on a weekly basis. Since satellite-based AOD values are frequently missing due to the presence of clouds and in the mountainous areas, longitude, latitude, temperature, relative humidity, cloud fraction, and elevation were used to impute missing AOD values based on the previous study and our early work [19,42]. The formula of the first-stage random forest model is as follows:

$$\text{AOD}_{ij} = f(\text{Long}_i, \text{Lat}_i, \text{Temp}_{ij}, \text{Hum}_{ij}, \text{CF}_{ij}, \text{Elevation}_i) \quad (1)$$

where  $\text{AOD}_{ij}$  is the AOD value in grid  $i$  at day  $j$ ;  $\text{Long}_i$  and  $\text{Lat}_i$  are longitude and latitude in grid  $i$ ;  $\text{Temp}_{ij}$  is temperature in grid  $i$  at day  $j$ ;  $\text{Hum}_{ij}$  is relative humidity in grid  $i$  at day  $j$ ;  $\text{CF}_{ij}$  is the cloud fraction in grid  $i$  at day  $j$ ; and  $\text{Elevation}_i$  is elevation in grid  $i$ . The numbers of trees and  $m_{try}$  were set to 1000 and 2 (one-third of the number of predictors), respectively. The weekly models were evaluated by calculating the out-of-bag (OOB) coefficient of determination ( $R^2$ ) and root mean square error (RMSE).

The second-stage random forest model was utilised to estimate ground PM<sub>2.5</sub> concentrations. The predictors in the second-stage model were selected based on the previous studies and our early works [19,23,42,43]. All the predictors have plausible physical mechanisms on PM<sub>2.5</sub> concentrations. Temperature, relative humidity, boundary layer height, surface pressure, wind speed and direction, and precipitation significantly influence the distribution and formation of PM<sub>2.5</sub> [44]. Industrial areas, road networks, and human population are sources of PM<sub>2.5</sub>. Accessibility of data is also a consideration for including predictors. The other gaseous pollutants, such as nitrogen oxides and sulphur dioxides, are precursors of PM<sub>2.5</sub>, and we did not include them in the model since they are not available in grids without air quality monitoring stations. The formula of the second-stage random forest is as follows:

$$PM_{2.5\ ij} = f(AOD_{ij}, Days_j, Temp_{ij}, Hum_{ij}, BLH_{ij}, SP_{ij}, U10_{ij}, V10_{ij}, Prec_{ij}, NDVI_i, Indus\_area_i, Urban\_area_i, Elevation_i, Road\_len_i, Distoprim_i, Distohigh_i, Pop_i) \quad (2)$$

where  $PM_{2.5\ ij}$  is  $PM_{2.5}$  value in grid  $i$  at day  $j$ ;  $AOD_{ij}$  is AOD value in grid  $i$  at day  $j$ ;  $Days_j$  is the day of the year;  $Temp_{ij}$  is temperature in grid  $i$  at day  $j$ ;  $Hum_{ij}$  is relative humidity in grid  $i$  at day  $j$ ;  $BLH_{ij}$  is boundary layer height in grid  $i$  at day  $j$ ;  $SP_{ij}$  is surface pressure in grid  $i$  at day  $j$ ;  $U10_{ij}$  is zonal wind at a height of 10 m in grid  $i$  at day  $j$ ;  $V10_{ij}$  is meridional wind at a height of 10 m in grid  $i$  at day  $j$ ;  $Prec_{ij}$  is precipitation in grid  $i$  at day  $j$ ;  $NDVI_i$  is the normalised difference vegetation index in grid  $i$ ;  $Indus\_area_i$  is the area of industrial facilities in grid  $i$ ;  $Urban\_area_i$  is the urban and built-up area in grid  $i$ ;  $Elevation_i$  is elevation in grid  $i$ ;  $Road\_len_i$  is the length of roads in grid  $i$ ;  $Distoprim_i$  is the distance from the grid centroid to the nearest primary road in grid  $i$ ;  $Distohigh_i$  is the distance from the grid centroid to the nearest highway in grid  $i$ ; and  $Pop_i$  is the average population count in grid  $i$ . The descriptive statistics of predictors from 2011 to 2016 are presented in Table S2. Similar to the first-stage model, the number of trees was set to 1000, and  $m_{try}$  was tuned from 1 to 17.  $m_{try}$  was finally set to 15 for the second-stage model according to the highest OOB  $R^2$ . Ten-fold cross-validation was performed to evaluate the performance of the second-stage model. During the cross-validation process, the dataset was randomly divided into ten non-overlapping and equally sized partitions, and nine partitions and one partition were used as training and validating data, respectively. The cross-validation process was repeated ten times until all partitions had been used to validate the model performance. The cross-validation results were represented by cross-validation  $R^2$ , RMSE, percentage relative error (RE;  $(\text{mean}(\text{Observations}) - \text{mean}(\text{Predictions}))/\text{mean}(\text{Observations}) \times 100\%$ ), and mean absolute error (MAE). A tree-based ensemble model—eXtreme Gradient Boosting (XGBoost) [45] was performed to estimate ground  $PM_{2.5}$  concentrations and compared with the performance of random forest. The hyperparameters tuned to optimise the performance of XGBoost is shown in Table S3.

In addition to the ten-fold cross-validation, a subsite validation that used the grids overlapping with the AERONET sites as the validating data was conducted to check the model performance. Additionally, the ten-fold cross-validation results were stratified by seasons (spring: March–May; summer: June–August, autumn: September–November; winter: December–February) and quartiles of AOD coverage (below 25th percentile: <16.61%; 25th–50th percentile: 16.61–23.79%; 50th–75th percentile: 23.79–33.26%; and above 75th percentile:  $\geq 33.26\%$ ) to evaluate the model performance.

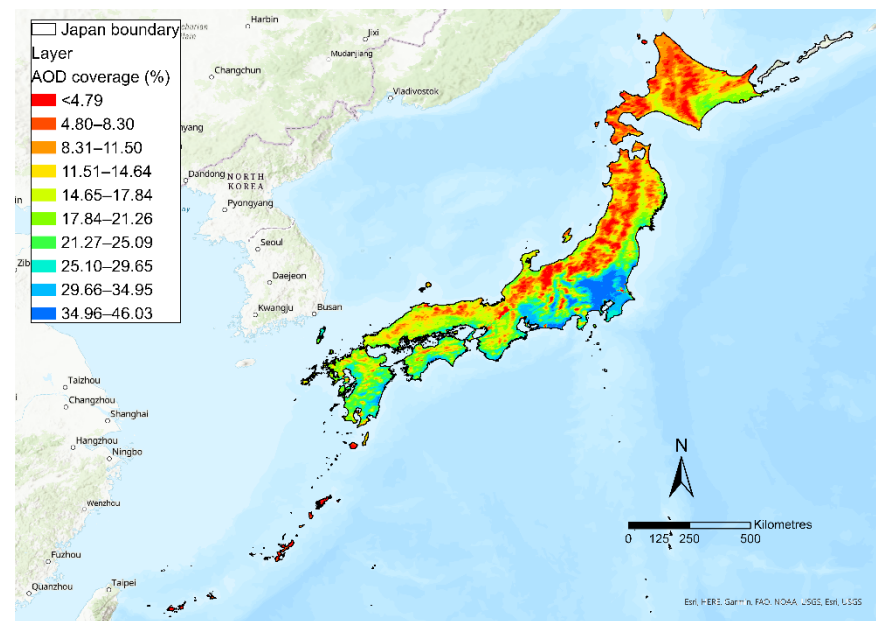
The relative importance of predictors in random forest model was determined by calculating permutation importance scores. Additionally, accumulated local effect (ALE) plots were used to explore the effects of predictors on predicted  $PM_{2.5}$  values.

All analyses were conducted using R software v4.1.0 (*caret*, *ranger*, *iml*, and *xgboost* packages; R Core Team, Vienna, Austria) and ArcGIS Pro 2.8.0 (ESRI, Redlands, California).

### 3. Results

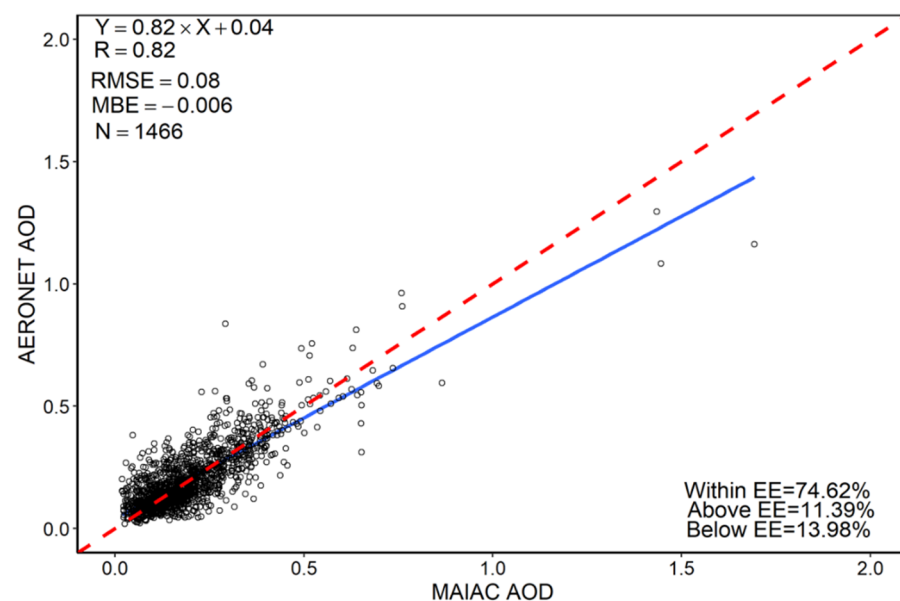
#### 3.1. Validation of MAIAC AOD with Ground Measurements over Japan

The overall mean coverage of MAIAC AOD in Japan was 15.48%. The highest coverage was in autumn (19.48%), followed by spring (17.94%), winter (14.05%), and summer (10.51%). The spatial distribution of MAIAC AOD coverage is shown in Figure 2. The mean coverage of AOD was higher in the Kanto, southern Chubu, and Kyushu regions; however, the mean coverage was extremely low (<4.79%) in the Hokkaido, Tohoku, and Okinawa regions (Figure 2).

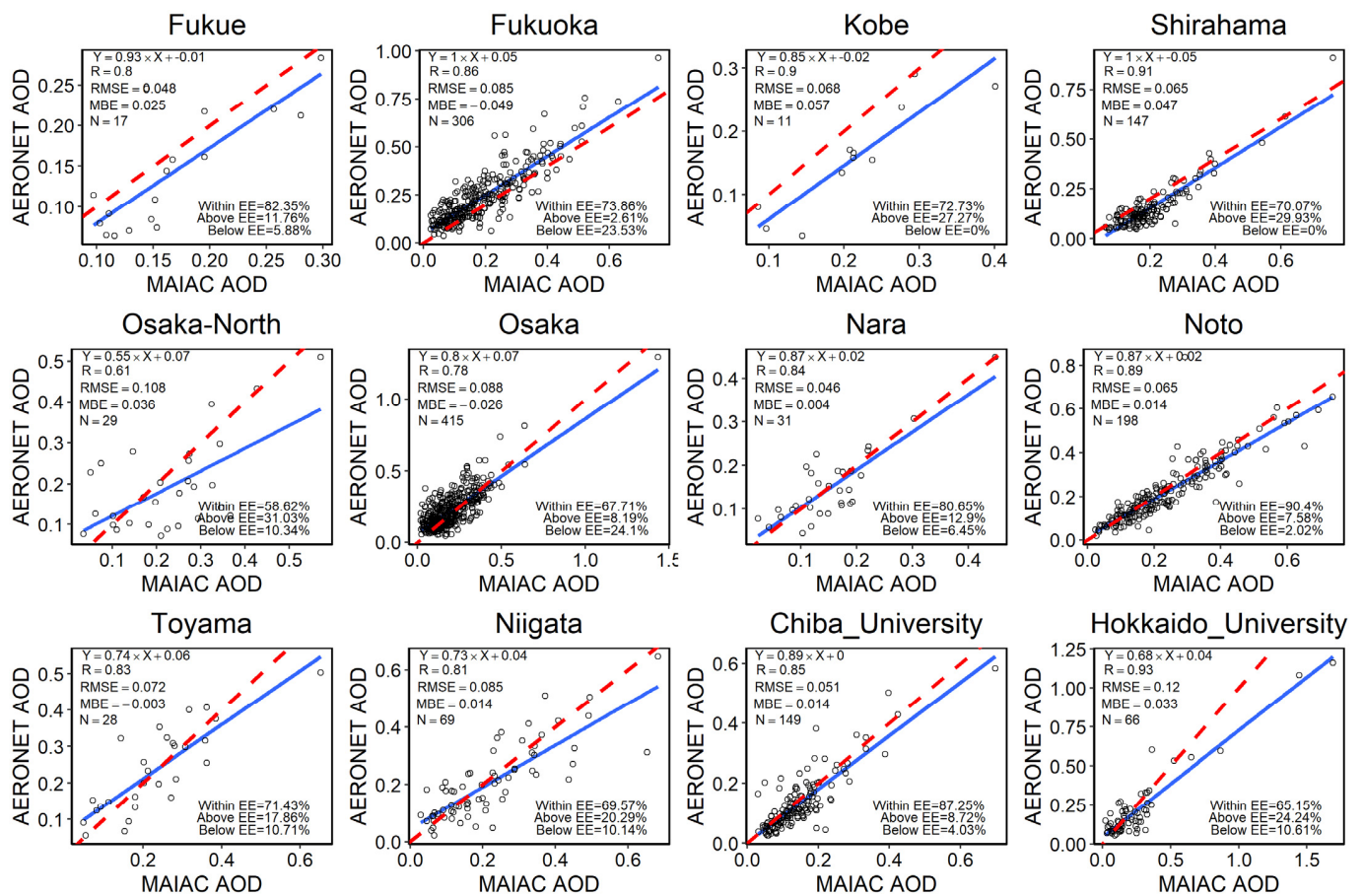


**Figure 2.** The mean coverage of Multi-Angle Implementation of Atmospheric Correction aerosol optical depth during 2011–2016 in Japan.

The descriptive statistics for MAIAC AOD and AERONET AOD are shown in Table S4. Overall, the 1-km-resolution MAIAC AOD was highly correlated with AERONET AOD ( $R = 0.82$ ), 74.62% of data falling within the EE (Figure 3). The MAIAC AOD slightly underestimated AOD values in Japan ( $MBE = -0.006$ ) (Figure 3). The performance of MAIAC AOD was highest in the Hokkaido University ( $R = 0.93$ ,  $RMSE = 0.12$ ,  $MBE = 0.033$ , and 65.15% of data within the EE), while the performance of MAIAC was lowest in the Osaka-North station ( $R = 0.61$ ,  $RMSE = 0.108$ ,  $MBE = 0.036$ , and 58.62% of samples within the EE) (Figure 4).



**Figure 3.** Comparison of daily AEROSOL ROBOTIC NETWORK (AERONET) aerosol optical depth (AOD; 9:00 a.m. to 3:00 p.m. local time) and Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD. The blue solid line shows the linear regression between MAIAC AOD and AERONET AOD. The red dashed line is the one-one line and the MAIAC AOD and AERONET AOD values are equal along the red dashed line.



**Figure 4.** Comparison of daily AERONET aerosol optical depth (AOD; 9:00 a.m. to 3:00 p.m. local time) and Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD at the 12 sites in Japan. The blue solid line shows the linear regression between MAIAC AOD and AERONET AOD. The red dashed line is the one-one line and the MAIAC AOD and AERONET AOD values are equal along the red dashed line.

### 3.2. Ground $PM_{2.5}$ Measurements

The overall average  $PM_{2.5}$  concentration in Japan during 2011–2016 was  $14.3 \pm 8.1 \mu\text{g}/\text{m}^3$  (mean  $\pm$  standard deviation [SD]). The average  $PM_{2.5}$  concentration was highest in spring ( $16.6 \pm 8.5 \mu\text{g}/\text{m}^3$ ), followed by summer ( $14.8 \pm 8.4 \mu\text{g}/\text{m}^3$ ), autumn ( $12.8 \pm 6.8 \mu\text{g}/\text{m}^3$ ), and winter ( $12.8 \pm 7.8 \mu\text{g}/\text{m}^3$ ) (Table 1).

**Table 1.** Descriptive statistics of in situ measurements of particulate matter with an aerodynamic diameter less than  $2.5 \mu\text{m}$  ( $PM_{2.5}$ ) in Japan during 2011–2016 (unit:  $\mu\text{g}/\text{m}^3$ ).

Period	Number	Mean	SD	Min	Q1	Median	Q3	Max	IQR
Overall	1,491,808	14.3	8.1	0.0	8.4	12.5	18.2	149.5	9.8
Spring	378,303	16.6	8.5	0.0	10.6	15.2	20.9	112.5	10.3
Summer	362,641	14.8	8.4	0.0	8.7	12.9	18.9	101.0	10.2
Autumn	368,560	12.8	6.8	0.0	7.9	11.5	16.2	97.2	8.3
Winter	382,304	12.8	7.8	0.0	7.3	10.8	16.3	149.5	9.0

Abbreviations: SD, standard deviation; Q1, 25th percentile; Q3, 75th percentile; Min, minimum; Max, maximum; IQR, interquartile range.

### 3.3. Model Performance of the First-Stage Model

The average OOB  $R^2$  and RMSE for the first-stage random forest model were 0.97 (range, 0.93–1.00) and 0.016 (unitless; range, 0.009–0.038), respectively. After imputing missing values, the mean MAIAC AOD was  $0.182 \pm 0.145$  (mean  $\pm$  SD) overall, and was highest



in spring ( $0.264 \pm 0.182$ ), followed by summer ( $0.206 \pm 0.158$ ), winter ( $0.134 \pm 0.087$ ), and autumn ( $0.128 \pm 0.079$ ) (Table 2). The seasonal trend of MAIAC AOD was consistent with in situ  $PM_{2.5}$  measurements.

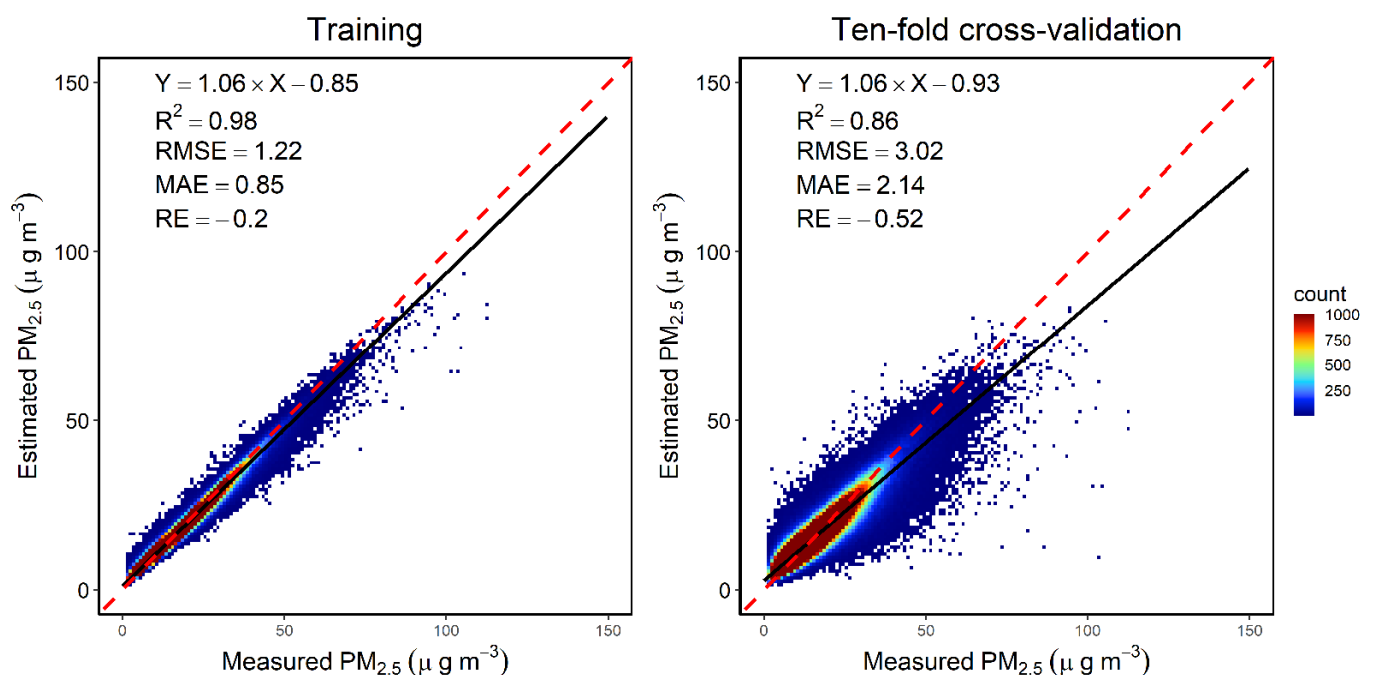
**Table 2.** Descriptive statistics of 1-km-resolution Multi-Angle Implementation of Atmospheric Correction (MAIAC) aerosol optical depth (AOD) during 2011–2016 across Japan (unit: unitless).

Period	Number <sup>a</sup>	Mean	SD	Min	Q1	Median	Q3	Max	IQR
Overall	129,090,082 (15.48%)	0.182	0.145	0	0.091	0.145	0.231	3.770	0.140
Spring	37,531,792 (17.94%)	0.264	0.182	0	0.148	0.226	0.330	3.180	0.182
Summer	22,073,935 (10.51%)	0.206	0.158	0	0.107	0.170	0.263	3.773	0.156
Autumn	40,183,589 (19.49%)	0.128	0.079	0	0.074	0.111	0.163	3.516	0.089
Winter	29,300,766 (14.05%)	0.134	0.087	0	0.076	0.115	0.170	1.750	0.094

<sup>a</sup> The number in brackets is the coverage rate. Abbreviations: SD, standard deviation; Q1, 25th percentile; Q3, 75th percentile; Min, minimum; Max, maximum; IQR, interquartile range.

### 3.4. Model Performance of the Second-Stage Model

A total of 1,488,612 matched pairs of MAIAC AOD and  $PM_{2.5}$  concentrations were available during 2011–2016 to develop the satellite-based  $PM_{2.5}$  model by random forest. The results of model training and 10-fold cross-validation are displayed in Figure 5. For model training, the satellite-based  $PM_{2.5}$  model had a low bias with a training  $R^2$  of 0.98, an RMSE of  $1.22 \mu\text{g}/\text{m}^3$ , an MAE of  $0.85 \mu\text{g}/\text{m}^3$ , and a RE of  $-0.20$ , which showed that the model fitted the measured  $PM_{2.5}$  values very well. For the 10-fold cross-validation, the cross-validation  $R^2$ , RMSE, MAE, and RE of the satellite-based model were 0.86,  $3.02 \mu\text{g}/\text{m}^3$ ,  $2.14 \mu\text{g}/\text{m}^3$ , and  $-0.52$ , respectively, which showed that the model was slightly overfitted. The training and 10-fold cross-validation of XGBoost are shown in Figure S2. The performance of random forest was better than XGBoost in this study.



**Figure 5.** Model training (left panel) and 10-fold cross-validation (right panel) of the satellite-based AOD- $PM_{2.5}$  model in Japan by using random forest. The red dashed line is the one-one line.

The spatial distributions of the residuals (differences between estimated  $PM_{2.5}$  and in situ measurements) are shown in Figure S2. Generally, the random forest model tended to slightly overestimate  $PM_{2.5}$  in the grids close to the shore (mean,  $0.1 \pm 0.5 \mu\text{g}/\text{m}^3$ ; range,  $-1.8$  to  $1.8 \mu\text{g}/\text{m}^3$ ). Nevertheless, the residuals on most grids (75%) were below 4% of the mean measurement of  $PM_{2.5}$  (Figure S2). Additionally, the observed and predicted

PM<sub>2.5</sub> showed a highly consistent trend during 2011–2016 according to the time series plot (Figure S3).

We used the grids overlapping with the AERONET sites to validate the second-stage random forest model, and the model performance was excellent at these sites with a validation  $R^2$  (RMSE, MAE, and RE) of 0.94 (1.78  $\mu\text{g}/\text{m}^3$ , 1.40  $\mu\text{g}/\text{m}^3$ , and 1.55, respectively) (Figure S5). Additionally, the training and cross-validation results stratified by season are shown in Figure S6. The model performances were similar across the four seasons, while the model performance in autumn was slightly worse than the other seasons (cross-validation  $R^2$ , RMSE, MAE, and RE of 0.83, 2.83  $\mu\text{g}/\text{m}^3$ , 2.03  $\mu\text{g}/\text{m}^3$ , and  $-0.63$ , respectively) (Figure S6). The model performance under different AOD coverages is presented in Figure S7. The model performance increased as AOD coverage increased, and the model performance at the grids with coverage rate  $\geq 33.26\%$  (cross-validation  $R^2$ , RMSE, MAE, and RE of 0.87, 2.74  $\mu\text{g}/\text{m}^3$ , 1.93  $\mu\text{g}/\text{m}^3$ , and  $-0.45$ , respectively) was slightly higher than the first, second, and third quartiles (Figure S7).

### 3.5. Important Predictors of PM<sub>2.5</sub> Concentrations

The top ten important predictors of ground PM<sub>2.5</sub> concentrations were day of the year (relative importance: 19.9%), temperature (12.3%), AOD (11.9%), relative humidity (11.7%), v10 wind (8.78%), u10 wind (7.55%), boundary layer height (7.48%), precipitation (5.58%), surface pressure (5.01%), and population density (1.83%) (Figure S8).

Increased AOD, surface pressure, population density, road length, and area of industrial facilities were positively associated with ground PM<sub>2.5</sub> concentrations (Figure S8). By contrast, increased relative humidity, precipitation, and elevation were negatively correlated with PM<sub>2.5</sub> concentrations (Figure S8). Day of the year, temperature, v10 wind, u10 wind, boundary layer height, NDVI, distance to the nearest highway, distance to the nearest primary road, and urban and built-up area showed non-linear relationships with ground PM<sub>2.5</sub> concentrations (Figure S8).

### 3.6. PM<sub>2.5</sub> Estimates

The overall average PM<sub>2.5</sub> estimate was  $12.5 \pm 5.5 \mu\text{g}/\text{m}^3$ . The PM<sub>2.5</sub> estimates were higher in southern Japan and in several metropolitan areas, including Tokyo, Nagoya, and Osaka, than in the northern region (Figure 6). The seasonal mean PM<sub>2.5</sub> estimates were highest in spring ( $14.7 \pm 6.0 \mu\text{g}/\text{m}^3$ ), followed by summer ( $13.2 \pm 5.5 \mu\text{g}/\text{m}^3$ ), autumn ( $11.1 \pm 4.6 \mu\text{g}/\text{m}^3$ ), and winter ( $11.1 \pm 4.8 \mu\text{g}/\text{m}^3$ ) (Table 3). The seasonal trends were consistent with those determined from in situ PM<sub>2.5</sub> measurements. The seasonal spatial distribution of estimated PM<sub>2.5</sub> is presented in Figure S9. The PM<sub>2.5</sub> concentrations showed a decreased trend from 2011 to 2016 (Figure 7), which is consistent with the estimation in a previous study [46].

**Table 3.** Descriptive statistics of estimated particulate matter with an aerodynamic diameter less than 2.5  $\mu\text{m}$  (PM<sub>2.5</sub>) during 2011–2016 in Japan (unit:  $\mu\text{g}/\text{m}^3$ ).

Period	Number	Mean	SD	Min	Q1	Median	Q3	Max	IQR
Overall	832,177,456	12.5	5.5	1.2	8.8	11.4	15.0	93.6	6.2
Spring	209,562,936	14.7	6.0	1.2	10.6	13.6	17.3	93.6	6.7
Summer	209,562,936	13.2	5.5	1.3	9.4	12.2	15.8	69.6	6.4
Autumn	207,285,078	11.1	4.6	1.4	8.0	10.1	13.0	77.0	5.0
Winter	205,766,506	11.1	4.8	1.4	7.9	10.1	13.0	88.1	5.1

Abbreviations: SD, standard deviation; Q1, 25th percentile; Q3, 75th percentile; Min, minimum; Max, maximum; IQR, interquartile range.

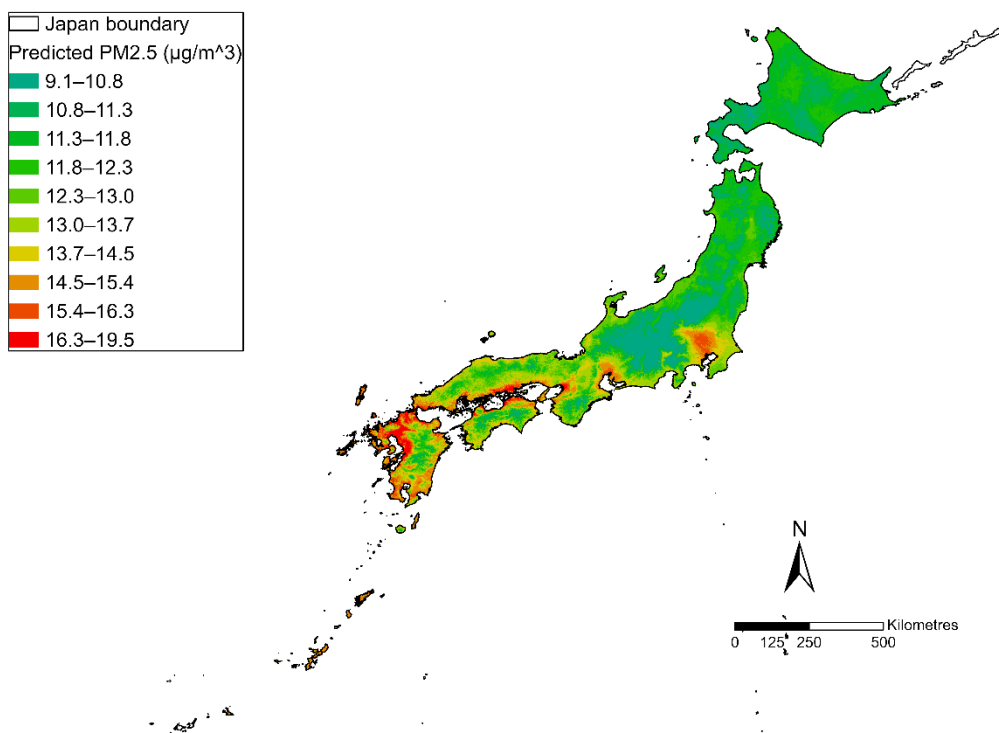


Figure 6. The overall average estimated PM<sub>2.5</sub> concentrations in Japan during 2011–2016.

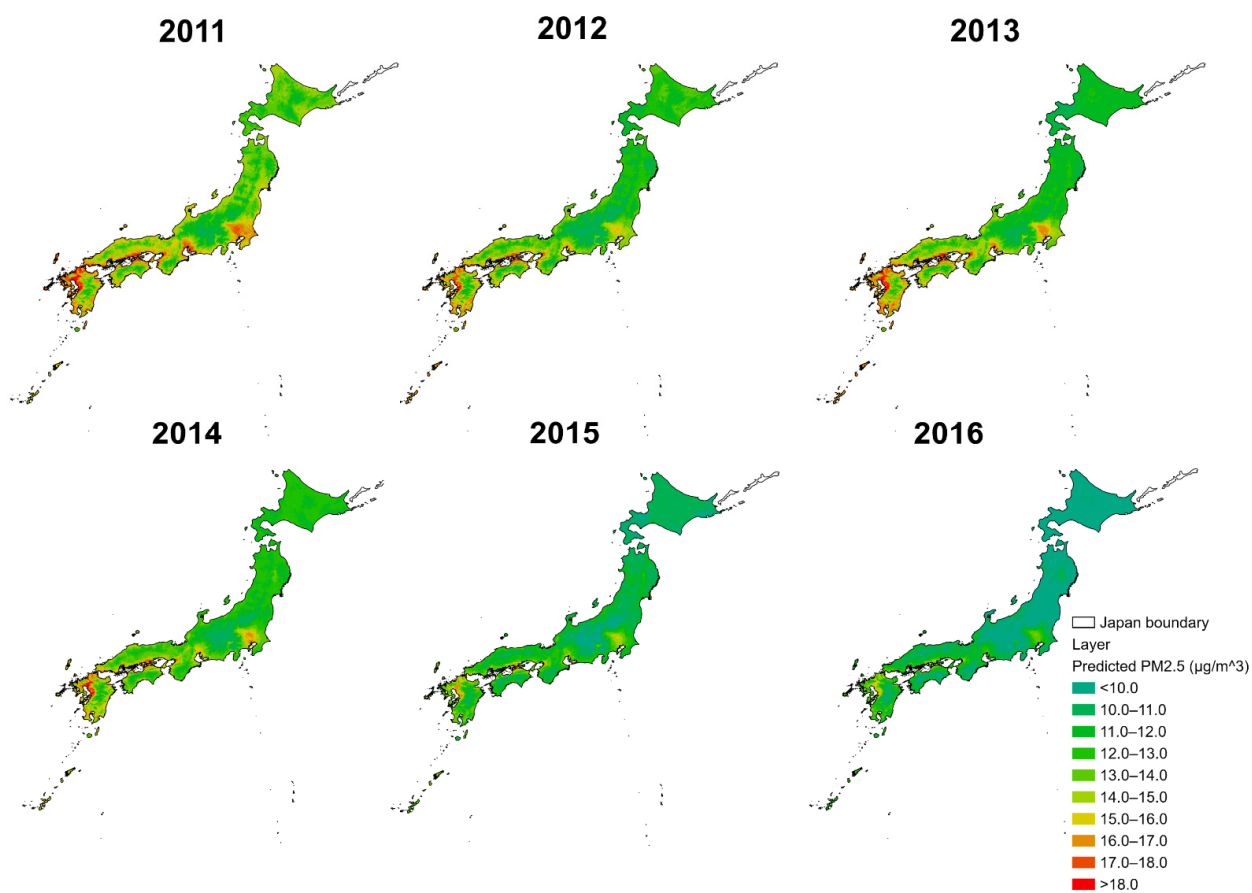


Figure 7. The annual distribution of estimated PM<sub>2.5</sub> in Japan from 2011 to 2016.

#### 4. Discussion

In the present study, we validated MAIAC AOD with ground measurements of AOD and developed a 1-km-resolution satellite-based model for daily PM<sub>2.5</sub> concentrations during 2011–2016 by using a random forest model in Japan. The MAIAC AOD showed a high retrieval accuracy in Japan ( $R = 0.82$ , with the EE: 74.62%), which was comparable to those conducted in South Asia (including Pakistan, India, and Bangladesh) ( $R = 0.882$ , within the EE: 72.22% and  $R = 0.887$ , within the EE: 73.5 for Aqua and Terra MAIAC AOD, respectively) [47] and better than in Central Asia ( $R = 0.730$ , within the EE: 58.7% for spring;  $R = 0.709$ , within the EE: 66.7% for summer;  $R = 0.729$ , within the EE: 62.4% for autumn, and  $R = 0.744$ , with the EE: 67.9% for winter) [38].

The satellite-based model achieved excellent performance with a training and 10-fold cross-validation  $R^2$  (RMSE, MAE, and RE) of 0.98 (1.22  $\mu\text{g}/\text{m}^3$ , 0.85  $\mu\text{g}/\text{m}^3$ , and  $-0.20$ ) and 0.86 (3.02  $\mu\text{g}/\text{m}^3$ , 2.14  $\mu\text{g}/\text{m}^3$ , and  $-0.52$ ), respectively. The top ten important predictors of the model were day of the year, temperature, MAIAC AOD, relative humidity, v10 wind, u10 wind, boundary layer height, precipitation, surface pressure, and population density. The overall average PM<sub>2.5</sub> estimate was  $12.5 \pm 5.5 \mu\text{g}/\text{m}^3$ , and the estimates were higher in southern Japan and metropolitan areas. The estimates showed that PM<sub>2.5</sub> concentrations in Japan decreased from 2011 to 2016. Long-range transportation is a main contributor of high PM<sub>2.5</sub> in western Japan during winter and spring. China implemented an Air Pollution Prevention and Control Action Plan in 2013 with a series of stringent clean air actions during 2013–2017, which led to PM<sub>2.5</sub> rapidly decreasing across China [48]. In addition, the domestic emissions, especially those derived from road transport, decreased in Japan [46]. Hence, the decrease in both foreign and domestic emission sources may cause a long-term decreasing trend in Japan.

Several studies developed nationwide PM<sub>2.5</sub> estimation models in Japan. Araki and colleagues modelled the spatial distributions of annual mean PM<sub>2.5</sub> concentrations using LUR and the regression-kriging method [49]. They included AOD at 500 nm from JAXA Satellite Measurements for Environmental Studies (JASMES) as a predictor in the models but excluded AOD during the backward variable selection process. Their final model retained the built-up area ratio, agricultural area ratio, population, distance to the nearest highway, distance to the coastline, precipitation, temperature, wind speed, and longitude as predictors. The leave-one-out cross-validation  $R^2$  (RMSE) was 0.71 (1.2  $\mu\text{g}/\text{m}^3$ ) and 0.81 (1.0  $\mu\text{g}/\text{m}^3$ ) for LUR and regression-kriging, respectively, after removing the spatial outlier [49]. In 2020, Araki and colleagues in Japan used an artificial neural network and included meteorological variables (precipitation, wind speed, temperature, and relative humidity) and land use data (built-up area ratio, green area ratio, population density, distance to coastline, and longitude) as predictors to estimate monthly PM<sub>2.5</sub> concentrations during 1987–2016. Their model achieved a spatial and temporal cross-validation  $R^2$  (RMSE) of 0.76 (1.9  $\mu\text{g}/\text{m}^3$ ) and 0.73 (2.0  $\mu\text{g}/\text{m}^3$ ), respectively [46]. The same group also applied the random forest model with ordinary kriging to estimate monthly PM<sub>2.5</sub> concentrations in Japan during 2010–2015 [50]. They used 5-fold cross-validation to validate the model performance and obtained a cross-validation  $R^2$  of 0.79 with RMSE of 2.1  $\mu\text{g}/\text{m}^3$  for all stations [50]. Their results showed that the top ten important predictors of PM<sub>2.5</sub> were suspended PM, ozone (O<sub>3</sub>), longitude, month, sulphur dioxides (SO<sub>2</sub>), temperature, relative humidity, latitude, nitrogen dioxides (NO<sub>2</sub>), and precipitation. Only a few studies attempted to estimate daily PM<sub>2.5</sub> concentrations in Japan. Shimadera and colleagues simulated daily air quality from April 2010 to March 2011 by using the Community Multiscale Air Quality Model (CMAQ) v5.0.2 with the Weather Research and Forecasting Model (WRF) v3.5.1. Their model performance was fair for PM<sub>2.5</sub> (with a Pearson's correlation coefficient of 0.75 and RMSE of 7.3  $\mu\text{g}/\text{m}^3$ ) [9]. We integrated MAIAC AOD with meteorological variables and land use data to develop the daily PM<sub>2.5</sub> estimation model in this study. Our model achieved a 10-fold cross-validation  $R^2$  of 0.86 with RMSE of 3.02  $\mu\text{g}/\text{m}^3$ . Our model performance is better than those in studies that merely relied on meteorological variables and land use data to estimate monthly or annual mean PM<sub>2.5</sub> concentrations. This shows

that satellite-based AOD can provide additional spatial and temporal information about daily PM<sub>2.5</sub> concentrations and improve the model performance.

Nearly all previous studies used ensemble models (i.e., random forest, gradient boosting, or XGBoost) to develop satellite-based PM<sub>2.5</sub> models, and only a few used a deep learning model or support vector machine (as summarised in Table S5). A study built a model in Great Britain, UK during 2008–2018 by using a random forest model and obtained a 10-fold cross-validation  $R^2$  (RMSE) of 0.767 (4.042  $\mu\text{g}/\text{m}^3$ ) [22]. A study in Guangdong–Hong Kong–Macao Greater Bay Area, China also developed models by random forest and achieved a 10-fold cross-validation  $R^2$  (RMSE) of 0.937 (3.527  $\mu\text{g}/\text{m}^3$ ), 0.905 (3.780  $\mu\text{g}/\text{m}^3$ ), and 0.884 (3.633  $\mu\text{g}/\text{m}^3$ ) for 2016, 2017, and 2018, respectively [51]. A recent study in Italy built models between 2013 and 2015 using an ensemble generalised additive model, and combined the results derived from a linear mixed-effect model, random forest, and XGBoost. They found that the cross-validation  $R^2$  (RMSE) was 0.79 (6.56  $\mu\text{g}/\text{m}^3$ ), 0.79 (5.29  $\mu\text{g}/\text{m}^3$ ), and 0.81 (6.34  $\mu\text{g}/\text{m}^3$ ) for 2013, 2014, and 2015, respectively [17]. Overall, our model performance is comparable to or better than those in these studies.

Our results showed that day of the year was the most important variable in the random forest model for estimating PM<sub>2.5</sub> (relative importance of 19.92%). Early studies that developed satellite-based AOD-PM<sub>2.5</sub> estimation models by using the linear mixed effect model needed to include day-specific intercepts and slopes to achieve high model performance because the relationship between AOD and PM<sub>2.5</sub> may change day-to-day due to differences in mixing height, relative humidity, particle composition, and vertical profiles [13]. Recent studies that applied the machine learning algorithm also found that day of the year (or Julian day) is an important variable for modelling PM<sub>2.5</sub> [22,52,53]. According to the ALE plot (Figure S8), the relationship between day of the year and PM<sub>2.5</sub> estimates displayed a bimodal pattern with two peaks between Julian day 0–100 (spring) and after Julian day 300 (winter). Day of the year can be a surrogate of seasonal effects. PM<sub>2.5</sub> concentrations in Japan displayed an obvious seasonal variation. Both foreign and domestic sources may cause the seasonal variations of PM<sub>2.5</sub> in Japan. Long-range transportation from Asia (yellow dust storm and anthropogenic aerosols) contributes to high PM<sub>2.5</sub> levels in western Japan, particularly in spring and winter [9,54]. Photochemical smog events frequently occurred in summer can contribute to high concentrations of PM<sub>2.5</sub> in summer [55]. Besides, increase in operation time of coal-burning power plants and heat appliance usage can cause raising PM<sub>2.5</sub> concentrations during winter [55].

AOD was the third most important variable in our model (relative importance of 11.89%). There was a strong and positive relationship between AOD and PM<sub>2.5</sub> according to the ALE plot (Figure S8). However, the results from previous studies are contradictory. A study conducted on an island, Great Britain of the UK, found that satellite-based AOD 0.47 and 0.55  $\mu\text{m}$  values retrieved from the MAIAC AOD product were not important predictors in a random forest model for estimating ground PM<sub>2.5</sub> concentrations, and the authors concluded that the contribution of satellite-based AOD in the model was very limited [22]. A study in Tehran, Iran also indicated that satellite-based AOD with an extremely high missing rate of 94.09% (3 km Terra Dark Target Algorithm AOD) did not significantly contribute to PM<sub>2.5</sub> estimations [56]. It should be noted that aerosol compositions (i.e., nitrate/sulphate ratio), presence of black carbon, and an arid environment with higher surface reflectance may cause poor correlations between satellite-based AOD and ground PM<sub>2.5</sub> concentrations [57]. Nevertheless, our results demonstrated that satellite-based AOD can provide valuable information for estimating ground PM<sub>2.5</sub> concentrations in Japan.

Meteorological variables had stronger influences on PM<sub>2.5</sub> estimates than population density and land use data according to the variable importance (Figure S8). A previous study also observed that forest area and population density have lower importance values than meteorological variables due to a lack of temporal variation [53].

Temperature was the second-most-important variable in our model (relative importance of 12.25%). We observed a U-shaped relationship between temperature and PM<sub>2.5</sub> concentrations (Figure S8). Temperature both positively and negatively influences PM<sub>2.5</sub>.

High temperature can increase  $PM_{2.5}$  concentrations by promoting photochemical reactions and accelerating formation of  $PM_{2.5}$  precursors and other secondary pollutants [44,58]. In winter, increased temperature may induce formation of a temperature inversion layer due to low surface temperature and result in accumulation of  $PM_{2.5}$  [44]. On the other hand, high temperature can enhance thermal activities, facilitate dispersion of  $PM_{2.5}$ , and increase evaporation loss of  $PM_{2.5}$  (loss of vapor, ammonium nitrate, and volatile and semi-volatile pollutants), causing decreases in  $PM_{2.5}$  concentrations [44]. In addition, low temperature can reduce atmospheric convection and lead to increases in  $PM_{2.5}$  concentrations [44]. We observed negative correlations of  $PM_{2.5}$  concentrations with relative humidity and precipitation (Figure S8). These results are unsurprising;  $PM_{2.5}$  concentrations may decrease as relative humidity increases beyond a threshold due to dry deposition [58]. Additionally, precipitation can decrease  $PM_{2.5}$  concentrations via wet deposition [58].

Boundary layer height was the seventh most important variable in our model (relative importance of 7.48%). Boundary layer height can affect the vertical distribution of PM; low boundary layer height leads to accumulation of  $PM_{2.5}$  near the earth's surface [44]. Tsai and colleagues in Taiwan reported that AOD normalised by haze layer height detected by ground-based LiDAR (i.e., the sum of boundary layer height and scaling height) remarkably improves the correlation between AOD and  $PM_{2.5}$  concentrations [59]. Several studies also included boundary layer height from reanalysed datasets as a predictor in satellite-based AOD- $PM_{2.5}$  models [30,60,61]. However, normalising AOD by boundary layer height or including boundary layer height as a predictor did not improve the performance of a linear mixed-effect model in our previous work [23]. It is possible that the linear mixed-effect model cannot capture the complex interactions between AOD and boundary layer height. Nonetheless, we applied a random forest model in this study, and boundary layer height was a critical variable in this model. This demonstrates that random forest is suitable for developing satellite-based AOD- $PM_{2.5}$  models, especially when predictors (AOD and meteorological variables in this study) are highly inter-correlated, and their relationships are not linear or monotonic.

This study has several strengths. First, we comprehensively validated the MAIAC AOD with the AERONET AOD across Japan. The results can provide a basic understanding of accuracy of MAIAC AOD in Japan. Second, this is the first study in Japan that used machine learning algorithms with 1 km MAIAC AOD to develop a high spatiotemporal resolution  $PM_{2.5}$  estimation model that can provide daily estimates of  $PM_{2.5}$  concentrations at a national scale. We assessed model performances of two machine learning algorithms—random forest and XGBoost for estimating  $PM_{2.5}$  concentrations and found that random forest had the best performance. Our cross-validation results showed that the random forest model can provide accurate estimates in areas without air quality monitoring stations from 2011 to 2016. The Japan Environment and Children's Study (JECS), a nationwide birth cohort study, recruited 103,099 participants from 15 regional centres during 2011–2014 [62], and the residential addresses of the participants are being geocoded. The daily  $PM_{2.5}$  estimates from the satellite-based model can be integrated with the geocodes of the JECS participants to analyse the acute and chronic effects of  $PM_{2.5}$  on children's health, and to assess the time windows when children are vulnerable. Third, this study provided ALE plots that can describe how predictors influence  $PM_{2.5}$  estimates in the model. Although ALE is not a new technology, it has rarely been used in previous studies. Here, we demonstrate the importance of ALE can help to quantify the effect of a specific predictor and its direction on  $PM_{2.5}$  estimates.

This study has some limitations. First, missing AOD values are a major limitation and challenge when building satellite-based models. AOD values can be missing due to the presence of clouds and high surface reflectance, and thus this problem is much worse for islands, which are frequently cloudy and rainy. In the present study, the average coverage rate of the 1-km MAIAC AOD in Japan was only 15.48%. In our previous study in Taiwan, the average coverage rate of the 10 km Dark Target algorithm AOD was 22.99% [23]. Schneider and colleagues reported that the coverage rate of the 1 km

MAIAC AOD in Great Britain, UK was 6–13% during 2008–2018 [22]. The coverage rates of AOD in island environments are frequently low based on these studies. Several methods, such as multiple imputation, random forest, and interpolation, were proposed to impute missing AOD values [19,53,63], but they may introduce uncertainties into the model. In 2014, Japan launched a geostationary satellite, Himawari-8, which can continuously provide AOD values every 10 min. A future study could use AOD values derived from this geostationary satellite to overcome the problem of missing AOD values. Second, our land-use data for industrial areas were from 2009 and thus not up-to-date. New data for industrial areas or points of interest in Japan are hard to access, which may constrain the estimation ability of the model. Third, emission inventory data and other criteria pollutants (i.e., nitrogen dioxides, sulphur dioxides, and ozone) were not included in our model. Interactions between pollutants can enhance the production of secondary pollutants, causing increases in PM<sub>2.5</sub> concentrations. A future study could include these data to improve the model performance.

## 5. Conclusions

This study applied a random forest model integrating satellite-based AOD, meteorological variables and land use data to generate full-coverage and daily PM<sub>2.5</sub> estimates at 1-km spatial resolution across Japan during 2011–2016. The model performance was excellent, with a 10-fold cross-validation  $R^2$  of 0.86 and an RMSE of 3.02  $\mu\text{g}/\text{m}^3$ , which showed that PM<sub>2.5</sub> estimates derived from the satellite-based model were reliable and accurate. According to the variable importance, the determinants of outdoor PM<sub>2.5</sub> in Japan were day of the year, temperature, satellite-based AOD, relative humidity, wind vectors, boundary layer height, precipitation, surface pressure, and population density. These estimates can be used to assess both the short-term and long-term effects of PM<sub>2.5</sub> on health outcomes in epidemiological studies.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/rs13183657/s1>, Table S1. Summary of data sources, coverage years, and spatial and temporal resolutions; Table S2. Descriptive statistics of 16 predictors from 2011 to 2016 (mean  $\pm$  standard deviation); Table S3. Hyperparameters tuned to optimise the performance of eXtreme Gradient Boosting (XGBoost); Table S4. Descriptive statistics for the Multi-Angle Implementation of Atmospheric Correction (MAIAC) aerosol optical depth (AOD) and AErosol RObotic NETwork (AERONET) AOD; Table S5. Comparison of model performances between this and other studies that applied machine learning algorithms to develop satellite-based PM<sub>2.5</sub> models; Figure S1. Locations of 12 AErosol RObotic NETwork (AERONET) stations in Japan (Hokkaido University, Niigata, Noto, Toyama, Chiba University, Osaka-North, Nara, Osaka, Kobe, Shirahama, Fukuoka and Fukue); Figure S2. Model training (left panel) and 10-fold cross-validation (right panel) of eXtreme Gradient Boosting (XGBoost) model for estimating PM<sub>2.5</sub> concentrations. The red dashed line is the one-one line; Figure S3. The spatial distribution of average residuals (differences between estimated PM<sub>2.5</sub> and in situ measurements) based on the random forest model; Figure S4. Time series variation of monthly average observed and predicted PM<sub>2.5</sub> concentrations during 2011–2016; Figure S5. Training (left panel) and validation (right panel) of the random forest model at the grids overlapping with the AERONET sites. The red dashed line is the one-one line; Figure S6. Model training (upper panel) and 10-fold cross-validation results (lower panel) of the random forest model stratified by four seasons. (a, e) Spring (March–May); (b, f) Summer (June–August); (c, g) Autumn (September–November); and (d, h) Winter (December–February); Figure S7. Model training (upper panel) and 10-fold cross-validation results (lower panel) of the random forest model stratified by quartiles of coverage. (a, e) the first quartile (coverage rate <16.61%); (b, f) the second quartile (coverage rate 16.61–23.79%); (c, g) the third quartile (23.79–33.26%); and (d, h) the fourth quartile ( $\geq$ 33.26%); Figure S8. Accumulated local effect plots for the effects of predictors on estimated PM<sub>2.5</sub> values. Numbers in brackets show the relative permutation importance (%); Figure S9. The seasonal spatial distribution of estimated PM<sub>2.5</sub> concentrations. (A) Spring (March–May); (B) Summer (June–August); (C) Autumn (September–November); and (D) Winter (December–February). References [64,65] are cited in the supplementary materials.

**Author Contributions:** Conceptualisation, C.-R.J., W.-T.C. and S.F.N.; methodology, C.-R.J. and W.-T.C.; formal analysis, C.-R.J.; resources, S.F.N.; data curation, C.-R.J.; writing—original draft preparation, C.-R.J.; writing—review and editing, W.-T.C. and S.F.N.; supervision, S.F.N.; funding acquisition, S.F.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of the Environment, Government of Japan.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** MODIS product was obtained from <https://ladsweb.modaps.eosdis.nasa.gov/search/> (accessed on 12 September 2021); AERONET AOD was downloaded from [https://aeronet.gsfc.nasa.gov/new\\_web/index.html](https://aeronet.gsfc.nasa.gov/new_web/index.html) (accessed on 12 September 2021); PM<sub>2.5</sub> from <https://www.nies.go.jp/igreen/index.html> (accessed on 12 September 2021); ECMWF-Int from <https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/> (accessed on 12 September 2021); land use data from [https://www.eorc.jaxa.jp/ALOS/en/lulc/lulc\\_index.htm](https://www.eorc.jaxa.jp/ALOS/en/lulc/lulc_index.htm) (accessed on 12 September 2021) and <https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-L05.html> (accessed on 12 September 2021); road network from [https://www.gsi.go.jp/kankyochiri/gm\\_japan\\_e.html](https://www.gsi.go.jp/kankyochiri/gm_japan_e.html) (accessed on 12 September 2021); population counts from <https://www.worldpop.org/geodata/listing?id=29> (accessed on 12 September 2021).

**Acknowledgments:** This research was funded by the Ministry of the Environment, Government of Japan. The findings and conclusions of this article are the sole responsibility of the authors and do not represent the official views of the Japanese government. The authors gratefully acknowledge NASA for providing MAIAC AOD and AERONET AOD data, National Institute for Environmental Studies for providing ground PM<sub>2.5</sub> measurements, the Japan Meteorological Agency and the European Center for Medium-Range Weather Forecasts (ECMWF) for meteorological variables, the JAXA Earth Observation Research Center and the Ministry of Land, Infrastructure, Transport and Tourism for providing land use data, the WorldPop for population density data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. US EPA Particulate Matter (PM) Basics. Available online: <https://www.epa.gov/pm-pollution/particulate-matter-pm-basics> (accessed on 12 September 2021).
2. Kim, K.H.; Kabir, E.; Kabir, S. A review on the human health impact of airborne particulate matter. *Environ. Int.* **2015**, *74*, 136–143. [[CrossRef](#)]
3. Lin, Y.T.; Shih, H.; Jung, C.R.; Wang, C.M.; Chang, Y.C.; Hsieh, C.Y.; Hwang, B.F. Effect of exposure to fine particulate matter during pregnancy and infancy on paediatric allergic rhinitis. *Thorax* **2021**, *76*, 568–574. [[CrossRef](#)] [[PubMed](#)]
4. Jung, C.R.; Chen, W.T.; Tang, Y.H.; Hwang, B.F. Fine particulate matter exposure during pregnancy and infancy and incident asthma. *J. Allergy Clin. Immunol.* **2019**, *143*, 2254–2262. [[CrossRef](#)] [[PubMed](#)]
5. Alexeeff, S.E.; Liao, N.S.; Liu, X.; Van DenEeden, S.K.; Sidney, S. Long-term PM<sub>2.5</sub> exposure and risks of ischemic heart disease and stroke events: Review and meta-analysis. *J. Am. Heart Assoc.* **2021**, *10*, 1–22. [[CrossRef](#)] [[PubMed](#)]
6. Tsai, T.-L.; Lin, Y.-T.; Hwang, B.-F.; Nakayama, S.F.; Tsai, C.-H.; Sun, X.-L.; Ma, C.; Jung, C.-R. Fine particulate matter is a potential determinant of Alzheimer’s disease: A systemic review and meta-analysis. *Environ. Res.* **2019**, *177*, 108638. [[CrossRef](#)] [[PubMed](#)]
7. Shi, L.; Zanobetti, A.; Kloog, I.; Coull, B.A.; Koutrakis, P.; Melly, S.J.; Schwartz, J.D. Low-concentration PM<sub>2.5</sub> and mortality: Estimating acute and chronic effects in a population-based study. *Environ. Health Perspect.* **2016**, *124*, 46–52. [[CrossRef](#)]
8. Li, P.; Sato, K.; Hasegawa, H.; Huo, M.; Minoura, H.; Inomata, Y.; Take, N.; Yuba, A.; Futami, M.; Takahashi, T.; et al. Chemical Characteristics and Source Apportionment of PM<sub>2.5</sub> and Long-Range Transport from Northeast Asia Continent to Niigata in Eastern Japan. *Aerosol Air Qual. Res.* **2018**, *18*, 938–956. [[CrossRef](#)]
9. Shimadera, H.; Kojima, T.; Kondo, A. Evaluation of Air Quality Model Performance for Simulating Long-Range Transport and Local Pollution of PM<sub>2.5</sub> in Japan. *Adv. Meteorol.* **2016**, *2016*. [[CrossRef](#)]
10. Eeftens, M.; Beelen, R.; Hoogh, K.d.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dèdelè, A.; Dons, E.; Nazelle, A.d.; et al. Development of Land Use Regression Models for PM<sub>2.5</sub>, PM<sub>2.5</sub> Absorbance, PM<sub>10</sub> and PM<sub>coarse</sub> in 20 European Study Areas; Results of the ESCAPE Project. *Environ. Sci. Technol.* **2012**, *46*, 11195–11205. [[CrossRef](#)]
11. Gupta, P.; Christopher, S.A.; Wang, J.; Gehrig, R.; Lee, Y.; Kumar, N. Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmos. Environ.* **2006**, *40*, 5880–5892. [[CrossRef](#)]
12. Xu, X.; Zhang, C.; Liang, Y. Review of satellite-driven statistical models PM<sub>2.5</sub> concentration estimation with comprehensive information. *Atmos. Environ.* **2021**, *256*, 118302. [[CrossRef](#)]
13. Kloog, I.; Koutrakis, P.; Coull, B.A.; Lee, H.J.; Schwartz, J. Assessing temporally and spatially resolved PM<sub>2.5</sub> exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos. Environ.* **2011**, *45*, 6267–6275. [[CrossRef](#)]



14. Hu, X.; Belle, J.H.; Meng, X.; Wildani, A.; Waller, L.A.; Strickland, M.J.; Liu, Y. Estimating PM<sub>2.5</sub> Concentrations in the Conterminous United States Using the Random Forest Approach. *Environ. Sci. Technol.* **2017**, *51*, 6936–6944. [CrossRef]
15. 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.; et al. Estimating ground-level PM<sub>2.5</sub> concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. *Remote Sens. Environ.* **2014**, *140*, 220–232. [CrossRef]
16. Beloconi, A.; Chrysoulakis, N.; Lyapustin, A.; Utzinger, J.; Vounatsou, P. Bayesian geostatistical modelling of PM<sub>10</sub> and PM<sub>2.5</sub> surface level concentrations in Europe using high-resolution satellite-derived products. *Environ. Int.* **2018**, *121*, 57–70. [CrossRef]
17. Shtein, A.; Kloog, I.; Schwartz, J.; Silibello, C.; Michelozzi, P.; Gariazzo, C.; Viegi, G.; Forastiere, F.; Karnieli, A.; Just, A.C.; et al. Estimating Daily PM<sub>2.5</sub> and PM<sub>10</sub> over Italy Using an Ensemble Model. *Environ. Sci. Technol.* **2020**, *54*, 120–128. [CrossRef]
18. Xie, Y.; Wang, Y.; Zhang, K.; Dong, W.; Lv, B.; Bai, Y. Daily Estimation of Ground-Level PM<sub>2.5</sub> Concentrations over Beijing Using 3 km Resolution MODIS AOD. *Environ. Sci. Technol.* **2015**, *49*, 12280–12288. [CrossRef] [PubMed]
19. Xiao, Q.; Wang, Y.; Chang, H.H.; Meng, X.; Geng, G.; Lyapustin, A.; Liu, Y. Full-coverage high-resolution daily PM<sub>2.5</sub> estimation using MAIAC AOD in the Yangtze River Delta of China. *Remote Sens. Environ.* **2017**, *199*, 437–446. [CrossRef]
20. You, W.; Zang, Z.; Zhang, L.; Li, Y.; Pan, X.; Wang, W. National-scale estimates of ground-level PM<sub>2.5</sub> concentration in China using geographically weighted regression based on 3 km resolution MODIS AOD. *Remote Sens.* **2016**, *8*, 184. [CrossRef]
21. Yang, L.; Xu, H.; Jin, Z. Estimating ground-level PM<sub>2.5</sub> over a coastal region of China using satellite AOD and a combined model. *J. Clean. Prod.* **2019**, *227*, 472–482. [CrossRef]
22. Schneider, R.; Vicedo-Cabrera, A.M.; Sera, F.; Masselot, P.; Stafoggia, M.; deHoogh, K.; Kloog, I.; Reis, S.; Vieno, M.; Gasparri, A. A satellite-based spatio-temporal machine learning model to reconstruct daily PM<sub>2.5</sub> concentrations across great britain. *Remote Sens.* **2020**, *12*, 3803. [CrossRef] [PubMed]
23. Jung, C.R.; Hwang, B.F.; Chen, W.T. Incorporating long-term satellite-based aerosol optical depth, localized land use data, and meteorological variables to estimate ground-level PM<sub>2.5</sub> concentrations in Taiwan from 2005 to 2015. *Environ. Pollut.* **2018**, *237*, 1000–1010. [CrossRef] [PubMed]
24. Levy, R.C.; Mattoo, S.; Munchak, L.A.; Remer, L.A.; Sayer, A.M.; Patadia, F.; Hsu, N.C. The Collection 6 MODIS aerosol products over land and ocean. *Atmos. Meas. Tech.* **2013**, *6*, 2989–3034. [CrossRef]
25. Lyapustin, A.; Wang, Y.; Korokin, S.; Huang, D. MODIS Collection 6 MAIAC algorithm. *Atmos. Meas. Tech.* **2018**, *11*, 5741–5765. [CrossRef]
26. Chu, Y.; Liu, Y.; Li, X.; Liu, Z.; Lu, H.; Lu, Y.; Mao, Z.; Chen, X.; Li, N.; Ren, M.; et al. A review on predicting ground PM<sub>2.5</sub> concentration using satellite aerosol optical depth. *Atmosphere* **2016**, *7*, 129. [CrossRef]
27. Brokamp, C.; Jandarov, R.; Hossain, M.; Ryan, P. Predicting Daily Urban Fine Particulate Matter Concentrations Using a Random Forest Model. *Environ. Sci. Technol.* **2018**, *52*, 4173–4179. [CrossRef]
28. Zhang, T.; He, W.; Zheng, H.; Cui, Y.; Song, H.; Fu, S. Satellite-based ground PM<sub>2.5</sub> estimation using a gradient boosting decision tree. *Chemosphere* **2021**, *268*, 128801. [CrossRef]
29. 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 PM<sub>2.5</sub> concentrations with missing-filled satellite data in China. *Atmos. Environ.* **2019**, *202*, 180–189. [CrossRef]
30. Di, Q.; Kloog, I.; Koutrakis, P.; Lyapustin, A.; Wang, Y.; Schwartz, J. Assessing PM<sub>2.5</sub> Exposures with High Spatiotemporal Resolution across the Continental United States. *Environ. Sci. Technol.* **2016**, *50*, 4712–4721. [CrossRef] [PubMed]
31. Wang, X.; Sun, W. Meteorological parameters and gaseous pollutant concentrations as predictors of daily continuous PM<sub>2.5</sub> concentrations using deep neural network in Beijing–Tianjin–Hebei, China. *Atmos. Environ.* **2019**, *211*, 128–137. [CrossRef]
32. BCJ Japan Third Mesh Data. Available online: [http://www.biodic.go.jp/kiso/col\\_mesh.html](http://www.biodic.go.jp/kiso/col_mesh.html) (accessed on 13 August 2020).
33. Goddard Space Flight Center Aeronet-Aerosol Robotic Network. Available online: <https://aeronet.gsfc.nasa.gov/> (accessed on 12 September 2021).
34. Sayer, A.M.; Hsu, N.C.; Bettenhausen, C.; Jeong, M.J. Validation and uncertainty estimates for MODIS Collection 6 “deep Blue” aerosol data. *J. Geophys. Res. Atmos.* **2013**, *118*, 7864–7872. [CrossRef]
35. NIES Environment Numerical Database. Available online: <https://www.nies.go.jp/igreen/index.html> (accessed on 13 August 2020).
36. Japan Meteorological Agency AMeDAS. Available online: <https://www.jma.go.jp/jma/en/Activities/amedas/amedas.html> (accessed on 1 June 2021).
37. Araki, S.; Yamamoto, K.; Kondo, A. Application of Regression Kriging to Air Pollutant Concentrations in Japan with High Spatial Resolution. *Aerosol Air Qual. Res.* **2015**, *15*, 234–241. [CrossRef]
38. Chen, X.; Ding, J.; Liu, J.; Wang, J.; Ge, X.; Wang, R.; Zuo, H. Validation and comparison of high-resolution MAIAC aerosol products over Central Asia. *Atmos. Environ.* **2021**, *251*, 118273. [CrossRef]
39. She, L.; Zhang, H.; Wang, W.; Wang, Y.; Shi, Y. Evaluation of the Multi-Angle Implementation of Atmospheric Correction (MAIAC) Aerosol Algorithm for Himawari-8 Data. *Remote Sens.* **2019**, *11*, 2771. [CrossRef]
40. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
41. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An introduction to Statistical Learning with Application in R*; Springer Nature: New York, NY, USA, 2013.
42. Wang, C.M.; Jung, C.R.; Chen, W.T.; Hwang, B.F. Exposure to fine particulate matter (PM<sub>2.5</sub>) and pediatric rheumatic diseases. *Environ. Int.* **2020**, *138*, 105602. [CrossRef] [PubMed]

43. Kloog, I.; Sorek-Hamer, M.; Lyapustin, A.; Coull, B.; Wang, Y.; Just, A.C.; Schwartz, J.; Broday, D.M. Estimating daily PM<sub>2.5</sub> and PM<sub>10</sub> across the complex geo-climate region of Israel using MAIAC satellite-based AOD data. *Atmos. Environ.* **2015**, *122*, 409–416. [[CrossRef](#)] [[PubMed](#)]
44. Chen, Z.; Chen, D.; Zhao, C.; Kwan, M.; Cai, J.; Zhuang, Y.; Zhao, B.; Wang, X.; Chen, B.; Yang, J.; et al. Influence of meteorological conditions on PM<sub>2.5</sub> concentrations across China: A review of methodology and mechanism. *Environ. Int.* **2020**, *139*, 105558. [[CrossRef](#)]
45. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
46. Araki, S.; Shima, M.; Yamamoto, K. Estimating historical PM<sub>2.5</sub> exposures for three decades (1987–2016) in Japan using measurements of associated air pollutants and land use regression. *Environ. Pollut.* **2020**, *263*, 114476. [[CrossRef](#)]
47. Mhawish, A.; Banerjee, T.; Sorek-Hamer, M.; Lyapustin, A.; Broday, D.M.; Chatfield, R. Comparison and evaluation of MODIS Multi-angle Implementation of Atmospheric Correction (MAIAC) aerosol product over South Asia. *Remote Sens. Environ.* **2019**, *224*, 12–28. [[CrossRef](#)]
48. Zhang, Q.; Zheng, Y.; Tong, D.; Shao, M.; Wang, S.; Zhang, Y.; Xu, X.; Wang, J.; He, H.; Liu, W.; et al. Drivers of improved PM<sub>2.5</sub> air quality in China from 2013 to 2017. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 24463–24469. [[CrossRef](#)]
49. Araki, S.; Shimadera, H.; Yamamoto, K.; Kondo, A. Effect of spatial outliers on the regression modelling of air pollutant concentrations: A case study in Japan. *Atmos. Environ.* **2017**, *153*, 83–93. [[CrossRef](#)]
50. Araki, S.; Hasunuma, H.; Yamamoto, K.; Shima, M.; Michikawa, T.; Nitta, H.; Nakayama, S.F.; Yamazaki, S. Estimating monthly concentrations of ambient key air pollutants in Japan during 2010–2015 for a national-scale birth cohort. *Environ. Pollut.* **2021**, *284*, 117483. [[CrossRef](#)]
51. Chen, W.; Ran, H.; Cao, X.; Wang, J.; Teng, D.; Chen, J.; Zheng, X. Estimating PM<sub>2.5</sub> with high-resolution 1-km AOD data and an improved machine learning model over Shenzhen, China. *Sci. Total Environ.* **2020**, *746*, 141093. [[CrossRef](#)]
52. Stafoggia, M.; Bellander, T.; Bucci, S.; Davoli, M.; Hoogh, K.d.; Donato, F.d.; Gariazzo, C.; Lyapustin, A.; Michelozzi, P.; Renzi, M.; et al. Estimation of daily PM<sub>10</sub> and PM<sub>2.5</sub> concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. *Environ. Int.* **2019**, *124*, 170–179. [[CrossRef](#)]
53. Zhang, R.; Di, B.; Luo, Y.; Deng, X.; Grieneisen, M.L.; Wang, Z.; Yao, G.; Zhan, Y. A nonparametric approach to filling gaps in satellite-retrieved aerosol optical depth for estimating ambient PM<sub>2.5</sub> levels. *Environ. Pollut.* **2018**, *243*, 998–1007. [[CrossRef](#)] [[PubMed](#)]
54. Coulibaly, S.; Minami, H.; Abe, M.; Hasei, T.; Sera, N.; Yamamoto, S.; Funasaka, K.; Asakawa, D.; Watanabe, M.; Honda, N.; et al. Seasonal Fluctuations in Air Pollution in Dazaifu, Japan, and Effect of Long-Range Transport from Mainland East Asia. *Biol. Pharm. Bull.* **2015**, *38*, 1395–1403. [[CrossRef](#)]
55. Nakata, M.; Sano, I.; Mukai, S. Air pollutants in Osaka (Japan). *Front. Environ. Sci.* **2015**, *3*, 18. [[CrossRef](#)]
56. Joharestani, M.Z.; Cao, C.; Ni, X.; Bashir, B.; Talebiesfandarani, S. PM<sub>2.5</sub> Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data. *Atmosphere* **2019**, *10*, 373. [[CrossRef](#)]
57. Engel-Cox, J.A.; Holloman, C.H.; Coutant, B.W.; Hoff, R.M. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmos. Environ.* **2004**, *38*, 2495–2509. [[CrossRef](#)]
58. Wang, J.; Ogawa, S. Effects of Meteorological Conditions on PM<sub>2.5</sub> Concentrations in Nagasaki, Japan. *Int. J. Environ. Res. Public Health* **2015**, *12*, 9089–9101. [[CrossRef](#)] [[PubMed](#)]
59. Tsai, T.C.; Jeng, Y.J.; Chu, D.A.; Chen, J.P.; Chang, S.C. Analysis of the relationship between MODIS aerosol optical depth and particulate matter from 2006 to 2008. *Atmos. Environ.* **2011**, *45*, 4777–4788. [[CrossRef](#)]
60. Just, A.C.; Wright, R.O.; Schwartz, J.; Coull, B.A.; Baccarelli, A.A.; Tellez-Rojo, M.M.; Moody, E.; Wang, Y.; Lyapustin, A.; Kloog, I. Using High-Resolution Satellite Aerosol Optical Depth To Estimate Daily PM<sub>2.5</sub> Geographical Distribution in Mexico City. *Environ. Sci. Technol.* **2015**, *49*, 8576–8584. [[CrossRef](#)] [[PubMed](#)]
61. Lv, B.; Hu, Y.; Chang, H.H.; Russell, A.G.; Bai, Y. Improving the Accuracy of Daily PM<sub>2.5</sub> Distributions Derived from the Fusion of Ground-Level Measurements with Aerosol Optical Depth Observations, a Case Study in North China. *Environ. Sci. Technol.* **2016**, *50*, 4752–4759. [[CrossRef](#)] [[PubMed](#)]
62. Michikawa, T.; Nitta, H.; Nakayama, S.F.; Yamazaki, S.; Isobe, T.; Tamura, K.; Suda, E.; Ono, M.; Yonemoto, J.; Iwai-Shimada, M.; et al. Baseline Profile of Participants in the Japan Environment and Children’s Study (JECS). *J. Epidemiol.* **2018**, *28*, 99–104. [[CrossRef](#)]
63. Yang, J.; Hu, M. Filling the missing data gaps of daily MODIS AOD using spatiotemporal interpolation. *Sci. Total Environ.* **2018**, *633*, 677–683. [[CrossRef](#)]
64. Chen, G.; Li, S.; Knibbs, L.D.; Hamm, N.A.S.; Cao, W.; Li, T.; Guo, J.; Ren, H.; Abramson, M.J.; Guo, Y. A machine learning method to estimate PM<sub>2.5</sub> concentrations across China with remote sensing, meteorological and land use information. *Sci. Total Environ.* **2018**, *636*, 52–60. [[CrossRef](#)] [[PubMed](#)]
65. Huang, K.; Xiao, Q.; Meng, X.; Geng, G.; Wang, Y.; Lyapustin, A.; Gu, D.; Liu, Y. Predicting monthly high-resolution PM<sub>2.5</sub> concentrations with random forest model in the North China Plain. *Environ. Pollut.* **2018**, *242*, 675–683. [[CrossRef](#)]