Science of the Total Environment 713 (2020) 136516

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Estimation of spatially continuous daytime particulate matter concentrations under all sky conditions through the synergistic use of satellite-based AOD and numerical models

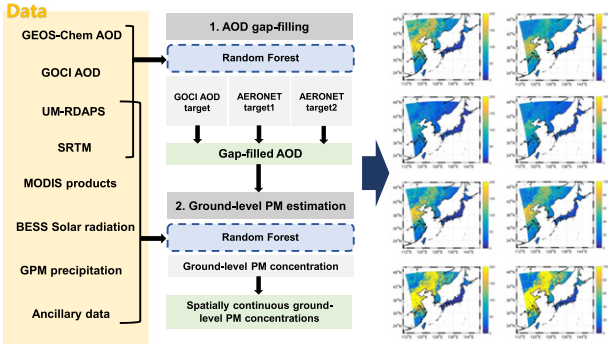
Seohui Park [a,1], Junghee Lee [a,1], Jungho Im [a,*], Chang-Keun Song [a], Myungje Choi [b], Jhoon Kim [c], Seungun Lee [d], Rokjin Park [d], Sang-Min Kim [e], Jongmin Yoon [e], Dong-Won Lee [e], Lindi J. Quackenbush [f]

[a] School of Urban & Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea
[b] Jet Propulsion Laboratory, NASA, Pasadena, CA 91109, USA
[c] Department of Atmospheric Sciences, Yonsei University, Seoul 03722, Republic of Korea
[d] School of Earth and Environmental Sciences, Seoul National University, Seoul 08826, Republic of Korea
[e] Environmental Satellite Centre, Climate and Air Quality Research Department, National Institute of Environmental Research, Incheon 22689, Republic of Korea
[f] Department of Environmental Resources Engineering, State University of New York, College of Environmental Science and Forestry, Syracuse, NY 13210, USA

HIGHLIGHTS

- Spatially continuous AOD and PMs were estimated using machine learning.
- Satellite-derive products and numerical model output were synergistically combined.
- AOD models resulted in R^2 ~ 0.74 when satellite-derived AOD did not exist.
- PM models with location information produced successful estimation results (R^2 ~ 0.9).
- The model performance of hourly PM concentrations varied little during the daytime.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 23 September 2019
Received in revised form 31 December 2019
Accepted 3 January 2020
Available online 8 January 2020

Editor: Jianmin Chen

Keywords:
Particulate matter
AOD
Satellite
Machine learning
Random Forest

ABSTRACT

Satellite-derived aerosol optical depth (AOD) products are one of main predictors to estimate ground-level particulate matter (PM_{10} and $PM_{2.5}$) concentrations. Since AOD products, however, are only provided under high-quality conditions, missing values usually exist in areas such as clouds, cloud shadows, and bright surfaces. In this study, spatially continuous AOD and subsequent PM_{10} and $PM_{2.5}$ concentrations were estimated over East Asia using satellite- and model-based data and auxiliary data in a Random Forest (RF) approach. Data collected from the Geostationary Ocean Color Imager (GOCI; 8 times per day) in 2016 were used to develop AOD and PM models. Three schemes (i.e. G1, A1, and A2) were proposed for AOD modeling according to target AOD data (GOCI AOD and AERONET AOD) and the existence of satellite-derived AOD. The A2 scheme showed the best performance (validation R^2 of 0.74 and prediction R^2 of 0.73 when GOCI AOD did not exist) and the resultant AOD was used to estimate spatially continuous PM concentrations. The PM models with location information produced successful estimation results with R^2 of 0.88 and 0.90, and rRMSE of 26.9 and 27.2% for PM_{10} and $PM_{2.5}$, respectively. The spatial distribution maps of PM well captured the seasonal and spatial characteristics of PM reported in the literature, which implies the proposed approaches can be adopted for an operational estimation of spatially continuous AOD and PMs under all sky conditions.

* Corresponding author at: School of Urban & Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea.
E-mail address: ersgis@unist.ac.kr (J. Im).
[1] The first two authors equally contributed to the paper.

https://doi.org/10.1016/j.scitotenv.2020.136516
0048-9697/

1. Introduction

Particulate matter (PM) in the atmosphere has been reported to be closely related to human health due to fine particle penetration into human organs or bloodstreams, pulmonary oxidative stress, and inflammatory response (Khaniabadi et al., 2017; Teoldi et al., 2017). Globally, the Western Pacific and South East Asia are the most affected regions by ambient air pollution (H. Xu et al., 2015; WHO, 2016; Lao et al., 2019). The annual averages of $PM_{2.5}$ concentrations in eastern China were 60–90 μg/m^3 from 2001 to 2006 (Van Donkelaar et al., 2010). Industrial and traffic emissions are major sources of $PM_{2.5}$ pollution in East Asia, and their seasonal variation is largely impacted by typical monsoon climate through aerosol transport (Huang et al., 2018; Sun et al., 2019).

Surface PM concentrations are usually measured at point-based ground-level stations, but for most regions monitoring sites are sparsely distributed with a focus on urban areas (Lee et al., 2011; Lu et al., 2019; Park et al., 2019). Studies have applied satellite-based PM modeling to produce spatially continuous surface PM concentrations over areas without ground-level monitoring sites (Kloog et al., 2015; Liu et al., 2009; Song et al., 2014; Van Donkelaar et al., 2014). Aerosol optical depth (AOD) is one of the most influential parameters for satellite-based PM modeling (Hu et al., 2017; Park et al., 2019), with AOD from polar-orbiting satellites being the most frequently contributed data for ground-level PM estimation (Guo et al., 2017; Lin et al., 2015; Liu et al., 2004; Meng et al., 2016; You et al., 2016a; You et al., 2016b). Some studies have utilized AOD products from geostationary satellite data whose temporal resolution is much higher than polar-orbiting satellite data (Chudnovsky et al., 2012; Park et al., 2019; J.-W. Xu et al., 2015).

Satellite-based AOD retrieval algorithms were developed under clear-sky conditions and require filtering cloudy pixels (Choi et al., 2016; Lee et al., 2010; Levy et al., 2013; Lyapustin et al., 2018; W. Zhang et al., 2018) since AOD-retrieval is not possible where there is cloud contamination (Sahoo and Patnaik, 2008). Christensen et al. (2017) summarized the challenges of passive satellite-based AOD retrievals under clouds (Christensen et al., 2017; Twohy et al., 2009; Várnai and Marshak, 2009). Previous studies have attempted to use interpolation techniques—typically kriging—to fill the spatial gaps of satellite-based AOD data (Lv et al., 2017; Xiao et al., 2017; Yang and Hu, 2018). The Aerosol Robotic Network (AERONET) acquires point-based measurements, which are frequently used as a reference for satellite-based AOD data, do not provide AOD information under cloudy conditions. Therefore, there is no way to verify AOD except under clear sky. However, ground-based PM concentrations are measured at stations constantly regardless of sky conditions, which provides an opportunity for producing spatially continuous surface PM concentrations.

While chemical transport models (CTM) also generate spatially continuous PM concentrations with high temporal resolution, they often produce biased results due to uncertainties from emissions, meteorological fields, and atmospheric chemical reactions (Geng et al., 2018; Lee et al., 2016; Park et al., 2011). Many studies have tried to combine CTM simulations with ground-based and satellite-derived observations, emission and meteorological parameters, and land use data to improve the performance of PM estimation (Geng et al., 2018; Lee et al., 2016; Van Donkelaar et al., 2016; J.-W. Xu et al., 2015). Previous studies used the ratio between AOD and $PM_{2.5}$ derived from the CTM to the satellite-derived AOD (J.-W. Xu et al., 2015). However, missing values in satellite-derived AOD still existed due to cloud cover. For spatially continuous PM concentrations, a seamless AOD distribution is required regardless of sky conditions. Recently, R. Zhang et al. (2018) and Tang et al. (2019) constructed two-stage random forest models for estimating ground $PM_{2.5}$ concentrations after filling gaps of satellite-retrieved AOD at local scale. The satellite-retrieved AOD data were used as a target variable for the gap-filled AOD modeling using the information of seasonality, meteorology, elevation, and vegetation index with auxiliary variables including population, land use, and road length.

This study aims to estimate spatially continuous PM concentrations using two random forest sub-models over East Asia covering eastern China, the Korean Peninsula, and Japan. The seamless AOD was modeled

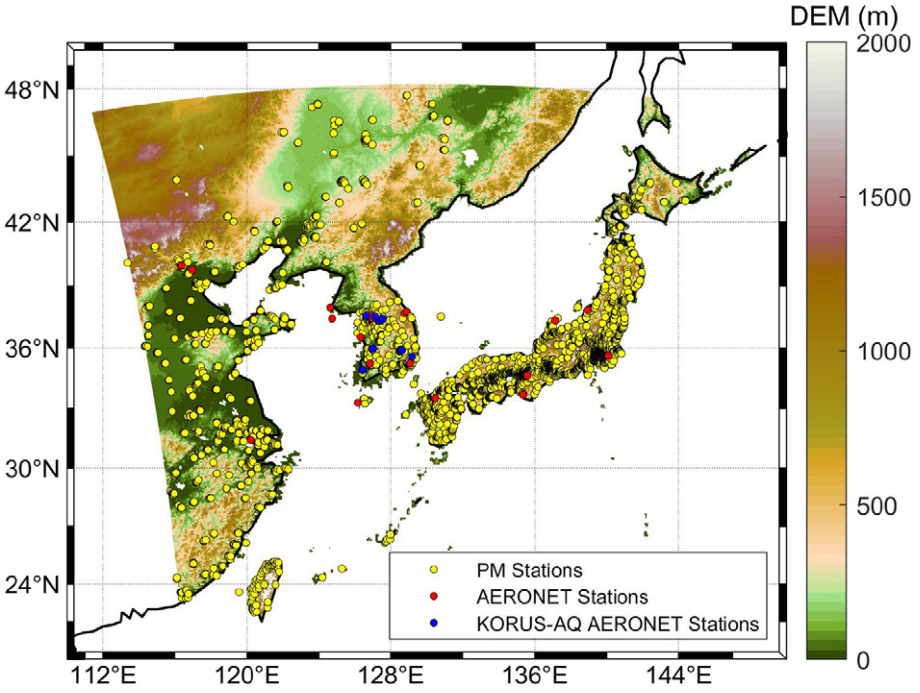


Fig. 1. Study area with AOD and PM reference stations. *Points in yellow, red, and blue indicate the locations of PM, AERONET, and KORUS-AQ AERONET stations respectively. The background image shows elevation for the study area, which corresponds to the GOCI coverage.

under two different conditions whether satellite-retrieved AOD exists or not. Three different schemes for AOD modeling were suggested considering target AOD data and the presence of satellite retrieved AOD data. Spatially continuous PM_{10} and $PM_{2.5}$ concentrations were then modeled using the best AOD model output among three different schemes. The contribution of input variables and spatiotemporal patterns for AOD and PMs were analyzed. This research is the first attempt to retrieve spatially continuous AOD with AERONET data where the satellite based AOD data do not exist. PM concentrations under all sky conditions were also produced mainly using satellite-derived products with model-based output in machine learning models.

2. Materials

2.1. Remote sensing data

Products from multiple satellite sensors were utilized to estimate PM_{10} and $PM_{2.5}$ concentrations in this study. The Geostationary Ocean

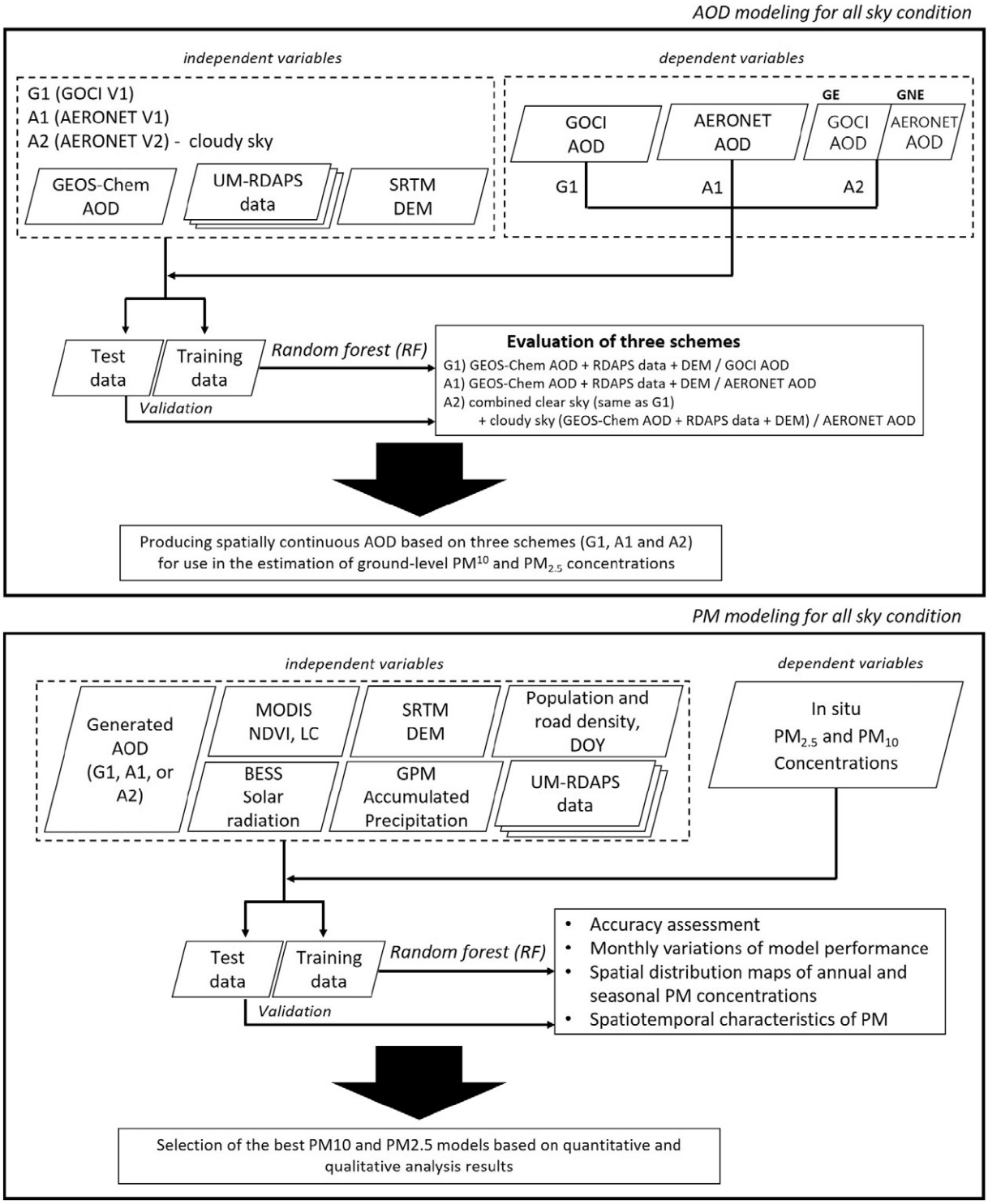


Fig. 2. Schematic process flow diagrams for estimating AOD and ground-level PM concentrations.

Table 1
List of input variables and their abbreviations for each model.

Data	Variable full name	Abbr.	AOD				$PM_{10}/PM_{2.5}$		
			G1	A1	A2		G1-PM	A1-PM	A2-PM
					GE[a]	GNE[b]			
GOCI	Aerosol optical depth	AOD					G1	A1	A2
GEOS-Chem	Aerosol optical depth	AOD	O	O	O	O	AOD	AOD	AOD
MODIS	Normalized difference vegetation index	NDVI					O	O	O
	Urban ratio	Urban_ratio					O	O	O
SRTM	Digital elevation model	DEM	O	O	O	O	O	O	O
GPM	Precipitation	Precip					O	O	O
UM-RDAPS	Air temperature	Tair					O	O	O
	Surface temperature	Tsuf	O	O	O	O			
	Dew-point temperature	Tdew	O	O	O		O	O	O
	Relative humidity	RH	O	O	O	O	O	O	O
	Maximum wind speed	MaxWS	O	O	O	O	O	O	O
	Accumulated 1-day MaxWS	1-MaxWS	O	O	O	O	O	O	O
	Accumulated 3-day MaxWS	3-MaxWS	O	O	O	O	O	O	O
	Accumulated 5-day MaxWS	5-MaxWS	O	O	O	O	O	O	O
	Accumulated 7-day MaxWS	7-MaxWS	O	O	O	O	O	O	O
	Visibility	Visibility	O	O	O	O	O	O	O
	PBLH	PBLH	O	O	O	O	O	O	O
	Surface pressure	Psuf	O	O	O	O	O	O	O
	Accumulated precipitation 3 h	AP3h	O	O	O	O			
BESS	Solar radiation	RSDN					O	O	O
Ancillary data	Road density	roadDens					O	O	O
	Population density	popDens					O	O	O
	Day of year	DOY	O	O	O	O	O	O	O
	Hour of day	HOD	O	O	O	O	O	O	O

$A1_{aug}$ and $A2_{aug}$ models trained with the augmented datasets used the same input variables with A1 and A2, respectively. The variables used for each model are represented with 'O' mark. The A2 model under GE condition is same as the G1.
[a] GE denotes where GOCI AOD exists.
[b] GNE denotes where GOCI AOD does not exist.

Color Imager (GOCI) is one of three payloads on the Communication, Ocean, and Meteorological Satellite (COMS) launched on 27 June 2010. GOCI observes the 2500 km × 2500 km region focusing on the Korean Peninsula, East China, Japan, and the surrounding seas providing full target area images with spatial resolution of 500 m at a 1-hour interval from 09:30 to 16:30 KST. This paper used aerosol products (i.e., AOD at 550 nm, Fine Mode Fraction at 550 nm, Single Scattering Albedo at 440 nm, Angstrom Exponent at 440–870 nm, Aerosol Type, and Dust Aerosol Index at 412–443 nm) produced by the GOCI Yonsei aerosol retrieval (YEAR) version-2 with 6 km × 6 km spatial resolution (Choi et al., 2018). In order to ensure the reliability of the GOCI AOD data, four filters (i.e. buddy check, local variance check, sub-pixel cloud fraction check, and diurnal variation check) proposed by Choi (Choi, 2017; Choi et al., 2018) were applied to the products. AOD values with adjacent clouds or low-quality flags, which might be influenced by undetected clouds or cloud edges (Wollner et al., 2014), were filtered out.

The Moderate Resolution Imaging Spectroradiometer (MODIS) sensor, with 36 spectral bands ranging from 0.4 μm to 14.4 μm, is onboard the Terra and Aqua satellites that cross the equator at around 10:30 and 13:30 local time, respectively. Sixteen-day normalized difference vegetation index (NDVI) products with 1 km resolution from Aqua (MYD13A2) and the annual land cover type product of 2013 with 500 m resolution (MCD12Q1) were downloaded from the National Aeronautics and Space Administration (NASA) Earthdata Search (https://search.earthdata.nasa.gov/). Urban area ratios within a 13 × 13 pixel neighborhood were calculated from the MODIS land cover type product following Park et al. (2019).

The Shuttle Radar Topography Mission (SRTM), which was onboard the STS-99 mission of the Space Shuttle Endeavour, used the single-pass interferometry by acquiring two different radar signals at the same time. Surface elevation was calculated from the difference between the signals from the two different antennas. The void-filled SRTM elevation product with 3-arcsecond resolution (approximately 90 m) was obtained from the US Geological Survey (USGS) EarthExplorer (https://earthexplorer.usgs.gov/).

The Global Precipitation Measurement (GPM) program was jointly developed by NASA and the Japanese Aerospace Exploration Agency (JAXA) and is managed by NASA's Goddard Space Flight Center. GPM obtains precipitation data with the Ku/Ka-band Dual-frequency Precipitation Radar (DPR) and a multi-channel GPM Microwave Imager (GMI). Half-hourly precipitation products with 0.1° spatial resolution (3IMERGHH) were downloaded from Goddard Earth Science Data and Information Service Center (GES DISC; https://mirador.gsfc.nasa.gov/) and transformed to the 24 h accumulated precipitation.

Table 2
Designations of AOD and PM models.

AOD models	PM models[b]
GOCI AOD (GA)	GA-PM_{org}[c,d]
	GA-PM_{lonlat}[e]
G1	G1-PM_{org}
	G1-PM_{lonlat}
A1	A1-PM_{org}
	A1-PM_{lonlat}
$A1_{aug}$[a]	$A1_{aug}$-PM_{org}
	$A1_{aug}$-PM_{lonlat}
A2	A2-PM_{org}
	A2-PM_{lonlat}
$A2_{aug}$	$A2_{aug}$-PM_{org}
	$A2_{aug}$-PM_{lonlat}

[a] The 'aug' subscript denotes AOD models with training data augmented by oversampling.
[b] Note that the PM models using AOD generated the best performing AOD model were discussed in the results.
[c] GA-PM models use GOCI-AOD as input, not simulated AOD. Thus, they work only for clear sky.
[d] 'org' denotes PM models without location information.
[e] 'lonlat' denotes PM models with location information.

Table 3
Validation results of AOD models using validation data.

AOD models	Slope	Intercept	R2	RMSE	rRMSE	MBE	MAE	IOA	NSE
G1	0.60	0.14	0.70	0.18	54.53%	0.01	0.12	0.89	0.69
A1	0.59	0.10	0.72	0.11	43.98%	0.00	0.07	0.89	0.70
$A1_{aug}$	0.73	0.09	0.74	0.10	40.70%	0.03	0.07	0.92	0.72
A2 for GNE	0.57	0.11	0.74	0.16	61.57%	−0.00	0.09	0.89	0.70
$A2_{aug}$ for GNE	0.74	0.13	0.74	0.15	52.26%	0.06	0.10	0.91	0.70

2.2. Model-based data

The GEOS-Chem, the Unified Model (UM)-Regional data Assimilation and Prediction System (RDAPS), and the Breathing Earth System Simulator (BESS) were employed in this study. AOD from GEOS-Chem and meteorological data from UM-RDAPS were used as input parameters to generate AOD over all-sky conditions. Meteorological data from UM-RDAPS and the shortwave radiation product from BESS were used with other satellite-based and ancillary data to estimate PM_{10} and $PM_{2.5}$.

This study used version 10-01 of GEOS-Chem, which is a global atmospheric chemistry three-dimensional model derived by using the meteorological data from the Global Forecast System produced by the US National Centers for Environmental Prediction. The MIX Asia Emissions Inventory was applied to provide the emission inputs needed to operate GEOS-Chem. The GEOS-Chem simulation covers eastern Asia, 70–150°E and 15–55°N, with horizontal resolution 0.25° × 0.3125°. The global GEOS-Chem simulation at 2° × 2.5° horizontal resolution was used to provide boundary conditions for the nested model (Choi et al., 2019; Lee et al., 2017; Park et al., 2003; Park et al., 2004).

UM-RDAPS is an operational regional numerical weather forecasting model of the Korea Meteorological Administration. Meteorological data from UM-RDAPS were used for estimating both AOD and PM. The spatial domain of UM-RDAPS is 77.38–176.56°E and 9.59–61.27°N. The initial and boundary conditions of RDAPS are based on UM-Global Data Assimilation and Prediction System (GDAPS) data with 25 km spatial resolution. About 100 meteorological parameters with spatial resolution of 12 km in 70 vertical layers are provided four times a day including 87-hour forecasts with a 3-hour interval. Air and surface temperature, dew-point temperature, relative humidity (RH), (accumulated) maximum wind speed, visibility, Planetary Boundary Layer Height (PBLH), 3-h accumulated precipitation, and surface pressure were used as input variables.

The BESS solar radiation products (i.e., shortwave radiation (SW), photosynthetically active radiation (PAR), and diffuse PAR) at 5 km spatial resolution are produced by integrating atmosphere and land products derived from MODIS in an atmospheric radiative transfer model with artificial neural networks (Ryu et al., 2018) (http://environment.snu.ac.kr/bess_rad/). The BESS SW products were used as predictors for estimating ground-level PM_{10} and $PM_{2.5}$.

Road density from the Global Roads Inventory Project (Meijer et al., 2018) (GRIP; available at www.globio.info), population density v4.11 from the Socioeconomic Data and Applications Center (SEDAC; https://sedac.ciesin.columbia.edu/), day-of-year (DOY), hour-of-day (HOD), and geographic coordinates (i.e. longitude and latitude) were used as additional input parameters only for PM estimation. Road density is provided as a global raster dataset at 5 × 5 arcminute resolution that sums the road length (distance) per each cell using a unit of a m km^{-2}. Population density was calculated for each administrative division with a unit of the number of people km^{-2}. DOY was converted to values ranging from −1 to 1 for a 1-year period using a sine function to consider seasonality, with 1 and −1 set as the middle of summer and winter, respectively (Stolwijk et al., 1999; Park et al., 2019). In addition, the sine function was used to convert the HOD to a value ranging from −1 to 1 for 24 h to account for the diurnal cycles with the Eq. (1).

$$HOD = \mathrm{Sin}((2 * \mathrm{pi} * UTC)/24) \quad (1)$$

where UTC is Coordinated Universal Time.

2.3. In-situ reference data

AERONET (https://aeronet.gsfc.nasa.gov/) is a global network of ground-based sun photometers that observe the physical and optical properties of atmospheric aerosols. Spectral AOD data measured in the range of 0.34–1.64 μm from AERONET have been utilized for verification of satellite-based AOD products because their uncertainty is low under clear sky conditions (Holben et al., 1998). Level 2.0 Version 2 direct sun algorithm data, automatically cloud cleared and manually inspected, were downloaded. For matching AERONET AOD with hourly GOCI AOD at 550 nm, the power law (Ångström, 1929) was applied to AOD values of neighboring wavelength and then averaged hourly. Data from 30 AERONET sites including 9 sites that were part of the Korea-United States Air Quality (KORUS-AQ) Field Study in Korea (Tables S1 and S2) were used (Fig. 1).

Hourly PM_{10} and $PM_{2.5}$ mass concentrations measured in μg m^{-3} in South Korea, eastern China, Taiwan, and Japan were acquired for the period from January to December of 2016. We obtained ground measurement data at 324 South Korean stations officially confirmed and provided by the AirKorea website (https://www.airkorea.or.kr/), at 1497 Chinese stations from National air quality data (http://beijingair.sinaapp.com/) measured by Beijing Municipal Environmental Protection Monitoring Center (BMEMC) using open Application Programming Interface (API), at 76 Taiwan stations from Taiwan Air Quality Monitoring Network (http://taqm.epa.gov.tw) from Environmental

Table 4
Prediction results of AOD models compared to the test dataset.

	AOD	Slope	Intercept	R2	RMSE	rRMSE	MBE	MAE	IOA	NSE
GE[a]	G1	0.47	0.17	0.39	0.16	64.49%	0.04	0.12	0.75	0.33
	$A1_{aug}$	0.65	0.12	0.65	0.12	47.89%	0.03	0.08	0.88	0.63
	GOCI	1.26	−0.05	0.84	0.12	49.59%	0.01	0.09	0.93	0.62
	GEOS-Chem	0.60	0.13	0.18	0.26	107.94%	0.04	0.17	0.62	−0.79
GNE[b]	G1	0.23	0.22	0.26	0.27	102.48%	0.02	0.16	0.59	0.26
	$A1_{aug}$	0.28	0.22	0.32	0.26	98.78%	0.03	0.16	0.64	0.30
	$A2_{aug}$	0.74	0.14	0.73	0.16	62.18%	0.07	0.11	0.90	0.68
	GEOS-Chem	0.34	0.18	0.14	0.33	126.60%	0.01	0.19	0.60	−0.13

[a] The GE denotes where GOCI AOD exists.
[b] GEN denotes where GOCI AOD dose note exist.

Table 5
Comparison of the model performance with other published studies.

Study	Target	Study area	Study period	Source of AOD	R^2	RMSE
Wang et al., 2017	$PM_{2.5}$	BTH[a] region in China	2015.07–2017.03	Himawari-8	0.86	24.3
Zeng et al., 2018	$PM_{2.5}$	Hubei in China	2017.01–2017.07	Himawari-8	0.82	30.08
Zang et al., 2018	PM_{10} $PM_{2.5}$ PM_1	China	2015.07–2016.12	Himawari-8	0.55 0.63 0.65	46.6 26.7 22.0
Present study	**PM_{10} $PM_{2.5}$**	**East Asia (GOCI coverage)**	**2016.01–2016.12**	**Estimated AOD (A2-AOD model)**	**0.88 0.90**	**26.9 15.77**

The bold is used to emphasize the results of this paper.
[a] BTH: Beijing–Tianjin–Hebei.

Protection Administration Executive Yuan R.O.C. (Taiwan), and at 2405 Japanese stations provided by National Institute for Environmental Studies, Japan (NIES; http://www.nies.go.jp). Of the 4302 total stations available in South Korea, China, and Japan, 3387 stations fell within the GOCI satellite coverage and were used to develop PM_{10} and $PM_{2.5}$ estimation models (Fig. 1).

3. Methods

Estimation of spatially continuous ground-level PM concentrations through the synergistic use of satellite data and numerical model output was conducted in two steps: estimating 1) seamless AOD and 2) ground-level PM_{10} and $PM_{2.5}$. Fig. 2 summarizes the process flow of this research. The extent of the study area was set based on the GOCI coverage area shown in Fig. 1. The entire input data were resampled every hour during the daytime from 09:00 to 16:00 KST (i.e. eight times per day), the retrieval time of the GOCI instrument, based on the grid of the GOCI YEAR version-2 aerosol data following Park et al. (2019). The station-based PM and AOD data were upscaled to the GOCI YEAR version-2 aerosol data grid, with a distance-weighted average algorithm applied if there were multiple stations in a single grid pixel, which is similar to Inverse Distance Weighted algorithm. The closer to the pixel center a station is, the more weight is assigned to the station (Meng et al., 2019). The spatial resolution of the outputs of

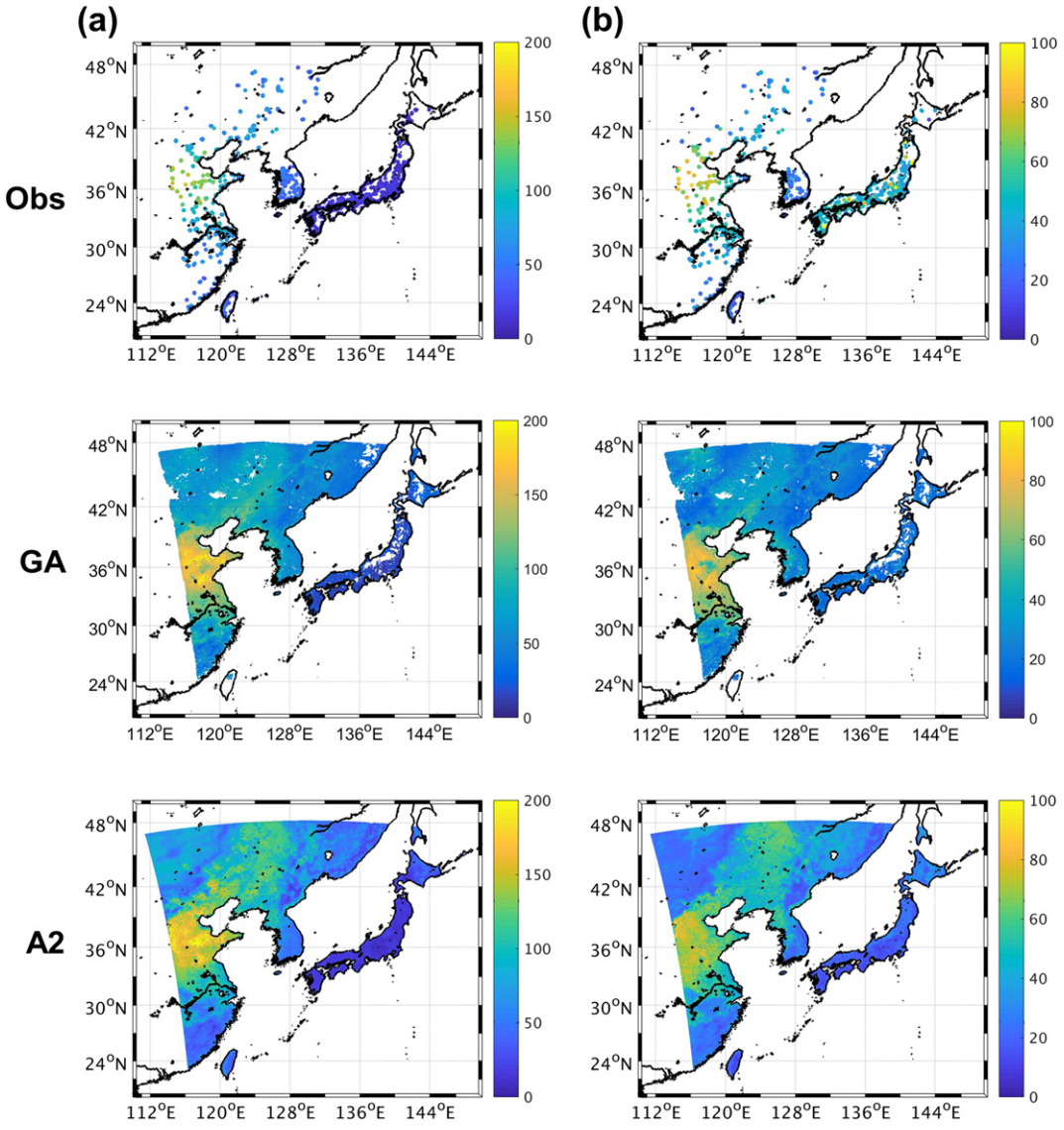


Fig. 3. Annual distribution of spatially continuous (a) PM_{10} and (b) $PM_{2.5}$ concentrations for observed (Obs) and model results (GA and $A2_{aug}$-PM).

AOD and PM models is 6 km × 6 km corresponding to the GOCI pixel grid.

3.1. Random Forest (RF)

Prior studies that estimated PM concentrations with machine learning techniques frequently used RF (Brokamp et al., 2018; Hu et al., 2017), which showed higher cross-validated accuracies than other machine learning techniques (Reid et al., 2015; Xu et al., 2018). RF is an ensemble model based on classification and regression trees (CART) whose multitude of trees are aggregated with majority voting and averaging in classification and regression tasks, respectively (Breiman, 2001). Individual tree models could overfit noise or biased samples in data, causing a large degree of variability among different data samples (a.k.a. high variance and low bias). An ensemble of individual tree models based on bagging and randomized node optimization (Breiman, 2001) can overcome the bias-variance problem of CART and smooth out the variance. In this study, RF was implemented using the *ranger* package, which is fast and particularly suited for high dimensional data, in the R statistical program (Wright and Ziegler, 2015). While some parameters—the number of trees (*num.trees*), the number of variables to possibly split in each node (*mtry*), minimal node size (*min.node.size*), and number of random splits (*num.random.splits*)—were optimized using a grid search algorithm, other parameters were set to default values (Table S3).

3.2. Estimation of all sky AOD

Spatially continuous AOD is needed to mitigate data gaps due to cloud. A total of 16 input variables with spatially continuous patterns were considered in the AOD estimation models using RF. Three schemes—G1, A1, and A2—with different target AOD and input variables were evaluated. The G1 scheme uses the GOCI AOD as the target variable, modeling with meteorological variables, elevation, DOY, and

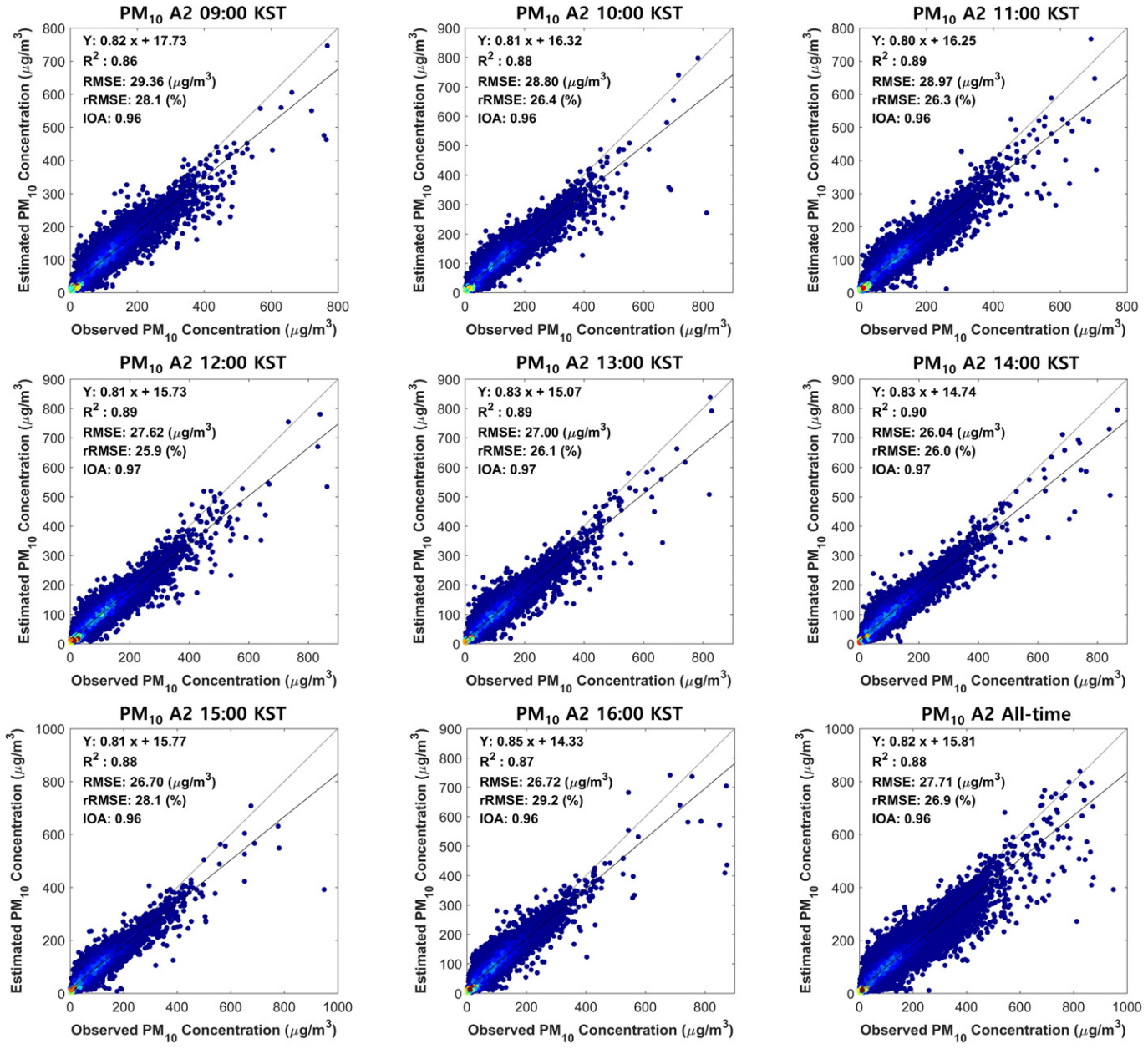


Fig. 4. Scatterplots comparing observed and estimated $A2_{aug}$-PM_{10} concentrations for different GOCI satellite operation times and for all-times combined. The warmer color presents a higher point density.

GEOS-Chem AOD as input variables. The stratified random subsampling was performed for the G1 dataset considering the range of GOCI AOD because too much data made it hard to build models due to the lack of memory. A balanced ratio between land and ocean samples was considered due to the distinctive aerosol sources between terrestrial (urban and rural) and maritime environments. The data were randomly divided into 80% for model calibration and 20% for model validation (Table S4).

AERONET AOD has a different filtering process when compared to GOCI AOD. AERONET Level 2.0 Version 2 provides data after automated cloud masking, manually quality-controlled data with pre- and post-field calibrations (Smirnov et al., 2000; Giles et al., 2019), while GOCI YEAR Version-2 filters out clouds, cloud shadows, bright surfaces, and high turbid water (Choi et al., 2018). In this study, AERONET AOD data were divided into two groups based on whether GOCI AOD data existed or not. We denoted the condition where both AERONET and GOCI AODs exist as "GE" and where GOCI AOD does not exist while AERONET AOD exists as "GNE". Scheme A1 was conducted under the GE condition. The input variables of scheme A1 are the same as those of scheme G1. Scheme A2 consists of models developed for GE and GNE. Under the GE condition, the G1 scheme was applied since the number of GOCI AOD data was much larger than that of AERONET AOD data. The input variables for the GNE condition were the same as those of scheme G1. Data for both GE and GNE conditions were randomly divided into 70% for model training, 20% for model validation, and 10% for model comparison of the three different schemes (G1, A1, and A2). The oversampling technique proposed by Park et al. (2019) was applied only on the training datasets of schemes A1 and A2 (GNE) for the moderate to high AERONET AOD ranges (i.e. 0.5–0.8, 0.8–1.0, 1.0–1.5, 1.5–2.0, and >2.0) due to the small number of data when compared to

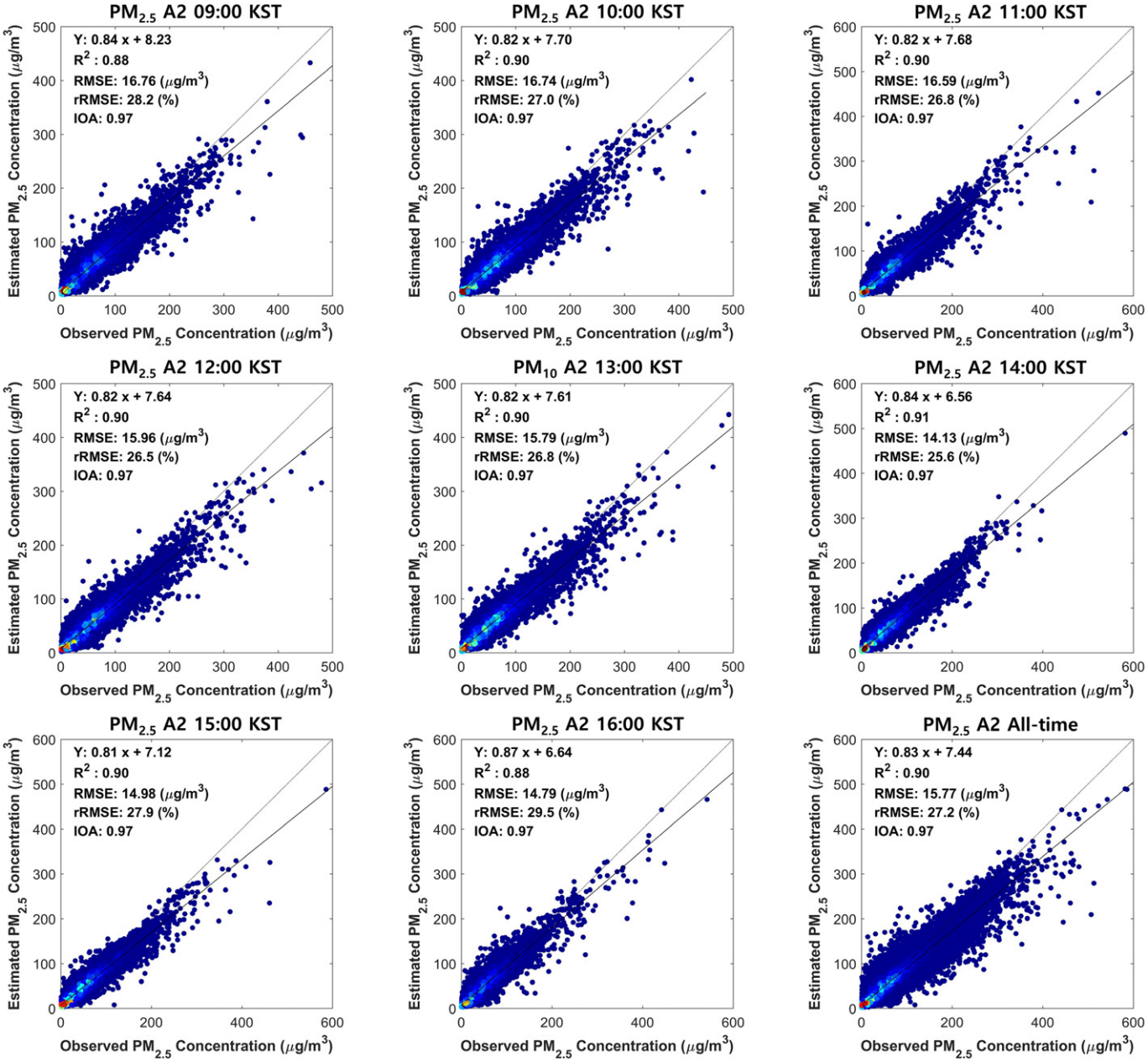


Fig. 5. Scatterplots comparing observed and estimated $A2_{aug}$-$PM_{2.5}$ concentrations for different GOCI satellite operation times and for all-times combined.

the low AERONET AOD range (i.e. 0.0–0.5). The AOD values nearby AERONET stations were assumed to be similar to the station AOD values. Thus, 5% random perturbation was applied to the AERONET AOD for the oversampled data with the input variables extracted from the pixels nearby the stations. The augmented A1 and A2 models trained with the oversampled training dataset were labeled $A1_{aug}$ and $A2_{aug}$ in this paper. A summary of predictor and target variables by scheme is presented in Table 1. The models developed were evaluated using the remaining 10% of AERONET AOD data (i.e. test dataset). The AOD model with the best performance among the G1, A1, and A2 schemes was then selected for developing PM estimation models.

3.3. Estimation of ground level PM_{10} and $PM_{2.5}$

The AODs estimated by three schemes were used with other variables to estimate spatially continuous ground-level PM_{10} and $PM_{2.5}$ concentrations (Tables 1 and 2) during daytime (09:00–16:00 KST). The data for PM estimations were randomly divided into 80% for model

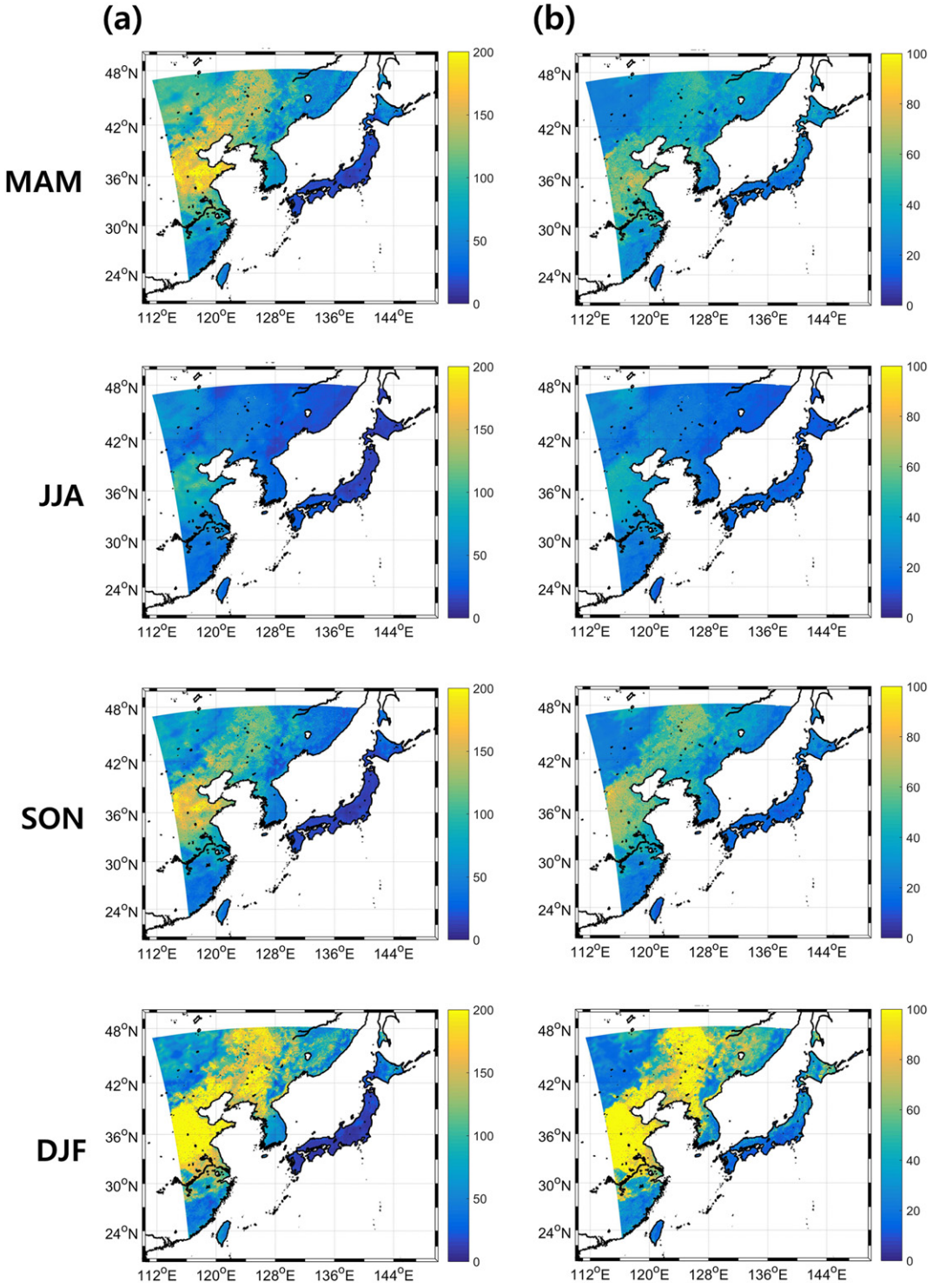


Fig. 6. Seasonal distribution of spatially continuous (a) $A2_{aug}$-PM_{10} and (b) $A2_{aug}$-$PM_{2.5}$. *MAM means March, April, and May. JJA indicates Jun, July, and August. SON means including September, October, and November. DJF is December, January, and February.

training and 20% for validation considering the seasonality and the level of PM concentrations. The PM models were developed based on the approach successfully evaluated in Park et al. (2019) with a total of 22 input parameters (Table 1). The ground PM sample size and concentrations vary by region (Fig. S1). PM concentrations measured in eastern China, South Korea, and Japan tend to be relatively high, moderate, and low, respectively. There is no dynamic variation of PM concentrations measured in Japan (concentrations were always <100 μg m^{-3} for both PM_{10} and $PM_{2.5}$) when compared to the other two countries. Therefore, PM models with location information (i.e. longitude and latitude) (PM_{lonlat}) were additionally developed to compare with the PM models without location information (PM_{org}).

Oversampling and subsampling techniques were applied to both PM_{org} and PM_{lonlat} estimation models due to unbalanced samples with dominance at low concentrations and low frequency at high concentrations (Park et al., 2019). In this study, we assumed that the PM concentrations nearby each station would be similar to the observed PM concentrations (i.e. target variable). Thus, input variables at nearby pixels from each station were extracted, and the target PM concentrations were randomly disturbed within 5% of the PM concentration at the station pixel. The subsampling technique is generally applied to a section having a relatively large number of samples (i.e. low concentrations) through simple random sampling. Intervals of 90 and 60 μg m^{-3} for PM_{10} and $PM_{2.5}$, respectively, were used to determine the sample sizes (i.e., 0–90, 90–190, …, 360–540, and >540 μg m^{-3} for PM_{10} and 0–60, 60–120, …, 360–480, >480 μg m^{-3} for $PM_{2.5}$). The oversampling and subsampling approaches proposed by Park et al. (2019) were applied for each interval. The number of samples before and after applying the oversampling technique for PM models is shown in Table S4. In this study, we compared the PM_{org} and PM_{lonlat} models through quantitative accuracy assessment and qualitative analysis of the spatial distributions of PMs. In addition, the performance of the proposed PM estimation models under all-sky and clear sky conditions was examined.

3.4. Model evaluation

The AOD and PM models developed were evaluated using seven statistical indicators—coefficient of determination (R^2), root mean square error (RMSE), relative RMSE (rRMSE), mean bias error (MBE), mean absolute error (MAE), index of agreement (IOA), and Nash-Sutcliffe model efficiency coefficient (NSE). IOA and NSE were adopted to compare the models between different input variables according to the schemes (Li et al., 2009; Saide et al., 2011). The best model was selected based on model performance evaluated through statistical indicators—the higher R^2, IOA, and NSE, and lower RMSE, rRMSE, MBE, and MAE.

In order to examine the reliability and stability of the models on the spatial and temporal domains, cross-validation was also conducted. Since there exist only 30 AERONET stations in the study area, leave-one-station-out cross-validation was conducted for the AOD models. Ten-fold cross-validation by PM stations was performed to see the spatial robustness of the PM models. In addition, to analyze the temporal reliability of the models, the hourly cross-validation (from 9 am to 4 pm) was conducted for both AOD and PM models.

4. Results and discussion

4.1. AOD estimation

The G1, A1, and A2 models were developed for filling AOD gaps using RF. The statistical results of model validation are summarized in Table 3 (refer to Table S6 for calibration results). Both $A1_{aug}$ and $A2_{aug}$ (GNE) showed higher performance than without augmentation with an increase of 0–0.03 in R^2, 0.02–0.03 in IOA and 0–0.02 in NSE, and a decrease of 3.28–9.31% in rRMSE, while the MBE and MAE showed little change between the two datasets. The augmented models contained an increased number of training samples (Table S4) that particularly helped to fit the high AOD values (over 0.8) when compared to original samples (not shown).

Three AOD models (i.e. G1, $A1_{aug}$ and $A2_{aug}$), GOCI AOD, and GEOS-Chem AOD were compared to the AERONET AOD test data (Table 4). Under the GE condition, the $A1_{aug}$ model showed higher performance than the GOCI AOD with an increase in R^2 of about 0.26 and decreases in RMSE ~ 0.04 and rRMSE ~ 16.60%. The performance of AOD models tends to decrease for the GNE condition than the GE condition. The $A2_{aug}$ model, which was developed separately under GE and GNE conditions, showed a better result (R^2 of 0.74, RMSE of 0.15, rRMSE of 52.26%, IOA of 0.91, and NSE of 0.70) under the GNE condition. The performance of the two models with same input variables (i.e., $A2_{aug}$ model for GNE and the $A1_{aug}$ model for GE) were similar. It appears necessary to build AOD separately for GE and GNE conditions because the GNE covers various sky conditions including the proximity of clouds and snows. Using the whole test dataset, the results showed R^2 values of 0.28, 0.36, and 0.68 for G1, $A1_{aug}$, and $A2_{aug}$ models, respectively.

Cross-validations by hour and station were also performed. The cross-validation results by hour (Table S7) showed similar performance with the validation results using the separate validation dataset. However, the cross-validation results by station (Table S8) showed lower performance than the validation results using the separate validation dataset. The low performance of the leave-one-station-out cross-validation is mainly because of the small number of AERONET stations (i.e. 30 stations) in our study area. However, the A2 and $A2_{aug}$ models under the GNE condition in Table S8 produced less erroneous results than the G1 model in Table 4.

The $A2_{aug}$ model consistently showed the best performance and was selected for subsequent PM estimation. Previous AOD gap filling studies used MODIS-derived AOD products (Xiao et al., 2017; W. Zhang et al., 2018; Zhao et al., 2019) as the target. The correlation between the modeled AOD and AERONET AOD showed R^2 values of 0.36–0.44 (Xiao et al., 2017; Zhao et al., 2019). The $A2_{aug}$ model showed a higher R^2 value than previous studies, but the G1 model presented a slightly lower R^2 value. However, considering the larger study area of East Asia, the G1 model also seems comparable with previous studies.

Fig. S2 summarizes relative variable importance identified by RF for AOD modeling by scheme. For all schemes, AOD (i.e. GEOS-Chem AOD) was identified as the most influential factor to the AOD models. Other than AOD, DOY and wind variables were commonly identified by RF as contributing variables for AOD modeling (Fig. S2). DOY is considered to represent the seasonality of AOD, while wind-based variables are related to accumulation of aerosols in the atmosphere. Fig. S3 shows the seasonal average of the AODs simulated by the different models. All AOD models tend to follow the same spatial patterns seen in the GEOS-Chem AOD. The $A2_{aug}$ model shows higher AOD values than the G1 and $A1_{aug}$ models since the range of the target AERONET AOD is wider for the GNE condition than the GE condition (not shown). However, the GNE condition cannot represent all sky conditions such as thick clouds, which is the limitation of our study.

4.2. PM_{10} and $PM_{2.5}$ estimation

Figs. S4 and S5 show scatterplots of the validation results of the PM_{org} and PM_{lonlat} models, respectively. All PM models showed relatively similar performance. The inclusion of location in the PM_{lonlat} models increased R^2 values from 0.84–0.86 to 0.89–0.90 and from 0.85–0.86 to 0.90, and NSE values from 0.83 to 0.88 and from 0.84 to 0.89, and reduced RMSE (rRMSE) ~5.77 μg m^{-3} (4.4%) and ~3.47 μg m^{-3} (6%) for PM_{10} and $PM_{2.5}$, respectively, compared to the PM_{org} models. The rRMSE values were reduced to <28% in all PM_{lonlat} models and the IOA values were over 0.96 for all models.

The hourly cross-validation produced R^2 (rRMSE) of 0.92–0.93 (24.9%–25.3%) and 0.93 (29.3%–29.6%) for PM_{10} and $PM_{2.5}$, respectively, which were similar to the validation results using the separate validation dataset (Table S9). The 10-fold cross-validation by station showed

a bit lower performance (R^2 ~ 0.63–0.77 and rRMSE ~32.5–35.8%) than the validation results using the separate validation dataset (Table S10). It is possibly because the stations were not regularly distributed throughout the study area, rather focusing on mega urban areas. Thus, the cross-validation results showed relatively low performance for sparsely distributed stations (not shown).

Consequently, this section focuses on discussing PM estimation with the PM_{lonlat} model using the $A2_{aug}$ scheme, as it produced the best performance among the AOD models investigated. The $A2_{aug}$-PM_{lonlat} models show good agreements compared to previous studies that estimated hourly PM concentrations using AOD derived from geostationary satellite sensors (Wang et al., 2017; Zang et al., 2018; Zeng et al., 2018) (Table 5). The literature using satellite-based AOD has usually investigated PM estimation under clear sky conditions. However, it should be noted that the models developed in this study provide PM concentrations in all-sky conditions.

Fig. S6 presents the monthly variations of the performance of $A2_{aug}$-PM_{lonlat} models under different sky conditions (i.e., all-sky, GNE, and GE) with statistical indicators of R^2, RMSE, and rRMSE for PM_{10} and $PM_{2.5}$. The PM_{lonlat} models using GOCI-based AOD (GA-PM_{lonlat} models) was also developed to compare with the $A2_{aug}$-PM_{lonlat} models. The summer season (i.e. Jun, July, and August), especially in August, showed the lowest performance resulting in R^2 values of 0.77 and 0.80 and rRMSE of 35.6% and 33.4% for PM_{10} and $PM_{2.5}$, respectively. Such poor model performance might be because the low PM concentrations were usually measured in the study area in the summer season (Fig. S6). The PM concentrations measured at stations for GE were generally low, which made the $A2_{aug}$-PM_{lonlat} models outperform the GA-PM_{lonlat} models for all months. Interestingly, there was a large difference between the GA-PM_{lonlat} models and the $A2_{aug}$-PM_{lonlat} models in August. This result implies that the range of PM concentrations, especially high concentrations, affects the model performance (Figs. S6–S7).

The relative variable importance identified by RF for estimating PM_{10} and $PM_{2.5}$ is summarized in Fig. S8. The estimated AOD ($A2_{aug}$) was identified as the most contributing variable for the PM_{org} models (Fig. S8(a), (c)), whereas lon (i.e. longitude) played the most important role in the PM_{lonlat} models (Fig. S8(b), (d)). Meteorological variables such as MaxWS (i.e. maximum wind speed), visibility, and RH (i.e. relative humidity) were commonly identified as contributing variables by both PM_{org} and PM_{lonlat} models to estimate spatially continuous PM_{10} and $PM_{2.5}$ concentrations. The importance of AOD and meteorological variables for PM estimation using the RF approach has consistently been reported in the literature (Hu et al., 2017; Park et al., 2019). Although the location information was a contributor to enhance the performance of the PM models, the longitude and latitude data could cause large uncertainty over areas where in-situ stations were sparsely distributed (e.g., North Korea, Inner Mongolia in China, and Primorsky krai province in Russia). Thus, the PM_{org} and PM_{lonlat} model results were combined using a Gaussian filter when creating the spatial distribution maps of PM concentrations.

The spatial distributions of annual PM concentrations combining the PM_{org} and PM_{lonlat} model results for $A2_{aug}$ and GA are depicted in Fig. 3, respectively. To ensure reliability, the PM distribution map from the GA model was generated using pixels with data available for >5% of 2016 (>16 days) (Park et al., 2019) (Fig. S9). PM in some areas of eastern China may have been overestimated due to the influence of latitude and longitude used in the PM_{lonlat} model, which would be mitigated by combining the PM_{org} and PM_{lonlat} models.

Figs. 4 and 5 show the $A2_{aug}$-PM_{10} and $A2_{aug}$-$PM_{2.5}$ models (PM_{lonlat} models) performance, respectively, for different hours during the GOCI satellite operation (09:00–16:00 KST). The $A2_{aug}$-PM_{lonlat} model showed good performance at different times with R^2 (RMSE) values of 0.86–0.90 and 0.88–0.91 (25.78–29.16 μg m^{-3} and 14.13–16.76 μg m^{-3}) and rRMSE of 25.9–29.36% and 25.6–29.5% for PM_{10} and $PM_{2.5}$, respectively. Both $A2_{aug}$-PM_{10} and $A2_{aug}$-$PM_{2.5}$ models resulted in similar model performance even at different time slots. The hourly distributions of PM_{10} and $PM_{2.5}$ concentrations on April 24 and 22, 2016, are also shown in Figs. S9 and S10, respectively. They well captured the distribution of the varied PM concentrations for different hours during the daytime.

Fig. 6 illustrates the spatial distribution of seasonal averaged PM_{10} and $PM_{2.5}$ concentrations. The spatial pattern shown, i.e. high concentrations over eastern China and low concentrations in Japan, and the seasonal patterns, i.e. high concentrations in Spring (MAM) and Winter (DJF) and low concentrations during Summer (JJA), agreed with observed PM concentrations (Choi et al., 2009; Park et al., 2019; Wei et al., 2019). The $A2_{aug}$-PM models also captured relatively high concentrations in the fall when anthropogenic emissions, such as heating, start in the northern part of the study area.

4.3. Novelty and limitations

In this study, we produced seamless AOD data and subsequently developed models for estimating ground-level PM concentrations over eastern Asia using a RF approach. Spatially continuous PM concentrations with high temporal resolution (eight times per day) were estimated over the extent of the geostationary satellite data (i.e. GOCI). The A2 and $A2_{aug}$ model-based AOD data showed consistent performance under both GE and GNE conditions, which benefits estimation of ground-level PM concentrations in all sky conditions. The hourly performance of the PM models during 09:00 KST to 16:00 KST was also robust and stable. The results imply that the proposed approaches can be adopted for an operational estimation of spatially continuous AOD and PMs. It should be highlighted that this study is the first attempt for developing models to produce AOD and PMs mainly using satellite data with numerical model output under all sky conditions over eastern Asia.

However, there are three limitations in our study. First, ground-level PM concentrations were modeled only on the land due to several land-based input variables and no in-situ PM data over the ocean. Ground-level PM concentrations over both land and ocean will be further investigated in future research. Second, ground reference data for both AOD and PMs are sparsely distributed in some areas, which could cause some spatial uncertainties where reference sites do not exist. Third, A2 and $A2_{aug}$ models trained under different GE and GNE conditions are still limited in reflecting cloudy conditions, which could propagate some errors for PM estimation under cloudy conditions.

5. Conclusion

The geostationary satellite sensor GOCI provides AOD products eight times per day, which enables to estimate ground-level PM concentrations with high temporal resolution. However, the satellite products have a disadvantage that they are not provided in areas where clouds, cloud shadows, and bright surfaces exist. This study suggests estimating spatially continuous PM_{10} and $PM_{2.5}$ concentrations under all sky conditions using a machine learning approach (i.e. RF) over eastern Asia including eastern China, the Korean Peninsula, and Japan. The seamless AOD produced using satellite- and model-based products was used to develop PM_{10} and $PM_{2.5}$ prediction models that work under all sky conditions. The $A2_{aug}$ scheme showed consistently good performance under both GE and GNE conditions when compared to other schemes for the AOD models. The performance of the PM estimation models varied little during all times a day (09:00–16:00 KST) showing R^2 values >0.87 and 0.88 and rRMSE <29.7% and 29.9% for PM_{10} and $PM_{2.5}$, respectively. The spatial and seasonal patterns of PM_{10} and $PM_{2.5}$ agreed well with the observed ground-level PM concentrations. Since land-based variables were used in this study, PM concentrations were not estimated for the ocean. With observation data measured in the ocean and additional feature selection processes, the proposed models will be improved in future research so to work for the ocean as well as the land.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study was supported by the National Institute of Environmental Research (NIER), Republic of Korea (NIER-2019-01-02-051), by the National Research Foundation of Korea (NRF-2017M1A3A3A02015981; NRF-2017M3D8A1092021), by the Korea Meteorological Administration (KMIPA 2017-7010), and by the Institute for Information & communications Technology Promotion (IITP) supported by the Ministry of Science and ICT (MSIT), Republic of Korea (IITP-2019-2018-0-01424).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.136516.

References

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Brokamp, C., Jandarov, R., Hossain, M., Ryan, P., 2018. Predicting daily urban fine particulate matter concentrations using a random forest model. Environmental Science & Technology 52 (7), 4173–4179.
Choi, M., 2017. Retrieval of Aerosol Optical Properties From Goci: Algorithm Improvement. Analysis and Application to PM. (Doctoral dissertation). Graduate School, Yonsei University, Seoul.
Choi, Y.S., Park, R.J., Ho, C.H., 2009. Estimates of ground-level aerosol mass concentrations using a chemical transport model with Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol observations over East Asia. Journal of Geophysical Research: Atmospheres 114 (D4).
Choi, M., Kim, J., Lee, J., Kim, M., Park, Y.-J., Jeong, U., ... Eck, T.F., 2016. GOCI Yonsei Aerosol Retrieval (YAER) algorithm and validation during the DRAGON-Ne Asia 2012 campaign. Atmospheric Measurement Techniques 9 (3), 1377–1398.
Choi, M., Kim, J., Lee, J., Kim, M., Park, Y.-J., Holben, B., ... Song, C.H., 2018. GOCI Yonsei aerosol retrieval version 2 products: An improved algorithm and error analysis with uncertainty estimation from 5-year validation over east Asia. Atmospheric Measurement Techniques 11 (1).
Choi, J., Park, R.J., Lee, H.-M., Lee, S., Jo, D.S., Jeong, J.I., ... Lee, M.-D., 2019. Impacts of local vs. trans-boundary emissions from different sectors on PM2.5 exposure in South Korea during the KORUS-AQ campaign. Atmospheric environment 203, 196–205.
Christensen, M.W., Neubauer, D., Poulsen, C.A., Thomas, G.E., Mcgarragh, G.R., Povey, A.C., ... Grainger, R.G., 2017. Unveiling aerosol–cloud interactions–part 1: Cloud contamination in satellite products enhances the aerosol indirect forcing estimate. Atmospheric Chemistry and Physics 17 (21), 13151–13164.
Chudnovsky, A.A., Lee, H.J., Kostinski, A., Kotlov, T., Koutrakis, P., 2012. Prediction of daily fine particulate matter concentrations using aerosol optical depth retrievals from the Geostationary Operational Environmental Satellite (GOES). J. Air Waste Manage. Assoc. 62 (9), 1022–1031.
Ångström, A., 1929. On the atmospheric transmission of sun radiation and on dust in the air. Geogr. Ann. 11 (2), 156–166.
Geng, G., Murray, N.L., Tong, D., Fu, J.S., Hu, X., Lee, P., ... Liu, Y., 2018. Satellite-based daily PM2.5 estimates during fire seasons in Colorado. Journal of Geophysical Research: Atmospheres 123 (15), 8159–8171.
Giles, D.M., Sinyuk, A., Sorokin, M.G., Schafer, J.S., Smirnov, A., Slutsker, I., ... Campbell, J.R., 2019. Advancements in the Aerosol Robotic Network (AERONET) version 3 database–automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements. Atmospheric Measurement Techniques 12 (1), 169–209.
Guo, Y., Tang, Q., Gong, D.-Y., Zhang, Z., 2017. Estimating ground-level PM2.5 concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. Remote Sens. Environ. 198, 140–149.
Holben, B.N., Eck, T.F., Slutsker, I., Tanre, D., Buis, J., Setzer, A., ... Nakajima, T., 1998. Aeronet—a federated instrument network and data archive for aerosol characterization. Remote Sensing of Environment 66 (1), 1–16.
Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L.A., Strickland, M.J., Liu, Y., 2017. Estimating PM2.5 concentrations in the conterminous United States using the random forest approach. Environmental Science & Technology 51 (12), 6936–6944.
Huang, T., Yu, Y., Wei, Y., Wang, H., Huang, W., Chen, X., 2018. Spatial–seasonal characteristics and critical impact factors of PM2.5 concentration in the Beijing–Tianjin–Hebei urban agglomeration. PLoS One 13 (9), e0201364.
Khaniabadi, Y.O., Goudarzi, G., Daryanoosh, S.M., Borgini, A., Tittarelli, A., De Marco, A., 2017. Exposure to PM10, NO2, and O3 and impacts on human health. Environ. Sci. Pollut. Res. 24 (3), 2781–2789.
Kloog, I., Sorek-Hamer, M., Lyapustin, A., Coull, B., Wang, Y., Just, A.C., ... Broday, D.M., 2015. Estimating daily PM2.5 and PM10 across the complex geo-climate region of Israel using MAIAC satellite-based AOD data. Atmospheric environment 122, 409–416.
Lao, X.Q., Guo, C., Chang, L.-y., Bo, Y., Zhang, Z., Chuang, Y.C., ... Lau, A.K., 2019. Long-term exposure to ambient fine particulate matter (PM2.5) and incident type 2 diabetes: A longitudinal cohort study. Diabetologia 62 (5), 759–769.
Lee, J., Kim, J., Song, C.H., Ryu, J.-H., Ahn, Y.-H., Song, C., 2010. Algorithm for retrieval of aerosol optical properties over the ocean from the geostationary ocean color imager. Remote Sens. Environ. 114 (5), 1077–1088.
Lee, H., Liu, Y., Coull, B., Schwartz, J., Koutrakis, P., 2011. A novel calibration approach of MODIS AOD data to predict PM2.5 concentrations. Atmospheric Chemistry & Physics 11 (15).
Lee, S., Song, C., Park, R., Park, M., Han, K., Kim, J., ... Woo, J.-H., 2016. GIST-PM-Asia v1: development of a numerical system to improve particulate matter forecasts in South Korea using geostationary satellite-retrieved aerosol optical data over Northeast Asia. Geoscientific Model Development 9 (1).
Lee, H.-M., Park, R.J., Henze, D.K., Lee, S., Shim, C., Shin, H.-J., ... Woo, J.-H., 2017. PM2.5 source attribution for Seoul in May from 2009 to 2013 using GEOS-Chem and its adjoint model. Environmental Pollution 221, 377–384.
Levy, R., Mattoo, S., Munchak, L., Remer, L., Sayer, A., Patadia, F., Hsu, N., 2013. The collection 6 MODIS aerosol products over land and ocean. Atmospheric Measurement Techniques 6 (11), 2989.
Li, H., Faruque, F., Williams, W., Al-Hamdan, M., Luvall, J., Crosson, W., ... Limaye, A., 2009. Optimal temporal scale for the correlation of AOD and ground measurements of PM2.5 in a real-time air quality estimation system. Atmospheric environment 43 (28), 4303–4310.
Lin, C., Li, Y., Yuan, Z., Lau, A.K., Li, C., Fung, J.C., 2015. Using satellite remote sensing data to estimate the high-resolution distribution of ground-level PM2.5. Remote Sens. Environ. 156, 117–128.
Liu, Y., Park, R.J., Jacob, D.J., Li, Q., Kilaru, V., Sarnat, J.A., 2004. Mapping annual mean ground-level PM2.5 concentrations using multiangle imaging spectroradiometer aerosol optical thickness over the contiguous United States. Journal of Geophysical Research: Atmospheres 109 (D22).
Liu, Y., Paciorek, C.J., Koutrakis, P., 2009. Estimating regional spatial and temporal variability of PM2.5 concentrations using satellite data, meteorology, and land use information. Environ. Health Perspect. 117 (6), 886–892.
Lu, L.L., Weng, Q.H., Guo, H.D., Feng, S.Y., Li, Q.T., 2019. Assessment of urban environmental change using multi-source remote sensing time series (2000–2016): a comparative analysis in selected megacities in Eurasia. Sci. Total Environ. 684, 567–577. https://doi.org/10.1016/j.scitotenv.2019.05.344.
Lv, B., Hu, Y., Chang, H.H., Russell, A.G., Cai, J., Xu, B., Bai, Y., 2017. Daily estimation of ground-level PM2.5 concentrations at 4 km resolution over Beijing-Tianjin-Hebei by fusing MODIS AOD and ground observations. Sci. Total Environ. 580, 235–244.
Lyapustin, A., Wang, Y., Korkin, S., Huang, D., 2018. MODIS collection 6 MAIAC algorithm. Atmospheric Measurement Techniques 11 (10).
Meijer, J.R., Huijbregts, M.A., Schotten, K.C., Schipper, A.M., 2018. Global patterns of current and future road infrastructure. Environ. Res. Lett. 13 (6), 064006.
Meng, X., Fu, Q., Ma, Z., Chen, L., Zou, B., Zhang, Y., ... Kan, H., 2016. Estimating ground-level PM10 in a Chinese city by combining satellite data, meteorological information and a land use regression model. Environmental Pollution 208, 177–184.
Meng, Y., Cave, M., Zhang, C., 2019. Comparison of methods for addressing the point-to-area data transformation to make data suitable for environmental, health and socio-economic studies. Sci. Total Environ. 689, 797–807.
Park, R.J., Jacob, D.J., Chin, M., Martin, R.V., 2003. Sources of carbonaceous aerosols over the United States and implications for natural visibility. Journal of Geophysical Research: Atmospheres 108 (D12).
Park, R.J., Jacob, D.J., Field, B.D., Yantosca, R.M., Chin, M., 2004. Natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: implications for policy. Journal of Geophysical Research: Atmospheres 109 (D15).
Park, R., Song, C., Han, K., Park, M., Lee, S.-S., Kim, S.-B., Shimizu, A., 2011. A study on the aerosol optical properties over East Asia using a combination of CMAQ-simulated aerosol optical properties and remote-sensing data via a data assimilation technique. Atmos. Chem. Phys. 11 (23), 12275–12296.
Park, S., Shin, M., Im, J., Song, C.-K., Choi, M., Kim, J., ... Lee, D.-W., 2019. Estimation of ground-level particulate matter concentrations through the synergistic use of satellite observations and process-based models over South Korea. Atmospheric Chemistry and Physics 19 (2), 1097–1113.
Reid, C.E., Jerrett, M., Petersen, M.L., Pfister, G.G., Morefield, P.E., Tager, I.B., ... Balmes, J.R., 2015. Spatiotemporal prediction of fine particulate matter during the 2008 northern california wildfires using machine learning. Environmental science & technology 49 (6), 3887–3896.
Ryu, Y., Jiang, C., Kobayashi, H., Detto, M., 2018. Modis-derived global land products of shortwave radiation and diffuse and total photosynthetically active radiation at 5 km resolution from 2000. Remote Sens. Environ. 204, 812–825.
Sahoo, T., Patnaik, S., 2008. Cloud removal from satellite images using Auto Associative Neural Network and Stationary Wevlet Transform. Paper Presented at the 2008 First International Conference on Emerging Trends in Engineering and Technology.
Saide, P.E., Carmichael, G.R., Spak, S.N., Gallardo, L., Osses, A.E., Mena-Carrasco, M.A., Pagowski, M., 2011. Forecasting urban PM10 and PM2.5 pollution episodes in very stable nocturnal conditions and complex terrain using WRF–Chem CO tracer model. Atmos. Environ. 45 (16), 2769–2780.
Smirnov, A., Holben, B., Eck, T., Dubovik, O., Slutsker, I., 2000. Cloud-screening and quality control algorithms for the AERONET database. Remote Sens. Environ. 73 (3), 337–349.
Song, W., Jia, H., Huang, J., Zhang, Y., 2014. A satellite-based geographically weighted regression model for regional PM2.5 estimation over the pearl river delta region in China. Remote Sens. Environ. 154, 1–7.

Stolwijk, A., Straatman, H., Zielhuis, G., 1999. Studying seasonality by using sine and cosine functions in regression analysis. J. Epidemiol. Community Health 53 (4), 235–238.
Sun, X., Luo, X.-S., Xu, J., Zhao, Z., Chen, Y., Wu, L., ... Zhang, D., 2019. Spatio-temporal variations and factors of a provincial PM2.5 pollution in eastern china during 2013–2017 by geostatistics. Scientific reports 9 (1), 3613.
Tang, D., Liu, D., Tang, Y., Seyler, B.C., Deng, X., Zhan, Y., 2019. Comparison of GOCI and Himawari-8 aerosol optical depth for deriving full-coverage hourly PM2.5 across the Yangtze River Delta. Atmos. Environ. 217, 116973.
Teoldi, F., Lodi, M., Benfenati, E., Colombo, A., Baderna, D., 2017. Air quality in the Olona Valley and in vitro human health effects. Sci. Total Environ. 579, 1929–1939. https://doi.org/10.1016/j.scitotenv.2016.11.203.
Twohy, C.H., Coakley Jr., J.A., Tahnk, W.R., 2009. Effect of changes in relative humidity on aerosol scattering near clouds. Journal of Geophysical Research: Atmospheres 114 (D5).
Van Donkelaar, A., Martin, R.V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., Villeneuve, P.J., 2010. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. Environ. Health Perspect. 118 (6), 847–855.
Van Donkelaar, A., Martin, R.V., Brauer, M., Boys, B.L., 2014. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. Environ. Health Perspect. 123 (2), 135–143.
Van Donkelaar, A., Martin, R.V., Brauer, M., Hsu, N.C., Kahn, R.A., Levy, R.C., ... Winker, D.M., 2016. Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. Environmental Science & Technology 50 (7), 3762–3772.
Várnai, T., Marshak, A., 2009. Modis observations of enhanced clear sky reflectance near clouds. Geophys. Res. Lett. 36 (6).
Wang, W., Mao, F., Du, L., Pan, Z., Gong, W., Fang, S., 2017. Deriving hourly PM2.5 concentrations from Himawari-8 AODS over Beijing–Tianjin–Hebei in China. Remote Sens. 9 (8), 858.
Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., Cribb, M., 2019. Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach. Remote Sens. Environ. 231, 111221.
Who, W.H.O, 2016. Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease.
Wollner, U., Koren, I., Altaratz, O., Remer, L.A., 2014. On the signature of the cirrus twilight zone. Environ. Res. Lett. 9 (9), 094010.
Wright, M.N., Ziegler, A., 2015. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in c++ and r (arXiv preprint arXiv:1508.04409).
Xiao, Q., Wang, Y., Chang, H.H., Meng, X., Geng, G., Lyapustin, A., Liu, Y., 2017. Full-coverage high-resolution daily PM2.5 estimation using MAIAC AOD in the Yangtze River Delta of China. Remote Sens. Environ. 199, 437–446.
Xu, H., Bi, X.-H., Zheng, W.-W., Wu, J.-H., Feng, Y.-C., 2015. Particulate matter mass and chemical component concentrations over four Chinese cities along the western pacific coast. Environ. Sci. Pollut. Res. 22 (3), 1940–1953.
Xu, J.-W., Martin, R., Van Donkelaar, A., Kim, J., Choi, M., Zhang, Q., ... Huang, L., 2015. Estimating ground-level PM2.5 in Eastern China using aerosol optical depth determined from the GOCI Satellite Instrument. Atmospheric Chemistry and Physics 15 (22), 13133–13144.
Xu, Y., Ho, H.C., Wong, M.S., Deng, C., Shi, Y., Chan, T.-C., Knudby, A., 2018. Evaluation of machine learning techniques with multiple remote sensing datasets in estimating monthly concentrations of ground-level PM2.5. Environ. Pollut. 242, 1417–1426.
Yang, J., Hu, M., 2018. Filling the missing data gaps of daily MODIS AOD using spatiotemporal interpolation. Sci. Total Environ. 633, 677–683.
You, W., Zang, Z., Zhang, L., Li, Y., Pan, X., Wang, W., 2016a. National-scale estimates of ground-level PM2.5 concentration in China using geographically weighted regression based on 3 km resolution MODIS AOD. Remote Sens. 8 (3), 184.
You, W., Zang, Z., Zhang, L., Li, Y., Wang, W., 2016b. Estimating national-scale ground-level PM25 concentration in China using geographically weighted regression based on MODIS and MISR AOD. Environ. Sci. Pollut. Res. 23 (9), 8327–8338.
Zang, L., Mao, F., Guo, J., Gong, W., Wang, W., Pan, Z., 2018. Estimating hourly PM1 concentrations from Himawari-8 aerosol optical depth in China. Environ. Pollut. 241, 654–663.
Zeng, Q., Chen, L., Zhu, H., Wang, Z., Wang, X., Zhang, L., ... Zhang, Y., 2018. Satellite-based estimation of hourly PM2.5 concentrations using a vertical-humidity correction method from Himawari-AOD in Hebei. Sensors 18 (10), 3456.
Zhang, R., Di, B., Luo, Y., Deng, X., Grieneisen, M.L., Wang, Z., ... Zhan, Y., 2018. A nonparametric approach to filling gaps in satellite-retrieved aerosol optical depth for estimating ambient PM2.5 levels. Environmental Pollution 243, 998–1007.
Zhang, W., Xu, H., Zheng, F., 2018. Aerosol optical depth retrieval over East Asia using Himawari-8/ahi data. Remote Sens. 10 (1), 137.
Zhao, C., Liu, Z., Wang, Q., Ban, J., Chen, N.X., Li, T., 2019. High-resolution daily AOD estimated to full coverage using the random forest model approach in the Beijing-Tianjin-Hebei region. Atmos. Environ. 203, 70–78.