SPATIOTEMPORAL RECOVERY OF HIMAWARI-8 HOURLY AEROSOL OPTICAL DEPTH PRODUCTS VIA THE NESTED BAYESIAN MAXIMUM ENTROPY METHOD

Xinghui Xia¹, Zhongmin Zhu², *, Tianhao Zhang¹, Gong Wei¹, Yuxi Ji¹

1State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, P.R. China - shirley_xia@whu.edu.cn
2College of Information Science and Engineering, Wuhan University, Wuhan, P.R. China - zhongminzhu@whu.edu.cn
3Electronic Information School, Wuhan University, Wuhan, P.R. China
4Wuhan Geomatics Institute, Wuhan, P.R. China

Commission III, WG III/6

KEY WORDS: AOD, Himawari-8, NBME, hourly sequence

ABSTRACT:

Satellite-derived aerosol optical depth (AOD) is an indispensable parameter when conducting studies related to atmospheric environment, climate change, and biogeochemical cycle. However, current satellite-derived AOD products are limited in related applications due to the large proportion of missing data, and the existed methods mainly concentrate on recovering AOD from polar-orbit satellite sensors. In order to solve these issues and take full use of the preponderance of geostationary satellite sensors with high frequency observation, we propose a spatiotemporal AOD recovery framework integrating multi-time scale AOD products based on the nested Bayesian maximum entropy methodology (NBME), aimed to obtain satellite-derived AOD datasets with low data missing and high accuracy. The experiment results show that the spatial coverage of AOD datasets increases from 20.5% to 70.0%, and the R² and RMSE of the recovered AOD against ground-based AERONET AOD are approximately 0.62 and 0.19, respectively. Moreover, the further simulated experiments indicate that the proposed method also performs better relatively when comparing with other popular recovery methods. Therefore, the proposed NBME recovery method can obtain a more convincing product both in applicable accuracy and visual quality.

1. INTRODUCTION

Atmospheric aerosols, generally solid and liquid particles from natural and man-made sources, are a primary uncertainty associated with climate change due to their direct effects of scattering and absorption of solar radiation and indirect effects that affect the radiative properties and lifetime of clouds (Kaufman et al., 1997; Kaufman et al., 2002; Gu et al., 2006; Gu et al., 2012; Intergovernmental Panel on Climate Change, 2013; Ramanathan et al., 2014; Zhao et al., 2019; Xia et al., 2022). Satellite-derived aerosol optical depth (AOD) is an effective and efficient parameter when conducting studies related to atmospheric environment, climate change, and biogeochemical cycle. Accurate understanding of the spatial distribution and dynamic changes of anthropogenic aerosol emissions is the cornerstone of the deteriorating atmospheric environment (Anderson et al., 2005; Andreae and Rosenfeld, 2008). Unfortunately, the relatively high data missing ratio of satellite-derived AOD due to the influence of satellite orbit interval, cloud occlusion and the limitations of retrieval algorithms, restricts the atmosphere related applications and researches to a certain extent (Hsu et al., 2012; Wang and Yuan et al., 2019). Therefore, AOD recovery in satellite-derived AOD is of great significance.

In recent years, scholars have undertaken a great deal of study into AOD recovery for different types of satellite sensors, which are primarily constrained to AOD products retrieved from polar-orbit satellites sensors. To our knowledge, the distribution of aerosol possesses highly spatial and temporal variability, even over a single day. Although polar-orbiting satellite sensors including but not limited to Moderate Resolution Imaging Spectroradiometer (MODIS), Multi-angle Imaging Spectroradiometer (MISR), and Sea-viewing Wide Field-of-view Sensor (SeaWiFS), could obtain all-weather and large-scale aerosol observation information, they could only achieve observations at most 1-2 times per day in the same area (Kokhanovsky, 2007). Nevertheless, apart from having the same observational advantage of polar-orbiting satellites sensors, due to the typical characteristics of high frequency observation up to minute, AOD retrieved by geostationary satellite sensors such as the Advanced Himawari Imager (AHI), can preferably meet the requirements of capturing the dynamic changes of daily aerosols.

In addition, the existing AOD recovery algorithms are generally divided into two categories. One type of the recovery approaches could only provide missing values at grids according to probability statistics information from the original satellite datasets, such as the linear or second-order polynomial functions (Mélin et al., 2007), the optimum interpolation (Xue et al., 2012; Xue et al., 2014), the empirical

* Corresponding author
orthogonal functions (Liu et al., 2015), the least-squares method (Gou et al., 2013) and the primitive machine learning method (Zhao et al., 2019). The other type of the AOD recovery method are based on geostatistical model derived from Tobler’s first law of geography and estimate the AOD at missing pixels by geostatistical information like spatial autocorrelation, consisting of universal kriging method (Chatterjee et al., 2010; Li et al., 2014), the geographically weighted regression (Wang et al., 2013) and their optimized ramification (Nguyen et al., 2012; Puttaswamy et al., 2014; Zhang et al., 2017). As scholars improve these methods by introducing temporal autocorrelation, the further methods like spatiotemporal interpolation (Yang et al., 2018) and Bayesian maximum entropy (Tang et al., 2016) evolve naturally, which have been proved to greatly ameliorates the completeness and quality of AOD products especially in cases where the original pixels are missing (Kang et al., 2010).

Therefore, we propose a spatiotemporal AOD recovery framework for geostationary satellite sensors, which is based on day-hour nested Bayesian maximum entropy methodology, and aim to take full advantage of the spatio-temporal autocorrelation between AOD datasets at multi-time scale and assimilate the highest quality AOD information by a nested nonlinear spatiotemporal geostatistical model. Furthermore, the proposed algorithm is evaluated from the perspectives of completeness and accuracy by several sets of experimental data.

2. DATA AND METHOD

2.1 Study Region and Data

The study region (106°E – 127°E, 21°N – 42°N) of this paper is mainly located in the East Asia. Compared to the other areas, the region has a relatively high population density and aerosol loading throughout the world, which also covers several major pollution-prone areas in China (Streets et al., 2009), such as the Beijing-Tianjin-Hebei, Yangtze River Delta, pearl River Delta and the Central Plains Economic Zone. Therefore, it is significant to generate the spatiotemporal continuous and precise AOD in this region.

In this study, the geostationary satellite aerosol products are AOD datasets at 500nm derived from Himawari-8/AHI, including Level 3 AOD (AOD_Merged) with the temporal and spatial resolutions of 1 hour and 5 km (ranging from 00:00 to 08:00 Coordinated Universal Time, UTC), and Level 3 AOD with the temporal and spatial resolutions of 1 day and 5 km, which are all obtained from JAXA (http://www.eorc.jaxa.jp/ptree) covering the whole year of 2016. The AERONET Level 2 AOD observation data at 500nm could be downloaded from http://aeronet.gsfc.nasa.gov, which is mainly used for constructing the soft data of satellite AODs and validate the accuracy of experimental results.

2.2 Nested AOD Recovery Methodology (NBME)

In terms of obtaining high-precision AOD estimation, the classic BME model is widely known for its powerful ability to fill the gaps of daily satellite-derived AOD, however, it doesn’t perform well enough in cases when there is too much overlap and redundancy of spatiotemporal information between the original data. Hence, we applied a nested BME model with day-, hour-level autocorrelation to eliminate this deficiency, which also can help take full advantage of every Himawari-8/AHI AOD image.

First of all, based on the available matchups for each hour (AOD_{x,y,Hour}), and each day (AOD_{x,y,Day}) in the experimental period, the coefficients of the linear model between hourly data and daily data could be obtained using least squares method. According to the obtained linear regression model, a corresponding ‘hourly’ data is reconstructed from the daily data, which also serves as input datasets for recovery framework.

Then the quantified global spatiotemporal trend of all pixels needs to be removing (Xia et al., 2022), which is a necessary pre-processing before analyzing the spatiotemporal autocovariance structure of input datasets. Meanwhile, the AOD residual of every pixel is further processed into soft data based on the equation (1):

$$AOD_{i,Hour} \sim N(AOD_{Hour} + \mu, \sigma^2)$$  \hspace{1cm} (1)

where $\mu$ and $\sigma^2$ is the mean and variance of random error between the satellite AOD data and the corresponding AERONET AOD, respectively. $AOD_{Hour}$ is the AOD residual of satellite AOD or reconstructed hourly AOD data, and $AOD_{i,Hour}$ is the Gaussian distribution probability soft data. After this, the relevant soft data of every estimation point is pre-processing before analyzing the spatiotemporal trend of all pixels.

According to the obtained linear regression model, a reconstructed hourly AOD data, and $AOD_{i,Hour}$ is the Gaussian distribution probability soft data. After this, the relevant soft data of every estimation point is pre-processing before analyzing the spatiotemporal trend of all pixels.

$$x_{AOD} = \int x_{AOD} f(x_{AOD}) dx_{AOD}$$  \hspace{1cm} (2)

Then the quantified global spatiotemporal trend of all pixels needs to be removing (Xia et al., 2022), which is a necessary pre-processing before analyzing the spatiotemporal autocovariance structure of input datasets. Meanwhile, the AOD residual of every pixel is further processed into soft data based on the equation (1):

$$AOD_{sof_{1,H}} \sim N(AOD_{Hour} + \mu, \sigma^2)$$  \hspace{1cm} (1)

where $\mu$ and $\sigma^2$ is the mean and variance of random error between the satellite AOD data and the corresponding AERONET AOD, respectively. $AOD_{Hour}$ is the AOD residual of satellite AOD or reconstructed hourly AOD data, and $AOD_{sof_{1,H}}$ is the Gaussian distribution probability soft data. After this, the relevant soft data of every estimation point is integrated by Bayesian maximum entropy on the basis of the spatiotemporal AOD residual model, expressed as equation (2).

$$x_{AOD} = \int x_{AOD} f(x_{AOD}) dx_{AOD}$$  \hspace{1cm} (2)

where $x_{Hour}$ and $x_{Day}$ represent the probabilistic Gaussian soft data derived from the Himawari-8/AHI AOD and the reconstructed hourly AOD, respectively. $f(x_{AOD}|x_{Hour}, x_{Day})$ is the posterior probability density function based on the spatiotemporal adjacent pixels, which can be expressed as equation (3) according to Bayesian rule (Christakos, 2002):

$$f(x_{AOD}|x_{Hour}, x_{Day}) = \frac{f(x_{Hour}, x_{Day}) f(x_{AOD}|x_{Hour}, x_{Day})}{\int f(x_{Hour}, x_{Day}) f(x_{AOD}|x_{Hour}, x_{Day}) dx_{Hour} dx_{Day}}$$  \hspace{1cm} (3)

where $x_{Hour}$, $x_{Day}$ is a vector of pixels $x_{Hour}, x_{Day} \rightarrow \ldots x_{Hour}, x_{Day}$ and $x_{AOD}$, $f_{sof}(x_{map})$ is the joint probability density function, which can be deduced by maximizing the entropy $E$ and introducing the constrain $G$ when variables are assumed to be continuous (Shannon, 1948; Jaynes, 1957). And $E$ is expressed as equation (4):

$$E = - \int dx_{map} f_{sof}(x_{map}) \log f_{sof}(x_{map})$$  \hspace{1cm} (4)

According to Lagrange multipliers $\lambda_i$, we can resolve the joint pdf $f_{sof}(x_{map})$:
\[ f_c(x_{\text{map}}) = \frac{1}{\int \exp \left( \sum_{a=1}^{n} \lambda_a G_a(x_{\text{map}}) \right) dx_{\text{map}}} \exp \left( \sum_{a=1}^{n} \lambda_a G_a(x_{\text{map}}) \right) \]

where \( G_a(x_{\text{map}}) \) is a vector of spatiotemporal covariance functions (Christakos, 2000). Therefore, the nonlinear mean estimation \( \mathbb{E}_{\text{AOD}} \) can be resolved by jointing the Equations (2) to (5). In the final output of the framework, \( \mathbb{E}_{\text{AOD}} \) will be merging with the spatiotemporal trend component of every pixel as the final AOD recovery values.

3. EXPERIMENT RESULT AND ANALYSIS

3.1 AOD Spatial Distribution and Temporal Variation

Figure 1 shows the AOD spatial distribution of the original satellite AOD data and the recovered AOD datasets in DOYs 43, 156, 212, 310, respectively. The selection principle of cases is that these four days are in different seasons, and the spatial absence of original satellite AOD in these days is distinct. The cases not only show that compared with the original AOD, the gaps of the recovered AOD datasets are significantly reduced, but also indicate that the recovered AOD datasets restore the real haze situation to a certain extent. To demonstrate the improvement in spatial coverage of AOD datasets after NBME, we quantitatively assessed the coverage by using the percentage of the valid AOD pixels over the whole study region.

![Figure 1](image1.png)

**Figure 1.** The AOD spatial distributions of (a) the original Himawari/AHI hourly AOD datasets and (b) AOD datasets after NBME in the study in DOYs 43, 156, 212, 310, respectively.

Figure 2 and figure 3 show the spatial distributions and temporal variation of the yearly averaged completeness of the original satellite AOD datasets and AOD datasets after NBME in 2016 (ranging from 03:00 to 05:00 UTC), respectively. Figure 2(a) shows that the yearly mean averaged coverage ratio of the original AOD datasets ranges from 0 to 44.4% (median: 21.4%), and Figure 2(b) shows that the yearly mean averaged coverage ratio of the recovered AOD datasets ranges from 21.6% to 89.6% (median: 59.7%), which both display the analogous spatial distribution characteristics. It indicates that the lowest coverage ratio always appears in inland of Western China, adjacent to the high-altitude areas like Yunnan–guizhou plateau. As shown in figure 3, we can conclude that the spatial coverage of AOD datasets after NBME is apparently higher than that of the original satellite AOD datasets, increasing from 20.5% to 70.1%. Moreover, over a quarter of days during the study period, the spatial completeness of the recovered AOD datasets is up to 80% and even 90% generally in March, whereas the relatively low completeness of the recovered AOD mainly occurs in winter, just a little over 40%. According to the figure 2(b), we can speculate that the lower coverage derives from the large-area snow/ice cover in high-altitude areas, especially in winter, where the missing AOD can’t be recovered when spatiotemporally adjacent AOD pixels lack.

![Figure 2](image2.png)

**Figure 2.** Spatial distributions of the yearly averaged completeness of (a) the original Himawari/AHI hourly AOD datasets and (b) AOD datasets after NBME

![Figure 3](image3.png)

**Figure 3.** Temporal variation of the yearly averaged completeness of the original Himawari/AHI hourly AOD datasets (green line) and AOD datasets after NBME (red line)

3.2 Assessment of the Accuracy of Recovered AOD

In order to validate the quality and accuracy of the recovered AODs, we collocate the original satellite AOD and the recovered AOD with AERONET AOD measurements. To ensure the comparability of matchups, we average AERONET AOD values within 3×3 pixels centered on the pixel where the AERONET site is located and within ±15 min of each satellite observation time. As shown as the following three equations (6), (7) and (8), the principals of correlation coefficient (R\(^2\)) of the linear regression, root mean square error (RMSE) and expected error (EE) are chosen to represent the validation results, where \( \text{AOD}_{\text{target}} \) is the original satellite AOD or AOD after NBME, and \( \text{AOD}_{\text{AERONET}} \) is the corresponding AERONET AOD data.
\[
R^2 = \frac{\left( \sum_i (\text{AOD}_{\text{target}} - \text{AOD}_{\text{AERONET}})(\text{AOD}_{\text{AERONET}} - \text{AOD}_{\text{original}}) \right)^2}{\left( \sum_i (\text{AOD}_{\text{target}} - \text{AOD}_{\text{AERONET}})^2 \right) \left( \sum_i (\text{AOD}_{\text{AERONET}} - \text{AOD}_{\text{original}})^2 \right)}
\]

(6)

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_i (\text{AOD}_{\text{target}} - \text{AOD}_{\text{AERONET}})^2}
\]

(7)

\[
\text{EE} = (1 \pm 0.05) \times \text{AOD}_{\text{AERONET}} \pm 0.05
\]

(8)

The overall validation results \((R^2: 0.62, \text{RMSE: 0.19})\) of AOD after NBME, which display a similar accuracy with those \((R^2: 0.62, \text{RMSE: 0.20})\) of the original AOD datasets. Moreover, nearly 61% AOD matchups fall within EE envelopes, increasing by 6% compared with the original hourly AOD datasets. As shown as figure 4(d), correlation coefficient \((R^2)\) is up to 0.60, and RMSE lays around 0.2, and approximately 60% of the matchups fall into the EE envelope, which prove that the recovered AODs performs well. By comparing the validation results of figure 4(a)-4(b) and 4(d), it indicates that the recovered AODs where the original AODs are missing are consistent with the AODs where the original AOD are available. Therefore, it could be concluded that the NBME recovery process by integrating the contribution of satellite-derived hourly and daily AOD products is practicable to achieve valid AOD products in not only completeness but also accuracy.

In addition, in order to obtain an adequate evaluation for the proposed recovery algorithm, the further simulated experiments are also implemented. The process of the simulated tests is as follows: at first, to obtain the target data, in the study area, suitable size area where adequately valid pixels exist of original data is randomly removed by masking; secondly, the target data serves as input data into the recovery process; finally, the recovery results of the artificially masked areas would be compared with the actual data. In our study, these simulated experiments are performed to validate the proposed recovery method by comparing with three other effective recovery methods: maximum likelihood method (MLE), universal kriging (UK), and geographically weighted regression (GWR). As shown as figure 5, two rows of panels show the AOD distributions of the groups of simulated experiments dated 27 April 2016 and 26 November 2016, respectively. Referring to the actual data, the results recovered by MLE are distorted where the valid original pixels are missing, and the results recovered by UK present an apparent spatial discontinuity effect for that only spatial autocorrelation is introduced. For the GWR algorithm, the results also exist partial spatial discontinuity and plenty of noise, thus it can be clearly seen that the results for NBME method are the closest to the actual data. From the above comparison, it indicates that the NBME recovery framework can obtain a more convincing product both in applicable accuracy and visual quality.
4. CONCLUSION

Current satellite-derived AOD products are mostly limited by spatial coverage in further atmospheric application, while the existing recovery algorithms mainly aim to the polar-orbit satellite-derived AOD products. In order to exert the function of geostationary satellite-derived AOD products in capturing aerosol dynamic changes, we propose the NBME recovery framework based on Bayesian maximum entropy methodology, and implement the framework in Himawari/AHI AOD datasets by integrating multi-time scale AOD products. Through the proposed recovery method, the spatial coverage of AOD dataset increases from 20.5% to 70.1%, and the overall accuracy of the recovered AOD datasets is basically consistent with the original datasets. Besides, through two groups of simulated experiments and by comparing with other popular recovery methods synthronically, we find that the proposed method performs better relatively, which indicates that the NBME recovery method can obtain high-accuracy products with applicable coverage and visual quality. Along with more joint observations of high temporal and spatial resolution satellites, the AOD sequence datasets recovered by NBME will provide more details of spatiotemporal variability of aerosol. In brief, this study not only yields possibilities for further application of high-resolution satellite AOD data in serving as a reliable reference for prevention and control of atmospheric pollution, but also strengthens the foundation for monitoring pollution in Southeast Asian region by remote sensing.

Acknowledgements

This study was supported financially by the National Natural Science Foundation of China (No. 42071353). We would like to express our gratitude to the JAXA for providing the Himawari/AHI products, and the Principle Investigators for establishing and maintaining the AERONET sites.

REFERENCES


