Remote Sensing of Environment xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Full-coverage high-resolution daily $PM_{2.5}$ estimation using MAIAC AOD in the Yangtze River Delta of China

Qingyang Xiao[a], Yujie Wang[b,c], Howard H. Chang[d], Xia Meng[a], Guannan Geng[a], Alexei Lyapustin[b,c], 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 Biostatistics and Bioinformatics, Emory University, Rollins School of Public Health, Atlanta, GA, USA*

ARTICLE INFO

Keywords:
$PM_{2.5}$
MAIAC
Chemical transport model (CTM)
Multiple imputation
Gap filling
Cloud fraction

ABSTRACT

Satellite aerosol optical depth (AOD) has been used to assess population exposure to fine particulate matter ($PM_{2.5}$). The emerging high-resolution satellite aerosol product, Multi-Angle Implementation of Atmospheric Correction (MAIAC), provides a valuable opportunity to characterize local-scale $PM_{2.5}$ at 1-km resolution. However, non-random missing AOD due to cloud/snow cover or high surface reflectance makes this task challenging. Previous studies filled the data gap by spatially interpolating neighboring $PM_{2.5}$ measurements or predictions. This strategy ignored the effect of cloud cover on aerosol loadings and has been shown to exhibit poor performance when monitoring stations are sparse or when there is seasonal large-scale missingness. Using the Yangtze River Delta of China as an example, we present a Multiple Imputation (MI) method that combines the MAIAC high-resolution satellite retrievals with chemical transport model (CTM) simulations to fill missing AOD. A two-stage statistical model driven by gap-filled AOD, meteorology and land use information was then fitted to estimate daily ground $PM_{2.5}$ concentrations in 2013 and 2014 at 1 km resolution with complete coverage in space and time. The daily MI models have an average R^2 of 0.77, with an inter-quartile range of 0.71 to 0.82 across days. The overall model 10-fold cross-validation R^2 (root mean square error) were 0.81 (25 μg/m^3) and 0.73 (18 μg/m^3) for year 2013 and 2014, respectively. Predictions with only observational AOD or only imputed AOD showed similar accuracy. Comparing with previous gap-filling methods, our MI method presented in this study performed better with higher coverage, higher accuracy, and the ability to fill missing $PM_{2.5}$ predictions without ground $PM_{2.5}$ measurements. This method can provide reliable $PM_{2.5}$ predictions with complete coverage that can reduce bias in exposure assessment in air pollution and health studies.

1. Introduction

Ambient air pollution, mostly $PM_{2.5}$ (fine particulate matter with an aerodynamic diameter of 2.5 μm or less), is responsible for > 3 million premature deaths per year around the world in 2010 (Lim et al., 2013). The highest per capita mortality is reported in the Western Pacific region where persistent high $PM_{2.5}$concentrations together with extremely high population density have raised serious public health concerns (Lelieveld et al., 2015). However, accurately assessing air pollution exposure in this region is challenging due to limited air pollution monitoring. To support exposure assessment for epidemiological studies and risk analysis, satellite aerosol optical depth (AOD) with global coverage, relatively high resolution, and a long data record has been employed to predict air pollution levels in the past decade (Kloog et al., 2012; Liu et al., 2005; Ma et al., 2016b). Previous studies indicated that satellite data can effectively extend ground air quality monitoring networks, but are challenged by non-random missingness due to cloud/snow cover, high surface reflectance, and extremely high aerosol loading that can be misclassified as cloud (Tao et al., 2012; Van Donkelaar et al., 2011). The non-random missingness in AOD retrievals may lead to bias in exposure assessment due to potential systematic differences in $PM_{2.5}$ concentrations when AOD is missing or retrieved. Zheng et al. (2016) reported that the accuracy of annual $PM_{2.5}$ predictions was lower than daily $PM_{2.5}$ predictions due to missingness in AOD, even after correcting annual $PM_{2.5}$ predictions with ground measurements. Other researchers raised concerns that large-scale seasonal missingness in satellite AOD will limit its usage in exposure assessment (Li et al., 2015; Ma et al., 2016a).

* 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).

http://dx.doi.org/10.1016/j.rse.2017.07.023
Received 17 March 2017; Received in revised form 29 May 2017; Accepted 21 July 2017

To improve the coverage of $PM_{2.5}$ predictions and reduce bias in exposure assessment, various gap-filling methods have been proposed recently. One strategy is to develop regional retrieval algorithms that are more suitable for local geographic conditions and atmospheric characters to retrieve more AOD pixels. For example, Li et al. (2012) improved the AOD retrieval algorithm and successfully retrieved AOD over bright targets in urban areas of north China during winter time where the MODIS Dark Target algorithm has failed. Van Donkelaar et al. (2011) relaxed the cloud screening criteria of MODIS Dark Target algorithm when studying the Moscow fire event in 2010, leading to a 21% increase in AOD coverage. Although this strategy can significantly increase the coverage and potentially improve the accuracy of satellite AOD, it is restricted to specific study regions and cannot fill missing AOD with true cloud coverage. Another strategy is to use spatial statistical models to estimate missing retrievals from the spatio-temporal autocorrelation of $PM_{2.5}$. For example, Just et al. (2015) used regional daily average $PM_{2.5}$ concentration and spatial smooth function to fill in missing $PM_{2.5}$ predictions. Kloog et al. (2012) used inverse probability weighting to address the non-random missingness when fitting prediction models, and then interpolated the missing $PM_{2.5}$ predictions using $PM_{2.5}$ predictions or measurements in surrounding grid cells with spatial smoothing. This method can improve the prediction coverage, but by relying on measurements from monitoring stations, it cannot fill missing data when predicting historical $PM_{2.5}$ concentrations before the establishment of air quality monitoring network, and it may exhibit poorer performance if the monitoring networks are sparse or when data over large geographical regions are missing. For example, in Southeastern China, monsoon season leads to several months of rainy and cloudy weather covering several provinces. Additionally, since the spatial pattern was normally fitted monthly or seasonally, it may underestimate the variance of $PM_{2.5}$. Moreover, complex cloud-aerosol interaction has been reported by previous studies (Alam et al., 2014; Myhre et al., 2007). Cloud cover is associated with meteorological conditions that affect aerosol production and deposition (Yu et al., 2015), thus $PM_{2.5}$ concentrations may not be spatially similar under versus outside a cloud and filling $PM_{2.5}$ concentrations from nearby predictions may introduce error.

In addition to remote sensing techniques, chemical transport models (CTM), such as GEOS-Chem (Bey et al., 2001) and CMAQ (Byun and Schere, 2006), have also been widely used to characterize atmospheric aerosol distribution, including $PM_{2.5}$ concentrations, $PM_{2.5}$ composition, and AOD. However, the accuracy of CTM simulations depend on the emissions inventory, meteorological input data as well as parameterization of chemical and physical processes included in the model (Stern et al., 2008). Previous studies reported that the prediction error of CTM model varied spatially and seasonally (Appel et al., 2008; Appel et al., 2012), and biased health effect estimates in epidemiological study (Butland et al., 2013). For example, Appel et al. (2012) reported that CMAQ overestimated $PM_{2.5}$ by > 30% over North America and underestimated $PM_{2.5}$ by up to 55% in winter in Europe. Quennehen (2015) evaluated seven models' performance in predicting ozone and aerosols over East Asia. They showed an overestimation in black carbon and sulfate aerosols in urban regions in China, as well as a general underestimation in scattering aerosols in the boundary layer, due to errors in emissions inventory and physical processing. Although fusing ground measurements and model simulations could improve prediction accuracy (Friberg et al., 2016), the error in CTM simulations may not be fully corrected in the fusion results. Moreover, although the spatial resolution of some CTM simulations can be as high as 4 km in regional studies, typical model simulations are at a relatively low spatial resolution (> 10 km), thus can hardly detect local-scale pollution variability that may be critical for some epidemiological studies (Punger and West, 2013).

In this study, we propose a method that brings together the emerging satellite aerosol product and CTM simulations by multiple imputation to fill missing AOD. Taking advantage of the high resolution and high accuracy of the latest MAIAC satellite AOD and the complete coverage of CTM AOD, our model provide high-accuracy $PM_{2.5}$ predictions with complete coverage at a 1-km resolution. By filling in the missingness in AOD rather than in $PM_{2.5}$ predictions, this model also reduced systemic prediction error by including the meteorology and land use information of all the grid cells in model development. This method is generalizable and can provide high-quality $PM_{2.5}$ predictions in other regions, especially in regions with large-scale missingness in AOD.

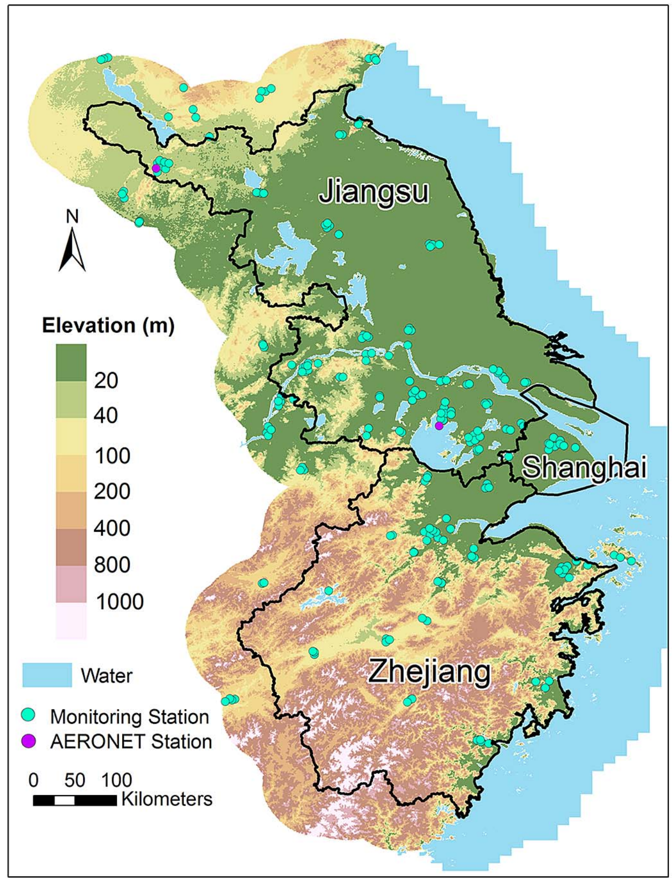

Fig. 1. Study region with a 50-km buffer, showing air quality monitoring stations and AERONET stations in the modeling region.

2. Data and methods

2.1. Study region

The study region (about 200,000 km^2) covers the Yangtze River Delta of China including Jiangsu Province, Zhejiang Province and Shanghai Metropolitan area (Fig. 1). It is one of the most populated regions on earth with approximately 156 million residents in 2010. This region is affected by summer monsoon with rainy and cloudy weather. A 50-km buffer was used in data collection and model development to ensure that gap-filled AOD and estimated $PM_{2.5}$ concentrations are of the same accuracy near the boundary as in the rest of the study domain.

2.2. Datasets

2.2.1. MAIAC AOD data

The latest AOD data retrieved by the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm from measurements of the Aqua (crossover at 1:30 pm local time) and Terra (crossover at 10:30 am local time) Moderate Resolution Imaging Spectroradiometer