A novel big data mining framework for reconstructing large-scale daily MAIAC AOD data across China from 2000 to 2020

Binjie Chen, Yang Ye, Cheng Tong, Jinsong Deng, Ke Wang, and Yang Hong

ABSTRACT

Satellite-based aerosol optical depth (AOD) retrieval products are essential in air pollution and climate change research. Unfortunately, cloud contaminations and unfavorable surface conditions result in a considerable proportion of missing AOD data. Numerous studies have been conducted to reconstruct large-scale AOD data gaps by utilizing adjacent spatiotemporal information or modeling AOD data via various external geographical data. However, the erratic variation of AOD and the inconvenience of external data weaken the accuracy and efficiency of reconstruction. To address these issues, a novel big data based iterative variation mining framework (IVMF), utilizing multi-spatiotemporal information on AOD variations, is proposed to reconstruct large-scale AOD data over China from 2000 to 2020. Simulated and real-data experiments are carried out to validate the effectiveness and robustness of the IVMF. Results show that the spatial patterns of the reconstructed AOD are consistent with those of the original AOD in the simulated experiments. The final reconstructed AOD data strongly correlate with in-situ AOD measurements in the real-data experiments (correlation coefficient of 0.91). After reconstruction, the average daily AOD coverage increases from 30.42% to 96.69% (a 218% increase). Besides, results reveal that central China exhibits severe AOD levels, while northwest China presents low AOD levels. Overall, the proposed IVMF can largely resolve the missing AOD data problem with outstanding accuracy and efficiency, and has great potential to be generalized to other regions and remote sensing products.

1. Introduction

Atmospheric aerosols are liquid or solid particles suspended in the ambient environment, which have significant climate effects through modifying atmospheric circulation and radiation balance of the Earth (Kaufman, Tanré, and Boucher 2002; Yu et al. 2020). Moreover, aerosol pollutions, especially fine particles with aerodynamic diameters less than 2.5 μm (PM2.5), have been associated with various adverse health effects (Jinnagara Puttaswamy et al. 2013; Murray et al. 2020). Therefore, understanding the spatiotemporal distributions of aerosols has been brought to the forefront of public attention (He et al. 2020). Aerosol optical depth (AOD), defined as the integral of the extinction coefficient of the aerosol in the vertical direction, is most typically used to reveal aerosol loading levels and distributions (Xie et al. 2019). Given the limitations of sparsely distributed ground-level monitoring networks, such as AErosol RObotic NETwork (AERONET), satellite-based AOD retrieval products are being increasingly applied as the primary sources of AOD data (Olcense, Palancar, and Toselli 2015; Zhang et al. 2020).

The Moderate Resolution Imaging Spectrometer (MODIS) on board the Terra satellite has been providing daily AOD retrieval products with near-global coverage since 26 February 2000 using the advanced Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm (Belle and Liu 2016; Lyapustin et al. 2018). However, because the retrieval of AOD relies strictly on clear skies and appropriate surface conditions, there are large data gaps in these daily AOD products, which hinders their application (Xiao et al. 2021). For example, approximately 80% of the MAIAC AOD retrievals in New York State were missing in 2016 (Pu and Yoo 2021). The daily mean spatial coverage of MAIAC AOD in north China was approximately 45% (Liang et al. 2018). Therefore, it is
critical to reconstructing the missing MAIAC AOD data to enhance the quality of the data and extend its availability (Wang et al. 2017).

To date, a variety of methods have been proposed to address the missing AOD data issue. The most common strategy is integrating multi-source AOD products, such as MODIS Terra and Aqua AOD, through linear regression to improve the AOD coverage (He, Gu, and Zhang 2020; Hu et al. 2014). Ma et al. (2014) fused MODIS Terra, Aqua and Multiangle Imaging SpectroRadiometer (MISR) AOD together to improve data coverage. Further, He and Huang (2018) combined four AOD products, including MODIS Terra/ Aqua 3-km DT AOD and 10-km DB AOD, to significantly increase the number of valid AOD pixels. Nonetheless, cloud contamination cannot be fully resolved by merging multiple AOD products and the uncertainties of these AOD products are hard to unify (Fu et al. 2020; Wang et al. 2019). Benefiting from auxiliary geographical data, especially AOD estimates and meteorological conditions from chemical transport models, statistical regression models have been developed to reconstruct AOD data with full coverage (Huang et al. 2018). Given the superiority of machine learning technologies, random forest and deep learning models have also been proposed to explore the relationship between AOD and external geographical variables, and these have helped obtain spatially continuous AOD data (Li et al. 2020a; Park et al. 2020; Zhang, Di, et al. 2018). However, these approaches are deeply reliant on external geographical data and tend to be inconvenient particularly for large areas. Moreover, the external data, such as outputs from chemical transport models often have relatively low spatial resolutions (>25 km), which are far coarser than that of MAIAC AOD (1 km). Thus, they tend to bias the reconstructed MAIAC AOD results (Li et al. 2020a).

Another branch of reconstruction methodology is geostatistical-based interpolation utilizing local spatiotemporal information. Following Tobler’s first law of geography (Tobler 1970), the most representative approach is the inverse distance weighting (IDW) method, which takes the distance between pixels to characterize their spatial correlation (Chen et al. 2019). To better determine the spatial correlation, kriging techniques have been applied to reconstruct AOD data, leading to improved performance (Jinnagara Puttaswamy et al. 2013; Xiao et al. 2021). Adjacent temporal information is also critical for interpolation. For example, Yang and Hu (2018) applied spatiotemporal kriging, considering both temporal and spatial autocorrelations, to improve AOD coverage from 14% to 67%. External data are not necessary for these interpolation approaches. Therefore, they are not subject to the low spatial resolutions of external data and are applicable when external data are unavailable, which improves their attractiveness for AOD reconstruction. However, kriging approaches require strict constraints of assumption, namely regional homogeneity of spatial variation (Li et al. 2018), and the variation of AOD is thought to be non-ignorable, especially for large heterogeneous areas with diverse aerosol models (Wang et al. 2019). Moreover, because these geostatistical approaches rely on adjacent spatiotemporal information to extend the AOD coverage, the high missing rate and erratic variation of AOD weaken the efficiency (missing rate after interpolation approximately 30–50%) and accuracy in large-scale interpolation (Chen et al. 2020). Hence, a new geostatistical framework is needed to remedy the issue of erratic AOD variation and to recover large-scale missing AOD data with high accuracy.

In the past half-century, China has experienced severe aerosol pollution resulting from both natural and artificial processes (Zhang et al. 2019). Natural processes, such as sand dust raised by strong winds in spring, can cause abrupt changes in the aerosol environment (Song et al. 2019a), while dramatic economic growth and energy consumption have resulted in gradual impacts on aerosol pollution levels (Zou et al. 2019). However, because of the wide geographic range, complex terrain, and diverse landscape, using satellite observation data to map the long-term distribution of aerosol levels in China is challenging (Li et al. 2017). Moreover, different development strategies and modernization patterns across China result in discrepancies in long-term aerosol variations (Wang et al. 2018b). Thus, there is an urgent need to find ways to map the long-term spatially continuous distribution of daily aerosol levels across China.

This study aims to present a novel big data based iterative variation mining framework (IVMF) to reconstruct MAIAC AOD data over the large heterogeneous region of China. Specifically, local multi-spatiotemporal variation information is recognized and aggregated in a moving window to reconstruct MAIAC AOD data. The iteration strategy is further utilized to improve the data
coverage on a large scale. In addition, considering the spatial heterogeneity and temporal dependency, respectively, the spatial and temporal similarity are designed to enhance the accuracy of results. Finally, 1-km spatial resolution daily MAIAC AOD data across China from 2000 to 2020 are generated and validated against simulated and real-data experiments.

2. Datasets and methodology

2.1 Datasets

2.1.1 MAIAC AOD data

MAIAC is the latest algorithm to perform aerosol retrievals and atmospheric correction from MODIS observations (Lyapustin et al. 2018). It decouples aerosol and surface contributions using time series data, and the combination of time series and spatial analysis helps improve the retrieval accuracy over bright surfaces (Lyapustin et al. 2011a, 2011b). Compared to previous MODIS AOD products, the MAIAC AOD products over China perform better on retrieval accuracy and spatial resolution (Zhang et al. 2019). In this study, we acquired daily 1-km Terra MAIAC AOD products from the NASA Level-1 and Atmosphere Archive & Distribution System Distributed Active Archive Center website (https://ladsweb.modaps.eosdis.nasa.gov/). We extracted AOD values at 550 nm from 26 February 2000 (the first day MAIAC AOD data were available) to 31 December 2020. Data cleaning was conducted based on the quality assurance (QA) flags provided by MAIAC AOD products (QAcloud_mask = clear). Moreover, AOD values greater than 3 were excluded according to the reported range of valid AOD values (Li et al. 2020a, 2020b).

2.1.2 AERONET data

AERONET (https://aeronet.gsfc.nasa.gov/) is a global network providing in-situ AOD measurements using sun photometers. It has been widely served as “ground truth” in satellite AOD retrieval validation (Li et al. 2020a, 2020b). The version 3 AERONET AOD data of level 1.5 (cloud-screened and quality controlled) were used to validate the accuracy of the reconstructed AOD. A total of 19 sites across China were considered (Figure 1) and the details are listed in Table S1. Since AERONET did not measure AOD at 550 nm, AERONET AOD at 550 nm was interpolated using the Ångström exponent and AERONET AOD measured in the 670 nm and 440 nm channels to meet the consistency with MAIAC AOD (Li and Zhang 2019). AERONET AOD greater than 3 was also excluded. After data preprocessing, there were 17,746 valid measurements from 2000 to 2020. To match the in-situ AERONET AOD and MAIAC AOD results, MAIAC

![Figure 1. Geographic distribution of the AERONET sites considered in this paper.](image-url)
AOD pixels within the spatial window of 25 × 25 km, centered on the AERONET sites, were averaged, and AERONET measurements in the temporal window of ± 60 minutes, centered on the satellite overpass time, were averaged (Liu et al. 2019).

2.2. Methodology

As depicted in Figure 2, a big data based iterative variation mining framework (IVMF) was proposed to reconstruct daily MAIAC AOD data gaps across China from 26 February 2000 to 31 December 2020. A spatial window (centered on each target pixel) and a temporal window (centered on the target day) were jointly applied to combine the spatiotemporal variation information and attain the reconstructed result. Moreover, the spatial similarity among pixels across China and temporal similarity between days were characterized to further correct the reconstruction bias by fully using the 20 years of MAIAC AOD data. The procedures were iterated to reconstruct large-scale MAIAC AOD data until satisfied the constraints. Finally, simulated and real-data experiments were carried out to validate the final reconstructed daily MAIAC AOD results.

Figure 2. Flowchart of the proposed IVMF for MAIAC AOD reconstruction. $T_{p}$ represents the target pixel, $T_{d}$ represents the target day, $n$ is the increment of reconstructed AOD pixels after each iteration, Iter is the number of iterations.
### 2.2.1 Spatiotemporal variation mining

Spatiotemporal adjacent knowledge of the target pixel \( (T_p) \) on the target day \( (T_d) \) was the basis for filling the MAIAC AOD data gaps. However, the atmospheric environment is an unstable system and has intense spatiotemporal heterogeneity, which places restrictions on the practicality of traditional geostatistical-based interpolation methods (Kikuchi et al. 2018). Though the AOD variation is non-ignorable, even in the short-term (Wang et al. 2019), the local AOD variation should share similar patterns according to Tobler’s first law of geography (Tobler 1970). Therefore, it is possible to fill the AOD data gaps by mining the local AOD variation information. On the other hand, variations in topography, land use and land cover have shown significant impact on aerosol dispersion, dilution, and accumulation (Zhang, Di, et al. 2018). Thus, the AOD variation patterns are inevitably subject to spatial heterogeneity within the region. Previous studies employed distance- or NDVI-based interpolation to resolve this issue (Wang et al. 2019, 2018a). In this study, taking advantage of the 20 years of MAIAC AOD data, we proposed a hypothesis that the spatial similarity (long-term difference in variation pattern) between pixels, as identified by the Pearson correlation coefficient \((R)\), could constrain the effect of spatial heterogeneity. By searching for the surrounding common pixels \((C_{ps})\) on a given reference day \((R_d)\), and incorporating the calculated prior knowledge of spatial similarity, the missing \( T_p \) MAIAC AOD at \( T_d \) can be reconstructed as:

\[
\tau(T_p, T_d, R_d) = \tau(T_p, R_d) + \sum_{i=1}^{N} R_{tp,C_{pi}} \Delta(\tau(C_{pi}, T_d, R_d)) \]

\[
\sum_{i=1}^{N} R_{tp,C_{pi}} \Delta(\tau(C_{pi}, T_d, R_d)) + \sum_{i=1}^{N} R_{tp,C_{pi}} \Delta(\tau(C_{pi}, R_d, R)) - \tau(C_{pi}, R_d) \tag{1}
\]

where \( C_{pi} \) is the \( i^{th} \) surrounding pixels with valid MAIAC AOD at both \( T_d \) and \( R_d \) in the region; \( \tau(T_p, T_d, R_d) \) is the \( T_p \) MAIAC AOD at \( T_d \) to be reconstructed from information of \( R_d; \) \( \tau(T_p, R_d) \) is the \( T_p \) MAIAC AOD at \( R_d \) with a valid value; \( N \) is the total number of \( C_{ps} \) in the region; and \( R_{tp,C_{pi}} \) is the calculated spatial similarity between \( T_p \) and \( C_{pi}; \) \( \Delta(\tau(C_{pi}, T_d, R_d)) \) is the AOD variation of the \( C_{pi} \) between \( T_d \) and \( R_d; \) \( \tau(C_{pi}, T_d) \) and \( \tau(C_{pi}, R_d) \) are the \( C_{pi} \) MAIAC AOD at \( T_d \) and \( R_d \), respectively. Consequently, \( C_{ps} \) sharing long-term consistent variation patterns with \( T_p \) were assigned larger weights, while pixels having discrepant variation patterns with \( T_p \) were assigned smaller weights. These constraints mitigated the bias introduced by spatial heterogeneity. A typical example, illustrated in the Supplementary Material, confirmed the hypothesis. Because the \( C_{ps} \) far away from \( T_p \) had larger variation differences compared to \( T_p \), and there was a high computation burden of 9.6 million pixels across China, \( C_{ps} \) within a window of \( 11 \times 11 \) pixels centered on \( T_p \) were ultimately involved in reconstruction.

In summary, there were a total of three possible scenarios in the reconstruction process (Figure 3). (1) \( T_p \) at \( R_d \) was invalid. Under this circumstance, regardless of the \( C_{ps} \) found, no reconstructed result could be obtained because the term \( \tau(T_p, R_d) \) was unavailable. (2) \( T_p \) at \( R_d \) was valid, but no \( C_{ps} \) could be found. Since no useful spatiotemporal information existed to calculate the \( \Delta(\tau(C_{pi}, T_d, R_d)) \) term, \( \tau(T_p, T_d, R_d) \) could not be calculated. (3) \( T_p \) at \( R_d \) was valid, and the \( C_{ps} \) could also be found. This was the only applicable scenario in which the missing MAIAC AOD data could be reconstructed. After determining the \( C_{ps} \), the variations of these \( C_{ps} \) were obtained according to Eq. (2). Then Eq. (1) could be implemented to compute the reconstructed result of \( \tau(T_p, T_d, R_d) \). Finally, to ensure data quality, results outside the reported range of valid MAIAC AOD values (0–3) were discarded.

### 2.2.2 Multi-temporal ensemble

Multi-temporal information also plays an important role in missing data reconstruction (Long et al. 2020). Because we used the variation information between \( R_d \) and \( T_d \) to reconstruct the missing MAIAC AOD data, using MAIAC AOD data at only one \( R_d \) may fall into extreme situations and lead to unreliable results. Moreover, there could be continuous days with no reference data available restricting the reconstruction efficiency. Thus, a group of neighboring \( R_d \) was selected to reconstruct missing MAIAC AOD values in the IVMF. To be specific, adjacent weeks (7 days) before and after \( T_d \) (a total of 14 \( R_{ds} \)) were selected in the IVMF. Each \( R_d \) was processed following the above procedure to determine whether it was valid in reconstructing missing MAIAC AOD values (Figure 3).
In addition to the spatial uncertainty of MAIAC AOD variation patterns caused by spatial heterogeneity, AOD variations also show different patterns among different $R_{ds}$, resulting in differences in temporal dependency. To tackle this issue, we applied the ensemble strategy and introduced temporal similarity (TS) in the IVMF. Ensemble strategies are common approaches used in both big data regimes and remote sensing applications to aggregate estimations produced by different models and obtain improved results (Li et al. 2020b). Here, the AOD reconstructed from one $R_d$ can be treated as one output, and the ensemble strategy was employed to aggregate multi-temporal results from different $R_{ds}$. In the IVMF, TS was designed to represent the temporal dependency and treated as weight in the ensemble, as shown in Eq. (3):

$$ TS(T_d, R_d) = \frac{N_{C_{ps}}}{SD_{\Delta}} \quad (3) $$

where $N_{C_{ps}}$ is the total number of $C_{ps}$ in the region, and $SD_{\Delta}$ is the standard deviation of the calculated MAIAC AOD variations. The following aspects were considered in constructing the TS. (1) If the variations were stable across the region, then the reconstructed results were regarded as reliable. Standard deviation is a measure of the dispersion of data in statistics. A high $SD_{\Delta}$ indicates that the data points are spread far from the mean, revealing the instability of the variation. (2) If there were more pixels involved in the calculation, the information obtained increased and the uncertainty introduced by outliers decreased, indicating a more reliable result. (3) As a result, we assigned large weights to credible conditions with large number of $C_{ps}$ involved, and simultaneously, having small values of $SD_{\Delta}$. Conversely, conditions with few $C_{ps}$ and high values of $SD_{\Delta}$ were regarded as unreliable and assigned to small weights in the ensemble.

Finally, by involving TS as the temporal dependency to reduce the uncertainty, the multi-temporal reconstructed results calculated from the valid $R_{ds}$ were aggregated to generate the final result, expressed as:

$$ \tau(T_p, T_d) = \frac{\sum_{i=1}^{N} TS(T_d, R_{di}) \times \tau(T_d, T_p, R_{di})}{\sum_{i=1}^{N} TS(T_d, R_{di})} \quad (4) $$

where $\tau(T_p, T_d)$ is the final reconstructed $T_p$ MAIAC AOD result at $T_d$; $\tau(T_p, T_d, R_{di})$ is the $T_p$ MAIAC AOD at $T_d$ reconstructed from information of $R_{di}$; $N$ denotes the total number of valid $R_{di}$; and $TS(T_d, R_{di})$ represents the temporal similarity between $T_d$ and $R_{di}$.

### 2.2.3 Iterative reconstruction

Since the missing data often cover a large continuous area and it is irrational to account for a large surrounding area to calculate the variation information, we employed the successive iteration strategy to reconstruct the daily missing MAIAC AOD data across China. For each $T_d$, one iteration was referred to as the traversal computing of each
missing pixel. The following iteration was based on the results of the previous iteration. For each \( T_p \) at \( T_d \), the reconstructed MAIAC AOD value was updated during the iteration until the values converged. Here, convergence was defined as the change of reconstructed MAIAC AOD value being less than 1%. The stopping criteria of the IVMF were: (1) the increment of reconstructed AOD pixels (n) was less than 1% and 99% of the reconstructed MAIAC AOD pixels had converged; or (2) considering the trade-off between accuracy and efficiency, we constrained the maximum number of iterations (Iter) to 100.

2.2.4 Performance validation

To validate the performance of the IVMF, simulated and real-data experiments were carried out. Since there was no image with full MAIAC AOD coverage across China, we selected three images with the lowest amount of missing MAIAC AOD data across China in the simulated experiments (Table 1). The chosen images were 20 November 2001 (D1); 9 October 2013 (D2); and 11 November 2020 (D3). To demonstrate the capability and robustness of the IVMF, a manually masked region covering diverse landscapes and cities across China, with a relatively large area encompassing 1,000 pixels in width and 2,000 pixels in length (~2 million km²), was simulated for each chosen image. The original valid pixels in the region were applied to validate the reconstructed results. In the real-data experiments, the final reconstructed MAIAC AOD results and the original MAIAC AOD products were validated against in-situ AERONET measurements. The performance was quantitatively assessed with several metrics, including the coefficient of determination (\(R^2\)), correlation coefficient (\(R\)), root mean squared error (\(RMSE\)), index of agreement (\(IA\)), and expected error envelope (\(EE, \pm(0.05 + 20\%\)). The specific formulas of the above statistical metrics are provided in the Supplementary Material.

3. Results

3.1 Simulated experiments

As shown in Table 1, all the simulated experiments on the three selected images yielded satisfactory results, indicating the reliability and robustness of the IVMF. Specifically, D1 had the lowest \(R^2\), \(R\), and \(IA\) values, suggesting the worst performance among the three days. Conversely, since the AOD levels were the lowest at D1, the \(RMSE\) value (0.09) was the best. The highest \(R^2\), \(R\), and \(IA\) values approached 0.84, 0.93, and 92.05%, respectively, for D2. However, the most severe AOD levels at D2 resulted in the highest \(RMSE\) value (0.22) in the simulated experiments. D3 demonstrated the highest fraction within EE (76.72%), and performed satisfactorily on all the other four evaluation metrics. On average, the IVMF showed outstanding performance in reconstructing large areas of missing MAIAC AOD data, with \(R^2\), \(R\), \(RMSE\), and \(IA\) values of 0.71, 0.86, 0.14, and 85%, respectively, and the fraction within EE reaching 67.93%. The IVMF was more likely to overestimate AOD values as demonstrated by the fraction above EE (25.67%) being greater than the fraction below EE (6.40%).

To further illustrate the capability of the proposed IVMF, the reference MAIAC AOD, reconstructed MAIAC AOD, and the corresponding error maps are shown in Figure 4. It can be seen that all three reconstructed MAIAC AOD images were consistent with the corresponding reference MAIAC AOD images, especially the low AOD close to 0 at the left side of the simulated region. The overestimation of reconstructed MAIAC AOD in D1 was found mainly at the center of the simulated region, while the underestimated MAIAC AOD was on the right side of the region. The substantial number of pixels with overestimation resulted in the lowest \(R^2\) value for D1. For D2, the high AOD loadings on the

| Date       | Missing rate (%) | \(R^2\) | \(R\) | \(RMSE\) | \(IA\) (%) | Within EE (%) | Above EE (%) | Below EE (%) |
|------------|------------------|--------|------|---------|-----------|--------------|-------------|-------------|
| D1         | 11/20/2001       | 23.74  | 0.53 | 0.77    | 0.09      | 75.03        | 26.28       | 6.28        |
| D2         | 10/09/2013       | 28.51  | 0.84 | 0.93    | 0.22      | 92.05        | 35.40       | 4.97        |
| D3         | 11/11/2020       | 30.58  | 0.77 | 0.88    | 0.12      | 87.93        | 15.32       | 7.96        |
| Average    |                  | 27.61  | 0.71 | 0.86    | 0.14      | 85.00        | 25.67       | 6.40        |
right side of the region were reconstructed. However, the reconstructed high AOD loading area was larger than the reference image in the central portion, and was underestimated at the bottom. The AOD hot spots in D3 were largely captured in the reconstructed MAIAC AOD image, except for a small area in the middle of the region with underestimation. Meanwhile, because the simulated invalid region was relatively large, and the IVMF could only use spatial information outside the simulated region and temporal information from $R_{ds}$ to reconstruct the invalid MAIAC AOD. Hence invalid pixels remained after reconstruction. Overall, in the simulated experiments, the manually masked invalid region for three days with different AOD spatial patterns was largely reconstructed, indicating that the IVMF was reliable and robust in large-scale MAIAC AOD reconstruction.

### 3.2 Real-data experiments

To further evaluate the effect of the proposed IVMF, the final reconstructed MAIAC AOD results and original MAIAC AOD products were both validated against in-situ AERONET measurements in real-data experiments. Figure 4 shows the maps of the three images in the simulated experiments. The first row is the simulated images; the second and third rows are the reconstructed MAIAC AOD images and the reference MAIAC AOD images within the simulated invalid region. The last row is the corresponding reconstructed MAIAC AOD error images. The three columns represent D1, D2, and D3, respectively. Figure 5 presents density scatterplots of the validation results against AERONET measurements for (a) original MAIAC AOD products, and (b) final reconstructed MAIAC AOD results. The dashed lines are the EE envelope lines.
experiments (Figure 5). Over the period 2000 to 2020, a total of 17,746 measurements from 19 AERONET sites across China were available for validation. The original MAIAC AOD products had only 15,347 matching points with the AERONET measurements. After reconstruction, the number of matching points was improved to 17,680, with more than 99% of the available AERONET measurements matched. Meanwhile, the final reconstructed MAIAC AOD results showed satisfactory agreement with the AERONET measurements, with a high R(R^2) value of 0.91(0.82) and a low RMSE value of 0.22. Compared with the validation results of the original AOD, the performance slightly declined for the final reconstructed MAIAC AOD results. However, the number of available data points was significantly improved. 71.48% of the final reconstructed MAIAC AOD results met the EE envelope lines, and overestimation could be observed in both the original and final reconstructed MAIAC AOD, with the fraction above EE being higher than the fraction below EE. In general, the real-data experiments show that the proposed IVMF can maintain high quality MAIAC AOD results, while also significantly increasing the matching data points.

3.3 Efficiency of reconstruction

The efficiency of the reconstructing framework indicates the MAIAC AOD data coverage after reconstruction, which is also a critical issue in practice and needs to be addressed. In this part, daily coverage of the original MAIAC AOD and final MAIAC AOD results were first quantified. The daily coverage was calculated as the percentage of the valid MAIAC AOD pixels over land in the study area. As shown in the boxplot (Figure 6), the original MAIAC AOD had a relatively low coverage with an average value of 30.42%, and there were no days with coverage exceeding 80%. After reconstruction, the coverage of the final reconstructed MAIAC AOD was significantly improved and reached an average value of 96.69%, an increase of 218% when compared with that of the original MAIAC AOD. More importantly, nearly all the days had high coverage of over 90%, indicating the high reconstructing efficiency of the IVMF.

Furthermore, we assessed the MAIAC AOD coverage over the study area, expressed as the percentage of valid days from 2000 to 2020 for each pixel (Figure 7). As can be seen, the original MAIAC AOD demonstrated poor AOD coverage across China. All pixels had valid days making up less than 60% of the total days and approximately half of the pixels had a coverage of less than 20%, which severely limited the application of the data. Specifically, subregions around central, south, southwest and northeast China showed the lowest AOD coverage. As shown in Figure 7b, the AOD coverage was significantly improved after reconstruction. Nearly all pixels demonstrated a high AOD coverage of over 80%, with an average value of 96.76%. However, the AOD coverage after reconstruction was lower in northeast China than in other regions.

**Figure 6.** Boxplot of daily coverage of the original MAIAC AOD and final reconstructed MAIAC AOD results over the study period.
3.4 AOD spatiotemporal patterns

Figure 8 shows the averaged AOD levels from 2000 to 2020 before and after reconstruction. The overall spatial patterns of MAIAC AOD across China after reconstruction were largely consistent with the original AOD: high AOD loadings were mainly distributed around central China, while northwest China had a low AOD level. Moreover, after reconstruction, the spatial distribution of AOD was more coherent and the boundaries among regions with various AOD levels were more marked. For annual AOD variations (Figure S2), the highest AOD loadings occurred in 2011, wherein the annual mean AOD value over China was 0.42. The detailed annual AOD variations can be found in Supplementary Materials.

4. Discussion

4.1 Spatiotemporal variability analysis

This study proposed the novel big data mining IVMF to reconstruct large-scale missing MAIAC AOD data across China from 2000 to 2020. The original MAIAC AOD data had a relatively low coverage with an average value of 30.42%. Since the MAIAC AOD products were only available under clear skies, frequent rainy weather and cloud occurrences resulted in poor spatial continuity of MAIAC AOD in subregions around central, south, and southwest China (Qin et al. 2021). Moreover, unfavorable surface conditions, such as snow cover, and misclassifications of heavy aerosol layers as clouds can also lead to missing data even under clear skies (Song et al. 2019b). In northeast China, large amounts of AOD data were missing due to the perennial snow cover in winter (Qin et al. 2021). After reconstruction with the proposed IVMF, the large volume of missing AOD data was recovered, and the average AOD coverage approached almost 100%; this was satisfactory for further applications. Recall the maximum iteration number was constrained to 100 because of the computation burden, 100% coverage for all pixels cannot be guaranteed by the current study. In addition, because the IVMF mined local spatiotemporal variation information within the original MAIAC AOD products to
reconstruct the missing AOD data, the relatively low coverage of the original MAIAC AOD and insufficient surrounding information along the edge of the study area impeded the efficiency of reconstruction in northeast China.

Owing to less data availability, the original AOD products across China had high uncertainty in representing the accurate annual AOD levels. With significant improvements in AOD coverage after reconstruction, the annual AOD levels increased across China and showed more coherence in spatial distribution. Rapid urbanization and industrialization resulted in a large amount of anthropogenic aerosol emissions, leading to high AOD loadings around central China (Liu et al. 2019; Huang, Yan, and Zhang 2018). The AOD levels in the northwest and southwest China were lower due to less human activity and higher altitudes in these regions (Fang et al. 2020). The decreasing trend in AOD after 2011 could be related to the ecological civilization construction proposed by the Chinese government in 2012 (Lu et al. 2020). Various air pollution prevention and control action plans were promulgated, resulting in improved air quality and decreases in AOD loadings nationwide (Bai et al. 2019). Considering the satisfying validation results from both simulated and real-data experiments, the final reconstructed MAIAC AOD products are reliable in terms of high data coverage and reasonable spatiotemporal patterns in comparison to the original data.

4.2 Error analysis

This section investigated the errors induced in the IVMF when reconstructing the missing AOD data. It can be seen in the simulated and real-data experiments that the IVMF tended to overestimate the MAIAC AOD values (fraction above EE higher than fraction below EE). Figure S3 illustrates the reconstructed MAIAC AOD bias as a function of the original AOD in the simulated experiments. Overestimations were mainly found at low AOD values (0–0.2 for D1, 0–0.5 for D2, and 0–0.4 for D3), while underestimations were mostly presented when AOD values were high (>0.4 for D1 and D3, and >0.5 for D2). The number of underestimations for the high AOD condition was far less than the number of overestimations for low AOD, leading to the fraction above EE being higher than fraction below EE in the simulated experiments. However, the biases of the three days were all mainly distributed around 0 and shaped similar to a normal distribution, which is acceptable for reconstruction.

Validated against the AERONET measurements, the IVMF also overestimated at low AOD levels, especially at the levels of 0–0.5 (Figure 9). The same phenomenon was observed in the original MAIAC AOD products due to errors induced by aerosol models used in the MAIAC algorithm (Zhang et al. 2020; Filonchyk and Hurynovich 2020). Because the IVMF mines the local spatiotemporal variation information within the original MAIAC AOD products to reconstruct the missing AOD data, the overestimations in the original MAIAC AOD products resulted in overestimations in the final reconstructed MAIAC AOD. Regarding different AERONET sites, considerable overestimations were found at the Taihu site. Because there were water bodies around the Taihu site, and the water of Taihu Lake was polluted, the high uncertainty of surface reflectivity resulted in retrieving errors for the original MAIAC AOD in the Taihu site (Xie et al. 2019). After reconstruction, the errors were emphasized and the number of overestimations exceeded those within EE. The QOMS_CAS and NAM_CO sites had low R and R² values, but most of the AOD values fell within the EE envelope. These two sites were located at high altitudes regions (over 4.2 km), and the MAIAC algorithm designated AOD values of 0.014 to these sites using climatology parameters, and the reconstructed AOD values were also around 0.014 (Liu et al. 2019). Other sites all performed well in the real-data experiments, indicating the reliability of the original MAIAC AOD data and the effectiveness of the IVMF.

Different from traditional geostatistical-based interpolation, the iteration strategy was employed in the IVMF to reconstruct MAIAC AOD data gaps at a large scale and maximize the reconstruction efficiency. The primary concern in the iteration strategy was that the errors could be propagated during iteration, resulting in the inferior performance of the reconstruction. To better assess the uncertainty associated with the iteration strategy, we evaluated the model performance after each iteration in the simulated experiments, and set the maximum iteration number to 200 (Figure S4). Among the three days, the stopping criteria were satisfied for D1 after 114 iterations, while AOD values did not converge for D2 and D3 after 200 iterations. During the first 100 iterations, R², R, and IA values for the three
days declined gradually according to the same pattern. After that, $R^2$, $R$, and IA values increased slightly because the values gradually converged over iteration. RMSE values had the reverse patterns to that for $R^2$, $R$, and IA values. The fraction within EE showed different patterns among the three days. After iteration 100, the fraction within EE increased steadily for D1 and D3, while for D2 it decreased continuously from 60.28% to 59.63%. The slight decrease in the fraction within EE resulted from the extremely high AOD levels in D2, and the IVMF reconstructed a larger area with high AOD loadings. In general, the accuracy of the IVMF was inevitably influenced by the iteration strategy due to the propagation of errors, but the performance degradation remained acceptable considering the compromise between efficiency and accuracy.

4.3 Comparison to previous studies

Many researches have attempted to reconstruct missing AOD data gaps (Table 2). The spatiotemporal kriging method implemented by Yang and Hu (2018) reached a high reconstructed AOD coverage of 67.73%, but the study region was small (Beijing, China) and the AOD spatial resolution was low (10 km). On large scales, Lianfa et al. (2020a) developed a deep residual network to reconstruct weekly MAIAC AOD data from 2000 to 2016 over California. The reconstructed AOD showed good correlation with AERONET AOD. Chi et al. (2020) proposed a two-step model to recover missing AOD at 3-km spatial resolution from 2017 to 2019 in east Asia. The R value of reconstructed AOD and AERONET AOD was 0.87, and the daily AOD missing rate was reduced to 10%.

Table 2. Comparison of previous AOD reconstruction studies.

| Source                  | Study region | Study period | Temporal resolution | Spatial resolution | Reconstructed coverage | $R$ | $R^2$ | External data |
|-------------------------|--------------|--------------|---------------------|-------------------|------------------------|-----|-------|--------------|
| (Yang and Hu 2018)      | Beijing      | 2014         | Daily               | 10 km             | 67.73%                 | -   | -     | No           |
| (Li et al. 2020a)       | California   | 2000–2016    | Weekly              | 1 km              | 100%                   | 0.83| 0.69  | Yes          |
| (Chi et al. 2020)       | East Asia    | 2017–2019    | Daily               | 3 km              | 90%                    | 0.87| -     | Yes          |
| (Wang et al. 2019)      | Asia         | 2016         | Daily               | 10 km             | 45.9%                  | 0.87| -     | Yes          |
| This study              | China        | 2000–2020    | Daily               | 1 km              | 96.69%                 | 0.91| 0.82  | No           |

Note: – indicates a lack of data in the cited article.
after reconstruction. However, external data, such as meteorology and elevation data, were essential in the model. In contrast, Wang et al. (2019) proposed a large-scale AOD reconstruction framework with no external meteorology and elevation data. The validation results against AERONET AOD showed a high R value of 0.87, but the daily AOD coverage after reconstruction was only 45.9% and the AOD spatial resolution was 10 km. In this study, the real-data experiments showed that the reconstructed AOD was in good agreement with AERONET AOD (R = 0.91; $R^2 = 0.82$), and daily AOD coverage from 2000 to 2020 was improved from 30.42% to 96.69% after reconstruction. Compared to the aforementioned studies, our proposed IVMF had outstanding performance in accuracy and efficiency over a larger study area, longer period, and with a higher spatial resolution. Moreover, none of the aforementioned studies evaluated the reconstructed AOD in simulated experiments, which is essential in verifying the effectiveness of the missing data recovery method (Zhang, Yuan et al. 2018; Zeng et al. 2018). The simulated experiments of the three images were consistent with the reference images, with average $R^2$ and R values of 0.71 and 0.86, respectively. Therefore, requiring no external data, the proposed IVMF was both effective and reliable in reconstructing daily large-scale MAIAC AOD data gaps, and showed advantages over existing approaches.

4.4 Limitations and future work

Several limitations of our study should be noted. First, the spatial heterogeneity can change over time, especially over 20 years. This heterogeneity cannot be completely represented by only one value of spatial similarity (R) in the IVMF. However, the atmospheric environment is a complicated and unstable system, and it is hard to accurately represent the reality of the spatial heterogeneity of AOD at present. While our study estimated the spatial heterogeneity and reduced its effect to some extent, a better way to estimate the spatial heterogeneity still calls for further work. Second, the maximum iteration number was constrained to 100 because of the computation burden. Therefore, 100% coverage for all pixels cannot be guaranteed by the current study. However, AOD coverage with an average value of 96.76% is still satisfactory for further applications. In future work, more iterations could be executed to enlarge the efficiency of reconstructing and obtain spatially continuous AOD products. Finally, due to the vast geographical scope in China, the MODIS was unable to retrieve full MAIAC AOD data within one observation. Different MODIS observation times resulted in the abrupt change of AOD levels (a division line) over north China, both in the original MAIAC AOD products and the final reconstructed MAIAC AOD results. More research is needed to resolve the incompatibility in the AOD levels due to different observation times and to improve the data quality.

5. Conclusion

Satellite-based AOD retrieval products provide valuable opportunities to characterize large-scale aerosol distributions and air pollution levels. However, the high AOD missing rate due to cloud contamination has impeded their application, and thus, reconstructing the missing AOD data gaps is imperative and meaningful in practice. In this study, a novel big data based IVMF was proposed to reconstruct daily large-scale MAIAC AOD data gaps across China from 2000 to 2020. The validation results of simulated and real-data experiments indicated the effectiveness and robustness of the proposed framework. Compared to previous studies, our approach has advantages in the aspects of no external data requirement, relatively high reconstructing efficiency at large scales, and outstanding reconstructing accuracy. Besides, we reconstructed the daily 1-km MAIAC AOD over China for 20 years (from 26 February 2000 to 31 December 2020), resulting in a considerable increase in the daily AOD coverage from 30.42% to 96.69%; this will benefit further applications, such as the estimation of air pollution levels or climate change research. Furthermore, the IVMF can be generalized to other regions and different
remote sensing products, such as AOD data from other satellites, land surface temperature products, and other satellite-based observations of Earth properties.

**Acknowledgements**

The authors would like to express our gratitude to NASA and AERONET for providing the AOD data.

**Disclosure statement**

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

**Funding**

This work was supported by National Key Research and Development Program of China under [Grant No. 2020YFC1807501] and Key Research and Development Program of Zhejiang Province under [Grant No. 2022C03078].

**Data availability statement**

The data that support the findings of this study are openly available in Zenodo at 10.5281/zenodo.5520164. MAIAC AOD data can be accessed from https://ladsweb.modaps.eosdis.nasa.gov/search/. AERONET AOD data can be accessed from https://aeronet.gsfc.nasa.gov/.

**References**

Bai, K., M. Ma, N. B. Chang, and W. Gao. 2019. “Spatiotemporal Trend Analysis for Fine Particulate Matter Concentrations in China Using High-resolution Satellite-derived and Ground-measured PM2.5 Data.” *Journal of Environmental Management* 233: 530–542. doi:10.1016/j.jenvman.2018.12.071.

Belle, J., and Y. Liu. 2016. “Evaluation of Aqua MODIS Collection 6 AOD Parameters for Air Quality Research over the Continental United States.” *Remote Sensing* 8 (10): 815. doi:10.3390/rs8100815.

Chen, Z., J. Jin, R. Zhang, T. Zhang, J. Chen, J. Yang, C. Qu, and Y. Guo. 2020. “Comparison of Different Missing-Imputation Methods for MAIAC (Multiangle Implementation of Atmospheric Correction) AOD in Estimating Daily PM2.5 Levels.” *Remote Sensing* 12 (18): 3008. doi:10.3390/rs12183008.

Chen, Z., T. Zhang, R. Zhang, Z. Zhu, J. Yang, P. Chen, C. Qu, and Y. Guo. 2019. “Extreme Gradient Boosting Model to Estimate PM2.5 Concentrations with Missing-filled Satellite Data in China.” *Atmospheric Environment* 202: 180–189. doi:10.1016/j.atmosenv.2019.01.027.

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): 3786. doi:10.3390/rs12223786.

Fang, H., W. Qin, L. Wang, M. Zhang, and X. Yang. 2020. “Solar Brightening/Dimming over China’s Mainland: Effects of Atmospheric Aerosols, Anthropogenic Emissions, and Meteorological Conditions.” *Remote Sensing* 13 (1): 88. doi:10.3390/rs13010088.

Filonchyk, M., and V. Hurynovich. 2020. “Validation of MODIS Aerosol Products with AERONET Measurements of Different Land Cover Types in Areas over Eastern Europe and China.” *Journal of Geovisualization and Spatial Analysis* 4 (1). doi:10.1007/s41651-020-00052-9.

Fu, D., Z. Song, X. Zhang, X. Xia, J. Wang, H. Che, H. Wu, X. Tang, J. Zhang, and M. Duan. 2020. “Mitigating MODIS AOD Non-random Sampling Error on Surface PM2.5 Estimates by a Combined Use of Bayesian Maximum Entropy Method and Linear Mixed-effects Model.” *Atmospheric Pollution Research* 11 (3): 482–490. doi:10.1016/j.apr.2019.11.020.

He, Q., Y. Gu, and M. Zhang. 2020. “Spatiotemporal Trends of PM2.5 Concentrations in Central China from 2003 to 2018 Based on MAIAC-derived High-resolution Data.” *Environment International* 137: 105536. doi:10.1016/j.envint.2020.105536.

He, Q., and B. Huang. 2018. “Satellite-based Mapping of Daily High-resolution Ground PM2.5 In China via Space-time Regression Modeling.” *Remote Sensing of Environment* 206: 72–83. doi:10.1016/j.rse.2017.12.018.

He, L., L. Wang, B. Huang, J. Wei, Z. Zhou, and Y. Zhong. 2020. “Anthropogenic and Meteorological Drivers of 1980–2016 Trend in Aerosol Optical and Radiative Properties over the Yangtze River Basin.” *Atmospheric Environment* 223: 117188. doi:10.1016/j.atmosenv.2019.117188.

Hu, X., L. A. Waller, A. Lyapustin, Y. Wang, and Y. Liu. 2014. “10-year Spatial and Temporal Trends of PM2.5 Concentrations in the Southeastern US Estimated Using High-resolution Satellite Data.” *Atmospheric Chemistry and Physics* 14 (12): 6301–6314. doi:10.5194/acp-14-6301-2014.

Huang, K., Q. Xiao, X. Meng, G. Geng, Y. Wang, A. Lyapustin, D. Gu, and Y. Liu. 2018. “Predicting Monthly High-resolution PM2.5 Concentrations with Random Forest Model in the North China Plain.” *Environmental Pollution* 242 (Pt A): 675–683. doi:10.1016/j.envpol.2018.07.016.

Huang, Y., Q. Yan, and C. Zhang. 2018. “Spatial–Temporal Distribution Characteristics of PM2.5 In China in 2016.” *Journal of Geovisualization and Spatial Analysis* 2 (2). doi:10.1007/s41651-018-0019-5.

Jinnagara Puttaswamy, S., H. M. Nguyen, A. Braverman, X. Hu, and Y. Liu. 2013. “Statistical Data Fusion of Multi-sensor AOD over the Continental United States.” *Geocarto International* 29 (1): 48–64. doi:10.1080/10106049.2013.827750.

Kaufman, Y. J., D. Tanré, and O. Boucher. 2002. “A Satellite View of Aerosols in the Climate System.” *Nature* 419 (6903): 215–223. doi:10.1038/nature01091.
Kikuchi, M., H. Murakami, K. Suzuki, T. M. Nagao, and A. Higurashi. 2018. "Improved Hourly Estimates of Aerosol Optical Thickness Using Spatiotemporal Variability Derived from Himawari-8 Geostationary Satellite." *IEEE Transactions on Geoscience and Remote Sensing* 56 (6): 3442–3455. doi:10.1109/tgrs.2018.2800060.

Li, L., M. Franklin, M. Girguis, F. Lurmann, J. Wu, N. Pavlovic, C. Breton, F. Gilliland, and R. Habre. 2020a. "Spatiotemporal Imputation of MAIAC AOD Using Deep Learning with Downsampling." *Remote Sensing of Environment* 237. doi:10.1016/j.rse.2019.115184.

Li, L., M. Girguis, F. Lurmann, N. Pavlovic, C. McClure, M. Franklin, J. Wu, et al. 2020b. "Ensemble-based Deep Learning for Estimating PM2.5 Over California with Multisource Big Data Including Wildfire Smoke." *Environment International* 145:106143. doi:10.1016/j.envint.2020.106143.

Li, T., H. Shen, C. Zeng, Q. Yuan, and L. Zhang. 2017. "Point-Surface Fusion of Station Measurements and Satellite Observations for Mapping PM 2.5 Distribution in China: Methods and Assessment." *Atmospheric Environment* 152: 477–489. doi:10.1016/j.atmosenv.2017.01.004.

Li, X., and X. Zhang. 2019. "Predicting Ground-level PM2.5 Concentrations in the Beijing-Tianjin-Hebei Region: A Hybrid Remote Sensing and Machine Learning Approach." *Environmental Pollution* 249: 735–749. doi:10.1016/j.envpol.2019.03.068.

Li, L., J. Zhang, X. Meng, Y. Fang, Y. Ge, J. Wang, C. Wang, J. Wu, and H. Kan. 2018. "Estimation of PM2.5 Concentrations at a High Spatiotemporal Resolution Using Constrained Mixed-effect Bagging Models with MAIAC Aerosol Optical Depth." *Remote Sensing of Environment* 217: 573–586. doi:10.1016/j.rse.2018.09.001.

Liang, F., Q. Xiao, Y. Wang, A. Lyapustin, G. Li, D. Gu, X. Pan, and Y. Liu. 2018. "MAIAC-based Long-term Spatiotemporal Trends of PM2.5 In Beijing, China." *Science of the Total Environment* 616-617: 1589–1598. doi:10.1016/j.scitotenv.2017.10.155.

Liu, N., B. Zou, H. Feng, W. Wang, Y. Tang, and Y. Liang. 2019. "Evaluation and Comparison of Multimodel 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.

Long, D., L. Yan, L. Bai, C. Zhang, X. Li, H. Lei, H. Yang, et al. 2020. "Generation of MODIS-like Land Surface Temperatures under All-weather Conditions Based on a Data Fusion Approach." *Remote Sensing of Environment* 246:111863. doi:10.1016/j.rse.2020.111863.

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.

Lyapustin, A., J. Martonchik, Y. Wang, I. Laszlo, and S. Korkin. 2011b. "Multimodel Implementation of Atmospheric Correction (MAIAC): 1. Radiative Transfer Basis and Look-up Tables." *Journal of Geophysical Research* 116 (D3). doi:10.1029/2010jd014985.

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.

Lyapustin, A., Y. Wang, I. Laszlo, R. Kahn, S. Korkin, L. Remer, R. Levy, and J. S. Reid. 2011a. "Multimodel Implementation of Atmospheric Correction (MAIAC): 2. Aerosol Algorithm." *Journal of Geophysical Research* 116 (D3). doi:10.1029/2010jd014986.

Ma, Z., X. Hu, L. Huang, J. Bi, and Y. Liu. 2014. "Estimating Ground-level PM2.5 In China Using Satellite Remote Sensing." *Environmental Science & Technology* 48 (13): 7436–7444. doi:10.1021/es5009399.

Murray, C. J. L., A. Y. Aravkin, P. Zheng, C. Abbafati, K. M. Abbas, M. Abbas-Kangevari, F. Abd-Allah, et al. 2020. "Global Burden of 87 Risk Factors in 204 Countries and Territories, 1990-2019: A Systematic Analysis for the Global Burden of Disease Study 2019." *The Lancet* 396 (10258): 1223–1249. doi:10.1016/S0140-6736(20)30752-2.

Olcese, L. E., G. G. Palanca, and B. M. Toselli. 2015. "A Method to Estimate Missing AERONET AOD Values Based on Artificial Neural Networks." *Atmospheric Environment* 113: 140–150. doi:10.1016/j.atmosenv.2015.05.009.

Park, S., J. Lee, J. Im, C.-K. Song, M. Choi, J. Kim, S. Lee, et al. 2020. "Estimation of Spatially Continuous Daytime Particulate Matter Concentrations under All Sky Conditions through the Synergistic Use of Satellite-based AOD and Numerical Models." *Science of the Total Environment* 713:136516. doi:10.1016/j.scitotenv.2020.136516.

Pu, Q., and E. H. Yoo. 2021. "Ground PM2.5 Prediction Using Imputed MAIAC AOD with Uncertainty Quantification." *Environmental Pollution* 274: 116574. doi:10.1016/j.envpol.2021.116574.

Qin, W., H. Fang, L. Wang, J. Wei, M. Zhang, X. Su, M. Bilal, and X. Liang. 2021. "MODIS High-resolution MAIAC Aerosol Product: Global Validation and Analysis." *Atmospheric Environment* 264: 118684. doi:10.1016/j.atmosenv.2021.118684.

Song, Z., D. Fu, X. Zhang, X. Han, J. Song, J. Zhang, J. Wang, and X. Xia. 2019b. "MODIS AOD Sampling Rate and Its Effect on PM2.5 Estimation in North China." *Atmospheric Environment* 209: 14–22. doi:10.1016/j.atmosenv.2019.04.020.

Song, Y., B. Huang, Q. He, B. Chen, J. Wei, and R. Mahmood. 2019a. "Dynamic Assessment of PM2.5 Exposure and Health Risk Using Remote Sensing and Geo-spatial Big Data." *Environmental Pollution* 253: 288–296. doi:10.1016/j.envpol.2019.06.057.

Tobler, W. R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46: 234–240. doi:10.2307/143141.
Wang, Y., L. Chen, S. Li, X. Wang, C. Yu, Y. Si, and Z. Zhang. 2017. “Interference of Heavy Aerosol Loading on the VIIRS Aerosol Optical Depth (AOD) Retrieval Algorithm.” Remote Sensing 9 (4): 397. doi:10.3390/rs9040397.

Wang, S., X. Liu, X. Yang, B. Zou, and J. Wang. 2018a. “Spatial Variations of PM2.5 In Chinese Cities for the Joint Impacts of Human Activities and Natural Conditions: A Global and Local Regression Perspective.” Journal of Cleaner Production 203:143–152. doi:10.1016/j.jclepro.2018.08.249.

Wang, X., G. Tian, D. Yang, W. Zhang, D. Lu, and Z. Liu. 2018b. “Responses of PM2.5 Pollution to Urbanization in China.” Energy Policy 123:602–610. doi:10.1016/j.enpol.2018.09.001.

Wang, Y., Q. Yuan, T. Li, H. Shen, L. Zheng, and L. Zhang. 2019. “Large-scale MODIS AOD Products Recovery: Spatial-temporal Hybrid Fusion considering Aerosol Variation Mitigation.” ISPRS Journal of Photogrammetry and Remote Sensing 157:1–12. doi:10.1016/j.isprsjprs.2019.08.017.

Xiao, Q., G. Geng, J. Cheng, F. Liang, L. Rui, X. Meng, T. Xue, et al. 2021. “Evaluation of Gap-filling Approaches in Satellite-based Daily PM2.5 Prediction Models.” Atmospheric Environment 244:117921. doi:10.1016/j.atmosenv.2020.117921.

Xie, G., M. Wang, J. Pan, and Y. Zhu. 2019. “Spatio-temporal Variations and Trends of MODIS C6.1 Dark Target and Deep Blue Merged Aerosol Optical Depth over China during 2000–2017.” Atmospheric Environment 214:116846. doi:10.1016/j.atmosenv.2019.116846.

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. doi:10.1016/j.scitotenv.2018.03.202.

Yu, M., G. Tang, Y. Yang, Q. Li, Y. Wang, S. Miao, Y. Zhang, and Y. Wang. 2020. “The Interaction between Urbanization and Aerosols during a Typical Winter Haze Event in Beijing.” Atmospheric Chemistry and Physics 20 (16): 9855–9870. doi:10.5194/acp-20-9855-2020.

Zeng, C., D. Long, H. Shen, P. Wu, Y. Cui, and Y. Hong. 2018. “A Two-step Framework for Reconstructing Remotely Sensed Land Surface Temperatures Contaminated by Cloud.” ISPRS Journal of Photogrammetry and Remote Sensing 141:30–45. doi:10.1016/j.isprsjprs.2018.04.005.

Zhang, R., B. Di, Y. Luo, X. Deng, M. L. Grieneisen, Z. Wang, G. Yao, and Y. Zhan. 2018. “A Nonparametric Approach to Filling Gaps in Satellite-retrieved Aerosol Optical Depth for Estimating Ambient PM2.5 Levels.” Environmental Pollution 243 (Pt B):998–1007. doi:10.1016/j.envpol.2019.09.052.

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

Zhang, Q., Q. Yuan, C. Zeng, L. Xinghua, and Y. Wei. 2018. “Missing Data Reconstruction in Remote Sensing Image with a Unified Spatial–Temporal–Spectral Deep Convolutional Neural Network.” IEEE Transactions on Geoscience and Remote Sensing 56(8):4274–4288. doi:10.1109/tgrs.2018.2810208.

Zhang, Z., M. Zhang, M. Bilal, B. Su, C. Zhang, and L. Guo. 2020. “Comparison of MODIS- and CALIPSO-Derived Temporal Aerosol Optical Depth over Yellow River Basin (China) from 2007 to 2015.” Earth Systems and Environment 4 (3):535–550. doi:10.1007/s41748-020-00181-7.

Zou, B., J. You, Y. Lin, X. Duan, X. Zhao, X. Fang, M. J. Campen, and S. Li. 2019. “Air Pollution Intervention and Life-saving Effect in China.” Environment International 125:529–541. doi:10.1016/j.envint.2018.10.045.