Gap-filling MODIS daily aerosol optical depth products by developing a spatiotemporal fitting algorithm

Tao Zhang a, Yuyu Zhou a, Kaiguang Zhao b, Zhengyuan Zhu c, Ghassem R. Asrar d and Xia Zhao e

*Department of Geological and Atmospheric Sciences, Iowa State University 50011, Ames, IA, USA; aSchool of Environment and Natural Resources, Ohio Agricultural Research and Development Center, The Ohio State University 44691, Wooster, OH, USA; bDepartment of Statistics, Iowa State University 50011, Ames, IA, USA; cUniversities Space Research Association 21046, Columbia, MD, USA; dKey Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University 430079, Wuhan, Hubei, China

ABSTRACT
Aerosol loadings and their spatial distribution are among the most important atmospheric information needed for a range of applications such as air quality monitoring, climate research, and public health. A key measure of aerosol quantity is aerosol optical depth (AOD) and it has been routinely observed from space by Earth observing satellites/instrument, especially the Moderate Resolution Imaging Spectroradiometer (MODIS). Despite its global coverage and daily temporal resolution, the MODIS Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD product is fraught with missing values, severely limiting its use. A gap-filling method which is suitable for large-area application with high efficiency to obtain gapless AOD with reasonable spatial pattern and complete coverage is still lacking. Here, we proposed a novel spatiotemporal fitting algorithm to gap-fill the daily MODIS AOD product. Our algorithm is a multi-stage method aimed to address the non-stationary nature of AOD time series. First, the trend of daily AOD in a year in each pixel was fitted via smoothing splines and the residual was derived based on the original data and the trend. Second, the residual was spatially interpolated, leveraging the spatial correlation between the target pixel and the neighboring pixels. Third, the actual AOD was calculated as the sum of the trend and interpolated residual. We tested the algorithm against ground-based AOD data from 2011 to 2018 in China and further evaluated it via cross-validation at the global scale based on 10 selected MODIS tiles. Compared to the ground-reference AOD, the RMSE of our gapless datasets were 0.24 and 0.27 for Terra and Aqua, respectively; and the cross-validation showed a RMSE ranging from 0.045 to 0.055 (Terra) and 0.047 to 0.057 (Aqua) under different missing ratios. The novel gap-filling method outperforms the Interpolation-based Correlation Weighting (ICW) and Inverse Distance Weighting (IDW) algorithms in accuracy. Meanwhile, the gapless AOD using the novel algorithm shows lower accuracy than original MAIAC AOD, similar accuracy with the AOD from the Long-term Gap-free High-resolution Air Pollutants (LGHAP) concentration dataset, higher accuracy than the AOD from the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). Overall, the accuracy of gapless AOD using this algorithm meets the need of typical applications in relevant studies. The proposed algorithm is transferable to other regions, with the potential to be used even operationally and efficiently for generating accurate gapless global daily AOD datasets with the input of only MODIS MAIAC AOD data.

1. Introduction
Atmospheric aerosols are the solid and liquid particles suspended in the air, including dust, smoke, and haze (L. Li, Yang, and Wang 2014; J. Yang and Hu 2018). They play a critical role in regulating the Earth climate and environments, with enormous implications for air quality, public health, and ecosystem functioning (Yu et al. 2011). For example, the amount of aerosols determines how much solar radiation is scattered and absorbed by the atmosphere and also alters the cloud formation and their optical properties, thereby affecting temperature and precipitation (Zheng et al. 2020; Rosenfeld et al. 2008; Ramanathan et al. 2001). Aerosols, especially those fine particulate particles with diameters of less 2.5 μm, are known to spread biological organisms, reproductive materials, and pathogens as well as cause or enhance respiratory, cardiovascular, infectious, and allergic diseases (Pöschl 2005), which is concerning in developing countries such as China under rapid urbanization (Lu 2022).
et al. 2020; Gong et al. 2020; X. Li et al. 2020; Zhou et al. 2018). Meanwhile, aerosols also limit solar energy production (Lolli 2021; L. Zhang et al. 2021). Thus far, a key measure to quantify the loadings of aerosols is aerosol optical depth (AOD) – a spectral parameter indicating the amount of aerosols in an atmospheric column. Over large regions, AOD is generally derived and estimated from satellite observations.

Satellite-observed AOD is generally recognized as the best source of information for regional and global studies, especially compared to the relatively scarcity of measurements from ground-based stations, e.g. the AERosol ROBotic NETWORK (AERONET) (Holben et al. 1998). Examples of satellites capable of measuring AOD include the Advanced Very High Resolution Radiometer (AVHRR) (Hsu et al. 2017; Sayer et al. 2017), the Moderate Resolution Imaging Spectroradiometer (MODIS) (Remer et al. 2013, 2005), the Multi-angle Imaging Spectro Radiometer (MISR) (Kahn et al. 2010), and Ozone Monitoring Instrument (OMI) (Ahn, Torres, and Jethva 2014). Among the existing products, the MODIS-based AOD products are the most widely used because MODIS provides data with high spatial and temporal resolutions (i.e. 250 m, 500 m, and 1000 m with a repeat cycle of 1 to 2 days), globally (Gupta et al. 2016; Levy et al. 2013; Hsu et al. 2013; Lyapustin and Wang 2018). Despite their recognized values, satellite data have their own limitations. Probably the most critical limitation is the spatial and temporal gaps, due to either the limited capability of polar orbiting sensors especially over high-altitude and mountain terrains (Ningombam et al. 2021a; Cavazzani, Zitelli, and Ortolani 2015) or unfavorable environmental conditions such as clouds, high surface reflectance, extremely high aerosol concentrations (Levy et al. 2013; Wang et al. 2019a; Gupta et al. 2016). Gaps are also attributed to the lack of overlaps between adjacent satellite orbits (Sayer et al. 2015). A gapless remotely sensed AOD dataset is particularly desirable when assessing impacts of air pollution on human health. AOD is typically used as a proxy to empirically estimate atmospheric pollutants (Lin et al. 2016; Wei et al. 2019; Makar et al. 2018). Ignoring the missing values in AOD were found to cause large uncertainties in estimated concentrations of atmospheric pollutants (e.g. PM$_{2.5}$) (Ma et al. 2016; Q. Xiao et al. 2017). Specifically, over high-altitude and mountain terrain, there are a significant number of missing AOD data and the observed AOD data show a low accuracy because the occasionally occurred very low AOD values were within the retrieval errors (You et al. 2020; Ningombam et al. 2021a). Therefore, filling gaps in AOD data has been advocated as a remedy to further augment the potential of satellite-derived AOD products.

Many gap-filling algorithms have been proposed to improve AOD products. These algorithms can be generally divided into three groups (Table 1). The first group relies on fusing AOD datasets from different satellites to reduce the probability of gaps (Jinnagara Puttaswamy et al. 2014; Singh, Gautam, and Venkatachalam 2017; Tang, Bo, and Zhu 2016; Wei et al. 2019; Xu et al. 2015; D. Yang and Gueymard 2021; Go et al. 2020; Bai et al. 2022). The second group focus on training empirical models to build

| Groups                     | Methods                                      | Authors                                      |
|----------------------------|----------------------------------------------|----------------------------------------------|
| Data fusion-based          | Linear regression                            | Wei et al. 2019                              |
|                            | Bayesian                                     | Singh, Gautam, and Venkatachalam 2017       |
|                            | Bayesian maximum entropy (BME)               | Tang, Bo, and Zhu 2016                      |
|                            | Spatial statistical data fusion              | Jinnagara Puttaswamy et al. 2014            |
|                            | Maximum likelihood estimate (MLE)            | Xu et al. 2015                              |
|                            | Quantile regression (QR), QR neural network, QR forest | Go et al. 2020                            |
|                            | Tensor completion                            | D. Yang and Gueymard 2021                   |
| Collaborative variable-based| Random forest                                | Bai et al. 2022                             |
|                            | Multiple imputation                          | Zhao et al. 2019; Jiang et al. 2021         |
| Interpolation-based        | NDVI-based Weighted Linear Regression        | T. Zhang et al. 2017                        |
|                            | IDW                                          | L. Xiao, Lang, and Christakos 2018          |
| Hybrid                     | Ordinary Kriging                             | Yu et al. 2011                              |
|                            | Co-Kriging, Regression Kriging               | Y. Wu et al. 2011; Ruiz-Arias et al. 2013   |
|                            | Spatiotemporal kriging                       | Singh, Venkatachalam, and Gautam 2017       |
|                            | Linear regression & Ordinary Kriging         | J. Yang and Hu 2018                        |
|                            | Spatial-temporal hybrid fusion considering aerosol variation mitigation (ST-AVM) | Wang et al. 2019a                          |
|                            | Two-step model (TWS)                         | Chi et al. 2020                             |
statistical relationships between AOD and auxiliary data, and the missing AOD values are estimated from the fitted relationships based on auxiliary data (T. Zhang et al. 2017; Jiang et al. 2021; Q. Xiao et al. 2017; Zhao et al. 2019) by using algorithms such as random forest (Zhao et al. 2019; Jiang et al. 2021). The third group leverage spatial and/or temporal patterns of the original AOD datasets themselves to fill AOD gaps, based on spatial statistical modeling techniques such as Kriging (J. Yang and Hu 2018; Yu et al. 2011). There are also several studies that combined two or all of these three approaches (Lv et al. 2016; Wang et al. 2019a; Chi et al. 2020).

Despite the successes reported with the existing gap-filling algorithms, one challenge remaining is that algorithms with high efficiency for large-area applications to get gapless AOD with reasonable spatial pattern and complete coverage are still lacking. Of the three common approaches employed in the existing algorithms, merging multiple AOD datasets can usually reduce the gaps but cannot fill all the gaps (e.g. cloud coverage) completely (Wang et al. 2019a; Zhao et al. 2019). Fitting empirical relationships between AOD and auxiliary variables can fill all gaps, but it is computationally demanding and less practical for large scale applications, due to the large amounts of data required (Zhao et al. 2019). Also, quite often the fitted predictive models suffer from under-fitting or over-fitting (Wang et al. 2019a). Likewise, spatial interpolation methods (e.g. Kriging or spatiotemporal Kriging) alleviate the needs for auxiliary data and can fill all the missing values; however, the computation required is immense and these methods often produce unrealistic results or abnormal spatial patterns (e.g. overshooting) when the numbers of valid observations are small. Overall, these drawbacks make the existing algorithms less appealing for generating gapless datasets over extensive areas - a research opportunity that was pursued by this study.

Here, we proposed a gap-filling framework based on a novel spatiotemporal fitting algorithm, which belongs to the third group of gap-filling methods (i.e. spatial and/or temporal interpolation). Our algorithm leverages both the temporal and spatial patterns of observed AODs to overcome the commonly occurred issues in existing methods, i.e. poor efficiency and unrealistic results. Temporally, the algorithm explicitly addresses the nonlinearity and non-stationary (i.e. a non-constant trend in time) nature of AOD time series. Spatially, the algorithm considers the spatial auto-correlation. Briefly, we first used a smoothing spline function to fit the trend for individual pixel and derive the residual based on the original data and the trend. Then, we interpolated the residuals on a daily scale based on the correlation between target grids and their neighboring grids. Finally, the spline-fitted temporal means and the spatially-interpolated residuals were summed to recover the actual AOD. We tested this algorithm using ground-based AOD measurements in China and also performed cross-validation using 10 MODIS tiles randomly selected across the globe.

Figure 1. Study area and locations of AERONET sites (red points).
2. Study area and data

2.1. Study area

This study focused mainly on China (Figure 1), because it represents a populated region that has been facing severe air pollution events with very high atmospheric aerosol loading that affect all aspects of daily life (Wang et al. 2019b; N. Liu et al. 2019; Z. Zhang et al. 2019; Zhao et al. 2019). Regions like this, which are increasing globally, are in great needs for spatiotemporally continuous and gapless AOD datasets. It is also a geographically extensive area, large enough to test the general applicability of our algorithm. Another reason for choosing China is the availability of ground-based stations scattered across the country, providing high-fidelity independent AOD measurements to evaluate our gap-filling algorithm. Besides China, we also used a total of 10 MODIS tile globally, each approximately 1200 km × 1200 km in size, to test our algorithm via cross-validation. To improve spatial representativeness, the tiles were randomly selected across the globe with diverse landforms and land cover types, which included h08v05, h12v04, h13v10, h18v04, h20v11, h21v03, h22v06, h25v06, h27v05, and h30v11.

2.2. MODIS AOD products

The satellite AOD product we used is the MODIS Collection 6 Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD product (MCD19A2) that is based on 550 nm band with a spatial resolution of 1 km (Lyapustin et al. 2018). Due to the 1 km resolution, the MAIAC AOD data can reveal numerous hotspots with high AOD values, and this product is generally considered to be more accurate than other MODIS AOD products such as MODIS Dark-Target (DT) and Deep-Blue (DB) (N. Liu et al. 2019; Tao et al. 2019). We compiled all MODIS MAIAC daily gridded AOD products from both Terra and Aqua satellites in China, collected from the year 2011 to 2018, using Equation (1) for each grid.

\[ \text{Missing ratio} = 100 \times \frac{N_{\text{missing}}}{N_{\text{total}}} \]  

(1)

where \(N_{\text{missing}}\) and \(N_{\text{total}}\) are the number of days without valid observations and the number of days for the whole study period.

The overall missing ratio (i.e. fractions of gaps) is typically above 50% over most of China. The lowest missing values is 39% over the Gobi Desert of northern China due to low cloud fraction, and highest missing values is larger than 80% in southern China (mainly in high-altitude and mountain regions) (Figure 2, Table S1). The MAIAC AOD data of randomly selected 10 tiles (Section 2.1) for cross-validation purpose include tiles from Terra and Aqua satellites for the year 2005, 2010, and 2015. Different missing ratios were generated in these tiles in the cross validation (Section 3.3).

2.3. AERONET AOD observations

Ground reference data used to evaluate the gap-filled MODIS datasets were obtained from the AERONET. AERONET is a network of ground stations equipped with Sun photometers and is the most comprehensive datasets for point measurement of AOD across the globe. AOD measurements from AERONET were collected in the spectral region of 0.34–1.06 μm, approximately every ~15 min (https://aeronet.gsfc.

Figure 2. The multi-year average percent of missing observations in MODIS MAIAC AOD data from terra and aqua satellites in China from 2011 to 2018.
and multiple products were available from this network (Giles et al. 2019; Holben et al. 1998). We chose to use the Level 2.0 AOD data of Version 3.0, which is deemed to have low uncertainties. We interpolated AOD parameters at 550 nm using Ångström exponent in the two neighboring bands at 500 nm and 675 nm to make it comparable with MODIS MAIAC AOD (N. Liu et al. 2019). Within China, there exists a total of 23 AERONET sites; data for all the sites from 2011 to 2018 were compiled and matched with MODIS data for evaluation (Figure 1, Table 2).

2.4. Existing gapless AOD datasets

Existing gapless AOD datasets from the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) (Gelaro et al. 2017) and a long-term gap-free high-resolution air pollutants concentration dataset (LGHAP) (Bai et al. 2022) were used as reference datasets to be compared with our gapless AOD datasets using the novel gap-filling algorithm in this study. MERRA-2 is the first satellite reanalysis dataset to assimilate aerosol information of the earth system and the MERRA-2 AOD was also assimilated from multiple AOD datasets such as the Advanced Very High Resolution Radiometer (AVHRR), Multi-angle Imaging Spectroradiometer (MISR) and MODIS satellite AOD datasets, and AERONET ground-based observations (Shaheen, Wu, and Aldabash 2020; Yousefi et al. 2020; Sogacheva et al. 2020). In this study, the six-hourly [8:00–14:00 local time] instantaneous AOD (at wavelength 550 nm) data at a resolution of 0.5° latitude by 0.625° longitude in 2015 was used for comparison. The LGHAP AOD was generated by using multimodal aerosol data acquired from diverse sources (e.g. MERRA-2, MODIS, and AERONET) via a tensor flow based fusion method, which has higher accuracy than that of MERRA-2 AOD in China (Bai et al. 2022). It is the first long-term gap-free high resolution AOD dataset in China, which covers the period of 2000–2020 with daily 1 km resolution. The daily LGHAP AOD in 2015 was used for comparison in this study.

3. Methodology

The core of our proposed gap-filling method is a spatiotemporal fitting algorithm. The overall flowchart of the framework is presented in Figure 3, wherein MODIS MAIAC AOD data were firstly pre-processed by applying quality control flags and mosaicking AOD data from individual orbits into tiles. The spatiotemporal algorithm was then applied to the pre-processed MAIAC AOD image time series. First, the trend of daily AOD in a year in each pixel was fitted via smoothing splines and the residual was derived based on the original data and the trend. Second, the residual was spatially interpolated, leveraging the auto-spatial correlation between the target pixel and the neighboring pixels. Third, the actual AOD was calculated as the sum of the trend and interpolated residual. Details of each component of the framework are described below.

3.1. Data pre-processing

In pre-processing the MAIAC AOD data, we first selected AOD grids with high-quality assurance based on their quality control flag (i.e. the AOD_QA layers) (Lyapustin et al. 2018). Grids were considered to be suitable only if they satisfy all the following conditions: QA.CloudMask = Clear, QA.AdjacencyMask = Clear, and QA.LandorWater = Best quality/Climatology AOD (0.02). Specifically, in the third condition, we kept the default value of 0.02 in grids at high altitudes above 4.2 km (Land)/3.5 km, as the original MODIS MAIAC AOD product, to preserve the spatial continuity of
AOD. Air pollution and population density are very low in the high-altitude regions (mainly in Tibetan Plateau), leading to very low AOD. This filtering for high-quality data was applied to images of individual orbits; then, the resulting high-quality images were mosaiced into the standard MODIS tiles for each day. We used and average value when multiple measurements are available for a given grid.

3.2. Spatiotemporal fitting

3.2.1. Overview

The spatiotemporal fitting algorithm includes three steps (Figures 3 and 4). First, the original observation of a day for each grid was considered to include two components: an overall mean plus some daily fluctuations. The mean signal was fitted as a trend using a smoothing spline (Perperoglou et al. 2019), while the residual signal not modeled by the spline fit was considered as the daily fluctuations. The overall mean captures the temporal pattern of each grid, representing a stable part so that a smoothing fitting-model such as spline is a suitable tool for approximating it. The residual represents the random fluctuations in the observations. Of particular note, the residuals are “errors” from the view point of the spline smoothing, but they are actually true signals that will be further analyzed in the second step of proposed method. Therefore, the fitted overall mean generally had complete spatial coverage, and the gaps still remained in residual images. Second, the

Figure 3. Framework of the novel spatiotemporal gap-filling method for MODIS MAIAC AOD data from terra or aqua satellites.

Figure 4. An example of spatiotemporal fitting algorithm implemented on AOD data for the cross validation by excluding 75% valid data for the tile h27v05 (located in eastern China) on the day 187 in 2015.
residuals for individual days were interpolated based on the correlation between the target grid and its neighboring grids (Section 3.2.2). Finally, the sum of overall means and interpolated residuals was used to obtain the gapless AOD for each day.

In using the splines to fit the trends, we considered only those pixels that had a sufficient number of observations (e.g. >10) each year. Pixels with less than 10 observations, which are usually water or high mountain areas (Figure S2, Table S1), were treated separately by using an optional post-processing step. That is, we circularly used average values in moving windows with size of 9 by 9 grids to fill these gaps, since these regions are usually small in size and may not be of interest to the public.

### 3.2.2. Interpolation-based correlation weighting (ICW)

An Interpolation-based Correlation Weighting (ICW) technique, which was inspired by the Inverse Distance Weighting (IDW) interpolation method (Shepard 1968), was proposed to interpolate the residual for each day of a year. In the IDW method, the value of a target site (a site means a pixel or grid, specifically, a grid area was calculated for a pixel under MODIS projection) was estimated by calculating the weighted average of neighboring sites and the weights were calculated based on the distance between the target site and its neighboring sites, that is, the closer neighbors receive larger weights than that of further neighbors. In the ICW method, the value of target site \( S_0 \) at time \( t \) was estimated using Equation (2), i.e. the weighted average of the estimated values from the neighboring sites based on linear regression functions (Equation (4)). The weight between the target site and one of its neighboring sites was determined by Pearson’s correlation coefficient and was calculated based on the matched time series (days of a year) values of the two locations using Equation (3). The estimated value at the target site based on one of its neighboring sites was calculated using Equation (4), which was fitted using ordinary least square (OLS) method.

\[
V_{S_0}(t) = \sum_{i=1}^{n} w(S_0, S_i) \cdot V_{S_0}(S_i, t) \tag{2}
\]

\[
w(S_0, S_i) = \frac{\text{cor}(S_0, S_i)}{\sum_{i=1}^{n} \text{cor}(S_0, S_i)} \tag{3}
\]

\[
V_{S_0}(S_i, t) = a_i + \beta_i \cdot V_{S_0}(t) \tag{4}
\]

where \( V_{S_0}(t) \) is the estimated value of the target site at time \( t \); \( w(S_0, S_i) \) is the weight of the \( i \)-th neighboring site \( S_i \); \( V_{S_0}(S_i, t) \) is the estimated value of the target site at time \( t \) based on the \( i \)-th neighboring site; \( n \) is the number of neighboring sites; \( \text{cor}(S_0, S_i) \) is the Pearson’s correlation coefficient between \( S_0 \) and \( S_i \); \( a_i \) and \( \beta_i \) is the intercept and slope of the linear function between the target site and its \( i \)-th neighboring site, respectively; and \( V_{S_0}(t) \) is the value of the \( i \)-th neighboring sites at time \( t \).

### 3.2.3. Implementation of the ICW method

The ICW method was implemented as follows. First, each MODIS tile was divided into small blocks with the size of 10 by 10 grids and the block center grids were used as neighboring grids for interpolating missing residuals. That is, missing residuals in a block can be interpolated based on the values from the eight neighboring block center grids when the calculation extent is limited in a local window with the size of 3 by 3 blocks.

Second, in order to make sure that all the block center grids have valid data for the estimation of other grids, the missing values in the block center grids were interpolated based on IDW method. The steps used are as follows: (a) computing the average value of each block; (b) resampling the original MODIS tile (1200 by 1200 grids) to get a small image (120 by 120 grids) and the value of each grid in the new image is the average value of a block in the original MODIS tile; (c) interpolating missing values in the new image based on IDW method; (d) assigning interpolated values to the block center grids of the original MODIS tile when the original block center grid does not have valid data. These estimated values in the new image represent the average values of corresponding blocks without valid observations in the original MODIS tile, which are also the values of pixels in the block centers. In step (d), we set these estimated values in the block centers to ensure that all block centers have valid observations and block centers can be used as neighboring pixels.

Third, in order to reduce the possible boundary effects of the interpolated residuals between neighboring blocks, for each grid of a given block, only one of the neighboring block center grids which has the largest correlation coefficient with the target grid was used for estimation. This is because when all the grids
in a block were interpolated based on eight center grids of neighboring blocks, the boundary grids from different blocks were estimated based on different combination of block center grids, which may lead to systematic deviation. Therefore, Equations (2) and (4) can be simplified to Equation (5).

$$V_{S_0}(t) = a_m + \beta_m \cdot V_m(t)$$  \hspace{1cm} (5)

where $a_m$ and $\beta_m$ are the intercept and slope of the linear function between target grid and its neighboring grid with the maximum correlation coefficient with it, which is fitted using ordinary least square (OLS) method based on the matched time series (days of a year) values of the two locations; $V_m(t)$ is the value of the neighboring grid with the maximum correlation coefficient at time $t$.

Finally, all tiles were mosaiced as a region to avoid boundary effects between neighboring tiles, and residual of the block center grids were interpolated for each day.

### 3.3. Accuracy assessment

We evaluated the performance of the proposed gap-filling method in two ways. First is the evaluation of the gapless AOD data against the ground-based AOD from AERONET sites in China with ground measurements from 2011 to 2018. Second is the cross-validation based on randomly selected 10 MODIS tiles (Figure S1) at the global scale in representative years (2005, 2010, and 2015, which are equally distributed among the available years). In order to obtain a sufficient number of AOD pairs between MODIS retrievals and AERONET measurements, we matched the average AERONET measurement within ±60 minutes of MODIS overpass time and the average AOD value of 3 by 3 grids centered on the AERONET site (Wang et al. 2019b; Tang, Bo, and Zhu 2016). Thereafter, four statistical metrics including Correlation Coefficient (R), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Expected Error (EE) envelope were used to assess the accuracy of the proposed method, using Equations (6)–(8) (N. Liu et al. 2019). According to the definition of EE envelope in Equation (9), we also calculated the proportion within the EE envelope to evaluate the percentage of qualified data using Equation (10).

$$R = \frac{\sum_{i=1}^{N} (\text{AOD}_{\text{sat}} - \overline{\text{AOD}}_{\text{sat}}) \times (\text{AOD}_{\text{aero}} - \overline{\text{AOD}}_{\text{aero}})}{\sqrt{\sum_{i=1}^{N} (\text{AOD}_{\text{sat}} - \overline{\text{AOD}}_{\text{sat}})^2} \sum_{i=1}^{N} (\text{AOD}_{\text{aero}} - \overline{\text{AOD}}_{\text{aero}})^2}$$  \hspace{1cm} (6)

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\text{AOD}_{\text{sat}} - \overline{\text{AOD}}_{\text{aero}})^2}$$  \hspace{1cm} (7)

$$\text{MAE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left| \text{AOD}_{\text{sat}} - \overline{\text{AOD}}_{\text{aero}} \right|}$$  \hspace{1cm} (8)

$$\text{EE} = \pm (0.05 + 0.15 \times \text{AOD})$$  \hspace{1cm} (9)

$$\text{AOD} - \text{EE} \leq \text{AOD}_{\text{sat}} \leq \text{AOD} + \text{EE}$$  \hspace{1cm} (10)

where $\text{AOD}_{\text{sat}}$ and $\overline{\text{AOD}}_{\text{aero}}$ are the satellite AOD retrievals and ground measurements from AERONET sites, respectively; $\overline{\text{AOD}}_{\text{sat}}$ and $\overline{\text{AOD}}_{\text{aero}}$ are the corresponding averages; $N$ is the number of matches between the satellite observations and the ground AERONET measurements.

For cross validation, we selected 19 days (5% of 365 days) with the largest coverage of valid high quality observations in each year. That is, we calculated the ratio of valid observations for each day and rank the daily ratios for each year to find 19 days with maximum ratios. Then, for each selected day, we used the spatial patterns of missing grids from another day of the year to exclude specific percentage (i.e. 25%, 50%, and 75%) of valid data in the image. These excluded data were used for validation, and the rest valid data were used as the input of the gap-filling algorithm. Finally, under each exclusion scenario, the daily MAIAC AOD data of a year was gap-filled based on the proposed gap-filling method, and the estimations were compared to the original AOD values to evaluate the accuracy of proposed gap-filling method, and the R, RMSE, and MAE were selected as the accuracy indicators, where $\text{AOD}_{\text{sat}}$ and $\overline{\text{AOD}}_{\text{aero}}$ in Equations (6)-(8) were replaced by the estimated AOD and original MAIAC AOD, respectively.

We compared performance of the novel spatiotemporal gap-filling algorithm with the ICW (i.e. without considering on overall mean) and IDW algorithms.
using the above-mentioned cross validation method at the same spatial and temporal extent (i.e. 10 tiles in 2005, 2010, and 2015).

Moreover, we also compared the accuracy of the gapless AOD data with those of the LGHAP AOD and MERRA-2 AOD in 2015 using the ground-based AOD from AERONET sites in China. As these AOD datasets have different spatial and temporal resolutions, we aggregated them to the same spatiotemporal resolution (i.e. daily 100 km resolution) to do evaluation using the method in literature (Yousefi et al. 2020; Shaheen, Wu, and Aldabash 2020). For each day, our gapless AOD values from Terra and Aqua satellites were averaged at each grid, and then they were resampled to 100 km grids using the average value of 1 km grids within each 100 km grid. The daily LGHAP AOD data were also resampled to 100 km grids using spatial average. The six-hourly [8:00–14:00 local time] MERRA-2 AOD were aggregated using the temporal average values and then they were resampled to 100 km resolution using the nearest neighbour interpolation technique. The ground-based AOD from AERONET sites in the same time period (i.e. 8:00–14:00 local time) were also aggregated using the temporal average values to be used as ground truth for accuracy evaluation of these AOD datasets.

4. Results and discussion

4.1. Accuracy assessment using ground measurements

The gapless AOD data with both the original MAIAC and estimated AOD values show reasonable accuracies according to ground measurements from AERONET sites (Figure 5). As shown in Figure 5a, the gapless AOD data from Terra and Aqua satellites are in good agreement with AERONET measurements with the R values of 0.88 and 0.84, RMSE of 0.24 and 0.27, MAE of 0.13 and 0.16, the within EE point of 59.15% and 57.03%, respectively. Similar accuracies were also obtained in previous study by Chi et al. (2020), indicating the accuracies of our gapless AOD are within reasonably accepted ranges for scientific research and other applications. However, the method proposed by Chi et al. (2020) cannot be used on a global scale due to the time-consuming problem caused by the local window-based calculation, and limited spatial coverage of auxiliary data they used. On the contrary, the gap-filling method developed in this study is suitable for global scale gap-filling studies with high efficiency and without need for auxiliary data. We further explored the correlation between the estimated AOD and original MAIAC AOD and AERONET measurements, and their respective accuracies.

Figure 5. Evaluation of (a) gapless AOD, (b) estimated AOD, and (c) original MAIAC AOD from 2011 to 2018. The upper and lower figures are AOD data from terra and aqua satellites, respectively. The color of points represents their density, where the red color indicates the highest density and the blue color indicates lowest density. The red solid line represents the regression line, the dashed lines are the EE lines, and the black line is the 1:1 line.
(Figure 5b-c, Table 3). The results indicate that the accuracy of estimated AOD is lower than that of the original MAIAc AOD. Specifically, the RMSEs of 0.31 and 0.34 for estimated AOD from Terra and Aqua are higher than the 0.12 and 0.14 for the original MAIAc AOD; the MAE, R and within EE values of former are also lower than the latter. Moreover, the estimated AOD tended to underestimate the actual AOD compared to ground measurements from AERONET sites with the slope of 0.81 for both Terra and Aqua satellites (Figure 5b). Similar results were also found at most AERONET sites (Table 3). As shown in Table 3, the MAIAc AOD shows higher correlation with ground measurements from the AERONET than that of estimated AOD because the MAIAc AOD is the original MODIS AOD data while the estimated AOD is imputed based on valid values from MAIAc AOD. The estimated AOD could be interpolated from neighboring grids with lower/higher MAIAc AOD values, leading to possible underestimation/overestimation of some high/low AOD values in the missing grids. Specifically, correlation coefficients at Muztagh_Ata, NAM_CO, and QOMS_CAS sites are low, possibly due to low accuracy of AOD in high-altitude and mountain terrain, which has been found in literature (Ningombam et al. 2021b). As we kept default setting of 0.02 at high altitude region in the original MAIAc AOD product, the estimated AOD tended to underestimate the actual values when there are higher concentrations of air pollutants. This finding is mainly caused by insufficient satellite retrievals when aerosol loadings are extremely high or low, which has also been found in the study by using random forest model (Jiang et al. 2021).

Comparison of temporal pattern between satellite and ground AOD measurements at the representative AERONET sites indicates that the gapless AOD data can effectively capture the temporal variations of AOD which is promising for providing AOD information of locations without ground AOD measurements (Figure 6). Three representative sites (i.e. Beijing, XiangHe, and Taihu) with different overall missing ratios (52.2%, 54.7%, and 89.9%, respectively) were selected to explore the temporal pattern of the gapless AOD data against the ground-based measurements. The results illustrate that the gapless AOD data can generally capture the temporal pattern of ground measurements for different missing ratios of the original MAIAc AOD data (Figure 6), indicating the gapless AOD data can be used to effectively expand the spatial and temporal coverage of ground AOD measurements. However, there are still several days

| Site name          | RMSE  | Within EE (%) | Slope  | Overall missing ratio of original MAIAc AOD (%) |
|--------------------|-------|---------------|--------|-----------------------------------------------|
|                    | EST   | OB            | EST    | OB    | EST  | OB    | Terra | Aqua |
| Bangqiao           | 0.29  | 0.08          | 0.41   | 0.54  | 40.00| 70.00| 0.41  | 0.82 | 85.9 | 90.2 |
| Beijing            | 0.48  | 0.16          | 0.81   | 0.95  | 47.47| 77.65| 0.79  | 1.03 | 53.1 | 52.2 |
| Beijing_RADI       | 0.33  | 0.12          | 0.95   | 0.90  | 41.46| 81.18| 0.72  | 0.85 | 54.5 | 56.9 |
| Beijing-CAMS       | 0.45  | 0.15          | 0.80   | 0.96  | 43.49| 75.62| 0.84  | 1.07 | 52.2 | 51.7 |
| Chen-Kung_Univ     | 0.27  | 0.15          | 0.65   | 0.80  | 41.35| 60.36| 0.73  | 0.97 | 79.6 | 84.0 |
| Chiayi             | 0.37  | 0.09          | 0.53   | 0.92  | 42.38| 78.85| 0.61  | 0.97 | 82.2 | 90.1 |
| Douliu             | 0.42  | 0.13          | 0.42   | 0.88  | 36.33| 70.51| 0.53  | 0.88 | 82.2 | 89.3 |
| EPA-NCU            | 0.24  | 0.1           | 0.73   | 0.94  | 44.65| 68.54| 0.77  | 1.07 | 85.7 | 85.3 |
| Heng-Chun          | 0.18  | 0.05          | 0.79   | 0.95  | 51.57| 91.67| 0.80  | 1.06 | 93.6 | 96.2 |
| Hong_Kong_Sheung   | 0.27  | 0.08          | 0.68   | 0.95  | 39.04| 81.29| 0.80  | 0.98 | 85.0 | 85.7 |
| Kaohsiung          | 0.23  | 0.13          | 0.56   | 0.79  | 45.58| 62.07| 0.71  | 1.12 | 86.2 | 91.1 |
| Lulin              | 0.17  | 0.12          | 0.70   | 0.76  | 7.15 | 12.94| 1.82  | 2.2  | 82.3 | 93.8 |
| Muztagh_Ata        | 0.10  | 0.05          | 0.03   | 0.42  | 66.00| 83.67| 0.44  | 0.54 | 77.4 | 84.9 |
| NAM_CO             | 0.06  | 0.1           | 0.07   | -0.13 | 82.01| 82.69| 0.15  | 0.08 | 75.1 | 81.3 |
| NCU_Taiwan         | 0.23  | 0.08          | 0.76   | 0.98  | 36.21| 82.93| 0.84  | 1.07 | 85.6 | 85.2 |
| QOMS_CAS           | 0.05  | 0.03          | 0.24   | 0.01  | 85.20| 94.90| 0.23  | 0.27 | 46.6 | 52.6 |
| SACOL              | 0.22  | 0.12          | 0.60   | 0.92  | 48.73| 68.77| 0.8   | 0.93 | 62.0 | 64.3 |
| Taihu              | 0.41  | 0.31          | 0.75   | 0.95  | 26.52| 19.66| 1.09  | 1.25 | 89.9 | 89.9 |
| Taipei_CWB         | 0.28  | 0.1           | 0.66   | 0.90  | 39.02| 66.11| 0.65  | 1.01 | 87.2 | 90.8 |
| XiangHe            | 0.48  | 0.16          | 0.82   | 0.96  | 46.71| 78.26| 0.9   | 1.07 | 53.5 | 54.7 |
| Xinglong           | 0.32  | 0.09          | 0.78   | 0.97  | 46.25| 77.36| 1.02  | 1.19 | 61.1 | 66.0 |
| Xitian             | 0.34  | 0.09          | 0.46   | 0.88  | 22.00| 75.68| 1.06  | 1.06 | 82.3 | 79.6 |
| Zhongshan_Univ     | 0.31  | 0.08          | 0.72   | 0.96  | 33.75| 87.5  | 0.78  | 1.02 | 89.1 | 91.3 |

Note: EST represents estimated AOD, and OB represents observed MAIAc AOD (from both Terra and Aqua satellites). An overall missing ratio means the missing percentage of original MAIAc AOD during the study period within 3 by 3 grids centered at each AERONET site. An overall missing ratio can be calculated using Equation (1), where Nmiss and Ntot are the number of days without valid observations within 3 by 3 grids centered at each AERONET site and the number of days for the whole study period.
with estimated AOD values that are obviously lower than ground measurements when the original AOD product is largely or totally missing. The reduced accuracies of the estimated AOD in Figure 5 and Table 3 were mainly caused by these underestimations though they were only a small portion.

4.2. Accuracy assessment using cross validation

The gap-filling algorithm performs well according to cross-validation at 10 randomly selected tiles across the globe with three missing scenarios (Table 4). As shown in Table 4, the average RMSE of estimated AOD for Terra and Aqua ranges from 0.045 to 0.055 and 0.047 to 0.057, respectively, under different missing scenarios. The lowest RMSE value occurs in 2005 with 25% of valid observations excluded from Terra, while the highest RMSE value occurs in 2010 with 75% of valid observations excluded from Aqua, which varies within a small range of 0.041 – 0.062. The accuracies are also close to the values obtained from previous study by Chi et al. (2020), which a 10-fold cross validation was used and reaching RMSE of 0.05 in 2018.

The results of cross validation also indicated that as the number of missing values in the original MAIAC AOD increases, the accuracy of estimated AOD data tends to decrease, consistent with the results from comparison with AERONET measurements (Figure 6, Table 4). When the excluded ratio of valid values (percentage of excluded values from the valid observations) increases from 25% to 75%, the average RMSE value of AOD for Terra and Aqua increases from 0.045 to 0.055 and from 0.047 to 0.057, respectively (Table 4). Meanwhile, the standard deviation of RMSE also increases from 0.019 to 0.042 and from 0.021 to 0.062, respectively for Terra and Aqua, when the excluded ratio of valid values increases from 25% to 75%. This is also true for all testing years in Table 4. This finding is consistent with the results about lower accuracies of the estimated AOD than the original MAIAC AOD according to ground measurements in Section 4.1. These
Table 4. Cross validation of the proposed gap-filling algorithm in randomly selected 10 MODIS tiles for three different years, globally.

| Year  | Terra | AOD (±SD) | MAE (±SD) | RMSE (±SD) | 50% | 75% | 90% |
|-------|-------|-----------|-----------|------------|-----|-----|-----|
| 2005  | 0.805 (±0.050) | 0.030 (±0.056) | 0.011 (±0.061) | 0.775 (±0.0047) | 0.783 (±0.016) | 0.774 (±0.0072) | 0.766 (±0.0060) |
| 2010  | 0.833 (±0.030) | 0.022 (±0.023) | 0.019 (±0.024) | 0.815 (±0.0076) | 0.815 (±0.0076) | 0.774 (±0.0082) | 0.796 (±0.0060) |
| 2015  | 0.778 (±0.048) | 0.045 (±0.035) | 0.029 (±0.038) | 0.788 (±0.0069) | 0.788 (±0.0061) | 0.771 (±0.0070) | 0.786 (±0.0061) |

Note: SD means standard deviation.

findings indicated that the accuracy of gap-filled AOD could decrease with the increase of cloud contamination.

The estimated AOD shows a consistent spatial pattern with the original MAIAC AOD (Figure 7). Even 75% of valid observations were excluded from the original MAIAC AOD for cross validation in day 187 of 2015 (Figure 7), the estimated AOD and original AOD in the excluded area show similar spatial pattern. As shown in Figure 7, the estimated AOD has higher values in the center of the image than that on the boundary of the image, which is similar with the spatial pattern of the original AOD values. However, there are still different values at specific locations such as on the lower left of the image, where the estimated AOD values are higher than that of the original AOD values. Overall, most part of the original AOD image can be well filled using the gap-filling algorithm.

4.3. Spatial and seasonal patterns of gapless AOD

The gapless data capture well the spatial heterogeneity and seasonal variation of AOD in China, taking AOD from Aqua satellite as an example (Figure 8). Specifically, AOD values were high in the North China Plain for the entire year in 2015 as this region is highly populated and heavily polluted due to significant amount of emissions from industrial and agricultural activities as well as daily activities of citizens (Song et al. 2018; Filonchyk et al. 2019). The region of Tarim Basin experienced high AOD in Spring and Summer due to natural aerosols with the prevalence of desert dust emitted by the Taklimakan Desert during these seasons (Filonchyk et al. 2019). There were also high AOD values in southern China in Spring and Winter due to the occurrence of haze (X. Liu et al. 2016). The lowest AOD values were observed in the region of Tibetan Plateau because its high altitude prevented the aerosol penetration from other regions in China that suffered from heavy aerosol pollution (Filonchyk et al. 2019). Meanwhile, the original MAIAC AOD data may underestimate AOD values especially in eastern China, when they are temporally aggregated to seasonal scale, compared to the gapless AOD data (Figure 8), indicating the necessity of gap-filling on the original MAIAC AOD dataset.
Figure 7. Original AOD, gapless AOD, excluded AOD from the original MAIAC AOD (75% of the original data) for the cross validation and estimated AOD in the excluded area in the tile h27v05 (located at Eastern China) in day of 187 in 2015.

Figure 8. Seasonal average of gapless AOD, original MAIAC AOD, and their difference from aqua satellite in China in 2015.
4.4. Comparison with existing gap-filling methods

The accuracies of estimated AOD using the novel gap-filling algorithm are generally higher than those of the ICW and IDW algorithm (Tables 5 and 6). The new algorithm shows better performance than the ICW because it has considered overall mean to capture temporal trend of AOD using the smoothing spline function. The novel algorithm shows higher accuracies than those of the IDW because it considered spatiotemporal relationships between neighboring sites, which the IDW algorithm only considers spatial correlations of neighboring sites.

The new spatiotemporal gap-filling algorithm proposed in this study does not have the limitations in existing methods and can be used to develop full coverage of AOD data with reasonable quality over large areas. For the spatial coverage, the data obtained using existing methods still had some missing values when there are a large number of continuously missing data, even when several methods were combined to address this limitation (Chi et al. 2020; Wang et al. 2019a), but the proposed method in this study does not have this limitation. Our gap-filling framework can be implemented for over large areas with high computing efficiency. Moreover, we used the representative grids in local windows to interpolate residuals without the need/use of other auxiliary data. However, previous methods usually require auxiliary data such as AOD from other satellites and other related variables (Jiang et al. 2021; Zhao et al. 2019) or introduce local windows for calculation in each missing grid (Wang et al. 2019a) to fill gaps, which may require large amount of computing when applied over large areas.

4.5. Comparison with other seamless AOD datasets

The accuracy of our gapless AOD in this study performed well compared to existing two seamless AOD datasets, i.e. LGHAP AOD and MERRA-2 AOD. Figure 9 shows that the accuracies of our gapless AOD using the novel algorithm and the LGHAP AOD data have few differences and are higher than that of the MERRA-2 AOD data. Our gapless AOD using the novel algorithm only used the spatiotemporal information of the MAIAC AOD itself, while the LGHAP

| Year | Terra | Year | Aqua |
|------------------|-------|------------------|-------|
| 2005 | 0.035±0.024 | 2010 | 0.033±0.023 |
| 2010 | 0.034±0.025 | 2015 | 0.030±0.034 |
| 2015 | 0.045±0.024 | 2005 | 0.033±0.024 |
| 2005 | 0.051±0.023 | 2010 | 0.051±0.024 |
| 2010 | 0.055±0.024 | 2015 | 0.033±0.023 |

Note: SD means standard deviation.
Table 6. Cross validation of the IDW for the original AOD in randomly selected 10 MODIS tiles for three different years, globally.

| Satellite | Year | 25% | 25% | 50% | 75% |
|-----------|------|-----|-----|-----|-----|
| Terra     | 2005 | 0.059 (±0.034) | 0.079 (±0.036) | 0.095 (±0.038) | 0.144 (±0.041) |
|           | 2010 | 0.059 (±0.034) | 0.079 (±0.036) | 0.095 (±0.038) | 0.144 (±0.041) |
|           | 2015 | 0.059 (±0.034) | 0.079 (±0.036) | 0.095 (±0.038) | 0.144 (±0.041) |
| Aqua      | 2005 | 0.059 (±0.034) | 0.079 (±0.036) | 0.095 (±0.038) | 0.144 (±0.041) |
|           | 2010 | 0.059 (±0.034) | 0.079 (±0.036) | 0.095 (±0.038) | 0.144 (±0.041) |
|           | 2015 | 0.059 (±0.034) | 0.079 (±0.036) | 0.095 (±0.038) | 0.144 (±0.041) |

Note: SD means standard deviation.

AOD combined AOD datasets from multiple sources, the similar accuracies of these two datasets indicates that the good performance of the novel algorithm using limited input data.

Our gapless AOD data using the novel algorithm shows similar spatial patterns with those of the LGHAP AOD and MERRA-2 AOD for different seasons and more spatial details (Figure 10). Our gapless AOD and the LGHAP AOD show more spatial details than the MERRA-2 AOD due to higher spatial resolution. The LGHAP AOD tends to be smoother than our gapless AOD possibly due to information loss after aggregating multiple AOD datasets with different spatiotemporal resolutions.

4.6. Uncertainties and future work

The number of missing values in the original MAIAC AOD data is the main source of uncertainties in the gapless AOD. The number of missing values in time series of each grid also has impact on the temporal pattern of the fitted overall mean, which can lead to biases and reduce the accuracies of the gapless AOD (Figure 6). The percentage of missing values in data can largely affect the spatial pattern of the interpolated residuals and accordingly lead to different accuracies in the estimated AOD. The larger the number of missing values, especially if these missing values are consecutive, the lower the accuracy of the estimated AOD (Table 4, Figure 7). The reduced accuracy of estimated AOD can also be found in the results of some existing gap-filling techniques (Chi et al. 2020; Tang, Bo, and Zhu 2016). For example, RMSEs of original MODIS AOD and gap-filled AOD data are 0.13 and 0.23 by Chi et al. (2020), and 0.19 and 0.29 by Tang, Bo, and Zhu (2016), indicating lower accuracies of gap-filled AOD than that of the original AOD data due to uncertainties caused by missing values. Meanwhile, there are larger biases between the gap-filled AOD and the AERONET measurements than that of the original AOD data (Chi et al. 2020).

This study opens future research avenues to further improve the proposed gap-filling method. For example, clustering the pixels into several groups to fit overall mean values might help to mitigate the underestimation issue, therefore potentially improve the accuracy of gap-filled AOD when there are a large amount of missing values in the original MODIS MAIAC data. The proposed method can also be explored for
other MODIS like variables (e.g. land surface temperature) since this method does not depend on any hypothesis on the distribution of the data nor on other auxiliary data. Moreover, together with the sparse ground-level PM2.5 monitoring networks, the gapless AOD data from this study can help to build seamless PM2.5 concentration data for health-related studies (e.g. exposures to PM2.5).

5. Conclusions

In this study, we developed a novel gap-filling method that combines a smoothing spline function to capture the general spatiotemporal pattern of AOD, together with the Interpolation-based Correlation Weighting (ICW) method to gap-fill the missing data in the AOD product. Compared with existing gap-filling methods, the key improvement of the novel gap-filling framework is that it can be implemented over large areas with high computing efficiency and quality without the help of other auxiliary data. Using this method, we developed a 1 km spatial resolution daily gapless AOD dataset using the MODIS MAIAC AOD. The gapless AOD data can well capture the spatial heterogeneity and seasonal variations of AOD in China. The accuracy assessment of gapless AOD data showed good performance of the proposed spatiotemporal gap-filling...
method. According to AERONET ground measurements, the RMSE of gapless AOD datasets for Terra and Aqua were 0.24 and 0.27, and the within Expected Error (EE) envelope proportions were 59.15% and 57.03%, respectively. Cross validation under missing scenarios of 25%, 50%, and 75% of the original AOD data indicated that the range of average RMSE of estimated AOD for Terra and Aqua were 0.045 to 0.055 and 0.047 to 0.057, respectively. The accuracies of estimated AOD using the novel gap-filling algorithm are generally higher than those of the ICW and IDW. Moreover, our gapless AOD using the novel algorithm performed well in accuracy and showed similar but more spatial details compared to two existing seamless AOD datasets, i.e. LGHAP AOD and MERRA-2 AOD. The proposed method can be transferred to other regions, with the potential to be used even operationally and efficiently for generating the global gapless daily AOD dataset with the only input of MODIS MAIAC AOD data.

Acknowledgements

This research was funded by the College of Liberal Arts and Science’s (LAS) Dean’s Emerging Faculty Leaders award at the Iowa State University and the Iowa State University publication subvention grant.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Tao Zhang http://orcid.org/0000-0002-6681-7983
Yuyu Zhou http://orcid.org/0000-0003-1765-6789
Kaiguang Zhao http://orcid.org/0000-0001-5858-4967
Zhengyuan Zhu http://orcid.org/0000-0002-2266-0646
Ghassem R. Asrar http://orcid.org/0000-0002-1066-8384
Xia Zhao http://orcid.org/0000-0001-8467-9513

Data availability statement

The data that support the findings of this study are available on request from the corresponding author.

References

Ahn, C., O. Torres, and H. Jethva. 2014. “Assessment of OMI Near-UV Aerosol Optical Depth over Land.” Journal of Geophysical Research: Atmospheres 119 (5): 2457–2473. doi:10.1002/2013JD020188.
Bai, K., K. Li, M. Ma, K. Li, Z. Li, J. Guo, N. B. Chang, Z. Tan, and D. Han. 2022. “LGHAP: A Long-Term Gap-Free High-Resolution Air Pollutants Concentration Dataset Derived via Tensor Flow Based Multimodal Data Fusion.” Earth System Science Data 14 (2): 907–927. doi:10.5194/essd-14-907-2022.
Cavazzani, S., V. Zitelli, and S. Ortolani. 2015. “Long-Term Analysis of Clear Sky at Astronomical Sites: A Comparison between Polar and Geostationary Satellites.” Monthly Notices of the Royal Astronomical Society 452 (2): 2185–2194. doi:10.1093/mnras/stv1319.
Chi, Y., Z. Wu, K. Liao, and Y. Ren. 2020. “Handling Missing Data in Large-Scale Modis AOD Products Using a Two-Step Model.” Remote Sensing 12 (22): 1–24. doi:10.3390/rs12223786.
Filonchyk, M., H. Yan, Z. Zhang, S. Yang, W. Li, and Y. Li. 2019. “Combined Use of Satellite and Surface Observations to Study Aerosol Optical Depth in Different Regions of China.” Scientific Reports 9 (1): 6174. doi:10.1038/s41598-019-42466-6.
Gelaro, R., W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, et al. 2017. “The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2).” Journal of Climate 30 (14): 5419–5454. doi:10.1175/JCLI-D-16-0758.1.
Giles, D. M., A. Sinyuk, M. G. Sorokin, J. S. Schafer, A. Smirnov, I. Slutsker, T. F. Eck, et al. 2019. “Advancements in the Aerosol Robotic Network (AERONET) Version 3 Database - Automated near-Real-Time Quality Control Algorithm with Improved Cloud Screening for Sun Photometer Aerosol Optical Depth (AOD) Measurements.” Atmospheric Measurement Techniques 12 (1): 169–209. doi:10.5194/amt-12-169-2019.
Go, S., J. Kim, S. Park, M. Kim, H. Lim, J. Kim, D. Lee, and J. Im. 2020. “Synergistic Use of Hyperspectral Uv-Visible Omi and Broadband Meteorological Imager Modis Data for a Merged Aerosol Product.” Remote Sensing 12 (23): 1–34. doi:10.3390/rs12233987.
Gong, P., X. Li, J. Wang, Y. Bai, B. Chen, T. Hu, X. Liu, et al. 2020. “Annual Maps of Global Artificial Impervious Area (GAIA) between 1985 and 2018.” Remote Sensing of Environment 236 Elsevier: 111510. 10.1016/j.rse.2019.111510.
Gupta, P., R. C. Levy, S. Mattox, L. A. Remer, and L. A. Munchak. 2016. “A Surface Reflectance Scheme for Retrieving Aerosol Optical Depth over Urban Surfaces in MODIS Dark Target Retrieval Algorithm.” Atmospheric Measurement Techniques 9 (7): 3293–3308. doi:10.5194/amt-9-3293-2016.
Hsu, N. C., M. J. Jeong, C. Bettenhausen, A. M. Sayer, R. Hansell, C. S. Seftor, J. Huang, and S. C. Tsay. 2013. “Enhanced Deep Blue Aerosol Retrieval Algorithm: The Second Generation.” *Journal of Geophysical Research: Atmospheres* 118 (16): 9296–9315. doi:10.1002/jgrd.50712.

Hsu, N. C., J. Lee, A. M. Sayer, N. Carletta, S. H. Chen, C. J. Tucker, B. N. Holben, and S. C. Tsay. 2017. “Retrieving Near-Global Aerosol Loading over Land and Ocean from AVHRR.” *Journal of Geophysical Research: Atmospheres* 122 (18): 9968–9989. doi:10.1002/2017JD026932.

Jiang, T., B. Chen, Z. Nie, Z. Ren, B. Xu, and S. Tang. 2021. “Estimation of Hourly Full-Coverage PM2.5 Concentrations at 1-Km Resolution in China Using a Two-Stage Random Forest Model.” *Atmospheric Research* 248 February 2020 Elsevier: 105146. 10.1016/j.atmosres.2020.105146.

Jinnagara Puttaswamy, S., H. M. Nguyen, A. Braverman, X. Hu, and Y. Liu. 2014. “Statistical Data Fusion of Multi-Sensor AOD over the Continental United States.” *Geocarto International* 29 (1): 48–64. doi:10.1080/10106049.2013.827750.

Kahn, R. A., B. J. Gaitley, M. J. Garay, D. J. Diner, T. F. Eck, A. Smirnov, and B. N. Holben. 2010. “Multiscale Imaging Spectroradiometer Global Aerosol Product Assessment by Comparison with the Aerosol Robotic Network.” *Journal of Geophysical Research: Atmospheres* 115 (D23): D23209. doi:10.1029/2010JD014601.

Levy, R. C., S. Mattoo, L. A. Munchak, L. A. Remer, A. M. Sayer, F. Patadia, and N. C. Hsu. 2013. “The Collection 6 MODIS Aerosol Products over Land and Ocean.” *Atmospheric Measurement Techniques* 6 (11): 2989–3034. doi:10.5194/amt-6-2989-2013.

Li, X., P. Gong, Y. Zhou, J. Wang, Y. Bai, B. Chen, T. Hu, et al. 2020. “Mapping Global Urban Boundaries from the Global Artificial Impervious Area (GAIa) Data.” *Environmental Research Letters* 15 (9): 094044. doi:10.1088/1748-9361/ab9be3.

Li, L., J. Yang, and Y. Wang. 2014. “An Improved Dark Object Method to Retrieve 500m-Resolution AOT (Aerosol Optical Thickness) Image from MODIS Data: A Case Study in the Pearl River Delta Area, China.” *ISPRS Journal of Photogrammetry and Remote Sensing* Vol. 89: March 2014. Netherlands: International Society for Photogrammetry and Remote Sensing, Inc (ISPRS): 1–12. doi:10.1016/j.isprsjprs.2013.12.008.

Lin, C., Y. Li, A. K. H. Lau, X. Deng, T. K. T. Tse, J. C. H. Fung, C. Li, et al. 2016. “Estimation of Long-Term Population Exposure to PM2.5 For Dense Urban Areas Using 1-Km MODIS Data.” *Remote Sensing of Environment* 179 Elsevier B.V.: 13–22. 10.1016/j.rse.2016.03.023.

Liu, X., Q. Chen, H. Che, R. Zhang, K. Gui, H. Zhang, and T. Zhao. 2016. “Spatial Distribution and Temporal Variation of Aerosol Optical Depth in the Sichuan Basin, China, the Recent Ten Years.” *Atmospheric Environment* 147: 434–445. 10.1016/j.atmosenv.2016.10.008. Elsevier Ltd.

Liu, N., B. Zou, H. Feng, W. Wang, Y. Tang, and Y. Liang. 2019. “Evaluation and Comparison of Multiscale Implementation of the Atmospheric Correction Algorithm, Dark Target, and Deep Blue Aerosol Products over China.” *Atmospheric Chemistry and Physics* 19 (12): 8243–8268. doi:10.5194/ACP-19-8243-2019.

Lolli, S. 2021. “Is the Air Too Polluted for Outdoor Activities? Check by Using Your Photovoltaic System as an Air-Quality Monitoring Device.” *Sensors* 21 (19): 6342. doi:10.3390/s21196342.

Lu, X., S. Zhang, J. Xing, Y. Wang, W. Chen, D. Ding, Y. Wu, S. Wang, L. Duan, and J. Hao. 2020. “Progress of Air Pollution Control in China and Its Challenges and Opportunities in the Ecological Civilization Era.” *Engineering* 6 (12): 1423–1431. doi:10.1016/j.eng.2020.03.014.

Lv, B., Y. Hu, H. H. Chang, A. G. Russell, and Y. Bai. 2016. “Improving the Accuracy of Daily PM2.5 Distributions Derived from the Fusion of Ground-Level Measurements with Aerosol Optical Depth Observations, a Case Study in North China.” *Environmental Science & Technology* 50 (9): 4752–4759. doi:10.1021/acs.est.6b05940.

Lyapustin, A., and Y. Wang. 2018. “MODIS Multi-Angle Implementation of Atmospheric Correction (MAIAC) Data User ’ S Guide.” NASA: Greenbelt, MD, USA 6 (June): 1–19.

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

Ma, Z., X. Hu, A. M. Sayer, R. Levy, Q. Zhang, Y. Xue, S. Tong, B. Jun, L. Huang, and Y. Liu. 2016. “Satellite-Based Spatiotemporal Trends in PM2.5 Concentrations: China, 2004–2013.” *Environmental Health Perspectives* 124 (2): 184–192. doi:10.1289/ehp.1409481.

Makar, M., J. Antonelli, Q. Di, D. Cutler, J. Schwartz, Dominici, F. 2018. “Estimating the Causal Effect of Fine Particulate Matter Levels on Death and Hospitalization: Are Levels below the Safety Standards Harmful?” *Epidemiology* 28 (5): 627–634. doi:10.1097/ EDE.0000000000000690.

Ningombam, S. S., U. Chandra Dumka, S. Kalamani Mugil, J. Chandra Kunjial, R. K. Hooda, A. Sagar Gautam, and S. Tiwari. 2021a. “Impacts of Aerosol Loading in the Hindu Kush Himalayan Region Based on Merra-2 Reanalysis Data.” *Atmosphere* 12 (10): 1–18. doi:10.3390/atmos12101290.

Ningombam, S. S., H. J. Song, S. K. Mugil, U. C. Dumka, Larson, E. J. L., B. Kumar, and R. Sagar. 2021b. “Evaluation of Fractional Clear Sky over Potential Astronomical Sites.” *Monthly Notices of the Royal Astronomical Society* 507 (3): 3745–3760. Oxford University Press 10.1093/mnras/stab1971.

Perperoglou, A., W. Sauerbrei, M. Abrahamowicz, and M. Schmid. 2019. “A Review of Spline Function Procedures in R.” *BMC Medical Research Methodology* 19 (1): 1–16.BMC Medical Research Methodology 10.1186/s12874-019-0666-3.
Pöschl, U. 2005. “Atmospheric Aerosols: Composition, Transformation, Climate and Health Effects.” Angewandte Chemie - International Edition 44 (46): 7520–7540. doi:10.1002/anie.200501112.
Ramanathan, V., P. J. Crutzen, J. T. Kiehl, and D. Rosenfeld. 2001. “Atmosphere: Aerosols, Climate, and the Hydrological Cycle.” Science 294 (5549): 2119–2124. doi:10.1126/science.1064034.
Remer, L. A., Y. J. Kaufman, D. Tanré, S. Mattoo, D. A. Chu, J. V. Martins, Li, R.R., Ichoku, C., Levy, R.C., Kleidman, R.G., et al. 2005. “The MODIS Aerosol Algorithm, Products, and Validation.” Journal of the Atmospheric Sciences 62 (4): 947–973. doi:10.1175/JAS3385.1.
Remer, L. A., S. Mattoo, R. C. Levy, and L. A. Munchak. 2013. “MODIS 3 Km Aerosol Product: Algorithm and Global Perspective.” Atmospheric Measurement Techniques 6 (7): 1829–1844. doi:10.5194/amt-6-1829-2013.
Rosenfeld, D., U. Lohmann, G. B. Raga, C. D. O’Dowd, M. Kulmala, S. Fuzzi, A. Reissell, and M. O. Andreae. 2008. “Flood or Drought: How Do Aerosols Affect Precipitation?” Science 321 (5894): 1309–1313. doi:10.1126/science.1160606.
Ruiz-Arias, J. A., J. Duddia, V. Lara-Fanego, and D. Pozo-Vázquez. 2013. “A Geostatistical Approach for Producing Daily-Level-3 MODIS Aerosol Optical Depth Analyses.” Atmospheric Environment 79: 395–405. 10.1016/j.atmosenv.2013.07.002. Elsevier Ltd.
Sayer, A. M., N. C. Hsu, C. Bettenhausen, M. J. Jeong, and G. Meister. 2015. “Effect of MODIS Terra Radiometric Calibration Improvements on Collection 6 Deep Blue Aerosol Products: Validation and Terra/Aqua Consistency.” Journal of Geophysical Research: Atmospheres 120 (23): 12157-12174 10.1002/2015JD023878.
Sayer, A. M., N. C. Hsu, J. Lee, N. Carletta, S. H. Chen, and A. Smirnov. 2017. “Evaluation of NASA Deep Blue/SOAR Aerosol Retrieval Algorithms Applied to AVHRR Measurements.” Journal of Geophysical Research: Atmospheres 122 (18): 9945–9967. doi:10.1002/2017JD026934.
Shaheen, A., R. Wu, and M. Aldabash. 2020. “Long-Term AOD Trend Assessment over the Eastern Mediterranean Region: A Comparative Study Including A New Merged Aerosol Product.” Atmospheric Environment 238: April Elsevier Ltd: 117736: 117736. 10.1016/j.atmosenv.2020.117736.
Shepard, D. 1968. “Two- Dimensional Interpolation Function for Irregularly-Spaced Data.” Proceeding 23rd National Conference, New York, USA, 517–524. doi:10.1145/800186.810616.
Singh, M. K., R. Gautam, and P. Venkatachalam. 2017. “Bayesian Merging of MISR and MODIS Aerosol Optical Depth Products Using Error Distributions from AERONET.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (12): 5186–5200. doi:10.1109/JSTARS.2017.2734331.
Singh, M. K., P. Venkatachalam, and R. Gautam. 2017. “Geostatistical Methods for Filling Gaps in Level-3 Monthly Aerosol Optical Depth Data from Multi-Angle Imaging Spectroradiometer.” Aerosol and Air Quality Research 17 (8): 1963–1974. doi:10.4209/aaqr.2016.02.0084.
Sogacheva, L., T. Popp, A. M. Sayer, O. Dubovik, M. J. Garay, A. Heckel, Hsu, N. C., Jethva, H., Kahn, R. A., Kolmonen, P. and Kosmale, M. 2020. “Merging Regional and Global Aerosol Optical Depth Records from Major Available Satellite Products.” Atmospheric Chemistry and Physics 20 (4): 2031–2056. doi:10.5194/acp-20-2031-2020.
Song, Z., D. Fu, X. Zhang, Y. Wu, X. Xia, J. He, X. Han, R. Zhang, and H. Che. 2018. “Diurnal and Seasonal Variability of PM2.5 And AOD in North China Plain: Comparison of MERRA-2 Products and Ground Measurements.” Atmospheric Environment. 191: May Elsevier: 70–78. 10.1016/j.atmosenv.2018.08.012.
Tang, Q., Y. Bo, and Y. Zhu. 2016. “Spatiotemporal Fusion of Multiple-Satellite Aerosol Optical Depth (AOD) Products Using Bayesian Maximum Entropy Method.” Journal of Geophysical Research: Atmospheres 121 (8): 4034–4048. doi:10.1002/2015JD024571.
Tao, M., J. Wang, R. Li, L. Wang, L. Wang, Z. Wang, J. Tao, H. Che, and L. Chen. 2019. “Performance of MODIS High-Resolution MAIAC Aerosol Algorithm in China: Characterization and Limitation.” Atmospheric Environment 213 (388): 159–169.Elsevier 10.1016/j.atmosenv.2019.06.004.
Wang, Y., Q. Yuan, T. Li, H. Shen, L. Zheng, and L. Zhang. 2019a. “Large-Scale MODIS AOD Products Recovery: Spatial-Temporal Hybrid Fusion Considering Aerosol Variation Mitigation.” ISPRS Journal of Photogrammetry and Remote Sensing. 157: July Elsevier: 1–12. 10.1016/j.isprsjprs.2019.08.017.
Wang, Y., Q. Yuan, T. Li, H. Shen, L. Zheng, and L. Zhang. 2019b. “Evaluation and Comparison of MODIS Collection 6.1 Aerosol Optical Depth against AERONET over Regions in China with Multifarious Underlying Surfaces.” Atmospheric Environment 200 December 2018 Elsevier: 280–301. 10.1016/j.atmosenv.2018.12.023.
Wei, J., W. Huang, Z. Li, W. Xue, Y. Peng, L. Sun, and M. Cribb. 2019. “Estimating 1-Km-Resolution PM2.5 Concentrations across China Using the Space-Time Random Forest Approach.” Remote Sensing of Environment. 231: June Elsevier: 111221. 10.1016/j.rse.2019.111221.
Xiao, L., Y. Lang, and G. Christakos. 2018. “High-Resolution Spatiotemporal Mapping of PM2.5 Concentrations at Mainland China Using a Combined BME-GWR Technique.” Atmospheric Environment 173 June 2017 Elsevier: 295–305. 10.1016/j.atmosenv.2017.10.062.
Xiao, Q., Y. Wang, H. H. Chang, X. Meng, G. Geng, A. Lyapustin, and Y. Liu. 2017. “Full-Coverage High-Resolution Daily PM2.5 Estimation Using MAIAC AOD in the Yangtze River Delta of China.” Remote Sensing of Environment. 199: May Elsevier: 437–446. 10.1016/j.rse.2017.07.023.
Yu, H., J. Guang, Y. Xue, G. Leeuw, Che, Y. H., J. Guo, He, X. W., and T. K. Wang. 2015. “A Consistent Aerosol Optical Depth (AOD) Dataset over Mainland China by Integration of Several AOD Products.” Atmospheric Environment 114: 48–56. doi:10.1016/j.atmosenv.2015.05.023.

Yang, D., and C. A. Gueymard. 2021. “Probabilistic Merging and Verification of Monthly Gridded Aerosol Products.” Atmospheric Environment 247 September 2020 Elsevier Ltd: 118146: 10.1016/j.atmosenv.2020.118146.

Yang, J., and M. Hu. 2018. “Filling the Missing Data Gaps of Daily MODIS AOD Using Spatiotemporal Interpolation.” Science of the Total Environment 633: 677–683. 10.1016/j.scitotenv.2018.03.202. Elsevier B.V.

You, Y., T. Zhao, Y. Xie, Y. Zheng, J. Zhu, J. Xia, L. Cao, et al. 2020. “Variation of the Aerosol Optical Properties and Validation of MODIS AOD Products over the Eastern Edge of the Tibetan Plateau Based on Ground-Based Remote Sensing in 2017.” Atmospheric Environment 223 June 2019 Elsevier Ltd: 117257: 117257. 10.1016/j.atmosenv.2019.117257.

Yousef, R., F. Wang, G. Q. Ge, and A. Shaheen. 2020. “Long-Term Aerosol Optical Depth Trend over Iran and Identification of Dominant Aerosol Types.” Science of the Total Environment 722. doi:10.1016/j.scitotenv.2020.137906.

Yu, C., L. Chen, L. Su, M. Fan, and S. Li. 2011. “Kriging Interpolation Method and Its Application in Retrieval of MODIS Aerosol Optical Depth.” Proceedings - 2011 19th International Conference on Geoinformatics, Geoinformatics (IEEE), Shanghai, China, 1–6. doi:10.1109/Geoinformatics.2011.5981052.

Zhang, Z., W. Wu, M. Fan, J. Wei, Y. Tan, and Q. Wang. 2019. “Evaluation of MAIAC Aerosol Retrievals over China.” Atmospheric Environment 202 September 2018 Elsevier: 8–16. 10.1016/j.atmosenv.2019.01.013.

Zhang, L., X. Yi, M. Zhao, and Z. Gu. 2021. “Reduction in Solar Photovoltaic Generation Due to Aerosol Pollution in Megacities in Western China during 2014 to 2018.” Indoor and Built Environment 30 (8): 1286–1294. doi:10.1177/1420326X20939310.

Zhang, T., C. Zeng, W. Gong, L. Wang, K. Sun, H. Shen, Z. Zhu, and Z. Zhu. 2017. “Improving Spatial Coverage for Aqua MODIS AOD Using NDVI-Based Multi-Temporal Regression Analysis.” Remote Sensing 9 (4). doi:10.3390/rs9040340.

Zhao, C., Z. Liu, Q. Wang, J. Ban, N. X. Chen, and T. Li. 2019. “High-Resolution Daily AOD Estimated to Full Coverage Using the Random Forest Model Approach in the Beijing-Tianjin-Hebei Region.” Atmospheric Environment 203 July 2018 Elsevier: 70–78. 10.1016/j.atmosenv.2019.01.045.

Zheng, Z., C. Zhao, S. Lolli, X. Wang, Y. Wang, X. Ma, Q. Li, and Y. Yang. 2020. “Diurnal Variation of Summer Precipitation Modulated by Air Pollution: Observational Evidences in the Beijing Metropolitan Area.” Environmental Research Letters 15 (9): 094053. doi:10.1088/1748-9326/ab99fc.

Zhou, Y., X. Li, G. R. Asrar, S. J. Smith, and M. Imhoff. 2018. “A Global Record of Annual Urban Dynamics (1992–2013) from Nighttime Lights.” Remote Sensing of Environment 219: 206–220. doi:10.1016/j.rse.2018.10.015.