Remote Sensing of Environment 221 (2019) 665–674

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Impacts of snow and cloud covers on satellite-derived $PM_{2.5}$ levels

Jianzhao Bi[a], Jessica H. Belle[a], Yujie Wang[b,c], Alexei I. Lyapustin[b,c], Avani Wildani[d], Yang Liu[a,*]

[a] Department of Environmental Health, Emory University, Rollins School of Public Health, Atlanta, GA, USA
[b] Goddard Earth Sciences and Technology Center, University of Maryland Baltimore County, Baltimore, MD, USA
[c] NASA Goddard Space Flight Center, Greenbelt, MD, USA
[d] Department of Computer Science, Emory University, Atlanta, GA, USA

ARTICLE INFO

Keywords:
$PM_{2.5}$
AOD
MAIAC
Random Forest
Gap-filling
Snow cover
Cloud cover

ABSTRACT

Satellite aerosol optical depth (AOD) has been widely employed to evaluate ground fine particle ($PM_{2.5}$) levels, whereas snow/cloud covers often lead to a large proportion of non-random missing AOD. As a result, the fully covered and unbiased $PM_{2.5}$ estimates will be hard to generate. Among the current approaches to deal with the data gap issue, few have considered the cloud-AOD relationship and none of them have considered the snow-AOD relationship. This study examined the impacts of snow and cloud covers on AOD and $PM_{2.5}$ and made full-coverage $PM_{2.5}$ predictions with the consideration of these impacts. To estimate the missing AOD, daily gap-filling models with snow/cloud fractions and meteorological covariates were developed using the random forest algorithm. By using these models in New York State, a daily AOD data set with a 1-km resolution was generated with a complete coverage. The "out-of-bag" R^2 of the gap-filling models averaged 0.93 with an interquartile range from 0.90 to 0.95. Subsequently, a random forest-based $PM_{2.5}$ prediction model with the gap-filled AOD and covariates was built to predict fully covered $PM_{2.5}$ estimates. A ten-fold cross-validation for the prediction model showed a good performance with an R^2 of 0.82. In the gap-filling models, the snow fraction was of higher significance in the snow season compared with the rest of the year. The prediction models fitted with/without the snow fraction also suggested the discernible changes in $PM_{2.5}$ patterns, further confirming the significance of this parameter. Compared with the methods without considering snow and cloud covers, our $PM_{2.5}$ prediction surfaces showed more spatial details and reflected small-scale terrain-driven $PM_{2.5}$ patterns. The proposed methods can be generalized to the areas with extensive snow/cloud covers and large proportions of missing satellite AOD for predicting $PM_{2.5}$ levels with high resolutions and complete coverage.

1. Introduction

Fine particulate matter ($PM_{2.5}$) may be inhaled and deposit in alveoli, increasing the risk of cardiorespiratory disease (Bose et al., 2015; Burnett et al., 2014; Madrigano et al., 2013; Sorek-Hamer et al., 2016). Though air quality stations have been employed to measure $PM_{2.5}$ and its composition, their spatial coverage is insufficient to cover larger populations to study the impacts of $PM_{2.5}$ exposure on human health. Recently, satellite aerosol optical depth (AOD) with broad spatio-temporal availability has been widely applied in $PM_{2.5}$ exposure modeling (Hu et al., 2017; Kloog et al., 2012; Ma et al., 2016). With a 1-km resolution, Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD has been able to reveal pollution patterns in great detail and link between $PM_{2.5}$ and microenvironment more precisely (Hu et al., 2014; Lyapustin et al., 2011). High-resolution AOD is particularly critical for reflecting $PM_{2.5}$ pollution patterns in the regions dominated by local sources, *e.g.*, New York State in which residential wood combustions and on-road emissions are the primary $PM_{2.5}$ sources (Fine et al., 2002; Reff et al., 2009; Su et al., 2013).

Though satellite AOD can serve as a surrogate of ground $PM_{2.5}$, a non-linear relationship is shown between the two, which varies spatially and temporally (Paciorek and Liu, 2009). To reflect this relationship, various statistical models have been developed (Kloog et al., 2014; Liu et al., 2009; Xiao et al., 2017). For instance, Kloog et al. (2014) employed mixed models to estimate high-resolution $PM_{2.5}$ with an out-of-sample R^2 (coefficient of determination) of 0.88 in the Northeastern US. Xiao et al. (2017) built two-stage models to produce 1-km $PM_{2.5}$ estimates with complete coverage over the Yangtze River Delta (YRD) in China, which yielded cross-validation (CV) R^2s of ~0.8. In these parametric models, however, restrictive assumptions of independence and population distributions are needed, which can be a challenge for a complex mixture, *e.g.*, $PM_{2.5}$. In contrast, non-

* Corresponding author at: Emory University, Rollins School of Public Health, 1518 Clifton Rd. NE, Atlanta, GA 30322, USA.
E-mail address: yang.liu@emory.edu (Y. Liu).

https://doi.org/10.1016/j.rse.2018.12.002
Received 3 June 2018; Received in revised form 15 October 2018; Accepted 3 December 2018
Available online 13 December 2018

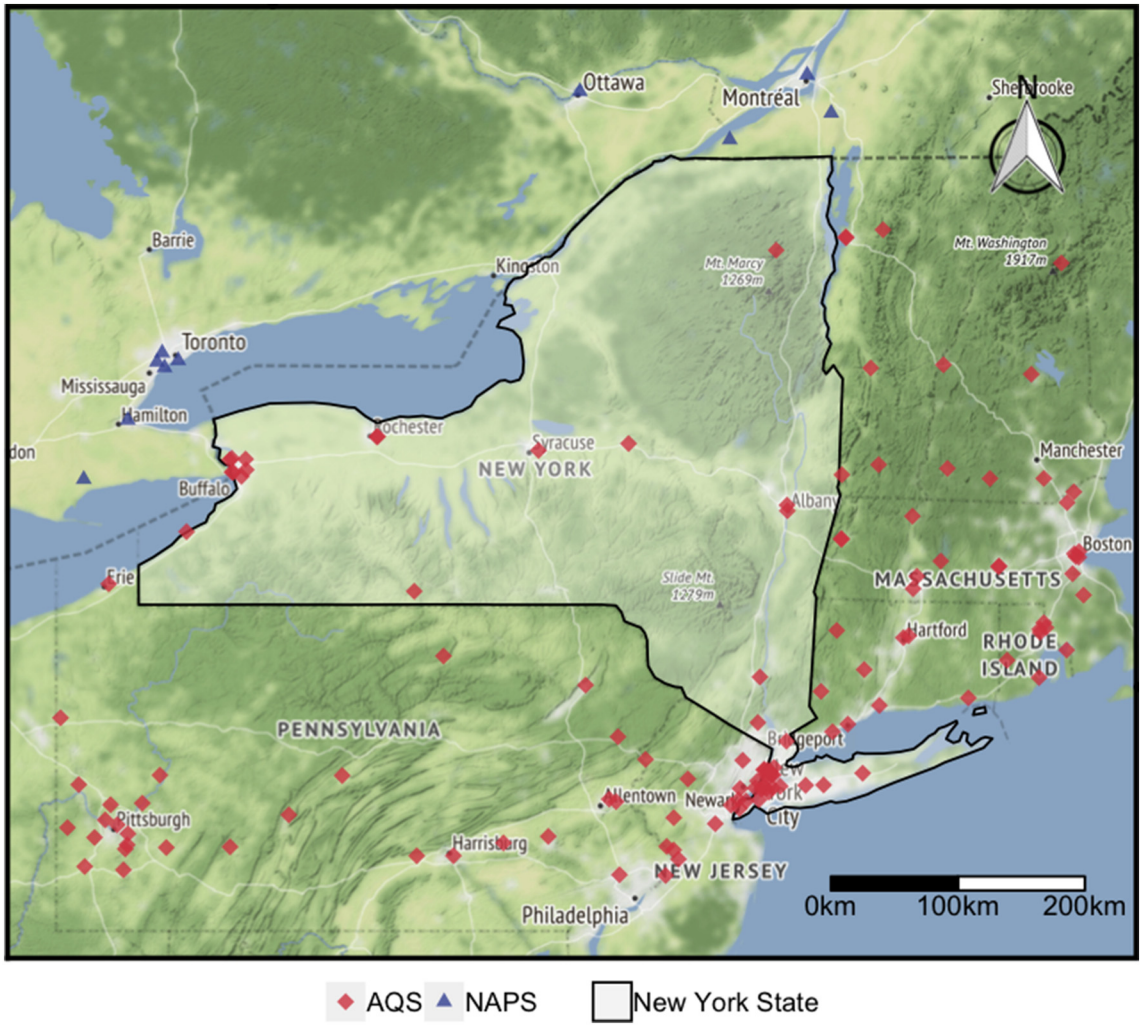

Fig. 1. Study areas. Latitude: [40.1°N, 45.6°N]; Longitude: [80.5°W, 71°W]. The areas outside New York State served as the buffer. Red diamonds are EPA AQS stations. Blue triangles are NAPS stations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

parametric models (*e.g.*, machine learning models) can capture the non-linear relationships and interactions between the variables with fewer *a priori* assumptions. Popular machine learning models for $PM_{2.5}$ prediction consist of Artificial Neural Networks (ANN) (Di et al., 2016; Gupta and Christopher, 2009; Zou et al., 2015) and Random Forests (RF) (Brokamp et al., 2018; Hu et al., 2017). To be specific, random forests are advantaged in the interpretability of modeling outcomes by their measures of variable importance (Friedman et al., 2001). Hu et al. (2017) developed a random forest model with satellite AOD and covariates, generating 12-km $PM_{2.5}$ predictions over the Contiguous US with a CV R^2 of 0.80 and a root-mean-square error (RMSE) of 1.78 μg/m^3.

The large proportion of non-random missing data is considered as a critical problem associated with the effective use of satellite AOD, which primarily results from cloud cover and high surface brightness from factors such as snow and ice (Chu et al., 2002; Hsu et al., 2013; Levy et al., 2007). This non-random missingness has significantly affected data availability and generated biases in the stages of data processing (Xiao et al., 2017). Several strategies have been developed to cope with this data gap problem. van Donkelaar et al. (2011) adopted adjusted cloud filtering standards in the Dark Target algorithm and achieved a 21% increase of available AOD during Moscow wildfire events in 2010. Kloog et al. (2011), based on universal kriging with daily mean $PM_{2.5}$ levels and random slopes, proposed a spatial smoothing model and generated full-coverage $PM_{2.5}$ predictions across the New England region of the US. Kloog et al. (2012) incorporated inverse probability weighting (IPW) in $PM_{2.5}$ regression to reduce the selection bias caused by AOD missingness. They achieved $PM_{2.5}$ predictions with higher reliability. However, it has been found that AOD and $PM_{2.5}$ levels were changed in the presence of cloud cover due to the shifted AOD/$PM_{2.5}$ physical characteristics under various meteorological conditions (Alam et al., 2014; Belle et al., 2017; Kang et al., 2015; Myhre et al., 2007). For instance, Belle et al. (2017) found that variations in cloud properties (*e.g.*, cloud optical depth and emissivity) were correlated with shifted levels of $PM_{2.5}$ and its components. Accordingly, the gap-filling process without considering cloud interactions may cause estimation biases. Xiao et al. (2017) showed that the cloud-AOD interactions can be partially explained by incorporating cloud features in AOD gap-filling processes. By using a multiple imputation model coupled with cloud fractions, they found a higher level of gap-filled AOD on cloudy days in the YRD of China. Similar to cloud, snow cover, with a large areal extent in high-latitudes of Northern Hemisphere (Robinson et al., 1993), has also resulted in a large proportion of missing AOD (Lyapustin et al., 2011; Xiao et al., 2016) and the changes of AOD/$PM_{2.5}$ levels (Chen et al., 2012; Emili et al., 2011; Green et al., 2015; Whiteman et al., 2014). Emili et al. (2011) found that due to the contamination of cloud and snow pixels, satellite AOD would be occasionally overestimated. Studies also suggested that in mountainous terrains, $PM_{2.5}$ levels were higher on snowy days because of more stagnant atmospheric conditions (Chen et al., 2012; Whiteman

et al., 2014). To our knowledge, no study has been conducted to deal with the snow-related AOD missingness in $PM_{2.5}$ modeling. Thus, incorporating snow features in AOD gap-filling processes may help to obtain more reliable $PM_{2.5}$ estimates in the regions with extensive snow cover.

In this case study, we incorporated satellite-retrieved snow and cloud fractions in an AOD gap-filling model to yield fully covered AOD data in New York State where there were extensive snow/cloud covers, especially in winter. Based on the gap-filled AOD, we generated high-resolution and fully covered $PM_{2.5}$ predictions. In the meantime, we examined the significance of the snow fraction in the AOD gap-filling and $PM_{2.5}$ prediction processing. We also identified the patterns of local $PM_{2.5}$ pollution in New York State using our high-resolution $PM_{2.5}$ predictions.

2. Data and methods

2.1. Study areas

New York State and its surrounding areas were selected as our study domain (Fig. 1). The surrounding areas served as the buffer to minimize the quality degradation on the edge of the domain and to ensure sufficient ground $PM_{2.5}$ stations within the study areas. These areas were suitable for examining the performance of our models since 1) the cloud cover and heavy snowfall led to a considerable amount of missing satellite AOD, particularly in winter (Belle et al., 2017) and 2) local $PM_{2.5}$ emission sources (*e.g.*, residential wood combustions and on-road emissions) were dominated in this state so that reliable predictions with a high spatial resolution were needed to reveal the spatial details of $PM_{2.5}$ pollution (Fine et al., 2002; NEI, 2014; Su et al., 2013). All analyses were based on the 1-km grid of MAIAC AOD, covering 474,392 grid cells in the entire areas.

2.2. $PM_{2.5}$ measurements

$PM_{2.5}$ measurements inside the US were provided by the United States Environmental Protection Agency (EPA) (https://www.epa.gov/). The measurements inside Canada were provided by the National Air Pollution Surveillance (NAPS) (http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx). NAPS has three types of samplers (Dichotomous, Partisol, and Speciation) for $PM_{2.5}$ measurements. Only Dichotomous and Partisol measurements were employed to ensure the data quality. For each ground station from AQS or NAPS, daily mean $PM_{2.5}$ levels were calculated and matched with AOD and other variables in the MAIAC grid cell. The entire areas covered 137 $PM_{2.5}$ ground stations (127 inside the US and 10 inside Canada; Fig. 1).

2.3. MAIAC AOD data

MAIAC is an advanced AOD data set based on Moderate Resolution Imaging Spectroradiometer (MODIS) Collection 6 Level 1B data. Using a time-series algorithm, the MAIAC data set has increased the accuracies of aerosol detections and optimized the spatial resolution to 1-km (Lyapustin et al., 2011). Since MAIAC is a novel algorithm for MODIS AOD, the quality assessment of this product is insufficient. A current validation over South America suggested that ~66% MAIAC AOD retrievals were within the expected error ($\pm (0.05 + 0.05 \times AOD)$) (Martins et al., 2017). Terra/Aqua satellites provide two MODIS data sets with different crossing times. In our $PM_{2.5}$ prediction models, Terra (descending node at 10:30 A.M. local time) and Aqua (ascending node at 1:30 P.M. local time) AOD served as two separate predictors to reflect diurnal changes of AOD.

2.4. MODIS cloud and snow fractions

Cloud and snow fractions as the percentages of cloud and snow in a grid cell are critical for the quantitative estimation of cloud- and snow-AOD interactions in the gap-filling processing. Cloud fractions were yielded from MODIS Level-2 Cloud product (MOD06_L2/MYD06_L2; https://modis.gsfc.nasa.gov/). In this product, the cloud fraction has been retrieved from infrared and visible wavelengths along with other physical and radiative cloud properties (Platnick et al., 2015). A global evaluation by Ackerman et al. (2008) suggested that MODIS cloud detection reached an agreement of ~85% with the ground lidar observations. Snow fractions were yielded from MODIS Snow Cover product (MOD10/MYD10; https://modis.gsfc.nasa.gov/). In this product, the snow fraction has been yielded from the Normalized Difference Snow Index (NDSI) (Hall and Riggs, 2011). MODIS Snow Cover product was found to have a high correlation ($r = 0.9$) with ground-truth observations (Salomonson and Appel, 2006). Terra and Aqua data sets were employed respectively and matched with the corresponding AOD data sets. Only daytime cloud/snow observations were adopted.

2.5. Meteorological data

Meteorological variables were obtained from the North American Regional Reanalysis (NARR) (http://www.emc.ncep.noaa.gov/) (Mesinger et al., 2006) and the North American Land Data Assimilation System (NLDAS) (http://ldas.gsfc.nasa.gov/nldas) (Mitchell et al., 2004). NARR has a resolution of 0.25° and a 3-h temporal interval. NLDAS has a resolution of 0.125° and a 1-h temporal interval. Since NLDAS has higher spatiotemporal resolutions, the parameters from this data set were used first. The parameters not provided by NLDAS were extracted from NARR. The meteorological parameters from NLDAS consisted of air temperature, humidity, surface pressure, precipitation, wind speed, potential evaporation, downward shortwave radiation as well as convective available potential energy (CAPE). The parameters from NARR covered planetary boundary layer height (HPBL) and visibility. To match the observation times of MAIAC AOD, we calculated the daily values of meteorological parameters by averaging the simulations from 9 A.M. to 3 P.M. (local time, *i.e.*, GMT 1400–2000 for NLDAS and GMT 1500, 1800, 2100 for NARR).

2.6. Land-use variables

The land-use parameters covered 1) the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation at a 1 arc-second (~30 m) resolution (https://asterweb.jpl.nasa.gov/gdem.asp), 2) LandScan ambient population in 2015 at a 900-m resolution (https://web.ornl.gov/sci/landscan/), 3) Normalized Difference Vegetation Index (NDVI) from MODIS vegetation indices at a 500-m resolution (MOD13/MYD13; https://modis.gsfc.nasa.gov/), and 4) distances to highways and major roads computed from ESRI StreetMap (Environmental System Research Institute, Inc., Redland, California).

2.7. Data matching

All data sets with various spatial resolutions were matched and fitted into the MAIAC AOD 1-km grid. For the data sets with spatial resolutions coarser than 1–km (*e.g.*, meteorological fields, cloud and snow fractions, and NDVI), the inverse distance weighting (IDW) interpolation was employed (Bartier and Keller, 1996). For the data sets with finer resolutions (*e.g.*, ASTER GDEM), the downsampling was conducted, *i.e.*, averaging neighboring pixels into one value. For the LandScan population, each data point was assigned to its nearest MAIAC grid cell.

2.8. AOD gap-filling model

Random Forest (RF) is an “ensemble learning” method yielding several decision trees and aggregating the regressing results from these

trees (Breiman, 2001; Liaw and Wiener, 2002). It adopts two types of bootstrap aggregating strategies (*i.e.*, a bootstrap training sample and a group of randomly selected independent variables) for each decision tree. These strategies allow the random forest to be a robust algorithm against overfitting (Breiman, 2001). A random forest model has two major parameters, *i.e.*, the number of decision trees in the forest (n_{tree}) and the number of independent variables randomly tried at each split (m_{try}). It also provides variable importance measures that inform variable weights and contribute to the interpretation of the model (Liaw and Wiener, 2002). Our AOD gap-filling model was based on the random forest algorithm, which is expressed as

$$\begin{aligned} &AOD_{st} \\ &= f(\text{cloud fraction}_{st}, \text{snow fraction}_{st}, \text{air temperature}_{st} \\ &\quad , \text{specific humidity}_{st}, \text{relative humidity}_{st}, \\ &\quad \text{previous day precipitation}_{st}, \text{elevation}_s, X_s, Y_s) \end{aligned} \tag{1}$$

where s denotes the location of a grid cell and t is the time of an observation. The dependent variable was the satellite AOD. Independent variables covered cloud and snow fractions, meteorological parameters (air temperature, specific humidity, relative humidity, and precipitation on the previous day), elevation, and spatial coordinates. The spatial coordinates were the real distances in kilometers to the central point of the study areas. By comparing the results with different settings, n_{tree} and m_{try} were set as 200 and 3 respectively to balance the prediction accuracy and computational efficiency.

The AOD gap-filling model (Eq. (1)) was fitted daily for both Terra and Aqua AOD data sets. Eq. (1) without the snow fraction (a.k.a. cloud-only gap-filling model) was also built to examine the impact of this variable on the gap-filled AOD. Since MAIAC AOD had a large proportion of missing data, we employed three rolling-day samples for the middle day's model. Out-of-bag (OOB) R^2 and RMSE were employed to assess the modeling performance. OOB R^2 is calculated from the predictions not in the bootstrap sample (a.k.a. “out-of-bag” sample) (Breiman, 2001). Since the mechanism of OOB is similar to cross-validation (CV), OOB R^2 is nearly equal to CV R^2 under large sample size (Friedman et al., 2001). RMSE is calculated by aggregating the errors of OOB predictions. Model-estimated “permutation accuracy importance” (Breiman, 2002) was used to suggest the variable importance. This importance measure is estimated in line with the fall of prediction accuracy after randomly permuting the OOB sample of the targeting variable (Liaw and Wiener, 2002).

2.9. $PM_{2.5}$ prediction model

The $PM_{2.5}$ prediction model also followed the random forest algorithm (Eq. (2)). The dependent variable was the $PM_{2.5}$ measurements from AQS and NAPS. Independent variables covered gap-filled Terra and Aqua AOD, meteorological parameters (air temperature, dew-point temperature, surface pressure, specific humidity, wind speed, visibility, planetary boundary layer height, potential evaporation, downward shortwave radiation, and CAPE), land-use parameters (population, distances to highways and major roads, elevation, and NDVI), and dummy variables for months and Julian days. Moreover, a convolutional layer of $PM_{2.5}$ levels was applied, representing the weighted averages of nearby $PM_{2.5}$ levels. This parameter helps to fully exploit the spatial autocorrelation of $PM_{2.5}$ and can significantly increase the accuracy of $PM_{2.5}$ prediction (Di et al., 2016; Hu et al., 2017). By comparing the results with different settings, n_{tree} and m_{try} were set as 500 and 7 respectively to achieve the best prediction accuracy. The $PM_{2.5}$ prediction model is expressed as

$$\begin{aligned} &PM2.5_{st} \\ &= f(\text{Terra AOD}_{st}, \text{Aqua AOD}_{st}, \text{air temperature}_{st} \\ &\quad , \text{dew point temperature}_{st}, \text{surface pressure}_{st}, \\ &\quad \text{specific humidity}_{st}, \text{wind speed}_{st}, \text{visibilty}_{st}, \text{HPBL}_{st} \\ &\quad , \text{potential evaporation}_{st}, \\ &\quad \text{downward shortwave radiation}_{st}, \text{CAPE}_{st}, \text{population}_s, \text{NDVI}_{st} \\ &\quad , \text{highway distance}_s, \text{major road distance}_s, \\ &\quad \text{elevation}_s, \text{PM2. 5 convolutional layer}_{st}, \text{month}_t, \text{day}_t) \end{aligned} \tag{2}$$

where s denotes the location of a grid cell and t denotes the time of an observation. The variable selection strategy followed Hu et al. (2017), in which the variables with low importance values were discarded (precipitation and relative humidity were discarded from the model). 10-fold CVs were employed to assess the mode performance, including overall, spatial, and temporal CVs (Xiao et al., 2017). The spatial CV generated validation samples in accordance with the locations of the $PM_{2.5}$ measurements. The temporal CV generated validation samples in line with the Julian days of the measurements. Besides the original model, Eq. (2) without both Terra and Aqua gap-filled AOD parameters (a.k.a. no-AOD prediction model) was built to examine the impact of the gap-filled AOD on the $PM_{2.5}$ predictions. Eq. (2) with AOD gap-filled with only considering cloud fractions (a.k.a. cloud-only prediction model) was also built to verify the importance of the snow parameter in the $PM_{2.5}$ prediction.

2.10. Comparison with another gap-filling method

To assess our gap-filling model, our full-coverage $PM_{2.5}$ predictions were compared to the gap-filled $PM_{2.5}$ generated by a previous gap-filling method proposed by Kloog et al. (2011) and Just et al. (2015) (Eq. (3)).

$$\sqrt{\text{PredPM}_{st}} = \alpha_0 + \alpha_1\sqrt{\text{MeanPM}_t} + s(XY)_s + \epsilon_{st} \tag{3}$$

where s denotes the location of a grid cell and t is the time of an observation. The dependent variable PredPM_{st} denotes the existing $PM_{2.5}$ predictions at location s on a day t. MeanPM_t denotes an average $PM_{2.5}$ measurement on a day t. $s(X,Y)_s$ is a spline surface of the coordinates of the grid cells. The reason for applying square root was to ensure the positive values of $PM_{2.5}$ (Just et al., 2015). We also tried calculating MeanPM_t within a 100-km buffer (Kloog et al., 2014). Yet the $PM_{2.5}$ distribution showed significant artificial patterns due to the sparse monitoring network so that the buffer was not adopted. The gap-filling model was fitted monthly to better reveal the temporal variations of $PM_{2.5}$ (Kloog et al., 2014). A 10-fold CV was employed to examine the modeling performance.

3. Results

3.1. Descriptive statistics for MAIAC AOD missingness

The Quality Assessment (QA) flags of MAIAC AOD were employed to infer the rates of missing AOD caused by cloud cover, snow cover and water/ice (Table S1). The average daily missing rates associated with cloud and snow are listed in Table 1 (the contribution of water/ice-related missing AOD was negligible in NYS). Since the missing AOD caused by snow cover was primarily in the first 15 weeks (105 days) of 2015, the average missing rates in this period are separately summarized in the table. This period is also referred to as the “snow season” hereinafter. The overall AOD missing rate was ~90% in 2015, and the cloud-related AOD missing rate was ~76%. Though the snow-related missingness only took up ~6% in this year, the percentage increased to ~20% in the snow season. There was a ~10% gap between the overall missing rate and the total rate caused by cloud and snow. This gap

Table 1
Average daily missing rates of MAIAC AOD caused by cloud and snow covers in NYS in 2015. The first 15 weeks of 2015 are defined as the "snow season". The ~10% gap between the overall missing rate and the total rate caused by cloud and snow resulted from the MAIAC pixels outside the sensor scanning range (*i.e.*, no measurement). The contribution of water/ice to the missing AOD was negligible in NYS.

AOD	Missing type	Mean	Median	25^{th}–75^{th} quantiles
Aqua	Overall	90.27%	96.54%	87.62%–99.84%
	Cloud-related	75.58%	79.47%	64.74%–91.48%
	Snow-related	6.14%	0.0005%	0%–2.48%
	Snow-related (first 15 weeks)	21.15%	14.03%	5.23%–31.18%
Terra	Overall	89.50%	96.54%	84.88%–99.75%
	Cloud-related	76.66%	81.32%	66.29%–92.12%
	Snow-related	5.55%	0.001%	0%–2.88%
	Snow-related (first 15 weeks)	19.03%	10.78%	5.00%–29.84%

resulted from the MAIAC pixels located in the areas outside the sensor scanning range (*i.e.*, no measurement). In general, though the missing AOD was primarily caused by cloud cover, snow-related AOD missingness remained a severe issue in the snow season.

3.2. AOD gap-filling by random forests

By Eq. (1), full-coverage AOD was yielded for both Terra and Aqua data sets. Daily RF models had a mean OOB R^2 of 0.93 (for both Terra and Aqua data sets) with interquartile ranges (IQR) from 0.91 to 0.95 for Terra and 0.90 to 0.95 for Aqua. The annual AOD distributions and the spatial differences between the gap-filled and original AOD are shown in Fig. 2. The overall and gap-filled AOD showed similar spatial patterns. The peak AOD levels appeared in the Allegheny Plateau in the Southern NYS. Some populated areas in NYS (*e.g.*, New York City) also had high-level AOD. For Terra, original and gap-filled AOD had the annual means of 0.11 (IQR: [0.10, 0.12]) and 0.26 (IQR: [0.25, 0.26]), respectively. For Aqua, original and gap-filled AOD had the annual means of 0.09 (IQR: [0.08, 0.10]) and 0.25 (IQR: [0.24, 0.25]), respectively. As Fig. 2(e) and (f) suggest, the gap-filled AOD was higher than the original AOD throughout the state. This outcome is consistent with Belle et al. (2017) and Xiao et al. (2017) who found clear evidence that the hygroscopic growth of aerosol droplets under cloudy and humid weather resulted in higher levels of AOD.

3.3. $PM_{2.5}$ prediction by random forests

3.3.1. Modeling performance

By Eq. (2), 1-km $PM_{2.5}$ predictions with full coverage were yielded. The prediction model showed an overall CV R^2 of 0.82 (Fig. 3(a)) with spatial and temporal CV R^2s of 0.74 and 0.81, respectively. The model had an RMSE of 2.16 $\mu g/m^3$, suggesting a good prediction accuracy. Fig. 3(b) shows the ranking of the variable importance. The $PM_{2.5}$ convolutional layer had the highest importance value, which is consistent with Hu et al. (2017) who confirmed an improvement of modeling accuracy by the $PM_{2.5}$ convolutional layer. Five of the top-seven important variables were land-use terms (population, distances to major roads, elevation, distances to highways, and NDVI). This outcome could reflect that the $PM_{2.5}$ sources in New York State were primarily local sources, and the transportation of regional pollution was relatively weak, which is consistent with the $PM_{2.5}$ emission inventory of New York State in 2014 (NEI, 2014). Though the gap-filled AOD parameters did not have high importance values, they still contributed to a changed spatial pattern of $PM_{2.5}$ estimates (as shown in Section 3.4.1). We also performed the $PM_{2.5}$ prediction for 2002–2012, finding that the modeling performance was stable in terms of CV R^2 and variable importance ranking (Table S2).

3.3.2. $PM_{2.5}$ predictions

Fig. 4(a) shows the annual $PM_{2.5}$ distribution in NYS in 2015 (the $PM_{2.5}$ distributions in different seasons are shown in Fig. S1). The $PM_{2.5}$ levels had a mean of 5.30 $\mu g/m^3$ with an IQR from 4.43 to 6.08 $\mu g/m^3$. The distribution showed clear patterns of higher $PM_{2.5}$ along with the roads and in the populated areas. These patterns are consistent with the $PM_{2.5}$ emission inventory of NYS (NEI, 2014), suggesting that the $PM_{2.5}$ sources were primarily residential wood combustions and on-road emissions. The highest $PM_{2.5}$ levels appeared in the large cities, *e.g.*, New York City, Long Island, and Buffalo. Other major cities, *e.g.*, Albany, Rochester, Yonkers, and Syracuse also had relatively high $PM_{2.5}$ levels. In contrast, the Adirondack Mountains in the Northeastern NYS with the lowest population density showed the lowest $PM_{2.5}$ levels.

Since New York State lies upon the portion of Northeast Appalachians, most of its areas, in particular Upstate New York, are dominated by the mountainous terrain. Our high-resolution $PM_{2.5}$ predictions showed obvious evidence of the correlation between elevation and $PM_{2.5}$ in NYS. This correlation can be partially reflected by the variable importance since the elevation was the 4^{th} important variable in the prediction model. Besides, we found obvious patterns of $PM_{2.5}$ accumulation in the valleys of the Allegheny Plateau in winter. An area in the Allegheny Plateau was selected to highlight the valley accumulation of $PM_{2.5}$. Fig. 5 shows the $PM_{2.5}$ distribution with contours (140-m intervals) in this area in winter (January, February, and December in 2015). Elevated $PM_{2.5}$ levels in the valleys were clearly shown. In these valleys, $PM_{2.5}$ levels were ~1–2 $\mu g/m^3$ higher than those in the surroundings. These results suggested that our 1-km $PM_{2.5}$ predictions could reflect small-scale $PM_{2.5}$ features driven by local geographical factors, which would be smeared with a coarser resolution.

3.3.3. Comparison with another gap-filling approach

To verify the effectiveness of our gap-filling model, we compared our $PM_{2.5}$ predictions to the $PM_{2.5}$ generated using a reported gap-filling method (Eq. (3)). For a fair comparison, Eq. (3) followed a subset of our $PM_{2.5}$ predictions derived from the original MAIAC AOD (*i.e.*, removing the $PM_{2.5}$ predictions generated from the gap-filled AOD). The ten-fold cross-validation R^2s of the monthly models had a mean of 0.69 ranged from 0.60 to 0.78, lower than the OOB R^2s of our gap-filling model. Fig. 4(b) shows the annual $PM_{2.5}$ distribution generated by Eq. (3). Even though the model had full-coverage predictions, the $PM_{2.5}$ were spatially over-smoothed. In contrast, since the major $PM_{2.5}$ sources were local sources in NYS (*i.e.*, residential wood combustions and on-road emissions) (NEI, 2014; Su et al., 2013), our $PM_{2.5}$ spatial patterns well captured these well-defined sources.

3.4. The importance of gap-filled AOD with snow cover parameter

3.4.1. Contribution of gap-filled AOD to $PM_{2.5}$ predictions

Since the gap-filled AOD parameters were not among the top-important variables (Fig. 3(b)), their contribution to our $PM_{2.5}$ predictions was examined to verify their validity in the prediction model. A prediction model without gap-filled AOD parameters was built (a.k.a. no-AOD prediction model), and the analysis of the $PM_{2.5}$ estimates emphasized on the snow season (first 105 days in 2015). Fig. 4(c) shows the spatial differences between full-model and no-AOD $PM_{2.5}$ in the snow season. Compared with the no-AOD $PM_{2.5}$, the full-model $PM_{2.5}$ had higher levels along with the roads and lower levels in the background areas (*e.g.*, the Adirondack Mountains). Hence, after the addition of the AOD parameters in the prediction model, the pattern of $PM_{2.5}$ sources was intensified, in particular on-road emissions. Quantitatively, the mean absolute difference between two $PM_{2.5}$ data sets was 0.13 $\mu g/m^3$ with an IQR from 0.06 to 0.18 $\mu g/m^3$. The maximum absolute difference reached 0.99 $\mu g/m^3$, suggesting a significant change in terms of a 105-day average. Also, the full-model $PM_{2.5}$ was compared with the $PM_{2.5}$ predictions derived from the original MAIAC AOD (Fig.

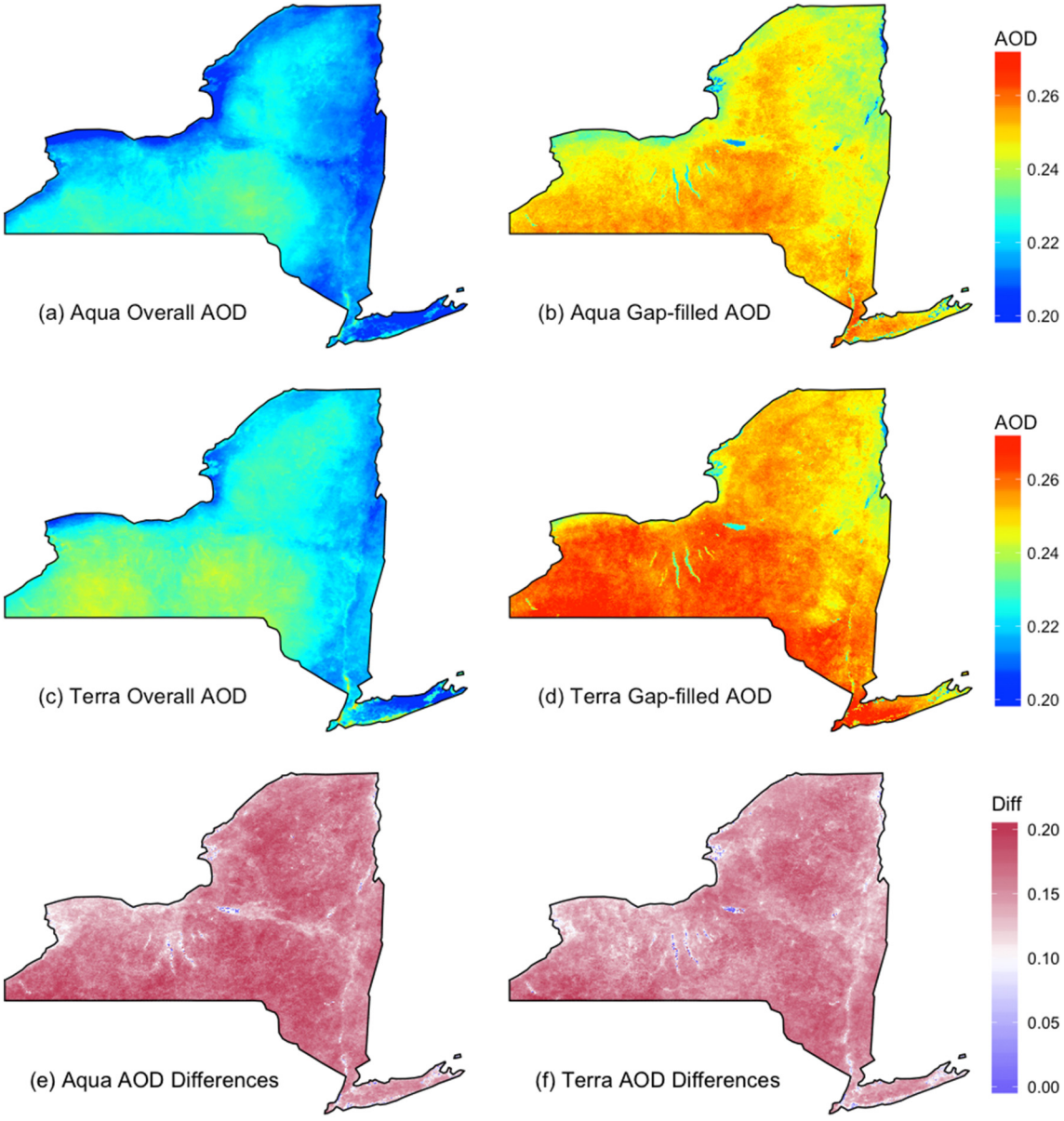

Fig. 2. AOD spatial distributions in 2015: (a) annual distribution of Aqua AOD (original and gap-filled AOD); (b) annual distribution of gap-filled Aqua AOD; (c) annual distribution of Terra AOD (original and gap-filled AOD); (d) annual distribution of gap-filled Terra AOD; (e) differences between gap-filled and original Aqua AOD (gap-filled minus original AOD); (f) differences between gap-filled and original Terra AOD (gap-filled minus original AOD).

S2). The significant differences between two data sets suggested the impact of not only the systematic changes of gap-filled AOD values, but also the significant increase of the sample size. These comparisons suggested that the pollution information incorporated in the gap-filled AOD significantly impacted the $PM_{2.5}$ prediction, leading to discernible changes in $PM_{2.5}$ patterns. The reason for the intensified pattern of on-road $PM_{2.5}$ pollution after the addition of the gap-filled AOD needs further studies.

3.4.2. *Influences of snow cover parameter on AOD and $PM_{2.5}$*

During and after the snow season, the rank change of the snow fraction's importance in the AOD gap-filling model indicated the significance of this variable in the gap-filling process. Among the 9 variables in the AOD gap-filling model, the average rank of snow fraction's importance was valued as 5 during the snow season and dropped to 8 after the snow season. In contrast, the rank of cloud fraction's importance was stable with an average of 2 throughout the year. As expected, this outcome indicated that the snow fraction was more important when it was snowy. Further, we built a cloud-only AOD gap-filling model (*i.e.*, the snow fraction was removed) to examine the impact of the snow fraction on the gap-filled AOD (the cross-validation performance listed in Table S3). We found that the mean absolute differences between the full-model AOD and the cloud-only AOD were 0.002 (IQR: [0.001, 0.003]) and 0.001 (IQR: [0.0004, 0.001]) for Terra and Aqua, respectively. During the snow season, the maximum absolute difference was 0.007 in terms of a 105-day average. Moreover, we also examined the impact of the snow fraction on the $PM_{2.5}$ predictions. By using the full-model and cloud-only AOD in the $PM_{2.5}$ prediction, two $PM_{2.5}$ data sets were yielded: 1) full-model $PM_{2.5}$ and 2) cloud-only $PM_{2.5}$. Fig. 4(d) shows the differences between the two in the snow season. Compared with the cloud-only $PM_{2.5}$, the full-model $PM_{2.5}$ tended to be higher in the regions with high $PM_{2.5}$ levels (*e.g.*, populated areas and major roads) and lower in backgrounds. Hence, the snow fraction improved the source patterns of $PM_{2.5}$. In the snow season, the mean absolute difference between two $PM_{2.5}$ data sets was 0.09 $\mu g/m^3$ (IQR: 0.02 to 0.12 $\mu g/m^3$), and the maximum absolute

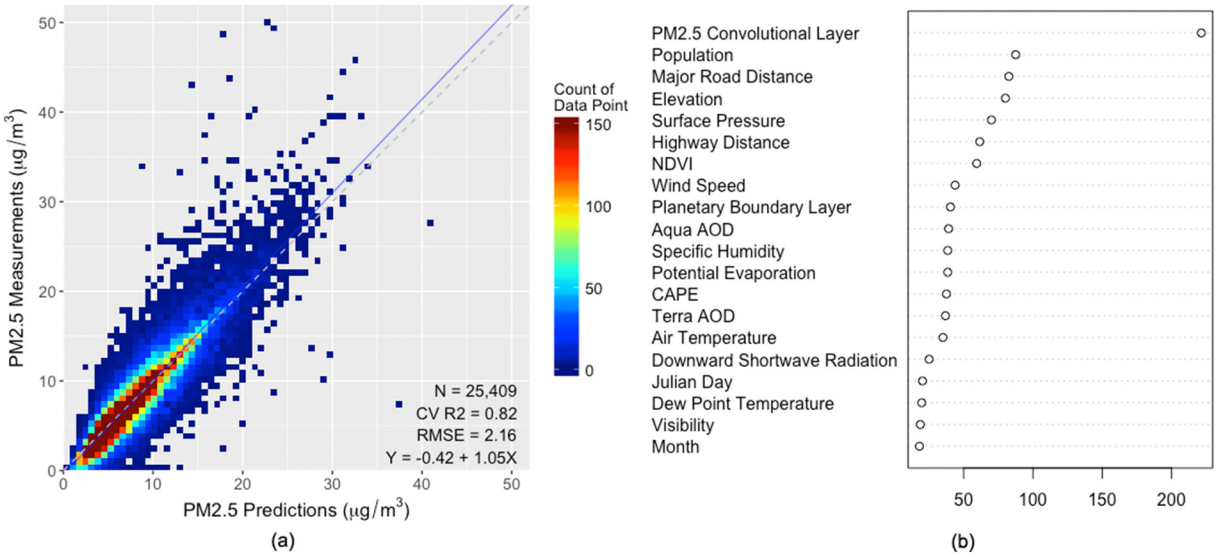

Fig. 3. $PM_{2.5}$ prediction modeling performance: (a) 10-fold cross-validation (CV) scatters with an R^2 of 0.82 and an RMSE of 2.16 μg/m^3; (b) variable importance ranking. The $PM_{2.5}$ convolutional layer had the highest value and five of the top-seven important variables were land-use terms.

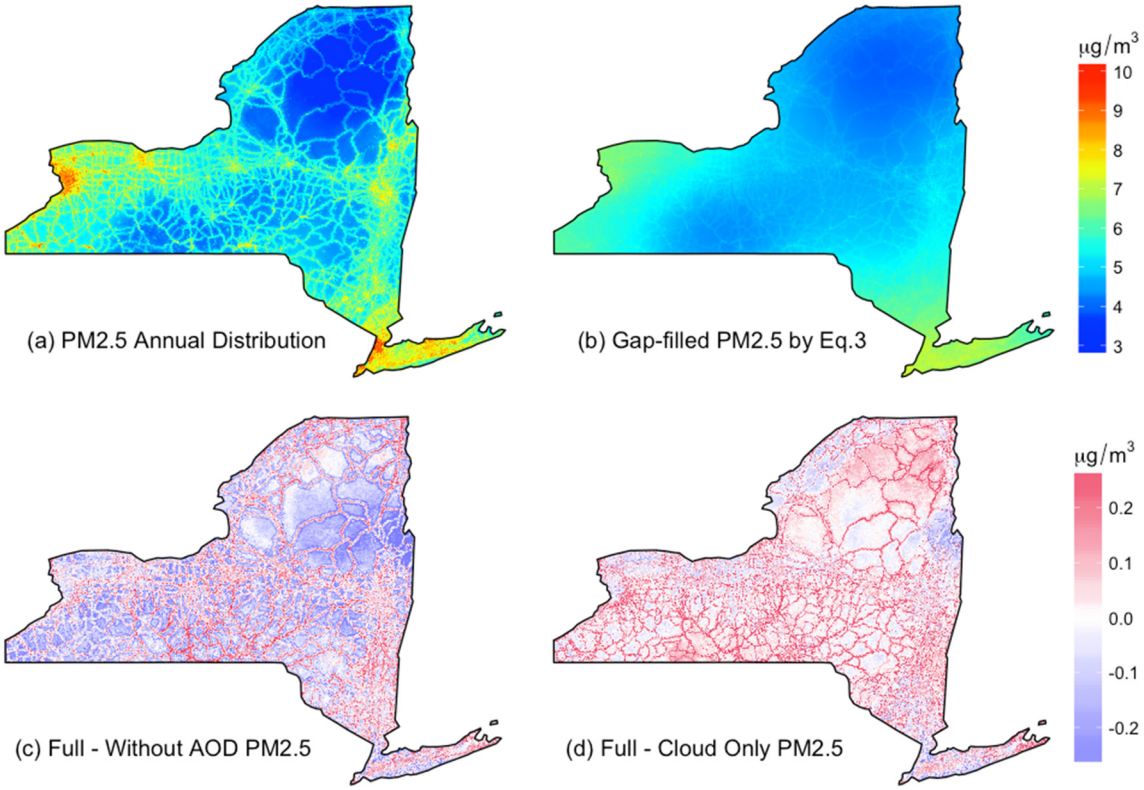

Fig. 4. $PM_{2.5}$ spatial distributions in 2015: (a) annual distribution of $PM_{2.5}$ with a 1-km resolution by Eq. (2); (b) gap-filled $PM_{2.5}$ by Eq. (3); (c) differences between full-model and no-AOD $PM_{2.5}$ in the snow season (full-model minus no-AOD $PM_{2.5}$); (d) differences between full-model and cloud-only $PM_{2.5}$ in the snow season (full-model minus cloud-only $PM_{2.5}$).

difference reached 0.99 μg/m^3. The mean difference was at a similar scale as the mean difference caused by the gap-filled AOD, which was 0.13 μg/m^3 in the snow season. This suggested that the snow cover might take up the majority of gap-filled AOD's impact on the $PM_{2.5}$ predictions. In brief, by influencing the gap-filled AOD, the impact of the snow cover was transferred to the $PM_{2.5}$ predictions, resulting in discernible changes in $PM_{2.5}$ patterns. The reason for the intensified source patterns after the addition of this parameter needs further studies.

4. Discussion

In this study, fully covered and high-resolution $PM_{2.5}$ levels were estimated based on the gap-filled MAIAC AOD in New York State in 2015. To the best of our knowledge, this is the first study applying both snow and cloud cover parameters in the gap-filling process to increase the spatiotemporal availability of satellite AOD and the accuracy of $PM_{2.5}$ estimation. Despite the ~90% AOD missingness in the region, our gap-filling models still had excellent performance with a mean OOB R^2

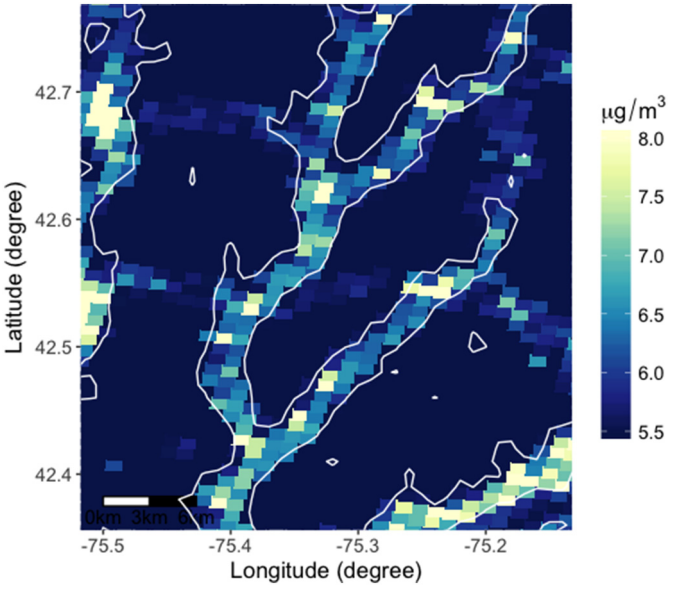

Fig. 5. $PM_{2.5}$ accumulation effect in the valleys of Upstate New York in winter. The areas are on the border of Chenango County and Otsego County (Latitude: [42.375°N, 42.75°N]; Longitude: [75.5°W, 75.15°W]). The widths of the valleys are ~3 km.

of 0.93. Xiao et al. (2017) adopted a multiple imputation model with similar predictors (except for the snow fraction) to conduct AOD gap-filling in the YRD of China, in which there was nearly no snow cover and the AOD missing rate was ~60%. They had a mean modeling R^2 of 0.77 with an IQR from 0.71 to 0.82. The improvement in our modeling performance reflected the advantages over the model of Xiao et al. (2017) and the potential of machine learning models dealing with the AOD gap-filling with complex snow/cloud-AOD interactions. Our gap-filled AOD was significantly higher than the original MAIAC AOD by an average of 0.15. This outcome is consistent with the findings of previous studies suggesting that the increased humidity caused by the cloud could lead to the aerosol hygroscopic growth (Belle et al., 2017; Kang et al., 2015; Xiao et al., 2017).

Our $PM_{2.5}$ prediction model showed a good performance with a CV R^2 of 0.82 and an RMSE of 2.16 μg/m^3. Hu et al. (2017), also applying the random forest algorithm to the $PM_{2.5}$ prediction, had a CV R^2 of 0.79 and an RMSE of 2.84 μg/m^3 in the Northeastern US (including New York and New England states). Compared to the model of Hu et al. (2017), our model achieved the major improvements of 1) conducting satellite AOD gap-filling and making full-coverage $PM_{2.5}$ predictions, 2) conducting $PM_{2.5}$ predictions at a higher 1-km spatial resolution, and 3) achieving similar modeling performance without the convolutional layers of land-use variables. Kloog et al. (2014) estimated high-resolution $PM_{2.5}$ with an "out-of-sample" R^2 of 0.88 and an RMSE of 2.33 μg/m^3 in the Northeastern US using a linear mixed effects model. However, their $PM_{2.5}$ gap-filling model, virtually a generalized additive model (GAM) with a spline surface (Eq. (3)), tended to over-smooth the $PM_{2.5}$ spatial details. In contrast, our $PM_{2.5}$ predictions showed strong local sources in the populated areas and major roads. 2014 National Emissions Inventory (NEI, 2014) suggested that the largest $PM_{2.5}$ emission source in New York State was the residential wood combustion, resulting in 17,916 tons of $PM_{2.5}$. This type of emission sources were primarily around the large and densely populated cities (*e.g.*, New York metropolitan area, Buffalo, Rochester, Syracuse and Albany) (Reff et al., 2009). Thus, the relationship between high $PM_{2.5}$ levels and high populations observed in this study was reasonable. Besides, the emission inventory (NEI, 2014) suggested that among the top-twelve $PM_{2.5}$ emission sources in New York State (out of 60 Emissions Inventory System (EIS) emission sectors), at least four of them had the direct relationship with on-road emissions. These sources (paved road dust, on-road diesel heavy-duty vehicles, unpaved road dust, and on-road non-diesel light-duty vehicles) led to 19,753 tons of $PM_{2.5}$ in 2014. These high-level $PM_{2.5}$ emissions could interpret the strong $PM_{2.5}$ signals along with the roads in our areas. Due to the 1-km resolution, our predictions could reflect small-scale $PM_{2.5}$ features driven by local topographic factors. In some mountainous areas of Upstate New York, 1–2 μg/m^3 higher $PM_{2.5}$ accumulated in valleys in winter was observed. This phenomenon is consistent with the findings of previous studies suggesting that the stagnant weather conditions in valleys caused by strong temperature inversions could lead to the increase in $PM_{2.5}$ levels (Gillies et al., 2010; Silcox et al., 2012; Wang et al., 2012).

This study considered snow cover in AOD gap-filling process by introducing the snow fraction in the gap-filling model. The significance of the snow fraction was partially reflected by its importance ranking in the gap-filling models. In the snow season (first 15 weeks of 2015) when the heaviest snowfalls appeared, the snow fraction was the 5th/6th important variable, compared with the rest of the year when its importance only ranked 8th/9th. To further examine the impact of the snow fraction on the $PM_{2.5}$ predictions, two prediction models without the gap-filled AOD or snow fraction were built. In the snow season, the gap-filled AOD alone caused the absolute changes of $PM_{2.5}$ by an average of 0.13 μg/m^3, and the snow fraction resulted in an average absolute change of 0.09 μg/m^3. Accordingly, the impact of the snow fraction might be responsible for the primary influence of gap-filled AOD on the $PM_{2.5}$ predictions in the snow season. For the spatial pattern, the snow fraction led to an increase of $PM_{2.5}$ estimates in the areas with large $PM_{2.5}$ sources (*e.g.*, populated areas and major roads). Previous studies found that the elevated $PM_{2.5}$ levels appeared to occur on snowy days due to more stagnant atmospheric conditions (Chen et al., 2012; Whiteman et al., 2014). The impact of snow on $PM_{2.5}$ patterns in this analysis could partially reflect this phenomenon. On the whole, by affecting the gap-filled AOD, the impact of the snow fraction was transferred to the $PM_{2.5}$ predictions, leading to discernible changes in $PM_{2.5}$ patterns. These analyses also suggested the advantages of estimating missing satellite AOD data for $PM_{2.5}$ prediction, instead of directly estimating the missing $PM_{2.5}$ levels. The former approach can, to a greater extent, draw upon the pollution characteristics incorporated in the satellite AOD and the additional information provided by AOD-related meteorological features. Though this study was limited to New York State, the methodology relative to the AOD gap-filling and $PM_{2.5}$ prediction is generalizable to other areas with extensive snow/cloud covers and large proportions of missing satellite AOD.

The major limitation of this study is that the physical characteristics of snow and cloud are insufficiently considered. The parameters used in this study only reflected the coverage of snow and cloud. Their different physical features, however, may cause changes in AOD and $PM_{2.5}$ levels (Belle et al., 2017). Accordingly, more snow/cloud characteristics (*e.g.*, cloud optical depth, cloud emissivity, surface albedo, *etc.*) can be applied to better interpret the interactions between snow/cloud and AOD/$PM_{2.5}$. With an increased number of snow/cloud parameters, a more suitable strategy to incorporate them, instead of additively applying them to the model, also deserves to be considered. Furthermore, the reason for snow/cloud-related AOD/$PM_{2.5}$ pattern changes needs further studies.

5. Conclusion

In this study, an AOD gap-filling model and a $PM_{2.5}$ prediction model based on the random forest algorithm were developed to estimate fully covered and high-resolution ground $PM_{2.5}$ in New York State in 2015. By introducing the MODIS snow/cloud fractions into the gap-filling process, a 100% gap-filled AOD data set was produced with an excellent modeling performance. The 1-km $PM_{2.5}$ predictions derived from the gap-filled AOD could reflect the detailed emission patterns and small-scale terrain-driven features. It is the first attempt where both

snow and cloud parameters are introduced into the AOD gap-filling process. Though we only applied fraction measures of snow and cloud, the importance of these parameters was still reflected, and the discernible interactions between snow/cloud and AOD/$PM_{2.5}$ were observed. It is necessary for future applications to adopt more physical characteristics of snow and cloud and to explore more suitable strategies to introduce these parameters into the gap-filling process. The methodology of this study can be generalized to other areas with extensive snow/cloud covers and large proportions of missing satellite AOD to estimate $PM_{2.5}$ exposures that previously could not be obtained. The improved $PM_{2.5}$ exposures with an increased sample size and a better quality are expected to contribute to downward epidemiological studies.

Acknowledgments

The work of J. Bi, J. Belle, and Y. Liu is partially supported by the NASA Applied Sciences Program (Grant # NNX14AG01G and NNX16AQ28G, PI: Liu). This work is also supported by the National Institute of Environmental Health Sciences of the National Institutes of Health under Award Number P30ES019776. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2018.12.002.

References

Ackerman, S., Holz, R., Frey, R., Eloranta, E., Maddux, B., McGill, M., 2008. Cloud detection with MODIS. Part II: validation. J. Atmos. Ocean. Technol. 25, 1073–1086.
Alam, K., Khan, R., Blaschke, T., Mukhtiar, A., 2014. Variability of aerosol optical depth and their impact on cloud properties in Pakistan. J. Atmos. Sol. Terr. Phys. 107, 104–112.
Bartier, P.M., Keller, C.P., 1996. Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW). Comput. Geosci. 22, 795–799.
Belle, J.H., Chang, H.H., Wang, Y.J., Hu, X.F., Lyapustin, A., Liu, Y., 2017. The potential impact of satellite-retrieved cloud parameters on ground-level PM2.5 mass and composition. Int. J. Environ. Res. Public Health 14.
Bose, S., Hansel, N., Tonorezos, E., Williams, D., Bilderback, A., Breysse, P., Diette, G., McCormack, M.C., 2015. Indoor particulate matter associated with systemic inflammation in COPD. J. Environ. Prot. 6, 566.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Breiman, L., 2002. Manual on Setting Up, Using, and Understanding Random Forests v3. 1. Statistics Department University of California Berkeley, CA, USA, pp. 1.
Brokamp, C., Jandarov, R., Hossain, M., Ryan, P., 2018. Predicting daily urban fine particulate matter concentrations using a random forest model. Environ. Sci. Technol. 52 (7), 4173–4179. https://doi.org/10.1021/acs.est.7b0538.
Burnett, R.T., Pope III, C.A., Ezzati, M., Olives, C., Lim, S.S., Mehta, S., Shin, H.H., Singh, G., Hubbell, B., Brauer, M., Anderson, H.R., Smith, K.R., Balmes, J.R., Bruce, N.G., Kan, H., Laden, F., Pruss-Ustun, A., Turner, M.C., Gapstur, S.M., Diver, W.R., Cohen, A., 2014. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. Environ. Health Perspect. 122, 397–403.
Chen, L.W.A., Watson, J.G., Chow, J.C., Green, M.C., Inouye, D., Dick, K., 2012. Wintertime particulate pollution episodes in an urban valley of the Western US: a case study. Atmos. Chem. Phys. 12, 10051–10064.
Chu, D.A., Kaufman, Y.J., Ichoku, C., Remer, L.A., Tanre, D., Holben, B.N., 2002. Validation of MODIS aerosol optical depth retrieval over land. Geophys. Res. Lett. 29.
Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., Schwartz, J., 2016. Assessing PM2.5 exposures with high spatiotemporal resolution across the continental United States. Environ. Sci. Technol. 50, 4712–4721.
Emili, E., Lyapustin, A., Wang, Y., Popp, C., Korkin, S., Zebisch, M., Wunderle, S., Petitta, M., 2011. High spatial resolution aerosol retrieval with MAIAC: application to mountain regions. J. Geophys. Res.-Atmos. 116.
Fine, P.M., Cass, G.R., Simoneit, B.R., 2002. Organic compounds in biomass smoke from residential wood combustion: emissions characterization at a continental scale. J. Geophys. Res. Atmos. 107.
Friedman, J., Hastie, T., Tibshirani, R., 2001. The Elements of Statistical Learning. Springer series in statistics, New York.
Gillies, R.R., Wang, S.Y., Booth, M.R., 2010. Atmospheric scale interaction on wintertime intermountain west low-level inversions. Weather Forecast. 25, 1196–1210.
Green, M.C., Chow, J.C., Watson, J.G., Dick, K., Inouye, D., 2015. Effects of snow cover and atmospheric stability on winter PM2.5 concentrations in Western U.S. valleys. J. Appl. Meteorol. Climatol. 54, 1191–1201.
Gupta, P., Christopher, S.A., 2009. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. J. Geophys. Res. Atmos. 114.
Hall, D.K., Riggs, G.A., 2011. Normalized-difference snow index (NDSI). In: Encyclopedia of Snow, Ice and Glaciers. Springer, pp. 779–780.
Hsu, N., Jeong, M.J., Bettenhausen, C., Sayer, A., Hansell, R., Seftor, C., Huang, J., Tsay, S.C., 2013. Enhanced deep blue aerosol retrieval algorithm: the second generation. J. Geophys. Res. Atmos. 118, 9296–9315.
Hu, X.F., Waller, L.A., Lyapustin, A., Wang, Y.J., Al-Hamdan, M.Z., Crosson, W.L., Estes, M.G., Estes, S.M., Quattrochi, D.A., Puttaswamy, S.J., Liu, Y., 2014. Estimating ground-level PM2.5 concentrations in the southeastern United States using MAIAC AOD retrievals and a two-stage model. Remote Sens. Environ. 140, 220–232.
Hu, X.F., 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. Environ. Sci. Technol. 51, 6936–6944.
Just, A.C., Wright, R.O., Schwartz, J., Coull, B.A., Baccarelli, A.A., Tellez-Rojo, M.M., Moody, E., Wang, Y.J., Lyapustin, A., Kloog, I., 2015. Using high-resolution satellite aerosol optical depth to estimate daily PM2.5 geographical distribution in Mexico City. Environ. Sci. Technol. 49, 8576–8584.
Kang, N., Kumar, K.R., Yin, Y., Diao, Y.W., Yu, X.N., 2015. Correlation analysis between AOD and cloud parameters to study their relationship over China using MODIS data (2003 − 2013): impact on cloud formation and climate change. Aerosol Air Qual. Res. 15 (958- +).
Kloog, I., Koutrakis, P., Coull, B.A., Lee, H.J., Schwartz, J., 2011. Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. Atmos. Environ. 45, 6267–6275.
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 spatiotemporal PM2.5 exposures in the mid-Atlantic states. Environ. Sci. Technol. 46, 11913–11921.
Kloog, I., Chudnovsky, A.A., Just, A.C., Nordio, F., Koutrakis, P., Coull, B.A., Lyapustin, A., Wang, Y.J., 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. Atmos. Environ. 95, 581–590.
Levy, R.C., Remer, L.A., Mattoo, S., Vermote, E.F., Kaufman, Y.J., 2007. Second-generation operational algorithm: retrieval of aerosol properties over land from inversion of moderate resolution imaging Spectroradiometer spectral reflectance. J. Geophys. Res. Atmos. 112.
Liaw, A., Wiener, M., 2002. Classification and Regression by RandomForest. 2. R news, pp. 18–22.
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, 886–892.
Lyapustin, A., Martonchik, J., Wang, Y.J., Laszlo, I., Korkin, S., 2011. Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. J. Geophys. Res.-Atmos. 116.
Ma, Z.W., Hu, X.F., Sayer, A.M., Levy, R., Zhang, Q., Xue, Y.G., Tong, S.L., Bi, J., Huang, L., Liu, Y., 2016. Satellite-based spatiotemporal trends in PM2.5 concentrations: China, 2004–2013. Environ. Health Perspect. 124, 184–192.
Madrigano, J., Kloog, I., Goldberg, R., Coull, B.A., Mittleman, M.A., Schwartz, J., 2013. Long-term exposure to PM2.5 and incidence of acute myocardial infarction. Environ. Health Perspect. 121, 192–196.
Martins, V., Lyapustin, A., de Carvalho, L., Barbosa, C., Novo, E., 2017. Validation of high-resolution MAIAC aerosol product over South America. J. Geophys. Res. Atmos. 122, 7537–7559.
Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E.H., 2006. North American regional reanalysis. Bull. Am. Meteorol. Soc. 87, 343–360.
Mitchell, K.E., Lohmann, D., Houser, P.R., Wood, E.F., Schaake, J.C., Robock, A., Cosgrove, B.A., Sheffield, J., Duan, Q., Luo, L., 2004. The multi-institution north American land data assimilation system (NLDAS): utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. J. Geophys. Res. Atmos. 109.
Myhre, G., Stordal, F., Johnsrud, M., Kaufman, Y.J., Rosenfeld, D., Storelvmo, T., Kristjansson, J.E., Berntsen, T.K., Myhre, A., Isaksen, I.S.A., 2007. Aerosol-cloud interaction inferred from MODIS satellite data and global aerosol models. Atmos. Chem. Phys. 7, 3081–3101.
NEI, 2014. 2014 National Emissions Inventory (NEI). (In).
Paciorek, C.J., Liu, Y., 2009. Limitations of remotely sensed aerosol as a spatial proxy for Fine particulate matter. Environ. Health Perspect. 117, 904–909.
Platnick, S., Ackerman, S., King, M., Meyer, K., Menzel, W., Holz, R., Baum, B., Yang, P., 2015. MODIS Atmosphere L2 Cloud Product (06_L2). NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA (doi, 10).
Reff, A., Bhave, P.V., Simon, H., Pace, T.G., Pouliot, G.A., Mobley, J.D., Houyoux, M., 2009. Emissions inventory of PM2.5 trace elements across the United States. Environ. Sci. Technol. 43, 5790–5796.
Robinson, D.A., Dewey, K.F., Heim, R.R., 1993. Global snow cover monitoring - an update. Bull. Am. Meteorol. Soc. 74, 1689–1696.
Salomonson, V.V., Appel, I., 2006. Development of the Aqua MODIS NDSI fractional snow cover algorithm and validation results. IEEE Trans. Geosci. Remote Sens. 44, 1747–1756.
Silcox, G.D., Kelly, K.E., Crosman, E.T., Whiteman, C.D., Allen, B.L., 2012. Wintertime PM2.5 concentrations during persistent, multi-day cold-air pools in a mountain valley. Atmos. Environ. 46, 17–24.
Sorek-Hamer, M., Just, A.C., Kloog, I., 2016. Satellite remote sensing in epidemiological studies. Curr. Opin. Pediatr. 28, 228–234.

Su, J.G., Allen, G., Miller, P.J., Brauer, M., 2013. Spatial modeling of residential woodsmoke across a non-urban upstate New York region. Air Qual. Atmos. Health 6, 85–94.
van Donkelaar, A., Martin, R.V., Levy, R.C., da Silva, A.M., Krzyzanowski, M., Chubarova, N.E., Semutnikova, E., Cohen, A.J., 2011. Satellite-based estimates of ground-level fine particulate matter during extreme events: a case study of the Moscow fires in 2010. Atmos. Environ. 45, 6225–6232.
Wang, S.Y., Gillies, R.R., Martin, R., Davies, R.E., Booth, M.R., 2012. Connecting sub-seasonal movements of the winter mean ridge in Western North America to inversion climatology in Cache Valley, Utah. J. Appl. Meteorol. Climatol. 51, 617–627.
Whiteman, C.D., Hoch, S.W., Horel, J.D., Charland, A., 2014. Relationship between particulate air pollution and meteorological variables in Utah's Salt Lake Valley. Atmos. Environ. 94, 742–753.
Xiao, Q., Zhang, H., Choi, M., Li, S., Kondragunta, S., Kim, J., Holben, B., Levy, R.C., Liu, Y., 2016. Evaluation of VIIRS, GOCI, and MODIS collection 6AOD retrievals against ground sunphotometer observations over East Asia. Atmos. Chem. Phys. 16, 1255–1269.
Xiao, Q.Y., Wang, Y.J., Chang, H.H., Meng, X., Geng, G.N., 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.
Zou, B., Wang, M., Wan, N., Wilson, J.G., Fang, X., Tang, Y.Q., 2015. Spatial modeling of PM2.5 concentrations with a multifactoral radial basis function neural network. Environ. Sci. Pollut. Res. 22, 10395–10404.