Continuous mapping of fine particulate matter ($PM_{2.5}$) air quality in East Asia at daily 6x6 km^2 resolution by application of a random forest algorithm to 2011-2019 GOCI geostationary satellite data

Drew C. Pendergrass[1], Daniel J. Jacob[1], Shixian Zhai[1], Jhoon Kim[2,3], Ja-Ho Koo[2], Seoyoung Lee[2], Minah Bae[4], and Soontae Kim[4]

[1]School of Engineering and Applied Sciences, Harvard University, Cambridge, Mass., USA
[2]Department of Atmospheric Sciences, Yonsei University, Seoul, South Korea
[3]Particulate Matter Research Institute, Samsung Advanced Institute of Technology (SAIT), Suwon, South Korea
[4]Department of Environmental and Safety Engineering, Ajou University, Suwon., South Korea

Correspondence to: Drew Pendergrass (pendergrass@g.harvard.edu)

Abstract. We use 2011-2019 aerosol optical depth (AOD) observations from the Geostationary Ocean Color Imager (GOCI) instrument over East Asia to infer 24-h daily surface fine particulate matter ($PM_{2.5}$) concentrations at continuous 6x6 km^2 resolution over eastern China, South Korea, and Japan. This is done with a random forest (RF) algorithm applied to the gap-filled GOCI AODs and other data and trained with $PM_{2.5}$ observations from the three national networks. The predicted 24-h $PM_{2.5}$ concentrations for sites entirely withheld from training in a ten-fold crossvalidation procedure correlate highly with network observations (R^2 = 0.89) with single-value precision of 26-32% depending on country. Prediction of annual mean values has R^2 = 0.96 and single-value precision of 12%. The RF algorithm is only moderately successful for diagnosing local exceedances of the National Ambient Air Quality Standard (NAAQS) because these exceedances are typically within the single-value precisions of the RF, and also because of RF smoothing of extreme $PM_{2.5}$ concentrations. The area-weighted and population-weighted trends of RF $PM_{2.5}$ concentrations for eastern China, South Korea, and Japan show steady 2015-2019 declines consistent with surface networks, but the surface networks in eastern China and South Korea underestimate population exposure. Further examination of RF $PM_{2.5}$ fields for South Korea identifies hotspots where surface network sites were initially lacking and shows 2015-2019 $PM_{2.5}$ decreases across the country except for flat concentrations in the Seoul metropolitan area. Inspection of monthly $PM_{2.5}$ time series in Beijing, Seoul, and Tokyo shows that the RF algorithm successfully captures observed seasonal variations of $PM_{2.5}$ even though AOD and $PM_{2.5}$ often have opposite seasonalities. Application of the RF algorithm to urban pollution episodes in Seoul and Beijing demonstrates high skill in reproducing the observed day-to-day variations in air quality as well as spatial patterns on the 6 km scale. Comparison to a CMAQ simulation for the Korean peninsula demonstrates the value of the continuous RF $PM_{2.5}$ fields for testing air quality models, including over North Korea where they offer a unique resource.

1. Introduction

Exposure to outdoor fine particulate matter ($PM_{2.5}$) is a global public health issue, accounting for 8.9 million deaths in 2015 [*Burnett et. al.,* 2018]. Beyond mortality, short-term exposure to elevated $PM_{2.5}$ levels is associated with numerous adverse health outcomes including increased hospital admissions for respiratory and cardiovascular issues [*Dominici et. al.*, 2006; *Wei et. al.*, 2019]. Long-term exposure is associated with neurodegenerative diseases such as dementia, Alzheimer's disease, and Parkinson's disease [*Kioumourtzoglou et. al.*, 2016]. High spatio-temporal monitoring of $PM_{2.5}$ concentrations to inform population exposure is important for both air quality regulation and epidemiological studies. Ground monitors can provide highly accurate measurements but have limited spatial coverage. Here we show how geostationary satellite observations of aerosol optical depth (AOD) over East Asia from the Geostationary Ocean Color Imager (GOCI) can be used with a random forest (RF) machine learning (ML) algorithm to provide continuous long-term reliable mapping of 24-h $PM_{2.5}$ at 6x6 km^2 spatial resolution.

The potential of satellites for high-resolution monitoring of $PM_{2.5}$ has long been recognized in the public health community [*Liu et al.*, 2004; *van Donkelaar et. al.*, 2006]. Satellites retrieve AOD by backscatter of solar radiation. The MODIS sensors launched in 1999 on the NASA Terra and Aqua satellites have been the main source of AOD data, with global coverage twice a day at up to 1 km resolution [*Remer et. al.*, 2005, 2013; *Lyapustin et. al.*, 2018]. Early approaches to relate AOD observations to surface $PM_{2.5}$ used chemical transport models (CTMs) to estimate local $PM_{2.5}$/AOD ratios [*Liu et al.,* 2004; *van Donkelaar et. al.*, 2006], with more recent studies adding ancillary satellite data on the vertical distribution of aerosol extinction [*Geng et. al.,* 2015*; van Donkelaar et. al.*, 2016; *van Donkelaar et. al.*, 2019]. Other approaches have used $PM_{2.5}$ network data to infer $PM_{2.5}$/AOD ratios [*Wang and Christopher*, 2003], with statistical models based on meteorological and land-use predictor variables to enable spatial extrapolation [*Gupta and Christopher*, 2009; *Liu et. al.*, 2009; *Kloog et. al.*, 2012; 2014]. More recently, non-parametric machine learning models have been developed to predict $PM_{2.5}$ from satellite AOD observations including neural networks [*Li et. al.,* 2017; *Zang et. al.*, 2019] and RFs [*Hu et. al.*, 2017; *Brokamp et. al*., 2018].

Geostationary satellites are now dramatically increasing the capability for mapping of $PM_{2.5}$ from space. The GOCI instrument launched in 2010 by the Korea Aerospace Research Institute (KARI) observes AOD eight times daily at 0.5x0.5 km^2 pixel resolution over eastern China, the Korean peninsula, and Japan [*Choi et. al.*, 2018]. The fine-pixel hourly information is intrinsically valuable and also facilitates cloud clearing [*Remer et al.*, 2012]. GOCI AOD data aggregated to 6x6 km^2 resolution have been used to estimate $PM_{2.5}$ in regional studies for eastern China and South Korea [*Xu et al.*, 2015; *Park et al*., 2019; *She et. al.*, 2020].

AODs cannot be observed under cloudy conditions, and AOD retrievals can also fail for other reasons including snow surfaces. Different methods have been used to fill the gaps and produce continuous data sets. Some studies use CTM AODs when satellite data are missing [*Hu et. al.,* 2017; *Stafoggia et. al.*, 2019]. Others use a statistical interpolation algorithm such as *Kianian et. al.* [2021], who combined a RF with the lattice kriging method to infer missing AOD over the US. Yet others first estimate $PM_{2.5}$ using available AOD observations, then infer missing $PM_{2.5}$ estimates using a separate gap-filling model [*Kloog et. al.*, 2014; *She et. al.,* 2020].

Here we apply a RF algorithm to 2011-2019 GOCI AOD data to construct a continuous dataset of 24-h $PM_{2.5}$ concentrations at 6x6 km^2 resolution for eastern China, South Korea, and Japan trained with the surface network data. Our AOD gap-filling strategy blends CTM information and statistical interpolation with a strategy determined by the RF algorithm. We characterize the error in the RF-produced dataset for both 24-h and annual $PM_{2.5}$ concentrations and demonstrate the ability of the dataset to capture spatial and day-to-day variability on urban scales. We exploit the continuity of the dataset to determine trends of $PM_{2.5}$ air quality in East Asia over the past half decade.

2 Data and methods

2.1 Datasets

GOCI AODs. GOCI is onboard the Korean Communication, Ocean, and Meteorological Satellite (COMS) that was launched by KARI in June 2010 [*Choi et. al.*, 2012; *Choi et. al.*, 2016]. The first ocean color imager placed in geostationary orbit, GOCI covers a 2,500x2,500 km^2 domain centered on the Korean peninsula at 36ºN and 130ºE with 0.5x0.5 km^2 pixels observed every hour from 00:30 to 07:30 UTC. AOD at 550 nm over land is retrieved using the GOCI Yonsei aerosol retrieval (YAER) V2 algorithm at an aggregated 6x6 km^2 spatial resolution and 1 h temporal resolution [*Choi et. al.*, 2018]. Aggregation filters out pixels affected by sunglint or clouds, as well as the darkest 20% and brightest 40% pixels within the 6x6 km^2 scene [*Choi et. al.*, 2018]. We further aggregate the 8x daily measurements of AOD into a daily (8-hr) mean for use in the RF.

Validation of the GOCI YAER V2 AOD with surface measurements from the AERONET surface network shows high correlation ($R = 0.91$), a root mean squared error (RMSE) of 0.16, and a mean bias (MB) of 0.01 [*Choi et. al.*, 2018]. GOCI YAER V2 also reports a Fine Mode Fraction (FMF) and a Multiple Prognostic Expected Error (MPEE) for the AOD but we find that they are not useful in our RF, as discussed later. For comparison, we also calculate a RF trained on the GOCI-AHI fusion AOD product of *Lim et. al.* [2021]. The Advanced Himawari Imager (AHI) instruments onboard the Himawari-8 and -9 geostationary meteorological satellites were launched in October 2014 and November 2016 respectively. AHI has a larger field of view than GOCI but a shorter record.

$PM_{2.5}$ network data. We use hourly $PM_{2.5}$ data from operational air quality networks in eastern China, South Korea, and Japan, and average them over 24 hours and over the 6x6 km^2 GOCI AOD grid to define targets for the RF algorithm. Data for eastern China are from the National Environmental Monitoring Center (https://quotsoft.net/air/) including 443 sites within the GOCI observing domain starting in May 2014 and increasing to 596 sites by 2019. Following *Zhai et. al.* [2019] we remove values with more than 24 consecutive repeats in the hourly timeseries as likely in error. Data for South Korea are from the AirKorea surface network of 123 sites (https://www.airkorea.or.kr/) starting in January 2015 and increasing to 298 sites by 2019. Data for Japan are from 1054 sites reported by the Japanese National Institute for Environmental Studies (NIES) for 2011-2017 (https://www.nies.go.jp/igreen/tj_down.html) and by the real-time Atmospheric Environmental Regional Observation System (AEROS) portal for 2018-2019 (Soramame; http://soramame.taiki.go.jp/DownLoad.php).

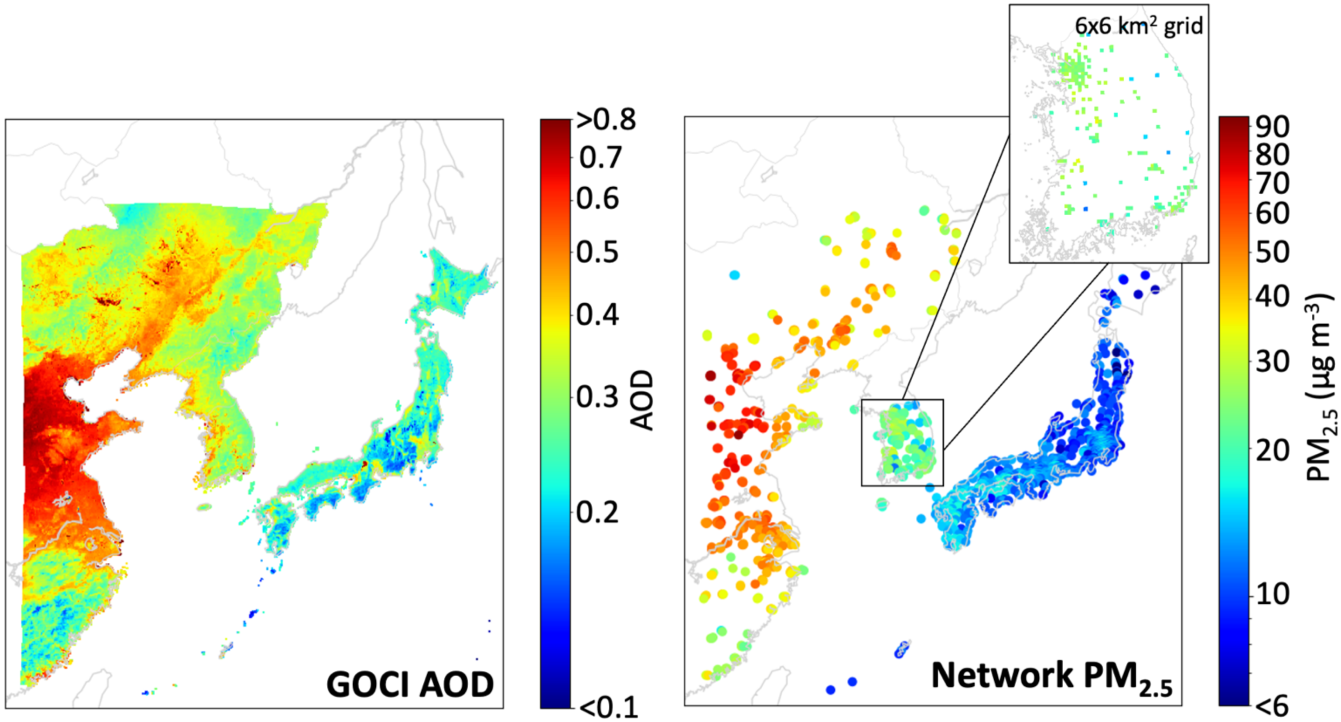

Figure 1: Mean aerosol optical depth (AOD) and $PM_{2.5}$ concentrations over the Geostationary Ocean Color Imager (GOCI) viewing domain, 2011-2019. The left panel shows mean GOCI AOD data on the 6x6 km^2 grid. The right panel shows the mean surface network $PM_{2.5}$ data for eastern China (starting in May 2014), South Korea (starting in January 2015), and Japan, using large data symbols for visibility. Zoomed inset for South Korea shows the surface network observations with symbols corresponding to the 6x6 km^2 grid of the GOCI data. Log scale used for colorbar.

Meteorological and geographical predictor variables. We use hourly meteorological data from the ERA5 global reanalysis, with resolution of 30x30 km^2 [*Hersbach et. al.*, 2020], as input predictor variables for the RF algorithm. For this purpose we aggregate the data to 24-h averages and allocate them to 6x6 km^2 GOCI grid cells by bilinear interpolation. We consider boundary layer height, 2-m air temperature and relative humidity (RH), 10-m meridional and zonal winds, and sea level pressure as potential meteorological predictor variables. We also include as geographical predictor variables latitude, year, day of year (1-366), and nation category (eastern China, South Korea, or Japan). We also considered 2015 population density [*CIESIN*, 2018] as a potential predictor variable but find that it is not useful as discussed in section 3.2.

Figure 1 shows the mean distributions of GOCI AOD and surface network $PM_{2.5}$ for 2011-2019 or for the more limited durations of their records (2014-2019 for eastern China $PM_{2.5}$, 2015-2019 for South Korea $PM_{2.5}$). The $PM_{2.5}$ networks are extensive but coverage is nevertheless sparse and often limited to large urban areas, as illustrated by the zoomed inset for South Korea. We find that only 1.0% of GOCI 6x6 km^2 grid cells have $PM_{2.5}$ observations in eastern China, 7.4% in South Korea, and 7.9%

in Japan. This geographic limitation in the $PM_{2.5}$ networks emphasizes the value of continuous coverage from the AOD data.

2.2 AOD gap-filling

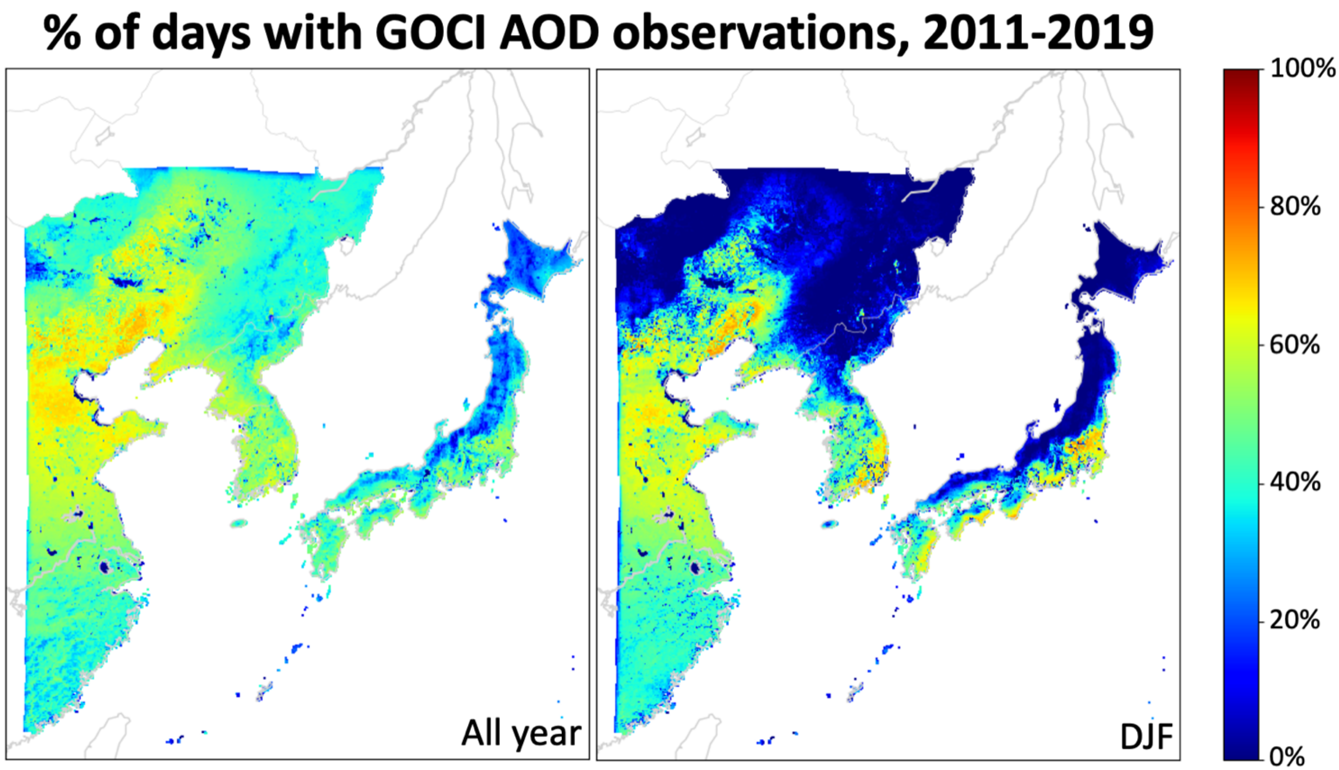

Figure 2: Percentage of days in 2011-2019 with at least one successful hourly retrieval of AOD on the 6x6 km^2 grid. The left panel shows year-round statistics while the right panel shows winter months (DJF) only.

Figure 2 shows the percentage of days with at least one successful hourly GOCI AOD retrieval on the 6x6 km^2 retrieval grid. There are substantial gaps in the record, mostly reflecting clouds and also snow cover in winter [*Choi et. al.*, 2018]. We seek to fill in these gaps to produce a continuous daily data set while accounting for the associated errors. We fuse two strategies according to the availability of nearby AOD retrievals: an inverse distance weighted (IDW) interpolation AOD_{IDW} of nearby retrievals [*Shepard*, 1968] and a bias-corrected monthly AOD_{GC} from the GEOS-Chem CTM:

$$AOD = \alpha\, AOD_{IDW} + (1-\alpha) AOD_{GC} \quad (1)$$

where α is a weighting factor that depends on the distance from nearest retrievals. GEOS-Chem is a widely used CTM for inferring $PM_{2.5}$ from satellite AOD data [*Liu et al.,* 2004; *van Donkelaar et. al.*, 2006; 2016; 2019; *Geng et. al.,* 2015]. Here we use scaled monthly mean GEOS-Chem AODs from a simulation by *Zhai et al.* [2021] for 2016 in East Asia with 0.5°x 0.625° resolution. That simulation reported a low mean bias relative to AERONET; we correct this for each year in the study period by using annual mean GOCI AODs on the 6x6 km^2 grid. In this way we obtain a spatial distribution of monthly mean AOD_{GC} values for 2011-2019 for use in equation (1).

We calculate the weighting factors α used in Equation (1) via the Gaspari-Cohn function, a fifth-order piecewise polynomial with a radial argument r [*Gaspari and Cohn*, 1999]. The Gaspari-Cohn function resembles a Gaussian distribution but with compact support, taking on a maximum value of 1 for $r = 0$ and a minimum value of 0 for $r \geq 2$. We define $r = l/c$ for a given 6x6 km^2 grid cell and day to be the distance l from the midpoint of the grid cell to that of the nearest observed grid cell, normalized by a spatial correlation length scale c determined from available AOD observations in and around that grid cell. We find that the value of c ranges from 110 km to 170 km over our domain.

2.3 Random forest algorithm

Table 1 lists the predictor variables included in the RF to infer 24-h $PM_{2.5}$ as dependent variable. RF is an ensemble machine learning method where many individual decision trees are fit to the training data and vote on an output value, with the average value taken as best estimate [*Breiman*, 2001].

Table 1. Random Forest predictor variables for 24-h $PM_{2.5}$[a]

GOCI gap-filled AOD observations[b]
8-h average AOD at 550 nm wavelength
α from Equation 1
Meteorology[c]
Boundary layer height (m)
10-m meridional wind (m s^{-1})
10-m zonal wind (m s^{-1})
2-m temperature (K)
2-m relative humidity[d] (%)
Sea-level pressure (Pa)
Metadata
Country dummy variables[e]
Latitude
Day of year
Year

[a]The RF algorithm predicts continuous 24-h $PM_{2.5}$ on a 6x6 km^2 grid for eastern China, South Korea, and Japan after training with $PM_{2.5}$ surface network data.
[b]8-hr average 550 nm AODs on the 6x6 km^2 grid retrieved with the YAER v2 algorithm [Choi et al., 2018]
[c] ECMWF ERA5 fields [*Hersbach et. al.,* 2020] at 30x30 km^2 spatial resolution and hourly temporal resolution, interpolated bilinearly to the GOCI grid and averaged over 24 hours.
[d] Estimated from temperature and dewpoint using the August-Roche-Magnus approximation [*Alduchov and Eskridge*, 1996].
[e]Three variables that, for each of eastern China, South Korea, and Japan, has value 1 if a grid cell is within those national borders and 0 otherwise.

Decision trees are fit recursively to the predictor variable. Suppose we have a collection of N data elements $i \in [1, N]$, denoted x_i, each composed of m predictor variables ($x_i \in \mathbb{R}^m$), and a corresponding list of N labels y_i that we would like to learn. In our case y_i denotes the observed $PM_{2.5}$ concentrations from the surface networks averaged on the 6x6 km^2 grid, and N denotes the number of these observations. The algorithm works by splitting the data into left and right subsets L and R at an optimum split point determined from the predictor variables in x_i [*Pedregosa et. al.*, 2011]. The optimum split point is defined as the one that minimizes the impurity G,

$$G(L,R) = \beta \cdot \mathrm{MSE}(L) + (1-\beta) \cdot \mathrm{MSE}(R) \tag{2}$$

where β represents the fraction of data in the subset L and MSE represents the mean squared error of each of the subsets,

$$\mathrm{MSE}(X) = \frac{1}{n}\sum_i (y_i - \bar{y})^2 \tag{3}$$

where $\bar{y}$ is the mean of the target labels within a given subset X and n is the number of elements in that subset. From there the same algorithm is recursively applied to the left and right subsets L and R until the tree is grown. We follow the advice of *Hastie et. al.* [2009] and grow trees until the data are fully classified (each leaf contains only one value).

Due to the recursive training structure, decision trees are sensitive to the data on which they are trained, because a change in one split point changes the composition of all its child nodes. Individual decision trees thus have high error variance but no inherent bias. It follows that averaging many individual and uncorrelated trees should yield a low variance, low bias prediction. We construct 200 trees in parallel and reduce correlation between them through a bagging procedure: for each of the 200 decision trees in the RF, sample the input data with replacement to form a new dataset of the same dimensions and then grow a decision tree from this bootstrapped data [*Breiman*, 2001]. Because of the high input sensitivity, a wide variety of decorrelated trees are grown. The predictions of each individual tree are averaged to yield the prediction of the RF. We fit our RF using the RandomForestRegression class in the Python module Scikit-learn [*Pedregosa et. al.*, 2011]. We attempted to further decorrelate trees by following *Breiman* [2001] and calculating split points of each individual tree using only a random subset of the *m* predictor variables; however, a sensitivity test we performed showed only minor differences with the base case and therefore we follow *Guerts et. al.* [2006] in considering all predictor variables in the training process.

We evaluate how the RF generalizes to predictions for the full 6x6 km^2 domain via a 10-fold crossvalidation. For each fold of the crossvalidation, we leave out 10% of $PM_{2.5}$ network sites (averaged on the 6x6 km^2 grid if needed) from each country. These 10% represent the test set; because we perform the validation ten times, each grid cell is in the test set exactly once. We compare predicted $PM_{2.5}$ to withheld observed $PM_{2.5}$ using four metrics: root mean square error (RMSE); the RMSE divided by mean observed $PM_{2.5}$ (relative RMSE, or RRMSE); the coefficient of variation (R^2); and the mean bias computed by averaging the difference between predicted and observed $PM_{2.5}$ (MB).

An outcome of interest is the ability of our predictions to capture exceedances of National Ambient Air Quality Standards (NAAQS). We categorize each prediction within the test sets into one of four classes: true positives (TP) where both predicted and observed $PM_{2.5}$ exceed the NAAQS threshold; true negatives (TN) where neither exceed the threshold; false positives (FP) where an exceedance is predicted but not observed; and false negatives (FN) where an exceedance is observed but not predicted [*Brasseur and Jacob*, 2017; *Cusworth et. al.*, 2018]. We use these classes to compute three overall prediction grades. The first, percent of detection (POD), gives the fraction of observed exceedances that were successfully predicted:

$$\text{POD} = \frac{\Sigma\,\text{TP}}{\Sigma\,\text{TP} + \Sigma\,\text{FN}} \quad \textbf{(4)}$$

The second, false alarm ratio (FAR), gives the fraction of predicted exceedances that did not occur:

$$\text{FAR} = \frac{\Sigma\,\text{FP}}{\Sigma\,\text{TP} + \Sigma\,\text{FP}} \quad \textbf{(5)}$$

The third, equitable threat score (ETS), compares how well the prediction does relative to random chance:

$$\text{ETS} = \frac{\Sigma\,\text{TP} - \beta}{\Sigma\,\text{TP} + \Sigma\,\text{FP} + \Sigma\,\text{FN} - \beta} \quad \textbf{(6)}$$

where β is the number of true positives obtained by random chance,

$$\beta = \frac{(\Sigma\,\text{TP} + \Sigma\,\text{FP}) \cdot (\Sigma\,\text{TP} + \Sigma\,\text{FN})}{\Sigma\,\text{TP} + \Sigma\,\text{TN} + \Sigma\,\text{FP} + \Sigma\,\text{FN}} \quad \textbf{(7)}$$

ETS is 1 for perfect prediction skill and 0 for no better or worse than chance.

Predictor variable selection is an important task in implementing a RF, as the addition of non-informative variables can decrease performance. Unlike linear regression which can naturally ignore unhelpful predictors, irrelevant data can by chance aid in minimizing impurity *G* at some stage in the optimization process making all subsequent splits suboptimal. The six meteorological variables given in **Table 1** are standard in AOD/$PM_{2.5}$ prediction [e.g. *Kloog et. al.,* 2014; *Li et. al.*, 2017], while the four spatio-temporal variables (location dummies, latitude, year, and day of year) and the retrieval gapfilling parameter α proved to be informative in sensitivity tests. In addition to the predictor variables in **Table 1**, we considered as additional variables the population density, the GOCI fine mode fraction (FMF), and the GOCI multiple prognostic expected error (MPEE), but we found that they worsened accuracy of the fit and so we did not retain them. Because population density worsened the fit we did not include other spatially varying but temporally fixed land-use variables such as road data. We also compared RFs trained on GOCI AOD and on GOCI-AHI fused AOD and found no significant difference in the fitting of $PM_{2.5}$. We therefore use the GOCI AOD product because of its longer record.

3 Results and discussion

3.1 Accuracy and precision of RF predictions

Figure 3 shows scatterplots, color-coded by count, comparing surface observations of 24-h and annual mean $PM_{2.5}$ to the predicted values in grid cells whose records are entirely withheld from training in the crossvalidation procedure. Predicted values for the annual mean are obtained by averaging the 24-h predictions. **Table 2** gives comprehensive statistics for East Asia and for each country. The 24-h predictions for East Asia have a negligible mean bias of 0.23 μg m^{-3} (annual, 0.22 μg m^{-3}), though the RF underpredicts $PM_{2.5}$ at the high tail of the distribution; we will return to that issue later in the context of NAAQS exceedances. Root mean square error (RMSE) between observed and predicted 24-h $PM_{2.5}$ is 8.8 μg m^{-3} s (annual, 3.3 μg m^{-3}) corresponding to a relative RMSE (RRMSE) of 37% (annual, 14%), as defined in section 2.3. The prediction captures 89% of the observed 24-h variance (R^2 = 0.89) and 96% of annual (R^2 = 0.96). These results compare favorably to previously reported inferences of 24-h and annual $PM_{2.5}$ at 1-10 km resolution from satellite data over China [*Hu et. al.*, 2019; *Xue et. al.*, 2019]. R^2 for annual mean $PM_{2.5}$ in South Korea is relatively low (0.41), which can be explained by the weak dynamic range of observed annual $PM_{2.5}$ in the country (**Figure 1**), as will be discussed later in this section. Our gap-filling strategy does not introduce bias for days without GOCI observations (and with AOD inferred instead from equation (1)), as the evaluation statistics for those days are similar to the whole population.

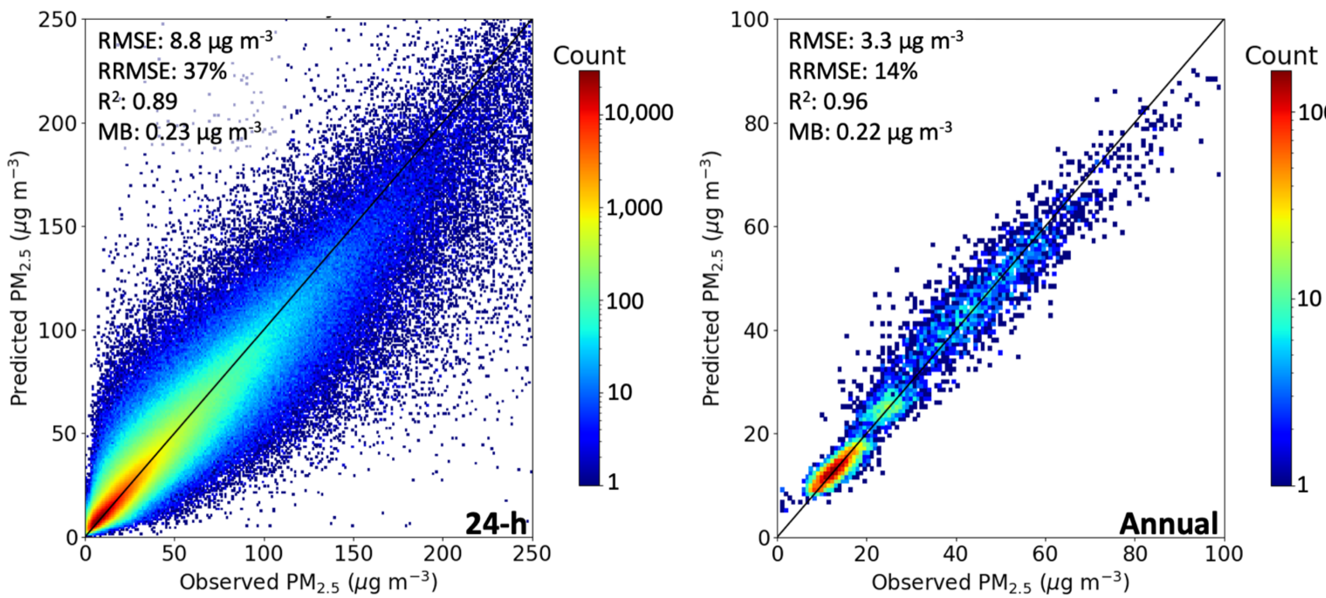

Figure 3: Ability of the random forest algorithm to predict 24-h and annual mean $PM_{2.5}$ in East Asia. Scatterplots depict the relationship between predicted and observed $PM_{2.5}$ at network sites withheld from training in the crossvalidation. The plots are two-dimensional histograms where pixel color corresponds to the count of observation/prediction correspondences within the corresponding bin on a logged scale. The identity line is plotted in black. For annual mean $PM_{2.5}$, grid cells with fewer than 80% of $PM_{2.5}$ observation days in a given year are removed to avoid biasing the average.

Table 2. Error statistics for fitting of $PM_{2.5}$ data by the RF algorithm[a]

	RMSE (μg m^{-3})	RRMSE	R^2	MB (μg m^{-3})
24-h $PM_{2.5}$				
Overall	8.8	37%	0.89	0.23
Eastern China	15	32%	0.85	0.49
South Korea	6.4	26%	0.82	0.16
Japan	3.6	27%	0.79	0.12
Annual $PM_{2.5}$				
Overall	3.3	14%	0.96	0.22
Eastern China	5.6	12%	0.86	0.53
South Korea	2.9	12%	0.41	0.24
Japan	1.6	12%	0.70	0.094

[a]Comparison statistics between predicted and observed $PM_{2.5}$ are for the grid cells in each of eastern China, South Korea, and Japan completely withheld from the RF training process in the crossvalidation procedure. Statistics shown are for root-mean-square error (RMSE), relative RMSE (RRMSE), coefficient of variation (R^2), and mean bias (MB).

One potential application of $PM_{2.5}$ monitoring from space would be to diagnose exceedances of national ambient air quality standards (NAAQS) at locations without network sites. **Table 3** shows the NAAQS for 24-h and annual $PM_{2.5}$ for the three countries and the ability of the RF algorithm to diagnose NAAQS exceedances in grid cells excluded from the training process in the crossvalidation procedure. 24-h exceedances correspond to the high tails of the distributions but annual exceedances are much more widespread. The POD column shows percent of true positives successfully detected, while the FAR shows the rate of false positives (defined in section 2.3). POD for 24-h exceedances ranges from 47%-78% by country (FAR: 16%-21%). PODs are higher for annual exceedances but that reflects the higher observed frequency of these exceedances. The ETS values ranging from 0.43-0.63 indicate that the model captures exceedances with much better skill than random guessing.

Table 3. Ability of the RF algorithm to diagnose exceedances of air quality standards[a]

	NAAQS (μg m^{-3})[b]	Exceedance frequency[c]		POD[d]	FAR[e]	ETS[f]
		Observed	RF			
24-h $PM_{2.5}$						
Eastern China	75	16%	15%	78%	16%	0.63
South Korea (old NAAQS)	50	5.9%	4.2%	57%	21%	0.47
South Korea (new NAAQS)	35	19%	17%	73%	20%	0.55
Japan	35	1.6%	0.91%	47%	17%	0.43
Annual $PM_{2.5}$						
Eastern China	35	77%	83%	97%	9.2%	0.54
South Korea (old NAAQS)	25	40%	44%	67%	39%	0.23
South Korea (new NAAQS)	15	100%	100%	100%	0%	NA
Japan	15	24%	20%	68%	20%	0.49

[a] Calculated using sites withheld from training in the crossvalidation procedure.

[b] National Ambient Air Quality Standards, specific to each country. We show results for the class 2 NAAQS in eastern China and for both pre-2018 ('old') and post-2018 ('new') NAAQS for South Korea because all observed grid cells exceed the new annual NAAQS of 15 μg m^{-3}.
[c] Percentage of site-days (24-h standard) or site years (annual standard) exceeding the NAAQS.
[d] Percent of detection (POD) defined as the percentage of exceedances successfully detected.
[e] False alarm ratio (FAR) defined as the percentage of predicted exceedances that did not occur.
[f] Equitable threat score (ETS) defined as the ability of the RF to predict exceedances beyond random chance.

The main difficulty for the RF algorithm to predict NAAQS exceedances is that many of those exceedances fall within the precision of individual predictions. This is illustrated in **Figure 4** with the cumulative probability density function (pdf) of the 24-h and annual mean $PM_{2.5}$ concentrations in eastern China, South Korea, and Japan, representing the same withheld data from the crossvalidation as in **Tables 2** and **3**. The 24-h RRMSE of 26-32% depending on country (**Table 2**) is shown as the grey envelope and is relatively flat across the distribution. Prediction of NAAQS exceedances within that uncertainty envelope is limited by the precision of the algorithm. All of the 24-h exceedances in Japan are within that envelope, as are most of the exceedances in eastern China and Korea. China has the largest fraction of exceedances beyond the RRMSE of the RF algorithm and therefore the best prediction success. An additional though smaller cause of bias is that the RF algorithm underestimates the high tail of the pdf, as is apparent in **Figure 4**, which explains in particular why we achieve a better FAR than POD for 24-h $PM_{2.5}$ in South Korea and Japan. Our worst NAAQS prediction performance is for annual $PM_{2.5}$ in South Korea for the old 25 μg m^{-3} standard, because most of the distribution is within the RRMSE envelope. Additionally, the already small dynamic range of observed annual $PM_{2.5}$ (black dots) is underestimated by the RF (blue dots). These culminate in an RF estimate with good RMSE but low R^2.

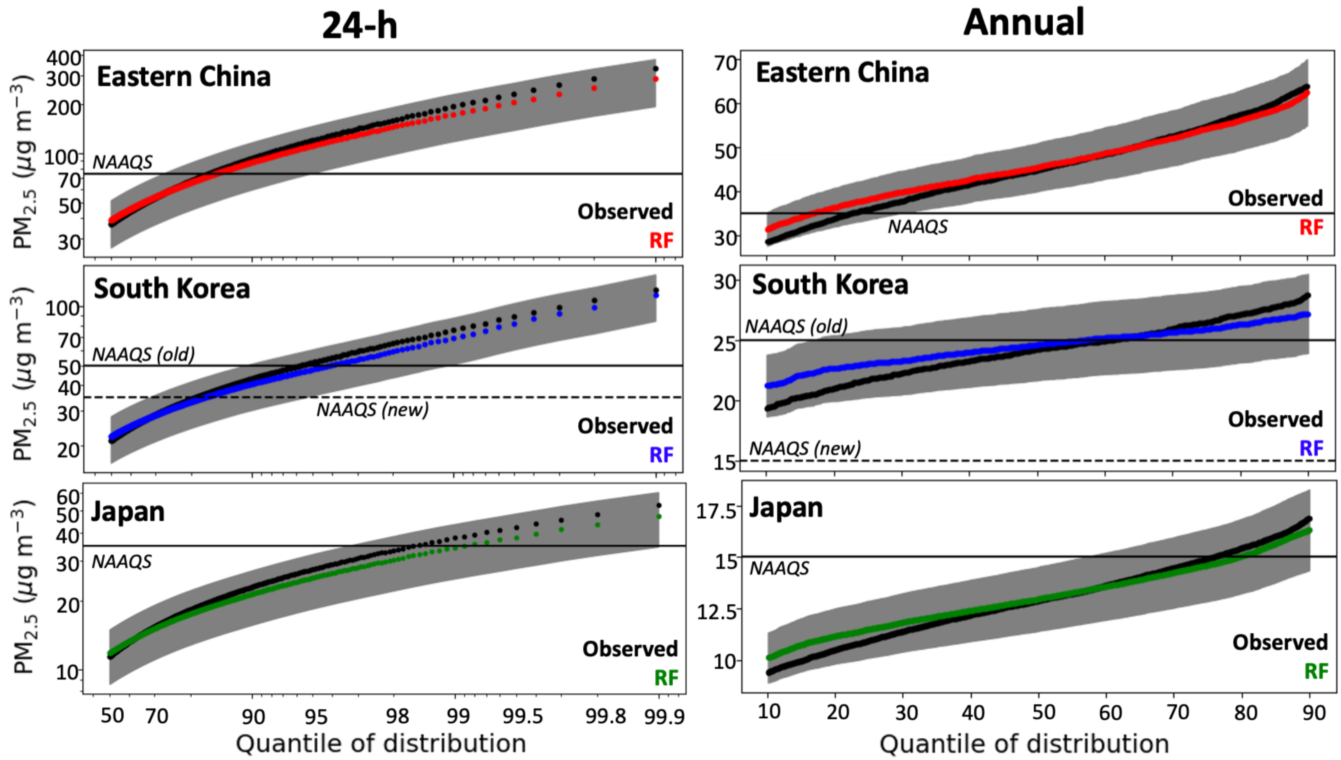

Figure 4: Cumulative probability density functions (pdfs) of 24-h and annual mean $PM_{2.5}$ concentrations in Eastern China, South Korea, and Japan. Observations (black) are compared to RF predictions (colored) taken from the crossvalidation. The grey envelope represents the relative root mean square error (RRMSE) of the RF algorithm as given in Table 2, measuring the predictive capability of the algorithm for individual events. The NAAQS for each country is shown as the horizontal line, with both the pre-2018 and post-2018 NAAQS shown for South Korea. Left panel scales are log-log while right-panel scales are linear. y-axis scales vary for the different countries.

We experimented with several modifications to the RF algorithm to improve prediction of NAAQS exceedances but with no success. These tests included training separate RFs for each of the three countries; training annual $PM_{2.5}$ predictions on annual (rather than 24-h) $PM_{2.5}$ data; directly predicting NAAQS exceedances by setting the learned label to be true if a day (year) is above the 24-h (annual) NAAQS for a given country; and applying different weights to the data so that the high tail is oversampled in the training process. None of these tests yielded significant improvements. Smoothing of the tails in RFs is a well-recognized problem [*Zhang and Lu*, 2012]. Following *Zhang and Lu* [2012] we attempted to train RFs to predict and correct the residuals but found this to be ineffective. Part of this tail smoothing could also result from the underlying GOCI AOD land product, which has a negative bias (-0.02) for high AODs and a positive bias (+0.02) for low AODs [*Choi et. al.*, 2018].

3.2 $PM_{2.5}$ temporal trends and spatial distributions

Figure 5 shows long-term trends of annual $PM_{2.5}$ for each country, as measured by the $PM_{2.5}$ network and as inferred from our RF algorithm for both areal and population-weighted means. We do not include RF $PM_{2.5}$ for years before the networks became available (and hence when the RF could be trained) because of concern over extrapolation bias. The $PM_{2.5}$ networks show decreasing trends in all three countries and these trends are consistent with the RF $PM_{2.5}$ for both areal and population-weighted means, demonstrating that the trends reported by the $PM_{2.5}$ networks are representative of the countries. However, the $PM_{2.5}$ networks in eastern China and South Korea underestimate the population-weighted means. Trends in South Korea and eastern China become flat between 2018 and 2019 (with a slight population-weighted increase in South Korea). This could possibly reflect interannual meteorological variability [*Zhai et al.*, 2019; *Koo et. al.*, 2020], but also an increase in oxidants producing secondary aerosol [*Huang et. al.*, 2021].

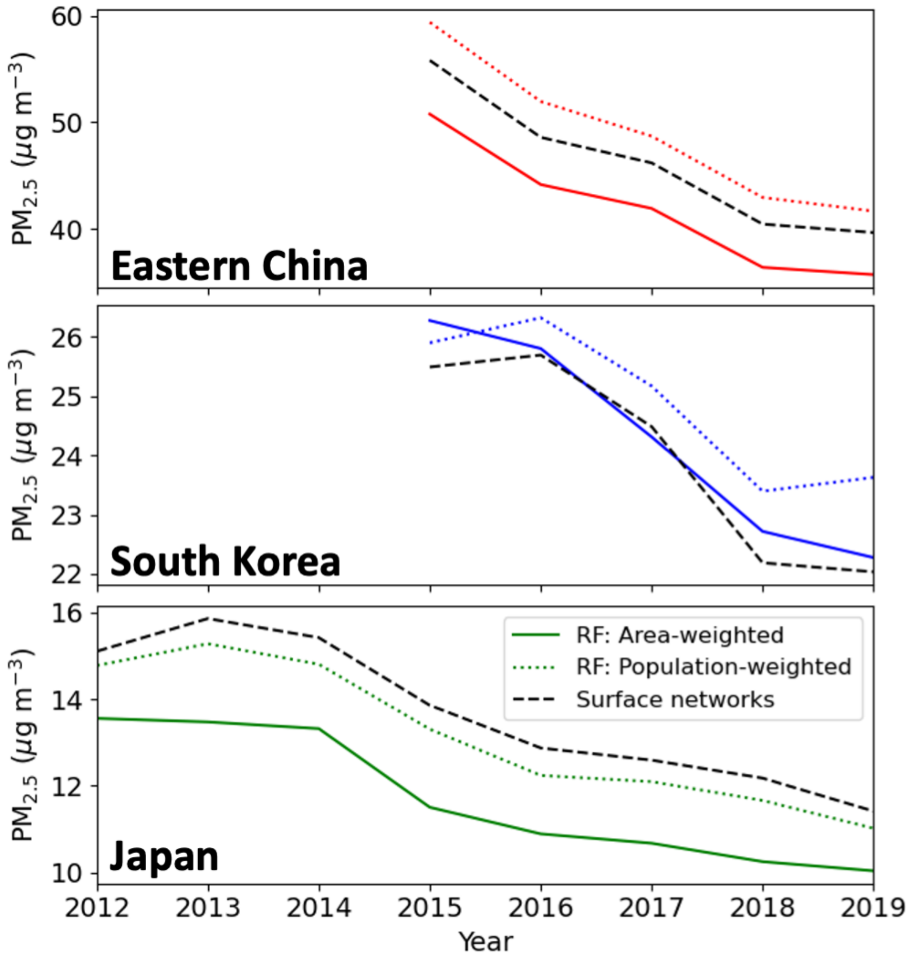

Figure 5: Trends in annual mean $PM_{2.5}$ concentrations for eastern China, South Korea, and Japan. Trends determined from the national surface $PM_{2.5}$ networks averaged over 6x6 km^2 grid cells, requiring at least 80% of data for all years plotted, are compared to trends inferred by the random forest (RF) algorithm with continuous temporal and spatial coverage on the 6x6 km^2 grid and weighted either by area or by population. Here we use an RF trained on all the data. Gridded population data are from CIESIN [2018]. The national $PM_{2.5}$ networks include 413 continuously observed grid cells in eastern China, 74 in South Korea, and 307 in Japan. Trends are initialized at the onset of the surface network for complete years of data; due to the unavailability of the early months of the year, 2011 is discarded for Japan and 2014 for eastern China.

Figure 6 shows the changes in annual mean $PM_{2.5}$ concentrations over South Korea between 2015 and 2019, as observed from the national network and as predicted by the RF. We focus on South Korea for discussion because it shows the advantages of satellite-based $PM_{2.5}$ in a region that already has good coverage. Continuous mapping from the RF algorithm enabled by the GOCI AODs adds enormous coverage to the sparse surface observations, including detection of $PM_{2.5}$ hotspots missing from the network such as the Iksan region on the west coast in 2015 that was subsequently added to the network by 2019.

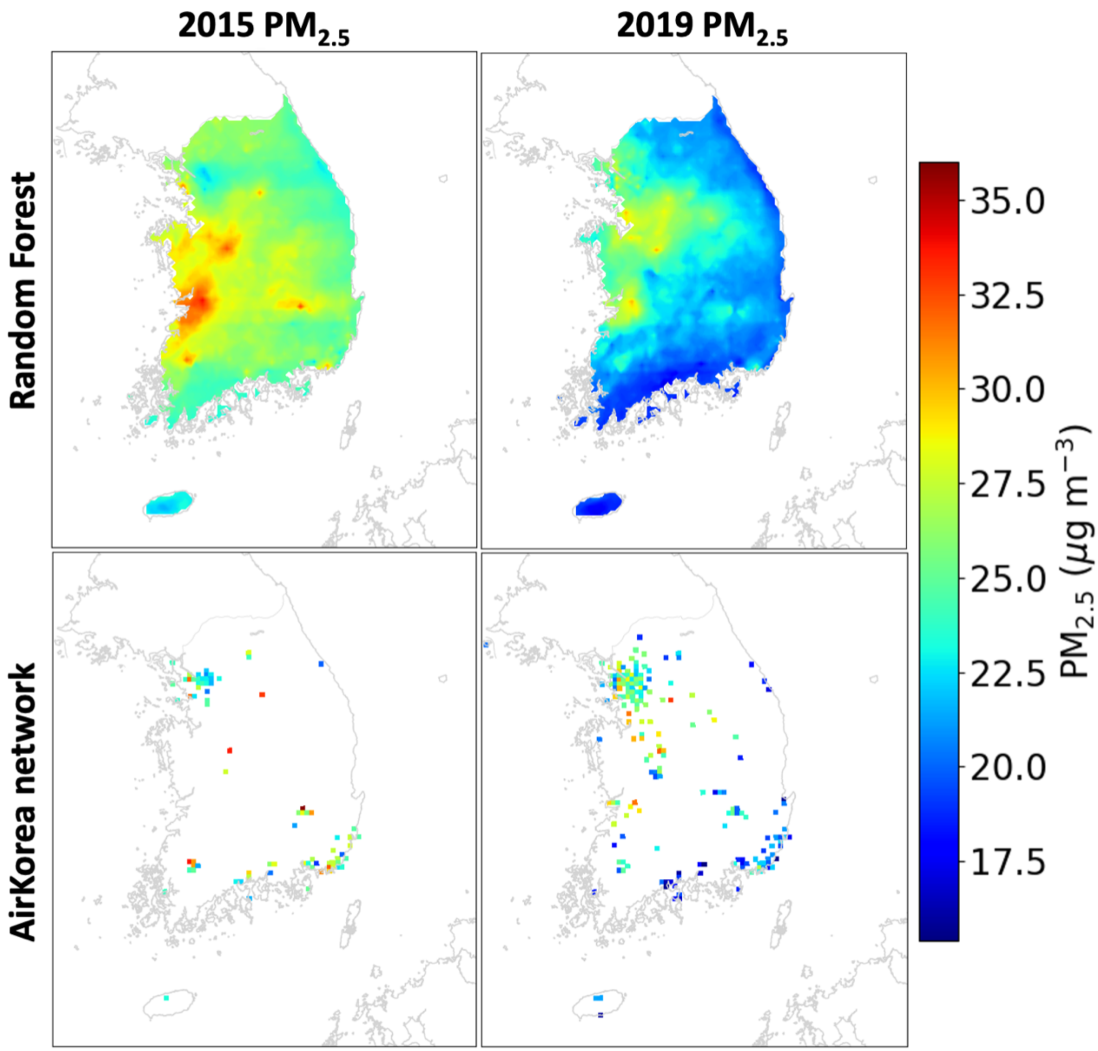

Figure 6: Annual mean $PM_{2.5}$ concentrations in South Korea in 2015 and 2019. RF predictions (top) inferred from an RF trained on all available data are compared to AirKorea network observations (bottom). Network observations are shown only if at least 80% of the year was observed.

Figure 7 depicts the relative 2015-2019 trends of $PM_{2.5}$ concentrations in South Korea derived from a linear regression applied to the annual RF $PM_{2.5}$ in each 6x6 km^2 grid cell. Such a spatially resolved trend analysis is uniquely enabled by the GOCI coverage. We find decreases across the country except in the Seoul Metropolitan area which mostly shows no significant trend except for a few pixels in Incheon. These results are consistent with the spatial patterns calculated from AirKorea data by *Yeo and Kim* [2019], who found 2015-2018 decreases in Incheon but not Seoul or the surrounding Gyeonggi province. Despite the insignificant changes in Seoul, substantial $PM_{2.5}$ decreases are found over other large urban areas including Busan, Ulsan, Daegu, and Gwangju. The three rapidly decreasing spots on the southern coast are Gwangyang, Sacheon, and Changwon, which house industrial complexes related to the South Korean shipbuilding industry that has recently declined [*Jung-a* 2016].

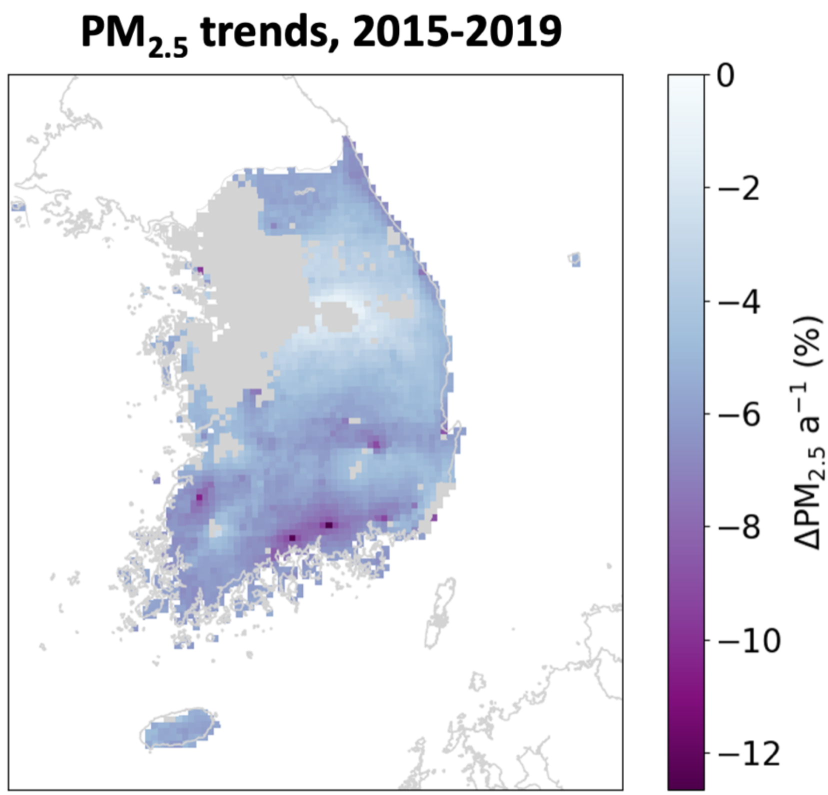

Figure 7: 2015-2019 trends per year in $PM_{2.5}$ concentrations across South Korea. The trends are obtained by ordinary linear regression of the annual mean RF $PM_{2.5}$ in each 6x6 km^2 grid cell with significant regression slopes ($\boldsymbol{p < 0.05}$), where the RF is trained on all the available data. Grid cells with insignificant trends are plotted in gray.

AOD and $PM_{2.5}$ in East Asia tend to have opposite seasonalities driven by boundary layer depth and RH [*Zhai et al.*, 2021]. **Figure 8** compares predicted and observed monthly mean $PM_{2.5}$ in the Beijing, Seoul, and Tokyo metropolitan areas, with predictions coming from withheld data in the 10-fold crossvalidation. Correspondence between modelled and observed $PM_{2.5}$ may be tighter than the nationwide annual means plotted in **Figure 5** because these urban areas are well-observed. We see that the RF algorithm fully captures the observed seasonality in $PM_{2.5}$, although some observed monthly spikes are underestimated. The Figure illustrates the lack of trend in the Seoul Metropolitan Area over

2015-2019 but also shows that winter and summer $PM_{2.5}$ in the region have opposite and roughly equal trends, with winter growing more polluted while summers become cleaner.

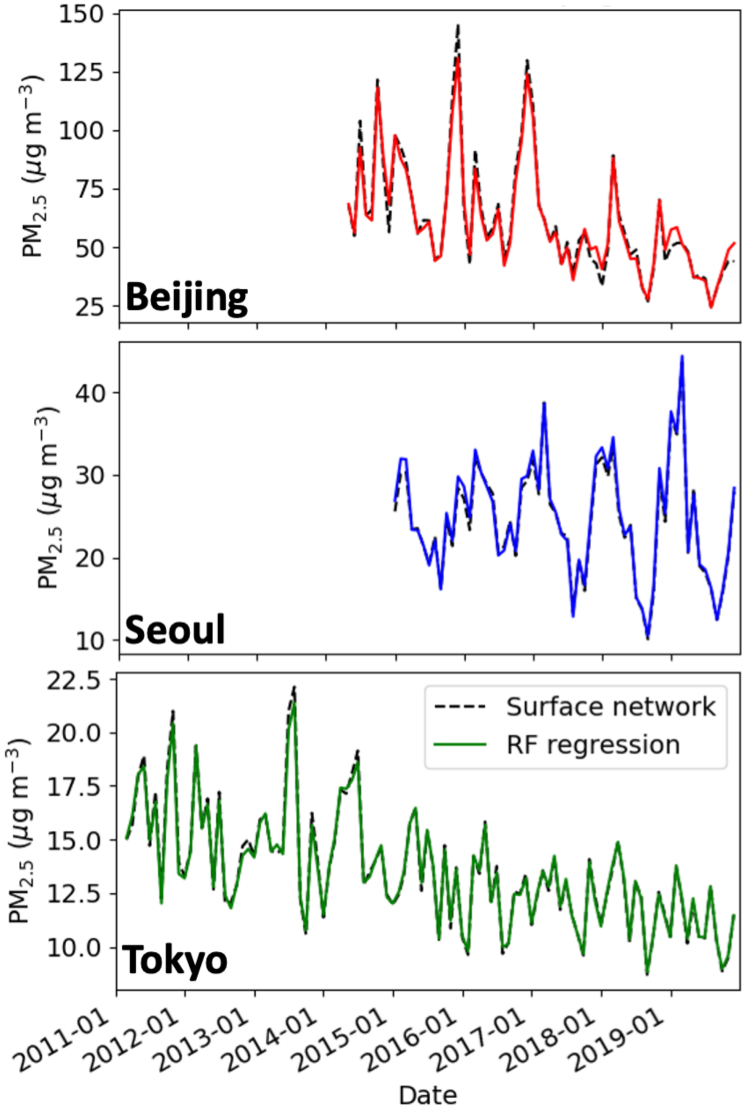

Figure 8: Monthly $PM_{2.5}$ concentrations in the Beijing Seoul and Tokyo metropolitan areas. Predictions from the RF algorithm for totally withheld sites in the crossvalidation are compared to network observations. Beijing is defined by the namesake province boundary, Seoul by the Seoul and Incheon boundaries, and Tokyo as Ibaraki, Saitama, Chiba, Tokyo, Kanagawa, and Yamanashi prefectures.

3.3 Urban-scale pollution events

We examine here the ability of the RF algorithm to capture the spatial and temporal variability of $PM_{2.5}$ pollution events on urban scales. **Figure 9** shows a predicted map of $PM_{2.5}$ — produced by a RF trained on all the data, with observed $PM_{2.5}$ overlaid — across the Seoul metropolitan area on May 24-29, 2016 corresponding to a severe pollution event sampled during the KORUS-AQ field campaign

[*Crawford et. al.*, 2021]. The dense $PM_{2.5}$ network for Seoul shows large variability at the sub 6x6 km^2 scale that the AOD data and thus this RF $PM_{2.5}$ product cannot resolve. However, the RF algorithm capture most of the variability in observed 24-h $PM_{2.5}$ concentrations aggregated on the 6x6 km^2 grid ($R^2 = 0.74$). The RF also captures successfully the day-to-day variability during the event.

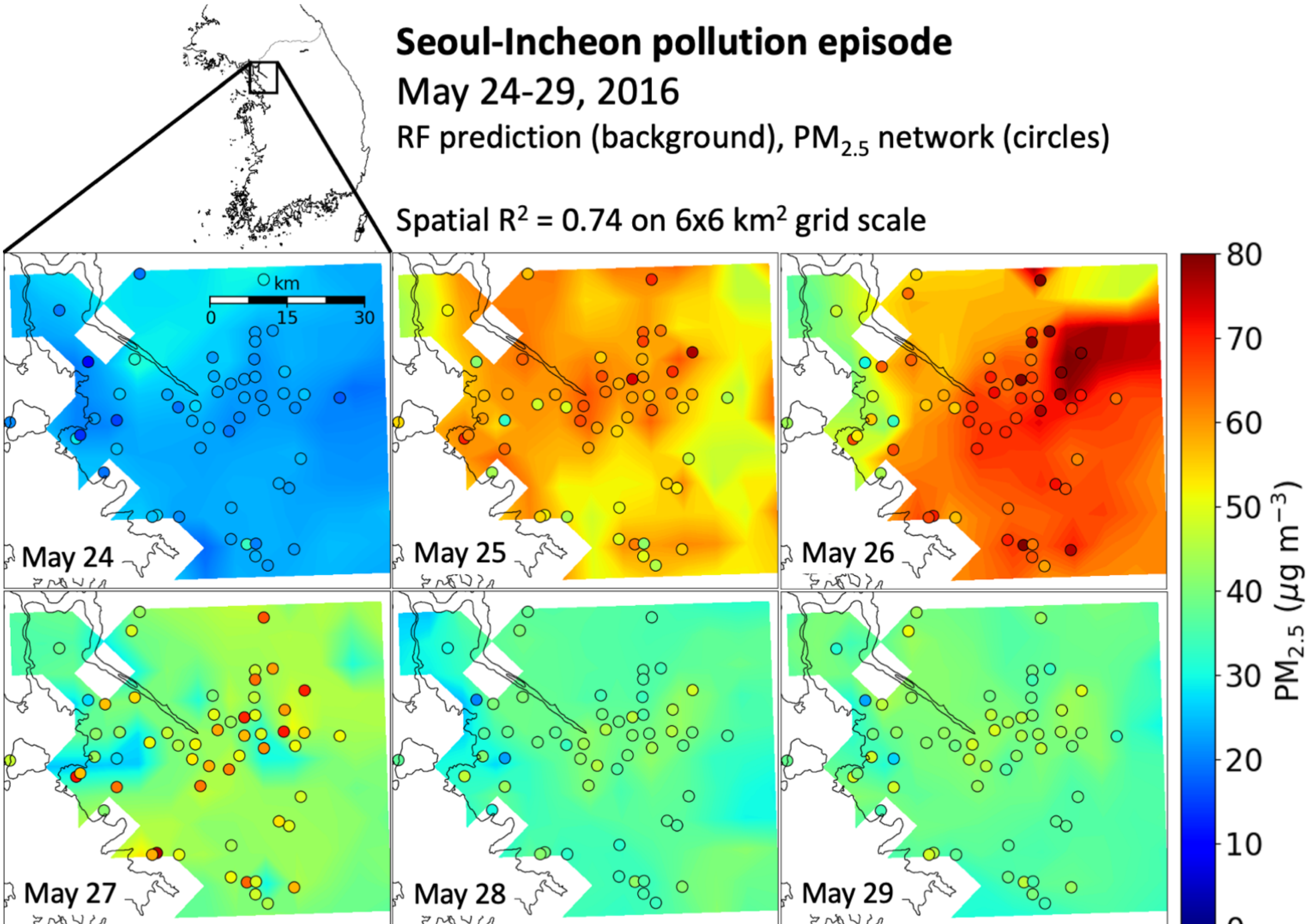

Figure 9: 24-h $PM_{2.5}$ concentrations during a pollution event in Seoul-Incheon (May 24-29, 2016). Predictions from the RF algorithm (background, on 6x6 km^2 grid scale) trained on all available data are compared to observations from the AirKorea surface network (circles).

Figure 10 shows an additional test of the RF algorithm with one of the most severe pollution events in the record, the December 16-21, 2016 Beijing winter haze episode. 24-h $PM_{2.5}$ concentrations exceeded 400 µg m^{-3} at some of the network sites. While there is a tight correspondence between the RF and observed 24-h $PM_{2.5}$ for Beijing grid cells (R^2 range: 0.74-0.99), the observations are on average 20 µg m^{-3} higher than the RF $PM_{2.5}$. The difference is most pronounced at the December 21 concentration peak which has mean observed value 396 µg m^{-3} to the predicted 348 µg m^{-3}. This reflects the RF smoothing of the high tail of the distribution as previously illustrated in **Figure 4**.

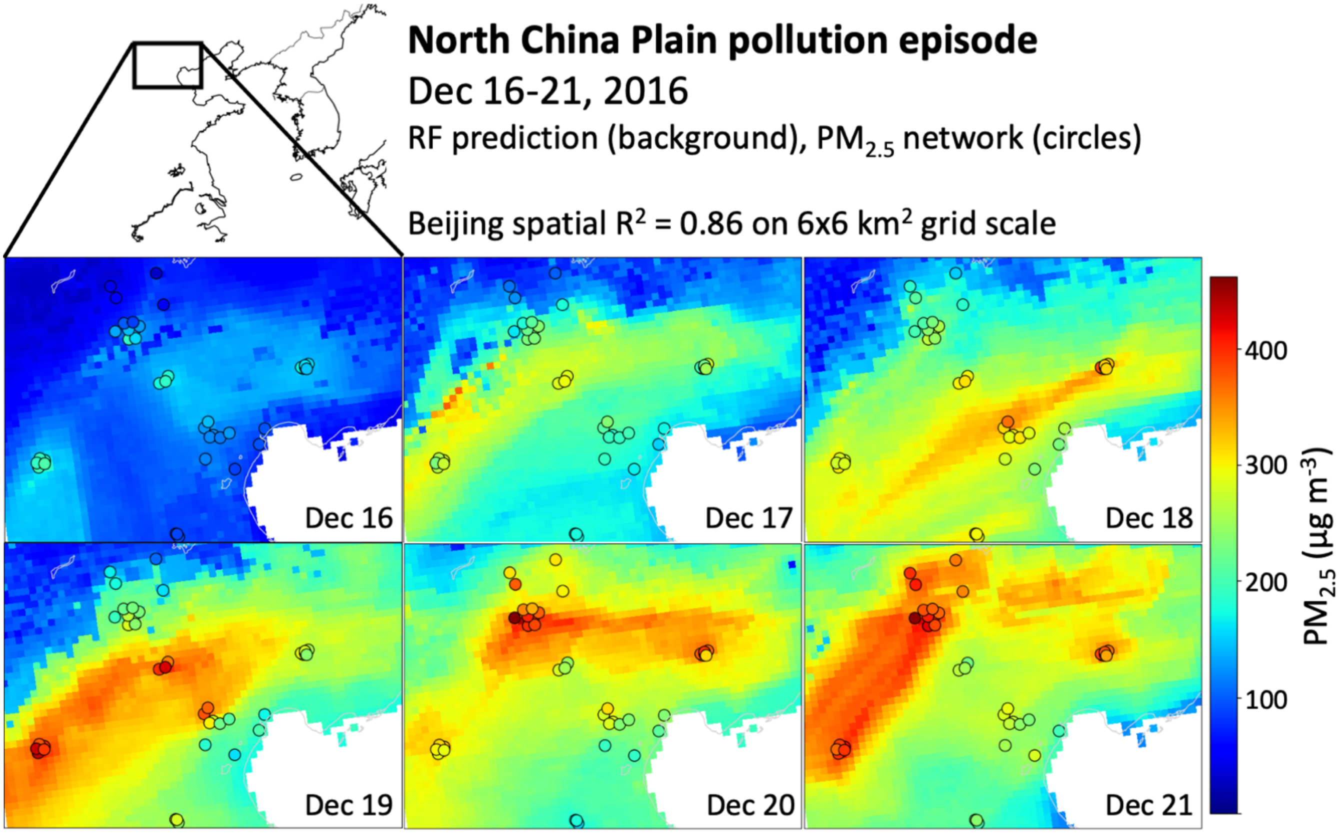

Figure 10: Same as Figure 9 but for a pollution event in Beijing on December 16-21, 2016.

3.4 Regional air quality model evaluation

Regional air quality model predictions of $PM_{2.5}$ are typically evaluated with observations from surface network sites, but the spatially continuous RF $PM_{2.5}$ fields offer more extensive coverage for model evaluation. We demonstrate this capability here with Community Multiscale Air Quality Modeling System (CMAQ version 4.7.1) simulations for the Korean peninsula including both South and North Korea at 9-km resolution [*Bae et al.*, 2018; *Bae et al.*, 2021]. There are no surface $PM_{2.5}$ data in North Korea to train the RF so we use the South Korea categorical variable to generate the RF $PM_{2.5}$ fields there.

The simulation for South Korea was conducted for 2015-2019 using emissions from the Clean Air Policy Support System (CAPSS) 2016 [*Choi et al.*, 2020] for South Korea and KORUSv5 [*Woo et al.*, n.d] for outside South Korea. The simulation for North Korea was conducted for 2016 using emissions from the Comprehensive Regional Emissions Inventory for Atmospheric Transport Experiment (CREATE) 2015 [*Woo et al.*, 2020] and CAPSS 2013. To prepare the boundary conditions, a coarse domain at 27-km horizontal grid resolution covering Northeast Asia was used.

Figure 11 illustrates the increased capability for model evaluation in South Korea enabled by the RF $PM_{2.5}$ fields. The bottom row shows the mean 2015-2019 $PM_{2.5}$ concentrations in CMAQ compared to the AirKorea network and to the RF, and the top row shows comparison scatterplots. The top left panel compares the CMAQ simulation to 2015-2019 mean $PM_{2.5}$ observations from the 398 AirKorea network sites. The top middle panel compares the RF $PM_{2.5}$ to the same AirKorea network data, showing excellent agreement. The RF-generated fields provide 1353 points for South Korea on the 9x9 km^2 CMAQ grid, and the top right panel shows the resulting increase in capability for evaluation of the CMAQ simulation. It shows in particular that CMAQ underestimates $PM_{2.5}$ in coastal environments, possibly because of unaccounted ship emissions.

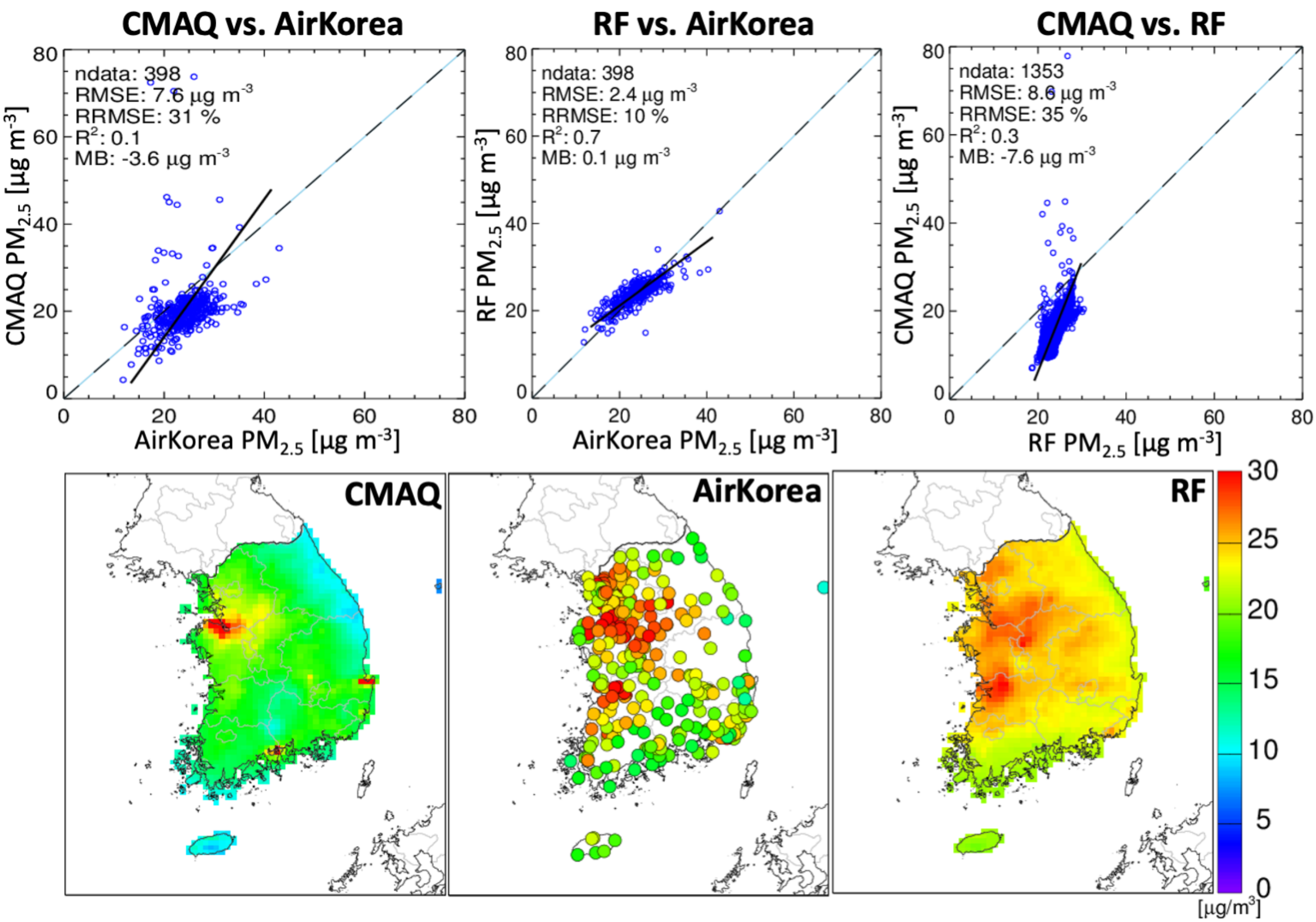

Figure 11: Mean $PM_{2.5}$ concentrations in South Korea in 2015-2019 as simulated by CMAQ, measured at the AirKorea sites, and represented by the RF. The top panels show scatterplots comparing the CMAQ and RF fields to the Air Korea measurements (398 sites), and CMAQ to the RF fields on the 9x9 km^2 CMAQ grid (1353 grid cells to compare). The bottom panels show maps of the mean 2015-2019 concentrations.

Figure 12 evaluates the CMAQ simulation with the RF $PM_{2.5}$ fields over North Korea. Unlike in South Korea, there are no observation sites in North Korea and RF $PM_{2.5}$ offers the only opportunity for local evaluation. CMAQ and RF $PM_{2.5}$ show dramatically different patterns. The highest $PM_{2.5}$ in CMAQ is in the Pyongyang capital region, while the RF shows highest values in the north-central region. The lack of reliable emission inventories for North Korea makes it difficult to arbitrate this difference. The RF is not trained for North Korea, which might lead to positive biases because the

AOD/$PM_{2.5}$ ratio modeled in the *Zhai et al.* [2021] GEOS-Chem simulation is higher over North Korea outside the mountainous east (range: 0.010-0.013 m^3 μg^{-1}) than over South Korea (0.008-0.010 m^3 μg^{-1}). However, the difference could also be explained by missing emissions in the inventory. Further evaluation could be done with border sites in South Korea and northeastern China. China MEE sites along the border are consistent with high $PM_{2.5}$ in north-central North Korea.

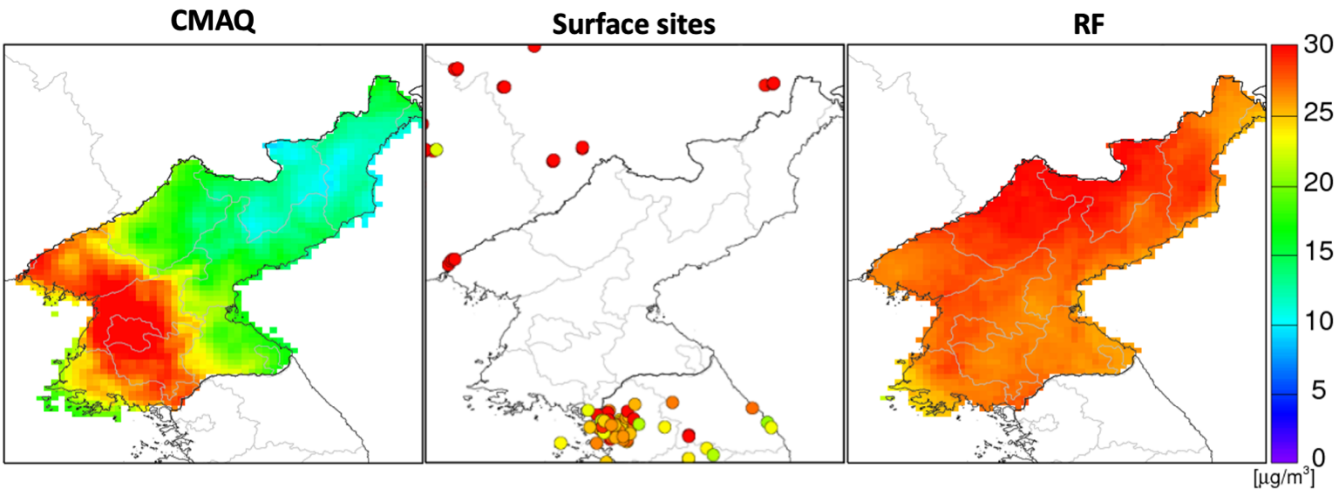

Figure 12: Mean $PM_{2.5}$ concentrations in North Korea in 2016 as simulated by CMAQ and as represented by the RF assuming South Korea as categorical variable. The middle panel shows observed $PM_{2.5}$ concentrations from the AirKorea and China MEE networks.

4 Conclusions

We used 2011-2019 geostationary aerosol optical depth (AOD) observations from the GOCI satellite instrument, in combination with a random forest (RF) machine learning algorithm trained on air quality network data, to produce a continuous 24-h $PM_{2.5}$ data set for eastern China, South Korea, and Japan at 6x6 km^2 resolution. The resulting gap-free product complements the air quality networks that cover only 1% of 6x6 km^2 grid cells in eastern China, 7% in South Korea, and 8% in Japan. It provides a general dataset for $PM_{2.5}$ mapping to serve regional pollution analysis, air quality monitoring, and public health applications.

We trained the RF algorithm on gap-filled AODs from the GOCI instrument and a suite of twelve meteorological, geographical, and temporal predictor variables. Gap filling of AODs was done by a weighted combination of nearest-neighbor data and chemical transport model fields, with the weight serving as an additional predictor variable. Testing of the RF algorithm by prediction of withheld network sites shows single-value precisions in each country of 26-32% for 24-h $PM_{2.5}$ and 12% for annual mean $PM_{2.5}$, with negligible mean bias. Accuracy statistics for $PM_{2.5}$ inferred on grid cells with no AOD retrieval (i.e. estimated using equation (1)) show similar accuracy statistics as the entire population, suggesting that the gap-filling procedure does not bias the results. The algorithm has only moderate success at predicting NAAQS exceedance events because most of these events are within the single-value precision, and also because of some smoothing of the extreme high tail of the $PM_{2.5}$ frequency distribution.

We compared the continuous 24-h RF $PM_{2.5}$ fields to spatial and temporal patterns observed at the national network sites. National trends of $PM_{2.5}$ inferred from the RF and weighted by area or population are consistent with those observed at network sites (2015-2019 in eastern China and South Korea, 2011-2019 in Japan), confirming that the trends observed at these sites are representative. However, the network sites in eastern China and South Korea underestimate population exposure. The RF $PM_{2.5}$ fields over South Korea show $PM_{2.5}$ hotspots missing in the early AirKorea network (2015) that are confirmed by subsequent addition of sites to the network (2019). The spatial distribution of RF $PM_{2.5}$ trends in South Korea shows decreases everywhere except in the Seoul metropolitan area where trends are flat. We show with time series in the capital cities (Beijing, Seoul, Tokyo) that the RF successfully captures the seasonality of $PM_{2.5}$ even though AOD and $PM_{2.5}$ have different and often opposite seasonalities.

We examined the ability of the RF algorithm to map air quality on urban scales by analysis of two multi-day pollution episodes in Seoul and Beijing. The algorithm captures the day-to-day temporal variability observed by the surface networks as well the spatial variability on the 6x6 km^2 scale. The highest $PM_{2.5}$ concentrations are underpredicted, which reflects the smoothing of the high tail of the distribution.

The continuous spatial coverage of $PM_{2.5}$ provided by the RF fields enables improved evaluation of the air quality models used in support of emission control policies. Comparison to a CMAQ simulation for South Korea in 2015-2019 reveals a large model underestimate in coastal environments undersampled by the AirKorea network. Comparison to a CMAQ simulation for North Korea in 2016, where the RF provides the only $PM_{2.5}$ data for model evaluation, shows drastically different patterns with the RF featuring high $PM_{2.5}$ throughout North Korea. The RF results in North Korea could be affected by errors due to lack of training data but they appear consistent with the $PM_{2.5}$ network observations at Chinese border sites.

Data availability 24-h 6x6 km^2 resolution daily $PM_{2.5}$ derived from the RF are made freely available on DataVerse at https://doi.org/10.7910/DVN/0L3IP7.

Author Contributions DP and DJJ designed the study. DP developed the RF and performed analysis. SZ, MB and SK ran and analyzed chemical transport model data. SL aided in satellite data processing. JK and JHK provided scientific interpretation and discussion. All authors provided input on the paper for revision before submission.

Competing interests The authors declare that they have no conflict of interest.

Acknowledgements This work was funded by the Samsung $PM_{2.5}$ Strategic Research Program and the Harvard-NUIST Joint Laboratory for Air Quality and Climate (JLAQC). GOCI data was provided by Korea Institute of Ocean Science and Technology (KIOST). DCP was funded by a US National Science Foundation Graduate Fellowship.

References

Alduchov, O. A., & Eskridge, R. E. (1996). Improved Magnus Form Approximation of Saturation Vapor Pressure. *Journal of Applied Meteorology*, 35(4), 601–609. https://doi.org/10.1175/1520-0450(1996)035<0601:IMFAOS>2.0.CO;2

Azuma, K., Kagi, N., Kim, H., & Hayashi, M. (2020). Impact of climate and ambient air pollution on the epidemic growth during COVID-19 outbreak in Japan. *Environmental Research*, 190, 110042. https://doi.org/10.1016/j.envres.2020.110042

Bae, M., Kim, H. C., Kim, B.-U., and Kim, S.: PM2.5 Simulations for the Seoul Metropolitan Area: (V) Estimation of North Korean Emission Contribution, J. Korean Soc. Atmos. Environ., 34, 294–305, https://doi.org/10.5572/KOSAE.2018.34.2.294, 2018.

Bae, M., Kim, B.-U., Kim, H. C., Kim, J., and Kim, S.: Role of emissions and meteorology in the recent PM2.5 changes in China and South Korea from 2015 to 2018, Environmental Pollution, 270, 116233, https://doi.org/10.1016/j.envpol.2020.116233, 2021.

Brasseur, G. P. and Jacob, D. J (2017). *Modeling of Atmospheric Chemistry*. Cambridge University Press.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

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. https://doi.org/10.1021/acs.est.7b05381

Burnett, R., et. al. (2018). Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proceedings of the National Academy of Sciences*, *115*(38), 9592–9597. https://doi.org/10.1073/pnas.1803222115

Center for International Earth Science Information Network – CIESIN – Columbia University. 2018. Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H49C6VHW

Chinese State Council (2013). Action Plan on Air Pollution Prevention and Control, available at: http://www.gov.cn/zwgk/2013-09/12/content_2486773.htm (last access: 12 April 2021).

Choi, J.-K., Park, Y. J., Ahn, J. H., Lim, H.-S., Eom, J., & Ryu, J.-H. (2012). GOCI, the world's first geostationary ocean color observation satellite, for the monitoring of temporal variability in coastal water turbidity. *Journal of Geophysical Research: Oceans*, 117(C9). https://doi.org/10.1029/2012JC008046

Choi, M., Kim, J., Lee, J., Kim, M., Park, Y.-J., Jeong, U., Kim, W., Hong, H., Holben, B., Eck, T. F., Song, C. H., Lim, J.-H., & Song, C.-K. (2016). GOCI Yonsei Aerosol Retrieval (YAER) algorithm and validation during the DRAGON-NE Asia 2012 campaign. *Atmospheric Measurement Techniques*, 9(3), 1377–1398. https://doi.org/10.5194/amt-9-1377-2016

Choi, M., Kim, J., Lee, J., Kim, M., Park, Y.-J., Holben, B., Eck, T. F., Li, Z., & 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), 385–408. https://doi.org/10.5194/amt-11-385-2018

Choi, M., Lim, H., Kim, J., Lee, S., Eck, T. F., Holben, B. N., Garay, M. J., Hyer, E. J., Saide, P. E., & Liu, H. (2019). Validation, comparison, and integration of GOCI, AHI, MODIS, MISR, and VIIRS aerosol optical depth over East Asia during the 2016 KORUS-AQ campaign. *Atmospheric Measurement Techniques*, 12(8), 4619–4641. https://doi.org/10.5194/amt-12-4619-2019

Choi, S., Kim, T., Lee, H., Kim, H., Han, J., Lee, K., Lim, E., Shin, S., Jin, H., Cho, E., Kim, Y., and Yoo, C.: Analysis of the National Air Pollutant Emission Inventory (CAPSS 2016) and the Major Cause of Change in Republic of Korea, 14, 24, 2020.

Crawford, J. H., Ahn, J.-Y., Al-Saadi, J., Chang, L., Emmons, L. K., Kim, J., Lee, G., Park, J.-H., Park, R. J., Woo, J. H., Song, C.-K., Hong, J.-H., Hong, Y.-D., Lefer, B. L., Lee, M., Lee, T., Kim, S., Min, K.-E., Yum, S. S., … Kim, Y. P. (2021). The Korea–United States Air Quality (KORUS-AQ) field study. Elementa: Science of the Anthropocene, 9(1). https://doi.org/10.1525/elementa.2020.00163

Cusworth, D. H., Jacob, D. J., Sheng, J.-X., Benmergui, J., Turner, A. J., Brandman, J., White, L., & Randles, C. A. (2018). Detecting high-emitting methane sources in oil/gas fields using satellite observations. *Atmospheric Chemistry and Physics*, 18(23), 16885–16896. https://doi.org/10.5194/acp-18-16885-2018

Dominici, F., Peng, R. D., Bell, M. L., Pham, L., McDermott, A., Zeger, S. L., & Samet, J. M. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. JAMA, 295(10), 1127–1134. https://doi.org/10.1001/jama.295.10.1127

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757. https://doi.org/10.1002/qj.49712555417

Geng, G., Zhang, Q., Martin, R. V., van Donkelaar, A., Huo, H., Che, H., Lin, J., & He, K. (2015). Estimating long-term PM2.5 concentrations in China using satellite-based aerosol optical depth and a chemical transport model. Remote Sensing of Environment, 166, 262–270. https://doi.org/10.1016/j.rse.2015.05.016

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Gräler, B., Pebesma, E., & Heuvelink, G. (2016). Spatio-Temporal Interpolation using gstat. *The R Journal*, 8(1), 204–218.

Gupta, P., & Christopher, S. A. (2009). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. Journal of Geophysical Research: Atmospheres, 114(D14). https://doi.org/10.1029/2008JD011496

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Random Forests. In *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (pp. 587–604). Springer. https://doi.org/10.1007/978-0-387-84858-7_15

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hu, H., Hu, Z., Zhong, K., Xu, J., Zhang, F., Zhao, Y., & Wu, P. (2019). Satellite-based high-resolution mapping of ground-level PM2.5 concentrations over East China using a spatiotemporal regression kriging model. Science of The Total Environment, 672, 479–490. https://doi.org/10.1016/j.scitotenv.2019.03.480

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. https://doi.org/10.1021/acs.est.7b01210

Huang, X., Ding, A., Gao, J., Zheng, B., Zhou, D., Qi, X., Tang, R., Wang, J., Ren, C., Nie, W., Chi, X., Xu, Z., Chen, L., Li, Y., Che, F., Pang, N., Wang, H., Tong, D., Qin, W., … He, K. (2021). Enhanced secondary pollution offset reduction of primary emissions during COVID-19 lockdown in China. *National Science Review*, 8(2). https://doi.org/10.1093/nsr/nwaa137

Jung-a, Song (2016). South Korean shipbuilders engulfed in crisis. *Financial Times*. https://www.ft.com/content/d74127ac-3140-11e6-8825-ef265530038e

Kianian, B., Liu, Y., & Chang, H. H. (2021). Imputing Satellite-Derived Aerosol Optical Depth Using a Multi-Resolution Spatial Model and Random Forest for PM2.5 Prediction. Remote Sensing, 13(1), 126. https://doi.org/10.3390/rs13010126

Kioumourtzoglou Marianthi-Anna, Schwartz Joel D., Weisskopf Marc G., Melly Steven J., Wang Yun, Dominici Francesca, & Zanobetti Antonella. (2016). Long-term PM2.5 Exposure and Neurological Hospital Admissions in the Northeastern United States. *Environmental Health Perspectives*, 124(1), 23–29. https://doi.org/10.1289/ehp.1408973

Koo, J.-H., Kim, J., Lee, Y. G., Park, S. S., Lee, S., Chong, H., Cho, Y., Kim, J., Choi, K., & Lee, T. (2020). The implication of the air quality pattern in South Korea after the COVID-19 outbreak. Scientific Reports, 10(1), 22462. https://doi.org/10.1038/s41598-020-80429-4

Kloog, I., Nordio, F., Coull, B. A., & Schwartz, J. (2012). Incorporating Local Land Use Regression And Satellite Aerosol Optical Depth In A Hybrid Model Of Spatio-Temporal PM2.5 Exposures In The Mid-Atlantic States. *Environmental Science & Technology*, *46*(21), 11913–11921. https://doi.org/10.1021/es302673e

Kloog, I., Chudnovsky, A. A., Just, A. C., Nordio, F., Koutrakis, P., Coull, B. A., Lyapustin, A., Wang, Y., & Schwartz, J. (2014). A new hybrid spatio-temporal model for estimating daily multi-year PM2.5 concentrations across northeastern USA using high resolution aerosol optical depth data. *Atmospheric Environment*, *95*, 581–590. https://doi.org/10.1016/j.atmosenv.2014.07.014

Li, T., Shen, H., Yuan, Q., Zhang, X., & Zhang, L. (2017). Estimating Ground-Level PM2.5 by Fusing Satellite and Station Observations: A Geo-Intelligent Deep Learning Approach. Geophysical Research Letters, 44(23), 11,985-11,993. https://doi.org/10.1002/2017GL075710

Lim, H., Choi, M., Kim, J., Kasai, Y., & Chan, P. W. (2018). AHI/Himawari-8 Yonsei Aerosol Retrieval (YAER): Algorithm, Validation and Merged Products. *Remote Sensing*, 10(5), 699. https://doi.org/10.3390/rs10050699

Lim, H., Go, S., Kim, J., Choi, M., Lee, S., Song, C.-K., & Kasai, Y. (2021). Integration of GOCI and AHI Yonsei aerosol optical depth products during the 2016 KORUS-AQ and 2018 EMeRGe campaigns. Atmospheric Measurement Techniques, 14(6), 4575–4592. https://doi.org/10.5194/amt-14-4575-2021

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). https://doi.org/10.1029/2004JD005025

Liu Yang, Paciorek Christopher J., & Koutrakis Petros. (2009). Estimating Regional Spatial and Temporal Variability of PM2.5 Concentrations Using Satellite Data, Meteorology, and Land Use Information. Environmental Health Perspectives, 117(6), 886–892. https://doi.org/10.1289/ehp.0800123

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

Park, S., Shin, M., Im, J., Song, C.-K., Choi, M., Kim, J., Lee, S., Park, R., Kim, J., Lee, D.-W., and Kim, S.-K. (2019). Estimation of ground-level particulate matter concentrations through the synergistic use of satellite observations and process-based models over South Korea, *Atmos. Chem. Phys*., 19, 1097–1113, https://doi.org/10.5194/acp-19-1097-2019.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830.

Phillips, Tom. Beijing smog: Pollution red alert declared in China capital and 21 other cities. (2016). *The Guardian*. http://www.theguardian.com/world/2016/dec/17/beijing-smog-pollution-red-alert-declared-in-china-capital-and-21-other-cities (last access: 13 April 2021)

Poulidis, A. P., Takemi, T., Shimizu, A., Iguchi, M., & Jenkins, S. F. (2018). Statistical analysis of dispersal and deposition patterns of volcanic emissions from Mt. Sakurajima, Japan. *Atmospheric Environment*, 179, 305–320. https://doi.org/10.1016/j.atmosenv.2018.02.021

Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., & Holben, B. N. (2005). The MODIS Aerosol Algorithm, Products, and Validation. *Journal of the Atmospheric Sciences*, 62(4), 947–973. https://doi.org/10.1175/JAS3385.1

Remer, L. A., Mattoo, S., Levy, R. C., Heidinger, A., Pierce, R. B., & Chin, M. (2012). Retrieving aerosol in a cloudy environment: Aerosol product availability as a function of spatial resolution. Atmospheric Measurement Techniques, 5(7), 1823–1840. https://doi.org/10.5194/amt-5-1823-2012

Remer, L. A., Mattoo, S., Levy, R. C., & Munchak, L. A. (2013). MODIS 3 km aerosol product: Algorithm and global perspective. *Atmospheric Measurement Techniques*, 6(7), 1829–1844. https://doi.org/10.5194/amt-6-1829-2013

She, Q., Choi, M., Belle, J. H., Xiao, Q., Bi, J., Huang, K., Meng, X., Geng, G., Kim, J., He, K., Liu, M., & Liu, Y. (2020). Satellite-based estimation of hourly PM2.5 levels during heavy winter pollution episodes in the Yangtze River Delta, China. *Chemosphere*, 239, 124678. https://doi.org/10.1016/j.chemosphere.2019.124678

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 23rd ACM National Conference*, 517–524. https://doi.org/10.1145/800186.810616

Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., de Hoogh, K., de' Donato, F., Gariazzo, C., Lyapustin, A., Michelozzi, P., Renzi, M., Scortichini, M., Shtein, A., Viegi, G., Kloog, I., & Schwartz, J. (2019). Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. Environment International, 124, 170–179. https://doi.org/10.1016/j.envint.2019.01.016

van Donkelaar, A. van, Martin, R. V., & Park, R. J. (2006). Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing. Journal of Geophysical Research: Atmospheres, 111(D21). https://doi.org/10.1029/2005JD006996

van Donkelaar Aaron, Martin Randall V., Brauer Michael, Kahn Ralph, Levy Robert, Verduzco Carolyn, & Villeneuve Paul J. (2010). Global Estimates of Ambient Fine Particulate Matter Concentrations from Satellite-Based Aerosol Optical Depth: Development and Application. Environmental Health Perspectives, 118(6), 847–855. https://doi.org/10.1289/ehp.0901623

van Donkelaar, A., et. al. (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. https://doi.org/10.1021/acs.est.5b05833

van Donkelaar, A., Martin, R. V., Li, C., & Burnett, R. T. (2019). Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. Environmental Science & Technology, 53(5), 2595–2611. https://doi.org/10.1021/acs.est.8b06392

Wang, J., and Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM2.5 mass: Implications for air quality studies. Geophysical Research Letters, 30(21). https://doi.org/10.1029/2003GL018174

Wei, Y., Wang, Y., Di, Q., Choirat, C., Wang, Y., Koutrakis, P., Zanobetti, A., Dominici, F., & Schwartz, J. D. (2019). Short term exposure to fine particulate matter and hospital admission risks and costs in the Medicare population: Time stratified, case crossover study. *BMJ*, 367, l6258. https://doi.org/10.1136/bmj.l6258

Woo, J.-H., Kim, Y., Kim, H.-K., Choi, K.-C., Eum, J.-H., Lee, J.-B., Lim, J.-H., Kim, J., and Seong, M.: Development of the CREATE Inventory in Support of Integrated Climate and Air Quality Modeling for Asia, Sustainability, 12, 7930, https://doi.org/10.3390/su12197930, 2020.

Woo, J-H, Kim, Y, Kim, J, Park, M, Jang, Y, Kim, J, Bu, C, Lee, Y, Park, R, Oak, Y, Fried, A, Simpson, I, Emmons, L, Crawford, J. n.d. KORUS Emissions: A comprehensive Asian emissions information in support of the NASA/NIER KORUS-AQ mission. Elementa: Science of the Anthropocene, in press.

Xu, J.-W., Martin, R. V., van Donkelaar, A., Kim, J., Choi, M., Zhang, Q., Geng, G., Liu, Y., Ma, Z., Huang, L., Wang, Y., Chen, H., Che, H., Lin, P., & Lin, N. (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. https://doi.org/10.5194/acp-15-13133-2015

Xue, T., Zheng, Y., Tong, D., Zheng, B., Li, X., Zhu, T., & Zhang, Q. (2019). Spatiotemporal continuous estimates of PM2.5 concentrations in China, 2000–2016: A machine learning method

with inputs from satellites, chemical transport model, and ground observations. *Environment International*, *123*, 345–357. https://doi.org/10.1016/j.envint.2018.11.075
Yeo, M. and Kim, Y. (2019). Trends of the PM2.5 concentrations and high PM2.5 concentration cases by region in Korea. *Particle and Aerosol Research*, 15(2), 45-56. http://dx.doi.org/10.11629/jpaar.2019.15.2.045 (in Korean).
Zang, L., Mao, F., Guo, J., Wang, W., Pan, Z., Shen, H., Zhu, B., & Wang, Z. (2019). Estimation of spatiotemporal PM1.0 distributions in China by combining PM2.5 observations with satellite aerosol optical depth. Science of The Total Environment, 658, 1256–1264. https://doi.org/10.1016/j.scitotenv.2018.12.297
Zhai, S., Jacob, D. J., Wang, X., Shen, L., Li, K., Zhang, Y., Gui, K., Zhao, T., & Liao, H. (2019). Fine particulate matter (PM2.5) trends in China, 2013–2018: Separating contributions from anthropogenic emissions and meteorology. Atmospheric Chemistry and Physics, 19(16), 11031–11041. https://doi.org/10.5194/acp-19-11031-2019
Zhai, S., Jacob, D. J., Brewer, J. F., Li, K., Moch, J. M., Kim, J., Lee, S., Lim, H., Lee, H. C., Kuk, S. K., Park, R. J., Jeong, J. I., Wang, X., Liu, P., Luo, G., Yu, F., Meng, J., Martin, R. V., Travis, K. R., … Liao, H. (2021). Interpretation of geostationary satellite aerosol optical depth (AOD) over East Asia in relation to fine particulate matter (PM2.5): Insights from the KORUS-AQ aircraft campaign and seasonality. Atmospheric Chemistry and Physics Discussions, 1–23. https://doi.org/10.5194/acp-2021-413
Zhang, G., & Lu, Y. (2012). Bias-corrected random forests in regression. *Journal of Applied Statistics*, 39(1), 151–160. https://doi.org/10.1080/02664763.2011.578621
Zhuang, J., Dussin, R., Jüling, A., & Rasp, S. (2020). JiaweiZhuang/xESMF: v0.3.0 Adding ESMF.LocStream capabilities (Version v0.3.0). Zenodo. https://doi.org /10.5281/ZENODO.1134365