LGHAP v2: A global gap-free aerosol optical depth and $PM_{2.5}$ concentration dataset since 2000 derived via big ~~earth~~ Earth data analytics

Kaixu Bai[1,2], Ke Li[1], Liuqing Shao[1], Xinran Li[1], Chaoshun Liu[1], Zhengqiang Li[3], Mingliang Ma[4], Di Han[1], Yibing Sun[1], Zhe Zheng[1], Ruijie Li[1], Ni-Bin Chang[5], and Jianping Guo[6]

[1] Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University, Shanghai 200241, China
[2] Institute of Eco-Chongming, 20 Cuiniao Rd., Chongming, Shanghai 202162, China
[3] State Environmental Protection Key Laboratory of Satellite Remote Sensing, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China
[4] School of Surveying and Geo-Informatics, Shandong Jianzhu University, Jinan 2501
01, China
[5] Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, FL, United States of America~~USA~~
[6] State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, China

Correspondence to: Kaixu Bai (kxbai@geo.ecnu.edu.cn) and Jianping Guo (jpguocams@gmail.com)

Abstract. The Long-term Gap-free High-resolution Air Pollutants concentration dataset (LGHAP) generated in our previous study ~~provides~~ provides spatially contiguous daily aerosol optical depth (AOD) and fine particulate matter~~s~~ ($PM_{2.5}$~~PMs~~) concentration~~ data~~ s at a 1-km grid resolution in China since 2000. This advancement empowered ~~some~~ unprecedented assessments of regional aerosol variations and ~~its~~ ~~their~~its influence~~impacts~~ on the ~~the~~ environment, health, and climate over~~in~~ the past ~~few~~ twenty years~~years~~. However, there is a need to ~~improve~~ enhance such a ~~MODIS-like gap-free~~ high ~~resolution~~ quality AOD and $PM_{2.5}$ concentration dataset with new robust features and extended spatial coverage. In this study, we present ~~the~~ version 2 of ~~such~~ a global-scale LGHAP dataset (LGHAP v2), which~~that~~ was generated using an improved big ~~earth~~Earth data analytics ~~approach~~ via a seamless integration of ~~distinct~~ versatile data science, pattern recognition, and ~~deep~~ machine learning methods. Specifically, ~~To better reconstruct the global AOD distribution from daily remotely sensed MODIS AOD imageries,~~ multimodal AODs and air quality measurements acquired from relevant satellites, ground monitoring stations, and numerical models ~~across the globe throughout the past two decades~~ were ~~firstly~~ harmonized by harnessing the capability of random forest-based data-driven models. ~~Then~~Subsequently, an improved tensor-flow-based AOD reconstruction algorithm was developed to weave the harmonized multi~~-~~source AOD~~s~~ products together for ~~gap-~~filling data gaps in Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD retrievals from Terra. The results of the ablation experiments demonstrated better performance of the improved tensor-flow-based gap~~-~~filling method ~~has a better performance~~ in terms of both convergence speed and data accuracy. Ground-based validation results indicated ~~a~~ good data accuracy of ~~the~~ this global gap-~~filled~~ free AOD dataset, with a ~~site-specific~~ correlation coefficient (R) ~~R~~ of 0.85 and~~-~~ root mean square error (RMSE)~~RMSE~~ of 0.14 compared to~~against~~ the worldwide AOD observations from AERONET, ~~which is better than~~outperform~~ed~~ing the purely reconstructed AODs (R = 0.83, RMSE = 0.15) ~~and~~ whereas slightly worse than ~~the~~raw ~~raw~~ ~~Multi-Angle Implementation of Atmospheric Correction (~~MAIAC~~)~~ AOD retrievals ~~from Terra~~ (R = 0.88, RMSE = 0.11). ~~Regarding~~For $PM_{2.5}$ concentration mapping, a~~A~~ novel deep~~-~~learning ~~model~~approach, termed as the ~~named as the~~ scene-aware ensemble learning graph attention network (SCAGAT), was ~~developed~~ hereby applied~~to enhance the estimation accuracy of global better predict PM2.5 concentrations across the globe~~through gap-free AOD data. ~~While~~~~By gaining a better~~ ~~By~~While ~~enhancing~~accounting for the ~~spatial~~ scene representativeness of data-driven models across regions, the SCAGAT algorithm performed ~~better~~ ~~superiorly in~~better during spatial extrapolation, largely reducing modeling biases over regions with limited

and/or even absent ~~even though the~~ in situ $PM_{2.5}$ concentration measurements~~ are limited or absent~~. The ~~sSite-specific~~ validation results indicated that the gap-free $PM_{2.5}$ concentration estimates exhibit higher prediction accuracies, with an R of 0.95 and an RMSE of 5.7 μg m^{-3}, compared to~~against the~~ $PM_{2.5}$ concentration measurements obtained from previous~~prior~~ly ~~held~~hold~~-~~out sites worldwide. Overall, ~~while~~ while leveraging state-of-the-art methods in data science and artificial intelligence, a ~~quality-enhanced~~quality –enhanced LGHAP v2 dataset was generated through big ~~E~~earth data analytics by cohesively weaving together multimodal AODs and air quality measurements from ~~different~~ diverse sources~~ together cohesively~~. The gap-free, high-resolution, and global coverage merits render the LGHAP v2 dataset an invaluable data~~-~~base to advance aerosol- and haze-related studies~~ and~~, as well as to trigger multidisciplinary applications for environmental management, health~~-~~-risk assessment, and climate change ~~analysis~~attribution. All gap-free AOD and $PM_{2.5}$ concentration grids in the LGHAP v2 dataset, as well as the data user guide and relevant visualization codes, are ~~shared online~~ publicly accessible at https://zenodo.org/communities/ecnu_lghap (Bai et al., 2023a).~~,~~ ~~with a data user guide and relevant visualization codes available at https://doi.org/10.5281/zenodo.10216396.~~

1

1. Introduction

Atmospheric aerosols, produced from either natural or anthropogenic emissions, have been proven to pose significant threats to human health, ambient environment, and climate (Up in the aerosol, 2022). The risks to public health from aerosol pollution are ~~clear~~evident, with about 4.2 million deaths per year attributable to the exposure of fine aerosol particles, as stated by the World Health Organization (WHO, 2022). With increased aerosol loading, aerosols can significantly impair atmospheric visibility because of~~due to~~ the hygroscopic effect, thereby reducing direct solar radiation on the Earth's surface (Liu et al., 2020; Wang and Yang, 2014; Wild et al., 2021; Yang et al., 2016). In addition to the evident influence~~impacts~~ on air quality (Li et al., 2017), atmospheric aerosols also have an important and complex influence on regional, and even global climate (Anon, 2022; Guo et al., 2016, 2019; Li et al., 2019; Yang et al., 2020; Zhao et al., 2020). Therefore, ~~an~~ accurate monitoring of the atmospheric aerosol loading is vital for improving our understanding of the human-driven ambient environment and exposure pathways in health~~-~~-risk assessment.

Aerosol optical depth (AOD), a measure of aerosols distributed within an air column from the Earth's surface to the top of the atmosphere, has been widely used as a key indicator of total atmospheric aerosol loading. ~~AOD observations from ground monitoring stations have long been recognized as the ground truth, and a few g~~Ground-based aerosol observing networks, ~~e.g.~~such as~~,~~ the internationally collaborated Aerosol Robotic Network (AERONET), China Aerosol Remote Sensing Network (CARSNET), and Sun~~-~~-Sky Radiometer Observation Network (SONET)~~,~~ ~~were had been established to provide global and/or regional aerosol measurements~~ have long served as the ground truth for AOD monitoring (Che et al., 2015; Giles et al., 2019; Li et al., 2018). However, the sparse distribution of ~~ground~~ aerosol monitoring stations ~~pose~~poses ~~as~~ significant challenge~~s~~ in~~to~~ gain~~ing~~ a ~~better~~ comprehensive understanding of the aerosol variations across the globe.

Satellite-based AOD ~~products~~ data ~~well~~ bridge this~~such a~~ gap by providing ~~spatially-resolved~~spatially resolved AOD retrievals with ~~a vast~~extensive spatial coverage. Over the past forty years, ~~A~~ a variety of space-borne instruments, e.g., Sea-V~~v~~iewing Wide Field-of-V~~v~~iew Sensor (SeaWiFS), Moderate Resolution Imaging Spectroradiometer (MODIS), Visible Infrared Imaging Radiometer Suite (VIIRS), and Polarization and Directionality of the Earth's Reflectances (POLDER), were~~had been~~ deployed onboard ~~different~~ various satellite platforms and launched into space ~~over the past forty years~~ (Wei et al., 2020). These versatile instruments provide ample AOD and aerosol property measurements, enabling ~~us~~ to map global AOD distribution with finer spatial resolutions~~ in a long run~~. Nonetheless, ~~satellite-based~~satellite-based AOD retrievals often suffer from excessive data gaps because of~~due to~~ extensive cloud cover~~s~~ and retrieval failures,~~,~~ significantly impairing the data application potential~~, of these spatially incomplete AOD imageries. Moreover, substantial data gaps in satellite-based~~

~~AOD products could as well as~~ and resulting in large uncertainties when assessing the influence of aerosol ~~impacts~~ on weather and climate.

A variety of gap-filling methods were developed and applied to reconstruct the missing values in the ~~satellite~~ remotely sensed satellite AOD images (Wei et al., 2020; Xiao et al., 2021). The simplest method is to fill in data gaps with valid observations from ~~other~~ alternative data sources, e.g., filling in data gaps in MODIS AOD images from Terra with AOD observations from Aqua (Bai et al., 2019; Sogacheva et al., 2020)~~,~~ or ~~simply to~~ fusing~~e~~ with AOD simulation outputs from numerical models (Xiao et al., 2021). Such a substitution method is straightforward and effective, particularly~~especially~~ in an era with big E~~e~~arth observation data. Nonetheless, cross-mission biases are always salient ~~among~~ between satellite-based retrievals, ~~stemmingacquired from different various platforms and/or instruments, are always salient~~ because of the~~due to~~ significant differences in ~~both~~ instrument~~s~~ properties and/or retrieval algorithms. Thus, b~~B~~ias correction is ~~thus~~ essential to reducing systematic biases (Bai et al., 2016b, 2016a)~~.,~~ , and ~~and d~~Distinct~~ifferent~~ methods~~,~~ such as linear regression and maximum likelihood estimation~~,~~ ~~were~~ are often applied ~~to account for~~address ~~cross-mission biases prior to~~before merging the ~~data~~for this purpose ~~merging~~ (Bai et al., 2016a, 2016b, 2019; Ma et al., 2016; Xu et al., 2015). More complex ~~data fusion~~ methods, like the Bayesian maximum entropy ~~(Tang et al., 2016; Wei et al., 2021b)~~, were also applied to fuse ~~the~~ AOD products even with ~~different~~ varying spatial resolutions (Tang et al., 2016; Wei et al., 2021b).

Another type of gap-filling method~~s~~ works, in ~~a~~ principle, to recover missing information via dominant pattern recognition and reconstruction over space and time, and the ~~data~~ Data ~~interpolating~~ INterpolating ~~empirical~~ Empirical ~~orthogonal~~ Orthogonal ~~functions~~ Functions (DINEOF) method is a representative one (Beckers and Rixen, 2003; Liu and Wang, 2019). Two similar methods were developed to fill~~-~~ ~~in~~ data gaps in the ground-measured $PM_{2.5}$ concentration time series and geostationary satellite-sensed AOD imag~~eri~~es (Bai et al., 2020; Li et al., 2022b). Similarly, Zhang et al. (2022) developed a spatiotemporal fitting algorithm to ~~gap-fill~~fill gaps in the daily MODIS AOD product~~, primarily~~ by predicting~~,~~ ~~with~~ AOD values ~~mainly predicted~~ based on annual trends and spatial residues inferred from neighboring pixels. Nonetheless, filling data gaps ~~are hardly to be properly reconstructed simply based on~~with a single data source is always challenging, particularly~~especially~~ for those with ~~excessive~~ extensive missing values (e.g., satellite-based AOD). ~~Retrieving the missing AOD information~~ ~~Leveraging~~Learning missing values from ~~diverse~~external ~~from diversified external data products~~information, ~~via various artificial intelligence learning algorithms, in artificial intelligence,~~ such as~~e.g.,~~ numerical AOD simulations (Li et al., 2020; Xiao et al., 2017) and ~~even~~ meteorological factors (Bi et al., 2019), was proven ~~to be~~ an effective and feasible way to~~for~~ improve the~~ing~~ spatial coverage of reconstructed AOD fields.

~~Given the powerful approximation capacity, the mMachine- learning method is anothers have been widely applied used approach forto downscalinge and bias-correcting numerical AOD simulations to match satellite AOD footprints, while data gaps in satellite-based AOD imageries were then filled with downscaled data (He et al., 2023; Wei et al., 2021a);. Given the powerful approximation capacity, machine- learning methods were extensively used for bias correction in gap-filling problems over recent years (Bai et al., 2022b, 2023b; He et al., 2023; Wang et al., 2022; Wei et al., 2021a; Xiao et al., 2021). MLeveraging machine- learning and t~~Tensor~~- completion~~flow-based method~~s~~, ~~i.e.,~~ a more complex big data analytics framework, ~~were~~as~~was developed~~ ~~used~~developed to ~~integrate~~integrate six satellite-based AOD datasets~~ and~~, numerical aerosol diagnostics, ~~as well as~~and in situ air quality measurements ~~(Bai et al., 2022a).~~ , while a machine-learning method, i.e., random forest, was applied for downscaling and bias-correction purposes (Bai et al., 2022a). ~~Based on~~Harnessing multimodal data fusion and missing value reconstruction capabilities ~~this data analytics approach~~, a long-term gap-free high-resolution MODIS-like AOD ~~and PM concentration~~ dataset (LGHAP version 1), was successfully ~~yielded~~generated ~~over~~in China, with~~The comparable~~ an overall data accuracy comparable ~~of reconstructed AODs well demonstrate the efficacy of this~~to raw satellite retrievals, from which gap-free $PM_{2.5}$ and PM_{10} concentrations were mapped on a daily basis ~~gap-filling approach, yielding a~~

~~long-term gap-free high-resolution MODIS-like AOD and PM concentration dataset (LGHAP version 1) in China~~. Despite the good performance, ~~Despite the good reconstruction performance~~additional~~Recent, additional~~~~further~~ investigations have ~~recently~~ recently proven ~~that~~ the critical importance of prior information ~~is vital for~~~~in~~on tensor-flow-based gap-filling ~~procedure~~, particularly~~especially~~ over areas with substantial missing values~~, and the reconstruction results would be prone to significant large uncertainty with few valid observations in the input tensor~~ (Bai et al., 2022a; Li et al., 2022a, 2022b). Moreover, the strategies of maintaining an invariant background filed and assigning equal weights ~~for~~ to different AOD inputs may ~~not only reduce~~slow down the convergence speed and~~but~~ degrade the reconstruction accuracy.

In this study, we present ~~An~~~~Leveraging an improved big Earth data analytics approach has generated —~~, a new global scale LGHAP dataset, referred to as ~~termed as~~ LGHAP v2 hereafter, ~~hereafter—, was hereby generated to~~which ~~furnishes~~extends ~~provide~~ daily ~~global~~ gap-free AOD and $PM_{2.5}$ concentrations from China to worldwide at a 1-km grid resolution ~~as of~~~~dating back to~~for the period of 2000 to 2021. To~~In order to~~ accommodate massive global ~~massive~~ Earth observations acquired from diverse ~~satellites, numerical models, and air quality monitoring stations~~sources, an improved big Earth data analytics approach was developed by harnessing several new algorithmic improvements ~~were applied~~ to enhance the tensor-flow-based AOD gap filling ~~approach~~. ~~These improvements, including an attention-reinforced tensor construction strategy and, an adaptive background information updating scheme, an optimized global data tile partition and rank updating, all aimed at improving convergence speed and mitigating modeling bias propagation in numerical reconstructed AOD diagnosticsfields.~~ Moreover, a novel deep-learning method, namely, ~~named as~~ the SCene-Aware ensemble learning Graph ATtention network (SCAGAT), was ~~developed~~ applied to fulfill far-more accurate ~~global~~ $PM_{2.5}$ concentration mapping across the globe, particularly over regions with limited air quality monitoring stations. ~~While b~~Benefiting from the customized algorithmic improvements and the ~~novel~~ innovative SCAGAT $PM_{2.5}$ concentration mapping ~~method~~approach, the LGHAP v2 dataset has not only ~~has an not only~~ extended ~~the~~ spatial coverage from China to ~~worldwideglobal scale~~worldwide~~, global boasting and but also~~ but also improved data accuracy ~~compared to LGHAP v1~~. ~~To our knowledge, this is the first publicly accessible and global long-term gap-free MODIS-like AOD and $PM_{2.5}$ concentration dataset with a daily 1-km resolution, which could be used to help deepen our understanding of global aerosol pollution variations as well as adverse impacts on public health and on the, ecosystem, weather, and climate. In the following sections 2 and 3, we provide provided a more detailed comprehensive description of the diversified data sources analyzed in this study, as well as the versatile artificial intelligent machine- learning and deep- learning methods used to manipulate big Earth observational data. In the subsequent sections 4 and 5, The Performance of algorithmic improvements as well as, the data accuracy of the global gap-free AOD and $PM_{2.5}$ concentration data, and the application potential of the LGHAP v2 dataset data were then comprehensively evaluated. To our knowledge,~~As a ~~the LGHAP v2 is the first~~ publicly accessible and global long-term gap-free MODIS-like AOD and $PM_{2.5}$ concentration dataset~~.~~, the LGHAP v2 servers as a promising data source to improve our understanding ~~This resource stands to~~ of global aerosol pollution dynamics~~, shedding light on~~and its ~~their~~ adverse impacts on public health, ecosystems, weather~~ patterns~~, and climate~~ change~~. ~~by comparing it to against the worldwide in situ AOD and $PM_{2.5}$ concentration measurements.~~

2. Data Sources

~~In the current~~~~this study~~Similar as our previous study, here we ~~still attempt~~ aim to synergistically integrate the big Earth data acquired from diverse sources to generate a global long-term gap-free AOD dataset with a daily 1-km resolution~~. Subsequently,, from which,~~ from which ~~s~~spatially contiguous $PM_{2.5}$ concentration estimates can be then derived using~~by~~ a more robust and accurate data-driven approach ~~way to minimize the gaps and maximize the prediction accuracy~~. ~~As shown in~~ Table 1 ~~illustrates~~describes~~,~~ ~~a~~ the ~~large~~ array~~variety~~ of big Earth data ~~were hereby~~ employed in ~~data production~~ this study, including gridded AOD products from six polar orbiting satellites, ~~as well as~~ numerically simulated MERRA-2 ~~AOD and~~ aerosol diagnostics, ~~eleven~~ ten meteorological reanalysis fields, and ~~six~~ datasets of in situ AOD and air pollutants~~s~~ concentrations measurements. Additionally, auxiliary ~~variables~~ parameters representing land use and land cover types, elevation, ~~and~~ population density, ~~;~~ as well as ~~aand~~ vegetation ~~index~~covers, were ~~used not only to help~~~~incorporated~~also employed as critical explanatory variables to harmonize ~~the~~ discrepancies among multimodal heterogeneous aerosol datasets ~~prior to data integration~~. Note the spatial and temporal resolution as well as the time period for each data product are different from that of the benchmark dataset, namely, the MAIAC AOD product, and a data homogenization method is therefore essential to account for such discrepancies to reduce possible bias propagation in the subsequent data fusion procedure. ~~andbut also to aid in the global $PM_{2.5}$ concentration mapping.~~

Table 1. Summary of the diverse big Earth data used in this study to ~~help~~ generate ~~a~~ global gap-free AOD ~~dataset~~ and $PM_{2.5}$ concentrations at a daily~~/~~ and 1-km resolution (LGHAP v2) from 2000 to 2021.

Category	~~Dataset~~Product	Temporal Resolution	Spatial Resolution	Time Period
AOD	MCD19A2 (MAIAC)	daily	1 km	2000–2021
	Terra/MISR	daily	4.4 km	2000–2021
	NPP/VIIRS	daily	5 km	2012–2021
	Envisat/AATSR	daily	10 km	2000–2012
	PARASOL/POLDER	daily	10 km	2005–2013
	SeaWiFS/OrbView-2	daily	10 km	2000–2010
	AERONET	hourly	N/A	2000–2021
Meteorological factors	Air temperature	hourly	0.25°	2000–2021
	U/V component of wind	hourly		
	Relative humidity	hourly		
	Surface pressure	hourly		
	Boundary layer height	hourly		
	Total column water vapor	hourly		
	Surface solar radiation downwards	hourly		
	Total precipitation	hourly		
	Instantaneous moisture flux	hourly		
	Visibility	3-hour	N/A	2000–2021
Air quality measurements	$PM_{2.5}$, PM_{10}, NO_2, SO_2, CO	hourly	N/A	2000–2021
Population	WorldPop	annual	1 km	2000–2020
Land cover	Impervious (GISA)	annual	30 m	2000–2020
	MCD12Q1	annual	500 m	2000–2021
NDVI	MOD13A3	monthly	1 km	2000–2021
Aerosol diagnostics	MERRA-2	hourly	0.5° × 0.625°	2000–2021
Elevation	SRTM DEM	N/A	90 m	N/A

2.1. Satellite-~~B~~based AOD P~~p~~roducts

The AOD retrievals, derived from MODIS sensor ~~observations~~ on board Terra ~~(AOD$_{Terra}$) with~~ using the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm (denoted as AOD$_{Terra}$ afterwards), ~~were hereby used~~~~served~~were hereby used as the benchmark ~~to~~ for generating~~e~~ the global long-term gap-free AOD dataset, given their finer spatiotemporal resolution and longer temporal coverage (Lyapustin et al., 2011, 2018; Mhawish et al., 2019). Previous studies have demonstrated the ~~a better~~superior quality of ~~the~~ AOD$_{Terra}$ ~~MAIAC AOD data~~ relative to other gridded AOD products (Chen et al., 2021; Martins et al., 2017; Qin et al., 2021) ~~and~~ in regard to~~, not only~~ data accuracy and~~but also~~ spatiotemporal completeness, even better than those retrieved with the well-known Dark Target and Deep Blue algorithms (Jiang et al., 2023; Liu et al., 2019). Figure S1 presents the spatial and temporal distribution of the coverage ratio of valid AOD$_{Terra}$ from 2000 to 2021 at each satellite footprint across the globe.

Satellite-based AOD retrievals from a few key instruments other than MODIS were also applied to support gap filling of AOD$_{Terra}$ and~~.~~ t~~T~~hey include: (1) ~~Visible Infrared Imaging Radiometer Suite (~~VIIRS~~,~~ on board Suomi-NPP~~)~~, (2) Multi-A~~a~~ngle Imaging SpectroRadiometer (MISR, on board Terra), (3) Advanced Along-Track Scanning Radiometer (AATSR, on board Envisat), (4) ~~POLarization and Directionality of the Earth's Reflectance (~~POLDER~~,~~ on board PARASOL~~)~~, and (5) ~~Sea-Viewing Wide Field-of-View Sensor (~~SeaWIFS~~,~~ on board SeaStar~~)~~. Meanwhile, MAIAC AOD data from MODIS on board Aqua were also applied as ~~the~~ an important complementary data~~ set~~ source~~ to support gap filling of AOD$_{Terra}$~~. Given their varied ~~the different~~ overpassing times and temporal spans, these multisensory AOD dataset can~~products~~ provide complementary observations to help reduce random errors ~~when~~ during the AOD data reconstruction~~ofng data gaps in AOD$_{Terra}$~~procedure because of~~due to~~ the known ~~increased~~ prior knowledge. ~~A brief summary~~More details of these AOD ~~products~~ ~~datasets~~products can be found in Bai et al. (2022a) and Wei et al. (2020).

2.2. Ground-B~~b~~ased AOD O~~o~~bservations and A~~a~~ir Q~~q~~uality M~~m~~easurements

2.2.1. AERONET AOD O~~o~~bservations

Ground-based AOD observations from AERONET have long been used as the ground truth ~~to~~ for validating~~e~~ AOD retrievals from other instruments, particularly~~especially~~ diverse satellite-based AOD retrievals. In this study, AOD observations from AERONET ~~(across the globe)~~ during the study period were employed as an independent data source to validate the data accuracy of the global gap-filled AOD dataset. To guarantee an adequate number of AERONET AOD samples, the Level 1.5 ~~(instead of~~~~rather than Level 2.0)~~ AOD observations instead of Level 2.0 were applied, though the latter has stricter screening criteria for quality control. For spatial registration, each AERONET AOD observation was spatially collocated with mean AOD values over grids within a 5~~0~~ × 5~~0~~ km window size. Figure S2 presents the spatial distribution of the AERONET sites ~~and the air quality monitoring stations that provideing the pivotal AOD and PM$_{2.5}$ concentration observations~~ used in this study.

2.2.2. Air Q~~q~~uality M~~m~~easurements

Concentrations of PM$_{2.5}$ and other relevant air pollutants, like NO_2, SO_2, PM$_{10}$, and CO, were acquired from a few environmental agencies and~~/or~~ monitoring centers, such as the United States Environmental Protection Agency, European Air Quality Portal, China National Environmental Monitoring Centre, Canada National Air Pollution Surveillance, and Japan National Institute for Environmental Studies, to name a few. Moreover, air quality measurements acquired from the World's Air Pollution Index, an open-source data hub, were included as well. ~~To guarantee uniformity and comparability of these ground-based data, we conducted necessary~~Given potential differences in measuring principles and quality control criteria, ~~preprocessing to~~we performed rigorous data cleaning measures to ~~standardize~~harmonize these multisource air quality measurements, ~~as well as~~including not only the removal of outliers but also~~converted the time-series to~~ an unification of time scales to~~a~~ daily average~~scale synchronized with satellite observations taken on the same dates~~. ~~The $PM_{2.5}$ concentrations were used as the learning target for global $PM_{2.5}$ concentration mapping.~~ Aiming to~~at~~ provide~~ing~~ critical ~~prior~~ information to facilitate the AOD gap-filling, ~~the~~ ground-based air quality measurements were ~~also~~ used as an important proxy for regional in situ AOD prediction, ~~benefitting from~~ largely because of the relatively dense distribution of air quality monitoring networks ~~as well as~~and ~~exploited~~ the ~~good~~ associations between aerosol loadings and regional air pollutant~~s~~ concentrations.

Atmospheric visibility, a common air quality indicator ~~that is~~ highly associated with aerosol loadings, was~~were~~ acquired from worldwide meteorological monitoring stations and used ~~as the critical predictor~~— ~~like~~similar to air pollutants ~~concentrations~~—to predict AOD over each monitoring site via data-driven modeling. Given the much denser distribution of ambient air quality and meteorological monitoring sites, as shown in Figure S2, ~~for the spatial distribution of global air quality and meteorological monitoring sites used in this study, as well as the good accuracy of site-based AOD predictions (Bai et al., 2022b; Li et al., 2022b),~~ a global virtual AOD monitoring network was in turn established, harnessing the associations between AOD and air quality relevant parameters.~~;~~ ~~This~~Such a virtual network provides~~ing~~ us with an unparalleled opportunity to improve AOD gap-filling accuracy and efficiency, particularly~~especially~~ ~~for in~~over regions ~~being disturbed by~~with massive data voids in satellite AOD ~~data voids~~imageries (Bai et al., 2022b; Li et al., 2022b).

2.3. Numerical S~~s~~imulations

2.3.1. MERRA-2 A~~a~~erosol D~~d~~iagnostics

~~Despite the coarse spatial resolution and large modeling bias, t~~The Modern-Era Retrospective Analysis for Research and Applications~~,~~ version 2 (MERRA-2) aerosol diagnostics, including total AOD and ~~chemical~~ aerosol components like black carbon, organic carbon, dust, and sulfate aerosols, were employed to provide prior information to advance AOD gap-filling. As ~~the~~ NASA's latest reanalysis for the satellite era, MERRA-2 is generated using the new~~ly~~ Earth system model, ~~of~~ Goddard Earth Observing System~~,~~ version 5 (GEOS-5), providing global simulations of a variety of geophysical and chemical variables on the Earth's surface. More detail~~ed descriptions~~ of the assimilation system and the data quality of MERRA-2 aerosol reanalysis can be found in ~~the literature, such as~~ Buchard et al. (2017) and Randles et al. (2017). By taking AOD_{Terra} ~~into account~~ as ~~a~~ the learning target, data-driven models were established to spatially downscale and bias-correct MERRA-2 AOD ~~background~~ field ~~to the level of AOD_{Terra}~~, with ~~with MERRA-2 aerosol diagnostics as well as~~ meteorological, geographical, and socioeconomic factors ~~used~~ used as covariates. This ~~The downscaling model not only improves the spatial resolution~~ and~~but also corrects large modeling biases in MERRA-2 AOD. Given the global complete coverage merit,~~ downscaled and bias-corrected MERRA-2 AOD ~~background~~ field, given its spatially contiguous coverage, ~~the downscaled gap-free AOD data~~ w~~ere~~as then used as critical ~~prior~~ information to facilitate the ~~AOD~~ gap-filling of AOD_{Terra}.~~, in particularly over regions lacking observational AOD.~~

2.3.2. ERA-5 R~~r~~eanalysis

As the latest atmospheric reanalysis produced by the European Center for Medium Weather Forecast, ERA-5 provides hourly estimates of a variety of atmospheric, terrestrial, oceanic, climatic, and meteorological variables. The data are provided for a ~~at about~~ 30 km grid resolution on the Earth's surface, ~~resolving~~delineating the atmosphere layer using 137 levels from the surface up to a height of 80 km, covering the period from January 1940 to the present (Hersbach et al., 2020). Atmospheric parameters, including surface pressure, air temperature, relative humidity, wind speed, total column water, total precipitation, surface solar radiation downward, instantaneous moisture flux, and boundary layer height, were ~~retrieved~~ acquired from ERA-5 and used as important modeling covariates~~, not only~~ in both data harmonization ~~models~~ and ~~to calibrate other AOD and~~

~~relevant data products to the level of AOD~Terra~, but also and, in global~~ $PM_{2.5}$ mapping models~~., to help approximate nonlinear associations between $PM_{2.5}$ and AOD.~~ A simple b~~B~~ilinear interpolation was applied to the ~~map~~ ERA-5 reanalysis data ~~down~~ to convert them to the AOD_{Terra} footprint resolution for spatial registration.

2.4. Auxiliary D~~d~~ata

Several socioeconomic and geographic factors were also applied as covariates to support ~~predictions of~~ AOD gap filling and $PM_{2.5}$ concentration ~~predictions~~mapping. Specifically, g~~G~~ridded population data from WorldPop were used to indicate the spatial distribution of residents, ~~which were applied~~serving as a critical proxy ~~of~~for anthropogenic ~~aerosol pollution~~air pollutants emission intensity. To ~~resolve~~ characterize the land~~-~~-use~~-~~-dependent aerosol emissions, land cover types and the vegetation index derived from MODIS~~-~~ ~~retrieval observations~~ products, along with~~as well as~~ the coverage ratio of ~~the~~ impervious surface calculated ~~at the AOD~Terra~ footprint~~from the land use dataset generated by Huang et al. (2022),~~-~~were also applied. The d~~D~~igital elevation data collected from the Shuttle Radar Topography Mission (SRTM) with a resolution of 1 arc-second were used to characterize the potential impact~~s~~ of topography on aerosol loadings.

3. Methods

3.1. Tensor-F~~f~~low-based AOD R~~r~~econstruction

3.1.1. Overview of AOD G~~g~~ap-F~~f~~illing M~~m~~ethod

Deriving spatially contiguous $PM_{2.5}$ concentrations from gap-filled AOD images has ~~been~~ proven more promising for a better ~~spatial~~ analysis of large-scale $PM_{2.5}$ distribution (Bai et al., 2022b). In this study, the big E~~e~~arth data analytics framework proposed in Bai et al. (2022a) was further adapted and improved for generating global gap-free AOD imag~~eri~~es to support various content-based mapping. As shown in Figure 1, ~~presents the workflow of~~ the improved big Earth data analytics framework ~~of the big Eearth data analytics for generating global gap-filled MODIS-like AOD maps. This framework~~ also consists of three primary data manipulation procedures, including: 1) machine~~-~~-learned multimodal data homogenization, 2) knowledge-reinforced AOD tensor compiling, and 3) tensor-flow-based AOD reconstruction, with algorithmic improvements primarily conducted in the latter two procedures. This improved big E~~e~~arth data analytics approach empowered us to weave together multimodal AODs and versatile big E~~e~~arth observations from ~~diversified~~ diverse sources, ~~together neatly~~ via a synergy of state-of-the-art machine- learning and tensor completion methods. Because ~~Since~~ the technical flow of this big E~~e~~arth data analytics framework was ~~elaborated~~previously detailed ~~on~~ ~~well elaborated~~ in Bai et al. (2022b), we hereby only provided an overview of this method while ~~emphasizing the newly~~describing more details of the newly developed algorithmic components in the following subsections.

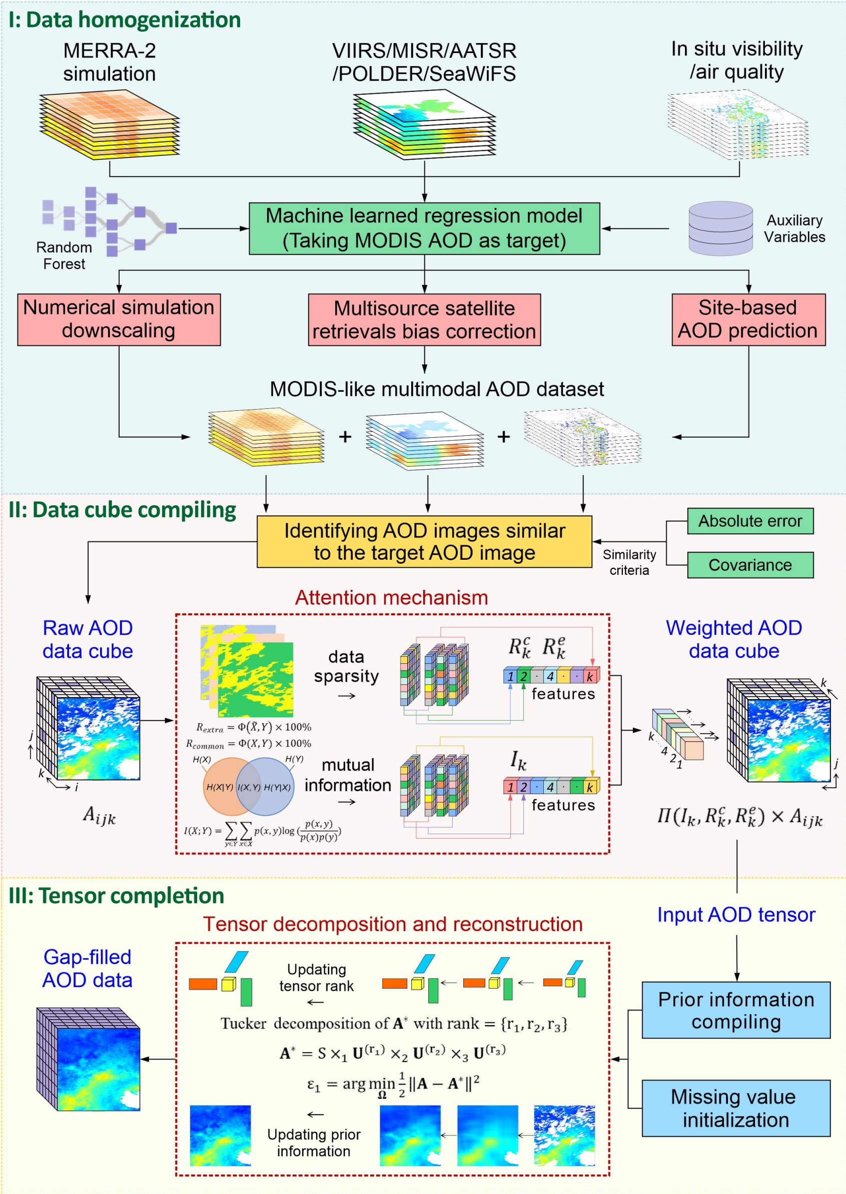

Figure 1. A schematic illustration of the ~~enhanced~~ improved big ~~earth~~Earth data analytics for generating the MODIS-like global gap-free AOD dataset.

The overall architecture of this big Earth data analytics framework was summarized as follows. M~~RLeveraging random forest-based regression models, m~~ultimodal AODs~~,~~ and relevant aerosol data acquired from different satellites, ground monitoring stations, and numerical models were first~~ly~~ harmonized to resemble the baseline dataset of ~~AOD from Terra retrievals (~~AOD$_{\mathrm{Terra}}$~~)Terra~~, aiming to~~at not only~~ minimize~~ing~~ both cross-sensor biases ~~arising from algorithmic differences~~ and spatial heterogeneities ~~and tobut also accounting for spatial heterogeneities because ofdue to different spatial resolutions~~. ~~This~~ Th~~e~~is data homogenization process is vital for the tensor-flow-based AOD gap-filling, because the bias-corrected and downscaled AOD estimates were critical inputs to form the AOD data cube. More details related to the multisource data homogenization were described in Text S1 in the supporting information. ~~The AOD data cube was then created based on homogenized data at each individual data tile. A proper AOD data cube compiling is undoubtedly essential for the tensor-flow-based AOD reconstruction.~~ To fill data gaps in each individual AOD$_{\mathrm{Terra}}$ image, an AOD data cube was then constructed~~, in our previous gap-filling framework,~~ by ~~simply~~ aggregating harmonized multisensory AOD data on the same date, along with historical AOD$_{\mathrm{Terra}}$ images resembling similar spatial patterns over the same region. Because of the~~Due to~~ excessive nonrandom missing values in the AOD$_{\mathrm{Terra}}$ imager~~ie~~s, both the downscaled MERRA-2 AOD grids and AOD estimates derived from air quality and visibility measurements were used conjunctively to identify ~~the historical~~ similar AOD_{Terra} imag~~eri~~es ~~with a similar spatial distribution~~from the historical image series. The selected historical AOD_{Terra} images and ~~bias-corrected~~bias corrected AOD images from other satellites on the same date were used individually ~~incorporated~~ as a slice of the tensor. Additionally, dispersed in situ AOD estimates and 5% of ~~the~~ randomly selected ~~AOD estimates from the~~ downscaled MERRA-2 AOD data were directly overlaid onto the corresponding AOD_{Terra} grids ~~where~~ without valid AOD retrievals ~~were not present~~absent. These implementations ~~not only~~ helped improve the gap-filling accuracy and~~but also~~ greatly boosted the convergence speed given the provision of prior knowledge.

High order singular value decomposition (HOSVD), an orthogonal Tucker decomposition method, was ~~finally~~ applied to each well-compiled AOD data cube for tensor-flow-based pattern recognition and ~~tensor~~ data completion. Data gaps within the input AOD tensor were first~~ly~~ filled with the spatial average of each individual AOD image to ~~initia~~initialize~~te~~ the tensor decomposition. The AOD tensor was then decomposed along ~~every~~ each two-dimension ~~of AOD tensor~~slice independently, and a new tensor was subsequently reconstructed based on the principal modes ~~learned along every each two-dimension slice of the tensor~~ via a low-rank approximation (i.e., generating an approximating matrix with reduced rank for compression). During this~~e tensor reconstruction~~ ~~process~~procedure, the AOD_{Terra} observations in the target image to be gap-filled were deemed ~~as the~~ hard data (i.e., true state and invariant throughout the tensor completion procedure) while multisensory AOD estimates and historical AOD_{Terra} images ~~were used~~served as ~~the~~ soft data (~~prior~~ supporting information and updated by iterates till convergence). By iteratively adjusting the dimension-varied ranks, the data values over grids to be gap-filled were updated and tuned to optimize both spatial homogeneity and information entropy concurrently (Bai et al., 2020, 2022a). ~~This~~ The tensor completion process continued till it reached~~ing~~ an ~~good~~ agreement (with a bias decay ratio < 0.1%) between the reconstructed values and the previously~~priorly~~ reserved AOD_{Terra} observations.

3.1.2. Algorithmic I~~i~~mprovements

To accommodate the massive data analytics for global-scale AOD gap-filling, ~~two~~ three major algorithmic enhancement modules were incorporated to help improve ~~the~~ reconstruction efficiency and accuracy, ~~focusing~~ with particular focus on the optimization~~ing~~ of data manipulation procedures in tensor-flow-based AOD gap~~-~~filling. ~~Instead of Rather than treating each slice of data in the raw AOD data cube equally, an attention mechanism was introduced to optimize the AOD tensor compiling, aiming at underscoring the importance of those AOD imageries with fewer data gaps while more closely resembling the target AOD~~~~Terra~~ ~~imagery during tensor-flow-based AOD reconstruction. Meanwhile, an adaptive prior information updating scheme was implemented to help mitigate the propagation of large modelingmodelling biases in numerical AOD diagnostics to the final reconstructed fields during the tensor reconstruction procedure. Moreover, the rank updating strategy was optimized to improve the computing efficiency in tensor completion.~~ A~~The a~~lgorithm 1~~below~~ presents the pseudo code of the optimized algorithm used for tensor-flow-based AOD reconstruction.

Algorithm 1. The pseudo code of the optimized algorithm used for tensor-flow-based AOD reconstruction.

Input: tensor $\mathbf{A} \in \mathbf{R}^{N_1 \times N_2 \times N_3}$ with $\boldsymbol{\Omega} = \{(i, j, k): A_{ijk} \text{ is observed}\}$, threshold T_1, T_2

Output: reconstructed entries $\mathbf{A}' = \mathbf{A}^*(:,:,k^t) \in \mathbf{R}^{N_1 \times N_2}$

Attention mechanism: $\omega_k = \Pi(MI_k, R_k^c, R_k^e)$

Initialize $A_{ijk}^* = \begin{cases} \omega_k \cdot A_{ijk} & (i,j,k) \in \boldsymbol{\Omega} \\ \sum_i \sum_j A_{ijk} & (i,j,k) \notin \boldsymbol{\Omega} \end{cases}$

for $r_3 = \frac{1}{3} N_3$ to 1 step –2 **do**

  $n_1 = n_2 = 0$

  while $\varepsilon_1 > T_1$ or $(n_1 < \frac{1}{3} N_1$ and $n_2 < \frac{1}{3} N_2)$ **do**

    $n_1 = n_1 + 1, n_2 = n_2 + 1$

    $r_1 = \frac{n_1 N_1}{75}, r_2 = \frac{n_2 N_2}{75}$

    $\mathbf{A}^* = \mathrm{HOSVD}(\mathbf{A}^*, \mathrm{rank} = \{r_1, r_2, r_3\})$:

    $\mathbf{A}^* = S \times_1 \mathbf{U}^{(r_1)} \times_2 \mathbf{U}^{(r_2)} \times_3 \mathbf{U}^{(r_3)}$

    $\varepsilon_1 = \arg\min_{\Omega} \frac{1}{2} \|\mathbf{A} - \mathbf{A}^*\|^2$

    $\mathbf{A}_{\Omega}^* = \mathbf{A}_{\Omega}$

    $\mathbf{A}_{\widetilde{\Omega}}^* = \omega_1 \mathbf{A}_{\widetilde{\Omega}}^* + \omega_2 \mathbf{A}_{\widetilde{\Omega}}$, $\widetilde{\boldsymbol{\Omega}}$ denotes background location

  end while

  if $\arg\min_{\Omega} \frac{1}{2} \|\mathbf{A} - \mathbf{A}^*\|^2 < T_2$ **then**

  break;

  end if

end for

3.1.2.1. Attention-R~~r~~einforced AOD T~~t~~ensor C~~c~~onstruction

In our previous study~~tensor completion framework as shown in Bai et al. (2022a)~~, b~~B~~oth the target data (i.e., AODTerra image ~~to be gap-filled~~) ~~as well as~~and soft data (i.e., AOD estimates from other data sources and historical AODTerra image~~rie~~s) ~~in the AOD tensor~~ were treated equally in the AOD tensor ~~during~~throughout the tensor decomposition and reconstruction process (Bai et al., 2022a) ~~in our previous tensor completion framework as shown in Bai et al. (2022a)~~. This~~e~~~~Such an~~ indifferent data treatment strategy ~~not only~~ neglected the information abundance of soft data and ~~but also~~ ~~ignored~~ the spatial similarity ~~of spatial patterns~~ between the soft and target data, leading the reconstructed field more likely to resemble the dominant patterns learned from imag~~eri~~es with fewer data gaps, rather than those~~images~~ with ~~—~~, ~~instead of~~~~rather than images with higher similarities — to the target data~~spatial patterns similar to the target image~~data~~. To account for this drawback, an attention mechanism was ~~implemented~~ hereby introduced to ~~weigh~~ assign different weights to each data slice ~~of data~~ in the input AOD tensor, aiming to~~at~~ improv~~e~~~~ing~~ the AOD reconstruction performance by learning from spatiotemporal features embedded in more relevant data fields instead of~~rather than~~ all the available data~~-~~.

As a widely used technique in deep~~-~~learning ~~regimes~~, the attention mechanism is a mimic of cognitive attention allowing the model to focus on specific parts of the input data, achieved by assigning higher weights to more crucial elements in ensemble learning. Regarding the tensor-flow-based AOD reconstruction task, ~~the~~ data slices with a higher similarity to the target image and fewer data gaps are supposed to ~~should~~ play ~~a~~ more important roles ~~be accorded more significances~~ than ~~those~~ less similar ones with extensive data gaps during tensor completion~~.~~. Three statistical metrics, ~~i.e.,~~ including mutual information (Shannon, 1948), spatial coverage ratio of common observations (R_{common}) between ~~each~~ soft data and hard data, and spatial coverage ratio of extra observations beyond common observations in soft data (R_{extra}), were calculated to determine the overall weight that should be assigned to each ~~data~~ slice of data in the input AOD tensor. Specifically, mutual information was applied to ~~gauge~~characterize the mutual dependenc~~y~~e between the target image and each slice of soft data ~~reference AOD images.~~, while c~~C~~ommon spatial coverage ratio was used to ~~quantify~~indicate the ~~reliability~~data amount for ~~of~~ mutual information calculation, ~~while~~and extra spatial coverage ratio ~~indicates~~ -was employed to depict additional information content that can be provided by ~~reference AOD images~~soft data. Equations (1–3) provide~~Below gives~~ the formulas to calculate these three statistical metrics.

$$MI(X,Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{1}$$

$$R_{common} = \Phi(X,Y) \times 100\% \tag{2}$$

$$R_{extra} = \Phi(\tilde{X},Y) \times 100\% \tag{3}$$

Note that~~where~~ X and Y refer to common observations in soft and hard data, respectively. The $\tilde{X}$ denotes extra observations in soft data.~~,~~ $p(x,y)$ is the joint probability mass function of X and Y, ~~and~~while $p(x)$ and $p(y)$ are the marginal distribution mass function of X and Y~~,~~, respectively. Additionally, $\Phi(X,Y)$ is the spatial coverage ratio of the common observations, and $\Phi(\tilde{X},Y)$ is the spatial coverage ratio of extra observations in the soft data. By multiplying these three normalized weights to the corresponding soft data, an attention-reinforced AOD tensor was constructed in turn, which was then used as the input data cube for tensor completion.

3.1.2.2. Adaptive P~~p~~rior I~~i~~nformation U~~u~~pdating

To facilitate the AOD gap-filling over regions with ~~abundant~~ substantial data gaps, in our previous method, ~~the~~ 5% random samples from the downscaled MERRA-2 AOD image (AOD$_{M2}$ hereafter) on the same date were used as prior information and directly ~~placed~~ overlaid ~~directly~~ onto grids without observational AOD (i.e., AOD$_{Terra}$ and site-based AOD estimates from air quality and visibility measurements). Although this ~~empowered us~~enabled to improve~~d~~ the convergence speed during tensor completion, the spatial patterns of the reconstructed field over regions with excessive data gaps were more likely to resemble the distribution of AOD$_{M2}$ ~~given~~~~due to~~ because of ~~the~~ ~~fixed~~these –unchanged ~~5% background~~prior information~~ values in the target AOD image~~~~an equal weight of the soft and hard data~~. ~~In other words, sparse observational AODs derived from air quality measurements played a relatively weak role in tensor completion when confronteding with AOD$_{M2}$.~~ In this~~such a~~ context, large modeling biases in AOD$_{M2}$ might be introduced into the final ~~reconstructed~~ reconstruction fields.

In this study, we introduced an adaptive prior information updating scheme to ~~help~~ mitigate potential bias propagation ~~from AOD$_{M2}$~~problem. ~~Differing from the strategy used in our previous method, t~~The main principle is to force t~~T~~he AOD prior information in the input AOD tensor ~~was also forced~~ to update ~~by~~ iterati~~ons~~vely throughout the tensor completion process, rather than maintaining them as invariant ~~as~~ ~~AOD$_{Terra}$~~ observations ~~throughout the tensor completion process~~. Specifically, random AOD$_{M2}$ samples were only used to ~~initiate~~ initialize the tensor construction, while weighted averages of the~~se~~ prior information and the corresponding reconstructed values were then used as new prior information for the next iteration. Meanwhile, the weights assigned to the reconstructed fields were gradually increased by iteration till convergence. The ~~ultimate goal~~goal was to improve the contribution of ~~reconstructed~~ reconstruction fields learning from actual observations while reducing the influence of ~~AOD$_{M2}$~~ background field. ~~Additionally, T~~the~~The~~ ablation experiments ~~also~~ demonstrated ~~that such a scheme is~~the effectiveness of this scheme in ~~mitigating bias propagation from AOD$_{M2}$, largely~~ improving the reconstruction performance over regions with limited observational data.

3.1.2.3. Optimized G~~g~~lobal D~~d~~ata T~~t~~ile P~~p~~artition and R~~r~~ank U~~u~~pdating

~~Given t~~The high ~~spatial and temporal~~spatiotemporal resolution of AOD$_{Terra}$ imag~~eri~~es presents a~~,~~ ~~performing global-scale AOD gap-filling is~~great ~~thus~~ ~~challenging~~ challenge in performing global-scale AOD gap-filling because of the~~due to~~ huge computational burden~~s~~. To improve~~ the~~ computational efficiency and to make the computing workload manageable, the following algorithmic ~~improvements~~ adjustments were ~~applied~~implemented. First~~ly~~, the continental ~~global~~ AOD$_{Terra}$ data ~~over landmass were~~worldwide were divided into 480 data tiles, with AOD gap-filling performed over each ~~data~~ tile independently. ~~The size of a tile was determined empirically after performing~~Through a set of gap-filling trials with ~~different~~ varying tile sizes, ~~and~~ a nominal tile size ~~of a tile~~ covering 700 × 700 pixels, refer to Figure S3 for the spatial distribution of the optimized data tiles, ~~(could be different over coastal regions)~~ was finally applied to balance the computing workload and reconstruction ~~the learning~~ accuracy~~. Figure S3 presents the spatial distribution of the optimized data tiles used in this study for global AOD gap-filling~~. Moreover, a 50-pixel overlap on the boundary of each tile was enforced, and an inverse distance weighting scheme was ~~finally~~ applied to these overlapped pixels when mosaicking the gap-filled tiles, aiming to eliminate the boundary effects between tiles toward a smooth distribution of AOD across the globe.

~~An optimized rank updating strategy was also proposed to improve the learning efficiency. In the tensor completion process~~, Since the tensor's decomposition and reconstruction processes in the tensor completion are driven by iteratively ~~updating~~ updated tensor ranks~~.~~, a~~A~~n optimized rank updating strategy was ~~also~~hereby proposed to improve the learning efficiency. ~~To improve the computational efficiency of global AOD gap-filling, we developed an optimized strategy to update the ranks between iterations.~~ Specifically, the ranks were updated in an ascending order along with the first and second dimensions in the inner loops to enhance the spatial details of reconstructed AOD fields. In contrast, the ranks were updated in a descending fashion along ~~with~~ the third dimension in the outer loop to aggregate the target AOD$_{Terra}$ image with the soft data in a low-rank approximation manner. This new rank updating strategy not only helps better resolve spatial details of AOD but also accelerate the convergence speed of tensor completion.

3.2. Global PM$_{2.5}$ C~~c~~oncentration ~~modeling~~Modeling

The sparse and uneven distribution of ground-based air quality monitoring stations poses significant challenges to global PM$_{2.5}$ concentration mapping, particularly~~especially~~ over regions ~~of~~ with fewer PM$_{2.5}$ concentration measurements (e.g., Africa and S~~s~~outh America in Figure S2). ~~AdditionallySo~~Nonetheless,~~Also,~~ how to reinforce the ~~scene~~spatial representativeness of data-driven models ~~when~~ to improve the spatial ~~extrapolating~~ extrapolation accuracy ~~them over~~~~across extensive spatial domainsce~~ is still elusive. ~~As a novel idea~~In this study, a recently developed deep learning method, namely, the scene-aware ensemble learning graph attention network (SCAGAT), was hereby ~~developed and~~ applied to better estimate global PM$_{2.5}$ concentrations from gap-filled AOD imageries~~ by accounting for the spatial scene representativeness of each data-driven model~~. Instead of~~Rather than~~ establishing a single ~~global~~ PM$_{2.5}$ estimation model using all available ~~data pairs~~data samples collected from worldwide monitoring stations, site-specific PM$_{2.5}$ estimation models were first~~ly~~ developed using ~~a~~ random forest over each air quality monitoring station with ~~long-term~~adequate PM$_{2.5}$ concentration measurements.

For a given grid, raw PM$_{2.5}$ concentration estimates were ~~then~~ estimated from a set of independent site-specific PM$_{2.5}$ estimation models~~;~~, of which should resemble similar geographic scene features as the given grid cell~~, ,~~ under the assumption that the relationship between AOD and PM$_{2.5}$ is similar over regions with an analogue environmental background. Nine distinct factors covering ~~geodetic~~ geographic location, land cover types, climate zones, AOD levels, and population density were utilized to characterize the scene attributes of each grid cell. Subsequently, a graph attention network was used to aggregate ~~these~~ raw PM$_{2.5}$ concentration estimates derived from site-specific models to ~~better predict the~~produce an ensemble ~~PM$_{2.5}$ concentration~~ estimate over the target grid cell.~~,~~ In the graph network, ~~Wwith~~ weights assigned to the adjacency matrix were determined in reference to the differences between nine different scene features, and the node bias was given as the testing accuracy of each site-specific PM$_{2.5}$ prediction model. ~~Figure S4 presents depicts the workflow of the proposed SCAGAT model for global PM$_{2.5}$ concentration mapping.~~ This ~~novel~~ innovative ensemble learning method enables us to better predict

$PM_{2.5}$ concentrations across the globe, particularly~~especially~~ over regions with ~~few~~ limited or even no~~ne~~ in situ $PM_{2.5}$ concentration measurements. Figure S4 depicts the workflow of the proposed SCAGAT model, and a~~A~~dditional~~More~~ details ~~regardingof the SCAGAT model~~ were introduced in Text S2~~ as part of the supplementary information.~~. For more detailed descriptions of this method, please refer to Li et al. (2024).

4. Results

4.1. Efficacy A~~a~~ssessment of A~~a~~lgorithmic E~~e~~nhancement M~~m~~odules

Ablation experiments were first~~ly~~ conducted to evaluate the accuracy improvement potential of each newly developed algorithmic enhancement module~~e~~. Three case studies were simulated by masking actual AOD_{Terra} retrievals with randomly selected cloud masks on different dates, and the methods reinforced with different enhancement modules were then applied to reconstruct the previously~~priorly~~ holdout AOD values. For inter~~-~~comparison, the AOD gap-filling framework developed in Bai et al. (2022a) was used as the benchmark method. As shown in Figure 2, the AOD distributions reconstructed ~~were reconstructed with~~using ~~a~~ ~~methods~~methods embedding attention mechanism and adaptive background information updating modules have smaller bias levels compared to ~~than~~ the benchmark method, which in turn justify the efficacy of these two new algorithmic enhancement modules. Given an equal weight of each slice of data in the input AOD tensor, the reconstructed data fields from the benchmark method were prone to resembling a mean state determined largely by the principal mode of the input tensor. In this context, peak ~~and/or low~~ values in the target image might be underestimated (or overestimated for low values) ~~if~~ because of relatively few soft data resembling similar patterns in the input tensor (e.g., Figure 2c).

~~With the involvement of the~~By incorporating the attention mechanism, each slice of data in the raw AOD data cube was adaptively weighted~~ adaptively~~, with ~~larger~~ greater weights given to ~~data slices~~those ~~not only~~ having ~~larger~~ broader spatial coverage and ~~but also~~ closer~~with~~ similarities to the target AOD_{Terra} image. This strategy is vital to reducing contributions from irrelevant data, particularly~~especially~~ when ~~facing~~ encountering ~~with~~ im~~un~~balanced data samples ~~in~~ within the raw AOD data cube, i.e., more irrelevant data and fewer similar imag~~eri~~es. Moreover, the importance of the target image was maximized during the tensor completion procedure by ~~giving~~ assigning ~~it~~ a 100% weight. Compared to the benchmark method, ~~peak and/or low~~extreme values in raw AOD_{Terra} images were better reconstructed using ~~by~~ the method embedding the attention mechanism. For instance, in Figure 2b, the benchmark method apparently overestimated low AOD values in the north, ~~in Figure 2b were apparently overestimated by the benchmark method,~~ whereas such ~~an effect~~a discrepancy was largely mitigated using methods involving the attention mechanism.

In contrast to the benchmark method ~~by using~~ which used an invariant background throughout the tensor completion process, an adaptive background updating scheme was ~~thus applied~~ incorporated here to ~~not only~~ accelerate the convergence speed and~~but also~~ mitigate possible error propagation arising from numerical simulations to the final ~~reconstructed~~ reconstruction fields. Compared to the benchmark method, a~~A~~s illustrated in Figure S5, the ~~enhanced method, involving~~ adaptive background updating module~~, indicated~~ enabled to ~~superior detection and resolution of~~reduce the adverse impact of manually added outliers in raw background fields,~~ compared to the benchmark.~~ ~~compared to the benchmark, the manually added outliers in raw background fields were better detected and reconciled by the improved method owing to the involvement of the adaptive background updating module,~~ th~~us~~ereby avoiding large error propagation from background fields into the reconstructed AOD data. Although t~~T~~he better quality of the reconstructed fields derived from the improved methods ~~well~~ demonstrates the efficacy of these two newly developed algorithmic enhancement modules~~.~~, ~~Nevertheless, as seencompared in Figure 2c,~~ the benefits ~~of these two enhancement modules were~~could be largely cancelled when ~~dealing~~ confront~~ed~~ing with images ~~with~~ containing excessive data gaps (e.g., Figure 2c~~,~~). ~~showing only a marginal improvement in accuracy improvement~~

~~relative~~compared ~~to the benchmark method.~~ The inherent reason could be attributed~~able~~ to few observational data in the target image for reference to leverage the attention mechanism to pinpoint similar AOD images from the historical data series.

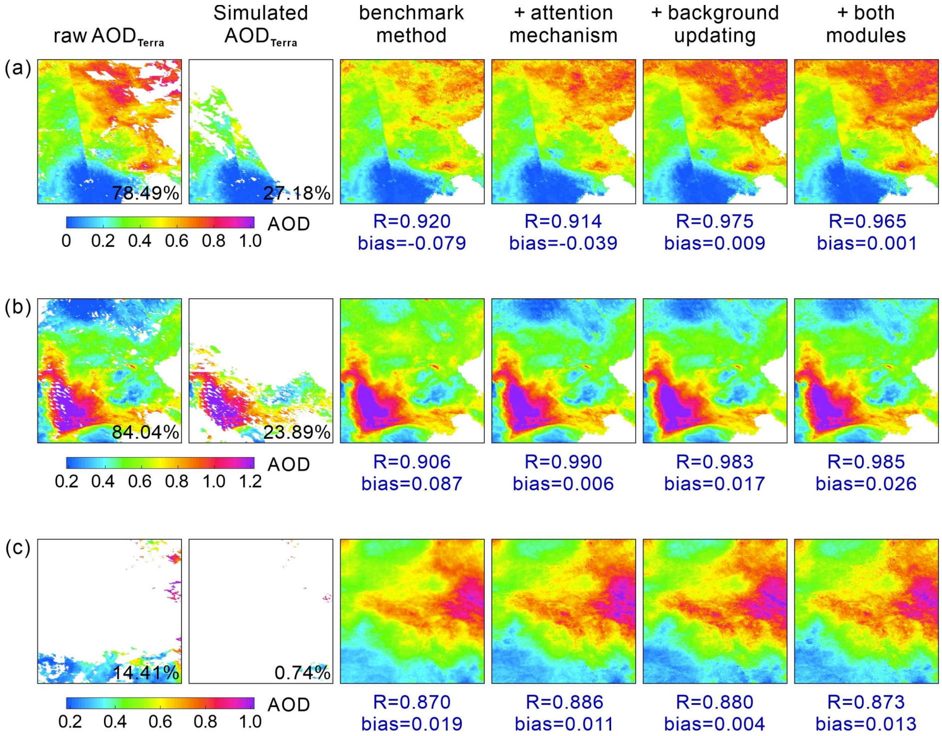

Figure 2. Performance evaluation of different algorithmic enhancement modules on the reconstructed AOD distribution. Raw AOD_{Terra} denotes the actual AOD retrievals from Terra, while simulated AOD_{Terra} refers to partially masked AOD_{Terra}. The benchmark method is the AOD gap-filling approach proposed in Bai et al. (2022a). The latter three columns present the reconstructed fields using the enhanced benchmark methods. The R and bias denote correlation coefficient and deviations between the ~~withheld~~holdout observed and reconstructed AOD data, respectively. The p~~P~~ercent numbers shown in the two left panels indicate a spatial coverage ratio of valid AOD retrievals over the selected scenes.

In Figure 3, we evaluated the impact~~s~~ of the missing rate of the target image ~~AOD_{Terra}~~ on the AOD gap-filling accuracy. By masking ~~raw~~ one truly observed AOD_{Terra} ~~retrievals~~ image with arbitrarily selected cloud masks, ~~the~~a series of ~~AOD_{Terra}~~ target images under different missing rates, as shown in the top panel of Figure 3, were ~~generated~~ simulated ~~and used as target images for~~for gap-filling trails~~(i.e., images as shown in the top panel of Figure 3)~~. ~~The results show~~As shown, ~~an~~the reconstructed fields fairly agreed with~~a strong good agreements between~~ the observed ~~and reconstructed~~ AOD fields, well resembling the actual AOD distribution over the outlined region, even in~~over~~ extreme situations with excessive data gaps, demonstrating an excellent performance of the proposed gap-filling method. As expected, the ~~reconstruction~~ accuracy of the reconstruction fields decreased along with an increase in the missing rate. For instance, when the missing rate was greater than 80%, the low values in the upper left ~~in~~ of the raw AOD_{Terra} image were not properly reconstructed~~when the missing rate was greater than 80%~~, largely because of the limited prior knowledge in the target image for use when constructing the raw AOD tensor. This effect also ~~highlighting~~ highlights the ~~vital~~ crucial importance of prior information on the gap-filling accuracy. Therefore, increasing prior information is the most promising way to improve the gap-filling accuracy ~~in gap-filling,~~ in particular for ~~those areas~~regions with substantial data gaps.

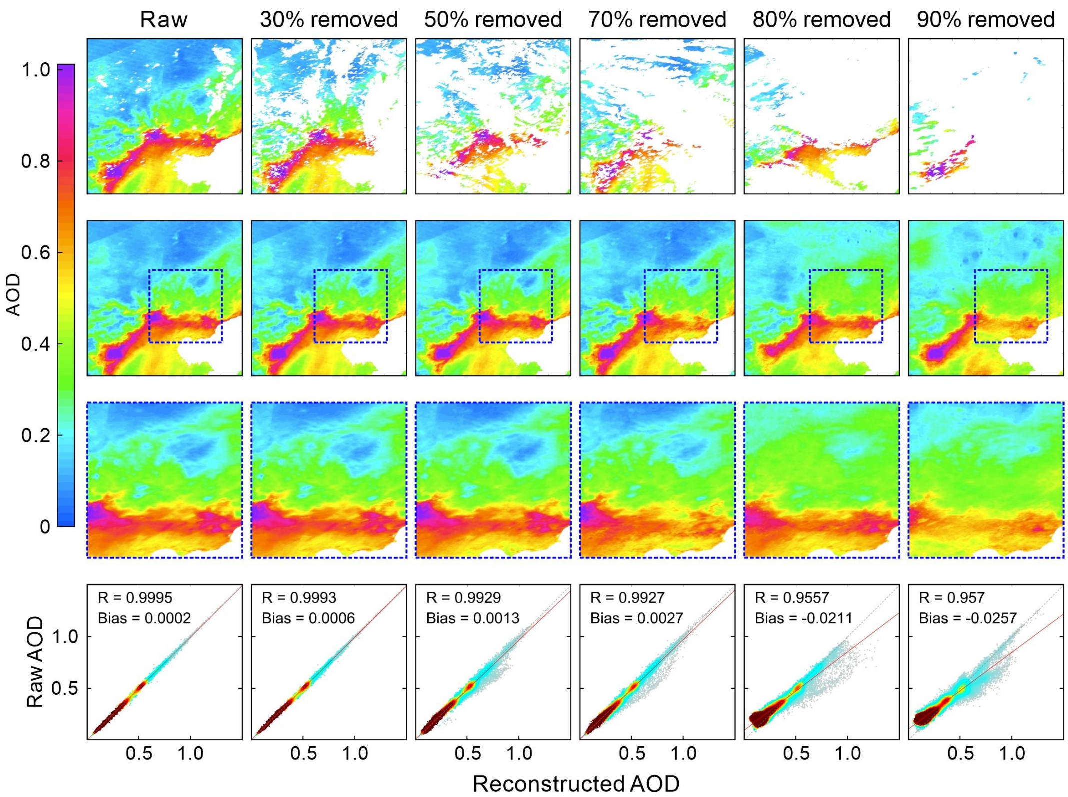

Figure 3. Impact~~s~~ of the missing rate on the AOD gap-filling accuracy. The n~~N~~umbers on the top indicate the percentage of removed AOD data in the raw AOD$_{Terra}$ image~~ (top panel)~~. The second row shows the distribution of the gap-filled AOD with zoomed~~-~~in maps present in the third row. The bottom panel ~~-~~presents scatter plots between the observed ~~AOD (withheld raw data)~~ and the reconstructed AOD ~~reconstructed from different inputs~~.

4.2. Data ~~a~~A~~a~~ccuracy of G~~g~~lobal G~~g~~ap-F~~f~~ree AOD in LGHAP v2

The gap-free AOD grids ~~dataset (~~in the LGHAP v2~~)~~ w~~as~~ere generated by filling in data gaps in AOD$_{Terra}$ images ~~the satellite-based~~ ~~MAIAC~~ ~~AOD retrievals (MCD19A2)~~ with reconstructed AOD estimates at each collocated footprint over land. In comparison~~By comparing~~ ~~against~~to the independent AOD observations from AERONET, the data accuracy of the gap-free AOD in the LGHAP v2 was comprehensively evaluated across the globe. Figures~~.~~ 4a–c present ~~a~~the spatial distribution of the site-specific correlation coefficient (R), root mean square error (RMSE), and bias between ~~the reconstructed~~ ~~reconstructed~~ AOD in the LGHAP v2 and ~~AOD~~ AERONET observations~~ from AERONET~~, respectively. Regardless of the uneven distribution of ground-based aerosol ~~monitoring~~ observing stations and ~~the difference~~ variations in data samples between sites, the ground validation results indicate a good agreement~~s~~ between the AOD in the LGHAP v2 and the AERONET observations, with ~~an average of~~ site-specific ~~correlation coefficient~~R of 0.76±0.14 and RMSE of 0.09±0.08 ~~at the~~on a global scale. ~~Meanwhile, the results indicate that~~Note site-specific data accuracy metrics ~~exhibit notable spatial heterogeneities~~ vary across ~~the globe~~regions, with larger biases mainly observed in the central and east Asia as well as in Africa—regions ~~frequently~~always~~, which where often~~ sufferings from high aerosol loadings.

Figures~~.~~ 4d–~~4~~i present scatter plots between the LGHAP v2~~gap-free~~ AOD and AERONET observations at six major continental regions. ~~The distinct accuracy metrics across regions also indicate significant spatial heterogeneities in the AOD data accuracy. When compared against~~to the ~~AOD observations from AERONET, t~~As shown, t~~T~~he reconstructed AOD estimates were prone to an underestimation of ~~underestimate~~ large AOD ~~observations~~ values (> 0.80) versus an~~ whereas~~ overestimation~~e~~ of low values (< 0.2) across these six regions. This~~Such an~~ effect is particularly common in ~~machine~~ machine-learning, largely because of~~due to~~ the imbalanced distribution of data values in the training samples (Johnson and~~&~~

Khoshgoftaar, 2019; Shi et al., 2022). ~~Likewise~~Similarly~~, the inherent~~ reason could be also applied fors ~~for~~ the ~~this effect in~~ tensor completion ~~might be identical, which could be largely attributable to~~ as ~~the principle of low-rank approximation to fulfilfulfillrequired for tensor reconstruction and the imbalanced (i.e., few extremes) AOD values (i.e., few extremes) in the input tensor. Consequently~~~~As a result,~~ the missed AOD extremes ~~could~~ may not be accurately~~were hardly to be~~ reconstructed to their nominal levels;~~.~~ i~~I~~nstead~~Rather~~, they tend ~~the reconstructed values were inclined~~ to resemble a mean state that was determined by principal modes via a low-rank approximation. ~~because of~~~~due to the imbalanced data distribution.~~

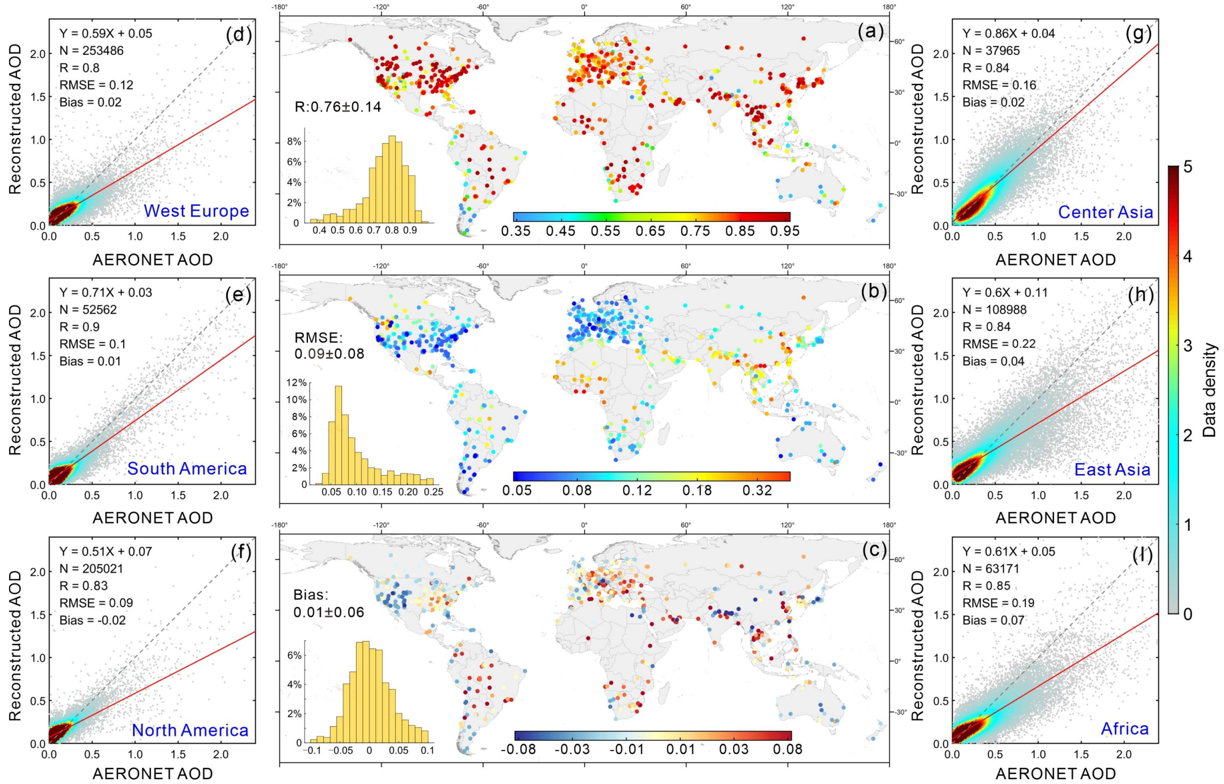

Figure 4. Data accuracy of daily gap-free AOD grids in the LGHAP v2 dataset compared~~by comparing~~ ~~against~~to ~~the~~ AOD observations from AERONET across the globe during 2000–2021. Note the AERONET AOD observations were independent data ~~from~~ and had been not used in the gap-filling process.

To further verify the data accuracy of the imputed AOD estimates, we ~~further~~ compared the ~~data accuracy of~~ gap-filled AODs in the LGHAP v2 dataset with two major gridded products~~, i.e.,~~ of ~~satellite-based~~ MAIAC ~~AOD retrievals from Terra~~ (AOD_{Terra}~~MCD19A2~~) and ~~downscaled MERRA-2 AOD (~~AOD_{M2}~~)~~. As shown in Table 2, the purely reconstructed AOD estimates have an~~a~~ R of 0.83 and an RMSE of 0.15 compared to~~against~~ the AERONET AOD observations at the global scale~~—~~, comparable to the data accuracy of AOD_{M2} (R = 0.83, RMSE = 0.14) but lower than that of AOD_{Terra} (R = 0.88, RMSE = 0.11). Nevertheless, the imputed AOD estimates achieved comparable data accuracies to~~as~~ AOD_{Terra} in Africa (R = 0.80, RMSE = 0.20) and Australia (R = 0.62, RMSE = 0.08), largely because of the availability of~~due to~~ abundant satellite-based AOD prior information~~retrievals over these two areas~~ (refer to the AOD coverage ratio shown in Figure S1) to facilitate AOD ~~gap-filling via~~ tensor completion. In contrast, the LGHAP v2~~imputed~~ AOD estimates in Europe and Asia have poorer data accuracies ~~with~~ relative to AOD_{Terra}, particularly~~especially~~ in Eastern Asia. The p~~P~~ossible reasons ~~for this~~ could be ascribed to ~~not only~~ extensive missing values, severe aerosol pollution levels, as well as significant spatial variations in aerosol loadings over these regions. Compared to AOD_{Terra}~~MAIAC AOD~~, the gap-filled AOD data tended to overestimate the AERONET AODs (17.59% versus 11.45% above the envelope of expected error), resulting in ~~the~~an even larger global mean AOD ~~values~~ (0.19~~) in the LGHAP v2 dataset than in the MAIAC AOD (~~ versus 0.17)~~.~~, implying a greater number of large AOD values were reconstructed in the imputed AOD estimates.

Moreover, the accuracy of ~~The gap-free AOD dataset (LGHAP v2) was generated by filling in data gaps in the satellite-based AOD retrievals (MCD19A2) with reconstructed AOD estimates at each collocated footprint over land. The gGround validation results indicate that the gap-filled AOD data in LGHAP v2 are in a good agreement with the AERONET AOD observations, with an R of 0.85 and an RMSE of 0.14 across the globe (Table 2), slightly worse than that of raw MCD19A2 (R = 0.88 and RMSE = 0.11) but higher than that of AOD_{M2} (R = 0.83 and RMSE = 0.14). This data~~LGHAP v2 AOD data ~~accuracy~~ outperforms that of the gap-filled AOD dataset (R^2 = 0.6031 and RMSE = 0.1350) generated by Guo et al. (2023), in which missing AODs in ~~MCD19A2~~ AOD_{Terra}~~MAIAC~~ were predicted ~~with~~ using ~~versatile~~ various proxy variables (e.g., meteorological factors and population density) via a random forest model.

Table 2. An intercomparison of AOD data accuracy between satellite-based retrievals (raw MAIAC AOD), numerical aerosol diagnostics (downscaled MERRA-2 AOD), purely reconstructed data, and the final gap-free product (LGHAP v2 AOD), by comparing AOD observations from AERONET across the globe during 2000−2021. Note the term "Purely Reconstructed AOD" refers to the imputed AOD estimates, while "LGHAP v2" refers to the gap-filled AOD dataset combining both satellite-based retrievals and purely reconstructed data. The expected error (EE) envelope for AOD over land was defined as ± (1.5 × $AOD_{AERONET}$ +~~±~~ 0.05).

AOD Dataset	Region	Mean AOD	Number of Monitors	Number of Samples	R	RMSE	Bias	Below EE (%)	Within EE (%)	Above EE (%)
MAIAC (AOD_{Terra})	Global	0.17	1,335	402,886	0.88	0.11	0.02	13.95	74.59	11.45
	North America	0.11	433	112,438	0.83	0.08	–0.01	4.62	80.93	14.44
	South America	0.11	81	28,265	0.94	0.07	0.02	14.17	75.85	9.97
	Europe	0.11	208	96,715	0.80	0.06	0.02	11.29	82.22	6.49
	Asia	0.31	321	90,821	0.90	0.14	0.02	18.79	68.22	12.99
	Africa	0.21	110	48,877	0.81	0.19	0.06	31.45	57.11	11.44
	Australia	0.09	28	12,427	0.62	0.07	–0.01	6.16	75.34	18.49
Downscaled MERRA-2 (AOD_{M2})	Global	0.18	1,335	811,438	0.83	0.14	0.02	11.76	78.98	9.26
	North America	0.12	433	216,264	0.80	0.09	0.00	5.71	86.22	8.07
	South America	0.13	81	49,721	0.90	0.11	0.02	12.87	81.64	5.49
	Europe	0.13	208	177,125	0.79	0.07	0.01	8.54	86.07	5.39
	Asia	0.29	321	175,781	0.78	0.24	0.06	22.54	65.14	12.32
	Africa	0.24	110	88,374	0.85	0.15	0.02	16.13	67.59	16.28
	Australia	0.10	28	21,051	0.76	0.06	–0.02	2.44	83.60	13.96
Purely Reconstructed AOD	Global	0.21	1,335	449,452	0.83	0.15	0.01	12.21	65.52	22.27
	North America	0.16	433	129,716	0.80	0.10	–0.02	5.23	67.52	27.25
	South America	0.17	81	30,073	0.88	0.11	0.00	10.51	67.11	22.38
	Europe	0.16	208	107,961	0.73	0.09	0.00	9.63	73.63	16.74
	Asia	0.33	321	107,876	0.81	0.24	0.03	18.64	56.60	24.76
	Africa	0.27	110	31,568	0.80	0.20	0.06	29.57	53.88	16.55
	Australia	0.13	28	9,628	0.62	0.08	–0.03	4.60	64.62	30.77
LGHAP v2	Global	0.19	1,335	756,166	0.85	0.14	0.01	12.96	69.44	17.59
	North America	0.13	433	216,055	0.82	0.09	–0.01	4.86	73.12	22.02
	South America	0.14	81	49,707	0.90	0.10	0.01	12.57	71.08	16.34
	Europe	0.13	208	176,959	0.76	0.08	0.01	10.24	77.40	12.36
	Asia	0.32	321	175,728	0.83	0.21	0.03	19.08	61.40	19.52
	Africa	0.23	110	75,110	0.81	0.19	0.06	29.61	56.64	13.75

Australia	0.11	28	21,048	0.63	0.08	–0.02	5.11	70.30	24.59

In Figure 5, we compared temporal variations in AOD between the LGHAP v2 dataset and ~~AERONET~~ ground-based observations at six AERONET ~~aerosol observing~~ sites with long-term ~~monitoring~~ records. Compared to discrete AOD observations from AERONET, the gap-free AOD time series accurately ~~well~~ reconstructed long-term variations of aerosol loading from 2000 to 2021 at these ~~six~~ monitoring sites, with R ranging from 0.83 to ~~–~~0.97 and RMSEs varying between 0.04 and 0.24. Note that t~~T~~he l~~L~~arge~~r~~ RMSEs observed at the Alta Floresta and Beijing sites are more likely ascribed to the reconstruction failures of ~~extreme~~ abnormal AOD peaks, largely because of very limited peak values for reference in the AOD tensor. Referring to histograms of AOD deviations between the LGHAP v2 and AERONET observations, more than 80% of AOD biases fell within the range of ~~were found to vary between~~ −0.1 ~~and~~ to 0.1, demonstrating a high accuracy of ~~the~~ gap-~~free~~ filled AOD in the LGHAP v2 dataset.

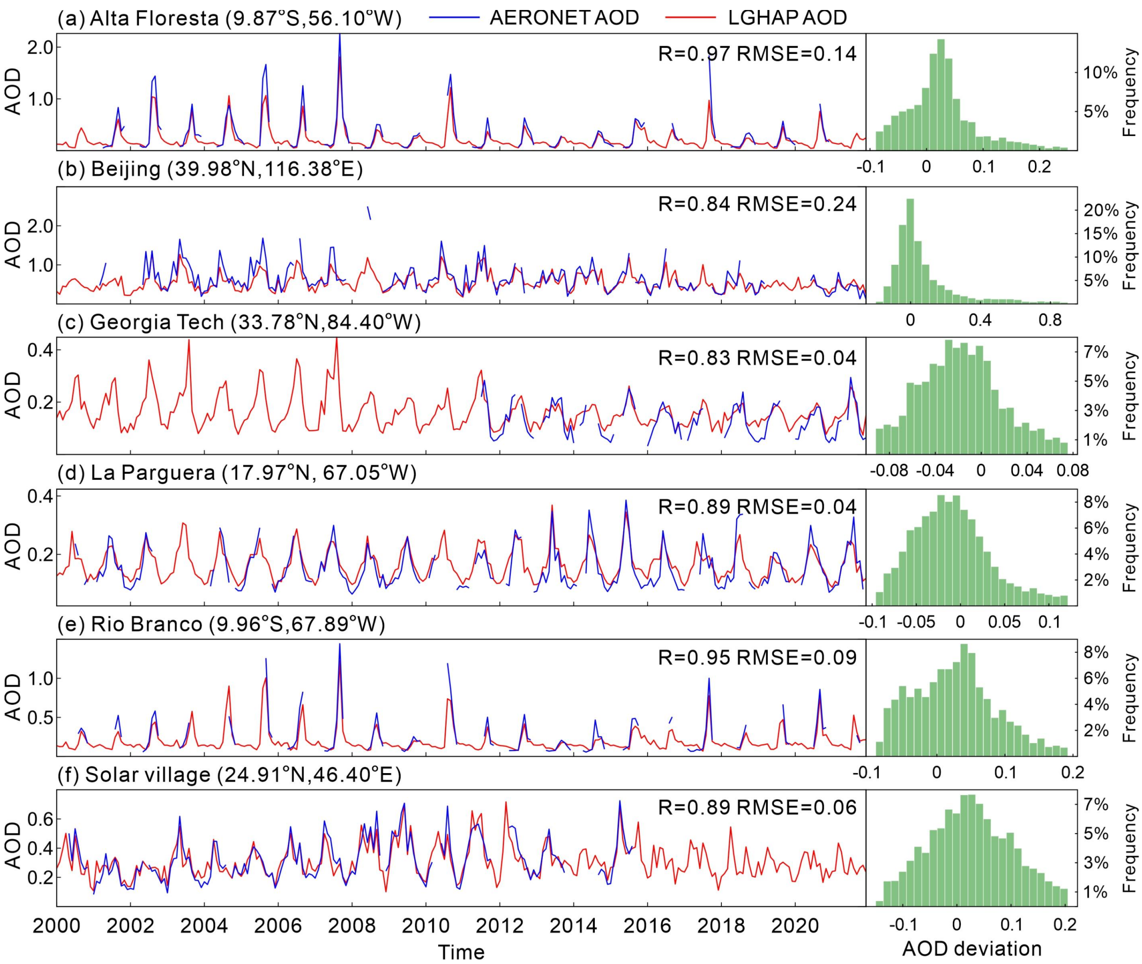

Figure 5. Temporal variations in the monthly AOD over six AERONET sites with long-term AOD observations ~~during~~ from 2000~~–~~ to 2021. The p~~P~~anels on the right present histograms of AOD deviations between the LGHAP v2 and AERONET observations at each individual site.

4.3. Data A~~a~~ccuracy of G~~g~~lobal G~~g~~ap-F~~f~~ree $PM_{2.5}$ C~~c~~oncentrations in LGHAP v2

Global gap-free $PM_{2.5}$ concentration estimates were ~~then~~ derived from gap-filled AOD images by taking advantage of the novel SCAGAT m~~odel~~ethod that was specifically developed ~~to fulfilfulfill the~~for global $PM_{2.5}$ concentration mapping. Additional~~More~~ details ~~related to the performance evaluation~~ of the SCAGAT ~~model~~ method were ~~described~~ provided in another ~~companion~~ study (Li et al., 2024), and here we ~~hereby~~ focused on the data accuracy of the global gap-free $PM_{2.5}$

concentration estimates. Figure 6 presents the validation accuracy of the daily gap-free $PM_{2.5}$ concentration estimates by comparing them ~~against~~to the ground-based $PM_{2.5}$ concentration records measured at 350 ~~independent (~~previously~~priorly~~ h~~e~~old-out~~)~~ ~~monitoring~~ sites. ~~The results~~As indicated, by accounting for spatial representativeness of the prediction models during the spatial extrapolation, ~~that~~ $PM_{2.5}$ concentration estimates derived from the SCAGAT model are in~~have~~ better ~~better~~ agreement~~s~~ with ground~~-measured~~ -based $PM_{2.5}$ concentration measurements~~s across the globe (~~, with an R ~~=~~ of 0.91 and an RMSE ~~=~~ of 9.5~~8~~7 μg m^{-3}~~)~~, ~~outperforming~~ surpassing the performance of our traditional ~~$PM_{2.5}$~~machine-learning~~ed~~ ~~prediction~~ models ~~without accounting for the spatial representativeness of the prediction models during the spatial extrapolation~~ (Bai et al., 2019, 2022a, 2023). Meanwhile, ~~As shown in Figure 6e, by taking advantage of the SCAGAT model, the $PM_{2.5}$ concentration estimates over China in LGHAP v2 have a higher data accuracy (R = 0.97, RMSE = 7.93 μg m^{-3}) than those in LGHAP v1 (R = 0.95, RMSE = 12.03 μg m^{-3}), neglecting a different number of validation samples.~~ ~~t~~The data accuracy was further improved by correcting mode~~l~~ling biases using sparsely distributed in-situ $PM_{2.5}$ concentration measurements via optimal interpolation, ~~where~~resulting in an improvement in~~with~~ R ~~improved~~ to 0.95 and a ~~reduction~~decrease in RMSE ~~was reduced down~~ to 5.7 μg m^{-3} (~~as shown in~~ Figure 6b). As shown in Figure 6e, ~~by leveraging the SCAGAT model,~~ the $PM_{2.5}$ concentration estimates over China in the LGHAP v2 have a higher data accuracy (R = 0.97, RMSE = 7.93 μg m^{-3}) than those in LGHAP v1 (R = 0.95, RMSE = 12.03 μg m^{-3})~~,~~. Figures~~s.~~ 6c–~~6~~d present a site-based distribution of R and RMSE for the LGHAP v2 $PM_{2.5}$ concentrations over each individual validation site. Compared to the United States of America and Europe, as ~~shown~~ depicted in Figures~~.~~ 6e–~~6~~g, larger $PM_{2.5}$ concentration biases were ~~more likely to be~~ observed in ~~Asia~~ China because of ~~due to the~~given higher $PM_{2.5}$ loadings therein.

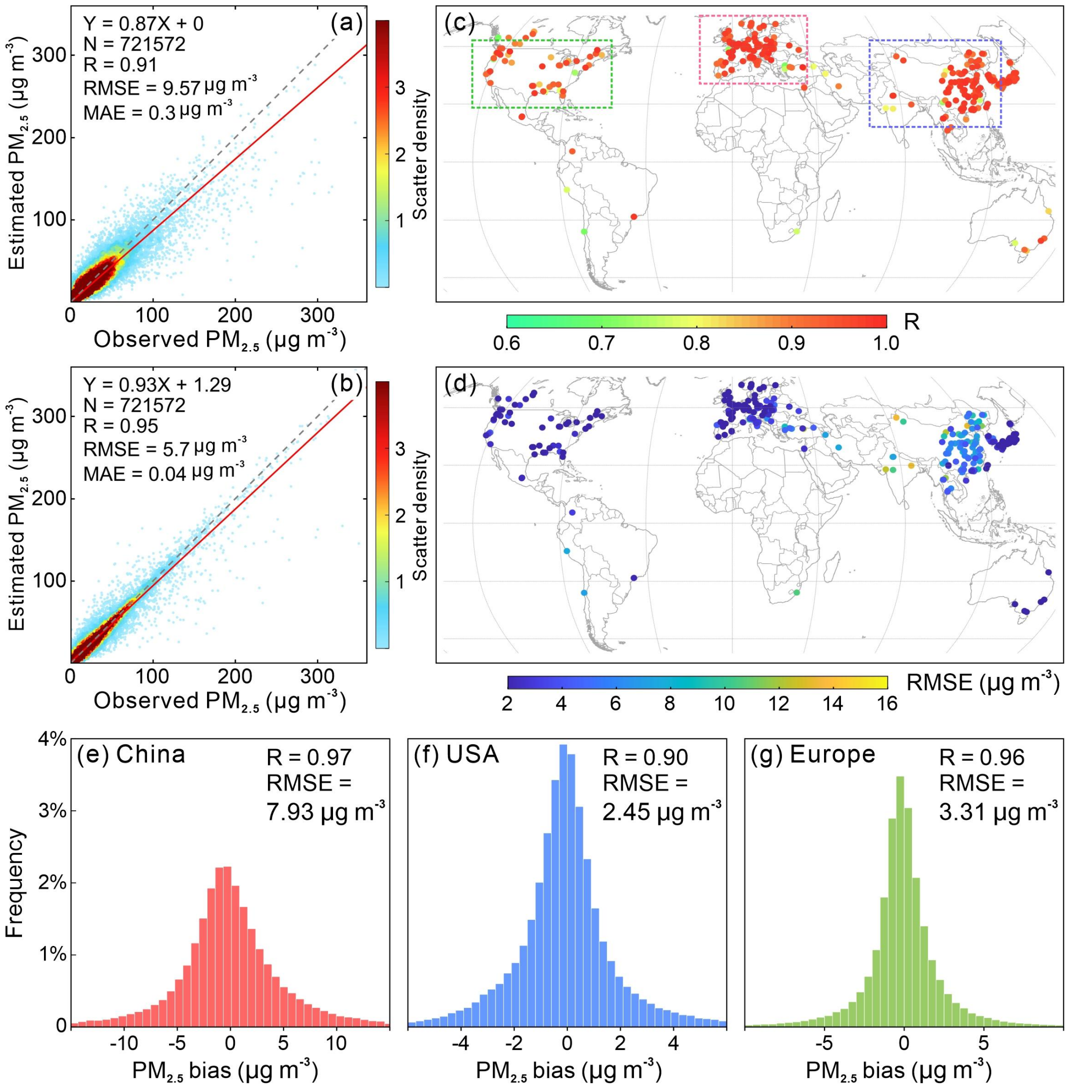

Figure 6. Site-based validation accuracy of $PM_{2.5}$ concentration estimates derived from gap-free AOD images using the proposed ~~SeGAT~~ SCAGAT method.~~.~~ (a) Scatter plots between $PM_{2.5}$ estimates derived from the ~~SeGAT~~ SCAGAT model and the withheld ~~ground-based~~ $PM_{2.5}$ concentration measurements. (b) Same as ~~Figure. 6~~(a) but for gap-free $PM_{2.5}$ estimates fusing ground measured $PM_{2.5}$ concentration measurements. (c–d) Site-based correlation coefficient and RMSE for LGHAP v2 $PM_{2.5}$ concentrations, respectively. (e–g) Histograms of LGHAP v2 $PM_{2.5}$ concentration bias over China, United States, and Europe, respectively. Note the ground-based $PM_{2.5}$ concentration data used here for validation ~~were held out priorly and used neither~~were ~~as not involved in~~ used neither in the model training nor in the data fusion procedures.

Table 3 presents the data accuracy of the gap-free $PM_{2.5}$ concentrations in the LGHAP v2 dataset during the period of 2000−2021 over nations with ~~adequate~~ sufficient records of ground-based ~~measurements of~~ $PM_{2.5}$ concentration measurements ~~records~~. It indicates that the data accuracy of $PM_{2.5}$ concentration estimates varied across regions, with R changing from 0.71 to 0.98 and RMSEs ranging between 1.15 and 32.69 μg m^{-3}. Regardless of the substantial differences in the total number of data pairs ~~across regions~~, larger RMSEs are mainly observed in regions like Mongolia (32.69 μg m^{-3}) and India (25.34 μg m^{-3}), which ~~where~~ often suffered from ~~high~~ severe $PM_{2.5}$ ~~loadings~~pollution episodes. The spatially varying accuracy metrics ~~between regions not only~~ highlight the great complexity in large-scale $PM_{2.5}$ modeling~~. This~~, which also ~~and~~~~but~~ underscores the critical importance of ~~considering~~accounting for spatial representativeness ~~via data-driven models,~~ when applying models over other regions for data extrapolation.

In Figure 7, we examined long-term variations in $PM_{2.5}$ concentrations in four different cities ~~during~~ from 2000~~–~~ to 2021. ~~The~~ A good agreement ~~between the LGHAP v2 PM2.5 concentration time series and~~with the ~~unseen (~~previously withheld~~)~~ ~~ground-based~~ $PM_{2.5}$ concentration measurements ~~confirms the significant~~demonstrated a high accuracy of the LGHAP v2 $PM_{2.5}$ concentration ~~dataset~~estimates. Compared to temporally discrete $PM_{2.5}$ concentration records measured by ground monitors, the gap-free LGHAP v2 $PM_{2.5}$ concentration time series enabled us to ~~examine~~ better understand the long-term variability of haze pollutions across the globe~~, benefiting from its given the gap-free merit~~. ~~Additionally, the agreement between the LGHAP v2 PM2.5 concentration time series and the unseen (previously withheld) ground-based PM2.5 concentration measurements confirm the significant accuracy of the LGHAP v2 PM2.5 concentration dataset. Therefore, this gap-free PM2.5 concentration dataset can be used with confidence when assessing long-term trends of haze pollution across the globe.~~ As shown, declining trends ~~in PM2.5 concentration~~ were observed in $PM_{2.5}$ concentrations as early as in 2006 in New York (United States), whereas apparent reductions were mainly observed ~~mainly~~ after 2012 in Jilin (China) and 2015 in Toyama (Japan). Overall, the gap-free and high accuracy merits render $PM_{2.5}$ concentrations in the LGHAP v2 dataset reliable data sources for assessing long-term trends of haze pollutions across the globe.

Table 3. The ~~d~~Data accuracy of gap-free $PM_{2.5}$ concentrations in the LGHAP v2 dataset ~~by~~ compared ~~ing~~ ~~to~~against ground-based ~~PM2.5 concentration data~~measurements in countries with ~~adequate~~ sufficient $PM_{2.5}$ ~~concentration measurements~~records. The N denotes the total number of $PM_{2.5}$ concentration data pairs for calculating R, RMSE, and bias.

Country	N	R	RMSE (μg m^{-3})	Bias (μg m^{-3})	Country	N	R	RMSE (μg m^{-3})	Bias (μg m^{-3})
China	3,113,160	0.97	8.27	0.36	Iran	67,434	0.74	10.14	−0.09
~~USA~~[NC1]United States	2,048,983	0.84	3.34	0.06	Brazil	50,252	0.81	5.63	0.78
Japan	1,810,436	0.96	1.82	0.07	Portugal	47,782	0.82	3.49	0.14
Canada	1,206,176	0.89	2.12	0.05	Hungary	41,524	0.92	4.59	−0.17
Korea	526,138	0.96	3.49	0.16	Sweden	40,839	0.91	1.61	−0.23
France	502,555	0.96	2.25	0.13	Norway	40,001	0.86	2.45	−0.07

Germany	472,103	0.97	1.94	0.04	Finland	38,884	0.93	1.15	−0.08
Italy	371,888	0.93	5.23	0.04	South Africa	35,314	0.71	10.84	−2.91
~~UK~~United Kingdom	309,181	0.94	1.95	0.11	Serbia	34,795	0.87	9.70	0.01
Spain	297,202	0.87	2.63	0.23	New Zealand	26,654	0.73	3.63	0.20
Czech Republic	209,274	0.97	3.38	0.24	Colombia	26,332	0.95	4.60	0.45
Australia	208,772	0.72	3.70	−0.03	Ukraine	22,692	0.84	5.79	−0.08
India	207,974	0.92	25.34	1.64	Bosnia-Herzegovina	20,297	0.94	12.08	1.59
Belgium	177,036	0.98	1.54	0.01	Greece	19,410	0.79	5.41	−0.10
Poland	175,782	0.95	5.03	0.52	Croatia	17,926	0.90	5.82	−0.44
Turkey	171,381	0.84	10.27	−0.99	Switzerland	14,719	0.75	3.98	−2.26
Austria	131,186	0.97	2.28	−0.14	Russia	14,357	0.84	4.06	0.58
Netherlands	119,047	0.97	1.72	−0.07	Estonia	13,793	0.91	1.48	0.19
Mexico	112,379	0.80	11.42	0.45	Lithuania	13,405	0.87	4.49	0.07
Chile	111,416	0.80	12.64	0.16	Ecuador	12,517	0.88	2.92	0.28
Slovakia	104,892	0.95	3.77	0.18	Vietnam	12,480	0.78	12.94	0.63
Thailand	82,206	0.89	13.21	1.25	Macedonia	10,416	0.92	10.81	2.17
Israel	68,012	0.83	5.08	0.32	Mongolia	9,926	0.91	32.69	−0.17

Figure 8 presents the temporal variations in the global annual mean $PM_{2.5}$ concentration distribution from 2000 to 2021. ~~First of all~~~~First~~As shown, the daily gap-free ~~merit of the~~ LGHAP v2 dataset ~~can~~ seamlessly supports the derivation of comparable annual mean $PM_{2.5}$ concentration maps between years, ~~as~~ and data gap related biases in raw AOD_{Terra} images were eliminated ~~because of~~~~due to the usage of daily gap-free $PM_{2.5}$ concentration data.~~. ~~However~~~~On the other hand, t~~Meanwhile, the quality-assured annual mean $PM_{2.5}$ concentration maps enable us ~~not only~~ to easily pinpoint the hotspot regions suffering from severe haze pollutions and~~but also~~ to ~~examine~~ analyze the long-term variability of global $PM_{2.5}$ concentrations ~~across the globe~~. Specifically, ~~As shown,~~ Mongolia, north India, eastern China, and central Africa were identified as four major regions with relatively high $PM_{2.5}$ loadings, in particular north India, becoming a hotspot region suffering from more severe $PM_{2.5}$ pollutions on the planet. Substantial $PM_{2.5}$ reductions were observed in eastern China ~~since~~ from 2014 onwards, with $PM_{2.5}$ concentrations reduced to ~~a~~ levels even comparable to countries in central Asia~~;~~.~~, and in turn however, north India was in turn a the hotspot region experiencing suffering from more severer $PM_{2.5}$ pollutions on the planet.~~

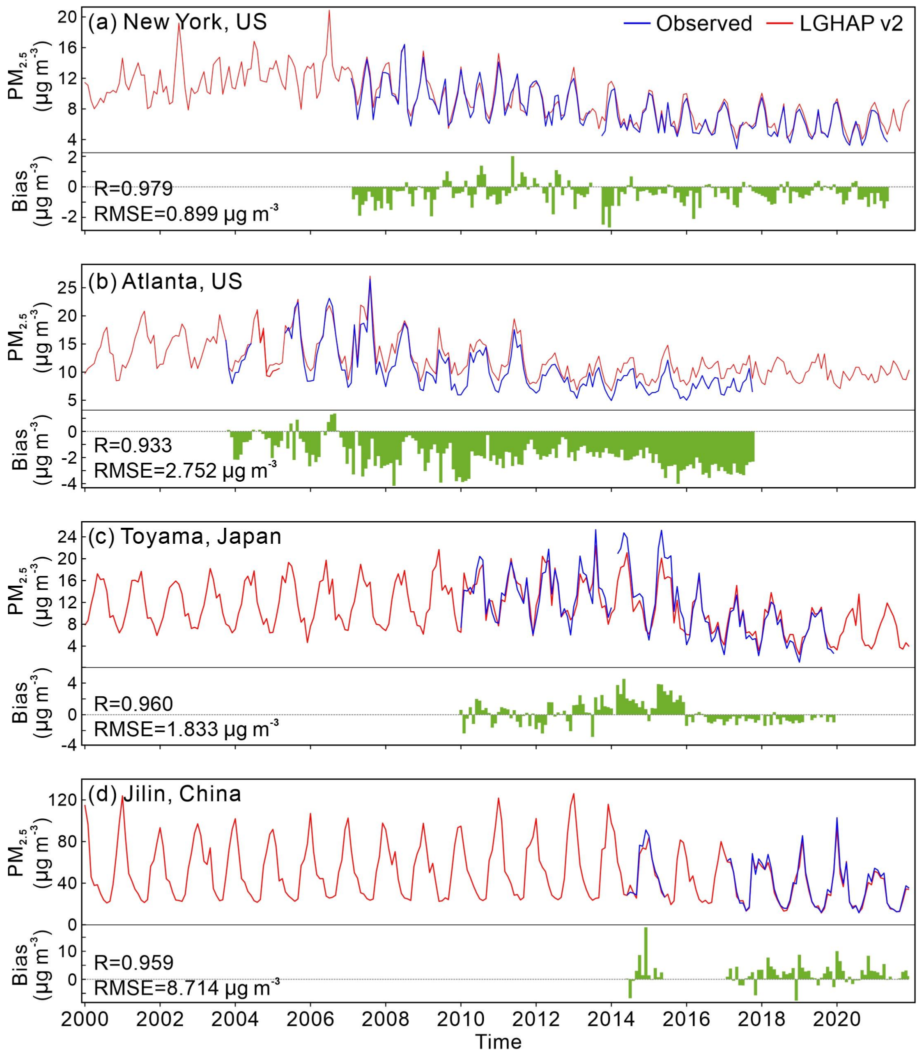

Figure 7. An inter-comparison of temporal variations in monthly mean $PM_{2.5}$ concentrations in four different cities between the LGHAP v2 and collocated ground-based $PM_{2.5}$ concentration measurements ~~during~~ from 2000~~–~~ to 2021.

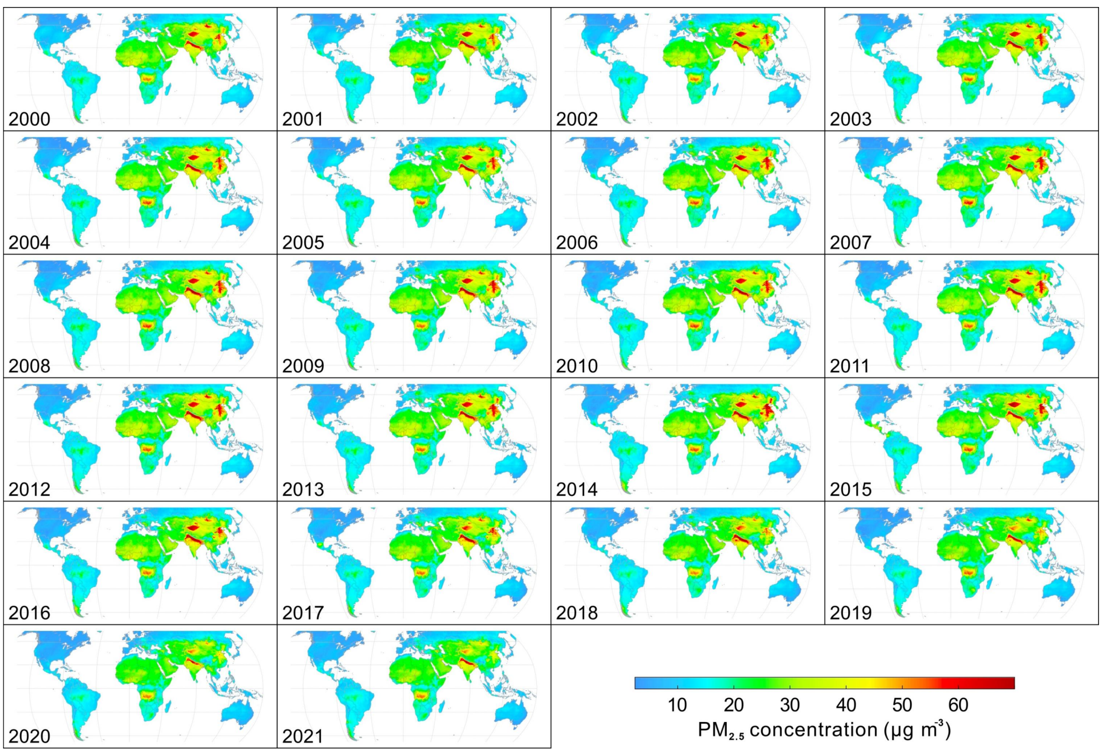

Figure 8. Spatial distribution of the global annual mean $PM_{2.5}$ concentrations derived ~~using~~ from the LGHAP v2 dataset ~~from~~ between 2000 ~~to~~ and 2021.

5. Discussion

Spatially contiguous AOD and $PM_{2.5}$ concentration grids are pivotal to regional air quality management, haze pollution exposure risk assessment, and aerosol radiative forcing diagnosis. By seamlessly gearing up state-of-the-art machine ~~-~~learning and tensor completion methods, a novel ~~framework of~~ big ~~earth~~Earth data analytics framework was developed to fulfill the generation of long-term high-resolution AOD and $PM_{2.5}$ concentration grids ~~as of 2000 in China~~ (LGHAP v1) in our previous study (Bai et al., 2022a). ~~Multimodal~~Specifically, multimodal AODs and rel~~ated~~evant air quality ~~measurements~~ data acquired from diverse satellites, numerical models, and ground monitoring stations were first~~ly~~ harmonized using random forest~~-based data-driven~~ models. Next, m~~M~~ultisource AOD data flows were ~~then~~ weaved neatly as the tensor inputs,~~,~~ ~~from and~~with ~~which~~ data gaps in daily MODIS AOD imag~~eri~~es ~~were~~ properly reconstructed via low-rank tensor completion. Finally, gap-free $PM_{2.5}$ concentration grids were mapped from gap-filled ~~AODs~~ AOD images using a random forest model ~~through machine-learned regression models~~. This big data analytics framework provided an effective solution to integrate multimodal ~~earth~~Earth observations from diverse~~stinct~~ sources to generate high-quality AOD and PM concentrations ~~data products~~ in China.~~, and the good data accuracies of these two gap-free datasets also well demonstrated the efficacy of this framework.~~

In this study, aiming to generate global gap-free AOD and $PM_{2.5}$ concentration grids, namely the LGHAP v2 dataset, the previous big ~~earth~~Earth data analytics framework ~~proposed in our previous study~~ was adopted but enhanced with several new features~~to generate global gap-free AOD and PM2.5 concentration grids, i.e., the LGHAP v2 dataset. Similarly, HOSVD was applied as the core method for tensor completion to fulfill the AOD gap-filling. Despite similar data manipulation procedures, s~~~~Several new algorithmic enhancement modules were also implemented~~, with particular focuses on ~~to~~ accommodating~~e~~ the rocketing data size and global scale modeling demand~~, aiming , not only to improve the computing efficiency and~~other than~~but also to reduce~~ reducing modeling biases. Specifically, an attention mechanism, inspired by deep-learning techniques, was hereby introduced to weight each data slice in the input tensor to account for the drawback induced by ~~Likewise, HOSVD was applied as the core method for tensor completion to fulfilfulfill the AOD gap-filling. Nonetheless, previous results indicated a potential drawback as an~~the equal weight ~~of each data slice in the AOD data cube rendered~~strategy, with~~, with l~~L~~arger weights ~~were~~ assigned to data slices ~~that better resembleding~~ with fewer data gaps and more similar to the ~~actual AOD distribution~~target image ~~on the target date with more valid observations~~. In ~~such a research context~~other words, both the spatial coverage ratio of valid observations in each soft data and the mutual information between the target and soft data ~~were used~~~~served as two relevant metrics~~were considered simultaneously to ~~help determine the~~ weight ~~assigned to~~ each data slice in the AOD tensor. A weighted AOD tensor was then calculated ~~and used as the input tensor data to compel guide~~ for tensor completion~~process, prioritizing focuseding on data slices more similar to likeclosely resembled the target image rather tha~~instead of~~n~~ using all the available ~~data~~information in the AOD tensor indifferently. ~~As demonstrated by~~ Although the ablation experiments shown in Figure 2~~,~~ have demonstrated the efficacy of ~~the AOD fields reconstructed from the~~this attention-reinforced tensor ~~better resembled the actual AOD distributions in the target MODISt AOD~~$_{Terra}$~~ images than those derived from the raw original AOD tensor without applying the attention mechanism~~construction strategy, the underlying philosophy, in particular the relative importance of mutual information and extra spatial coverage, has been not yet fully justified and assessed.

~~Meanwhile, a~~An adaptive background field updating scheme was also introduced to iteratively update prior information in the target AOD~~$_{Terra}$~~ images~~ during each iteration of tensor decomposition and reconstruction~~.~~, and the~~The ~~ultimate goal~~objectivegoal ~~was to mitigate the influence of prior information on the reconstruction accuracy, particularly reducing the~~

~~probability~~ risk of ~~possible propagation of large modelling biases from backgroundin AOD$_{M2}$ to the reconstructed AOD fields.~~ . Compared to the invariant prior information, adaptively updated prior information allowed for mitigating the influence of uncertainties in the prior information on the reconstruction accuracy, particularly large modeling biases from numerical ~~enabled us to not only improve the~~d ~~reconstruction efficiency~~ and~~but also significantly reduce~~d ~~the probability of large error propagation from numerical AOD~~ simulations. Despite these algorithmic improvements, ~~the inter-comparison results even indicated~~ a slightly reduced data accuracy of gap-filled AODs in China from the LGAHP v2 dataset was observed compared to those in the LGHAP v1 dataset. Further investigations revealed this was mainly due to the~~a~~ relatively poor data accuracy of the downscaled AOD$_{M2}$ data ~~becausesince a~~because a global-scale versus~~rather than~~ regional downscaling model was applied ~~to harmonize AOD$_{M2}$ in China~~. ~~This, in turn, underscores the vital importance of data cleaning procedures on reducing the bias levels of each supplementary data to manage the total error budget in the final analyzed data fields when performing big data analytics.~~ Nonetheless, benefiting from the adaptive background updating~~e~~ scheme, the modeling biases in AOD$_{M2}$~~background AOD fields~~ were effectively ~~mitigated~~suppressed in the final reconstructed AOD fields, evidenced by larger biases of AOD$_{M2}$ (R = 0.77, RMSE = 0.36) versus smaller biases of the purely reconstructed AOD (R = 0.82, RMSE = 0.26).

~~As illustrated in Figure 9, the~~The global gap-free and high-resolution benefits ~~gap-filled AOD grids with a daily 1-km resolution enable us~~ render the LGHAP v2 dataset a promising data source to ~~better~~ monitor global aerosol distribution and variations in space and time. ~~Aerosol~~ As illustrated in Figure 9, aerosol-related environmental disturbance episodes, such as sandstorms, wildfires, and haze pollution events, can be well indicated by local rising AODs ~~at the regional scale~~. More importantly~~st critically~~, the gap-filled AOD dataset provides us with an unprecedented opportunity to monitor aerosol loadings and variations even under cloud cover~~s~~, e.g., the haze pollution episodes over southern India and eastern China shown in Figures 9d and 9e.~~,~~ This is largely benefited from ~~the~~ the intelligent spatiotemporal pattern recognition ~~and learning~~, as well as the assimilation of air quality measurements from ground monitoring stations and numerical aerosol diagnostics. While this~~such a~~ global air quality mapping approach greatly facilitates the surveillance and management of air pollution around the world, the ~~high-resolution gap-free AOD and PM$_{2.5}$ concentration~~LGHAP v2 dataset would also ~~largely~~ significantly reduce ~~the uncertainty~~ uncertainties in the health-related aerosol exposure risk assessment results because of the gap-free and high-resolution advantages.

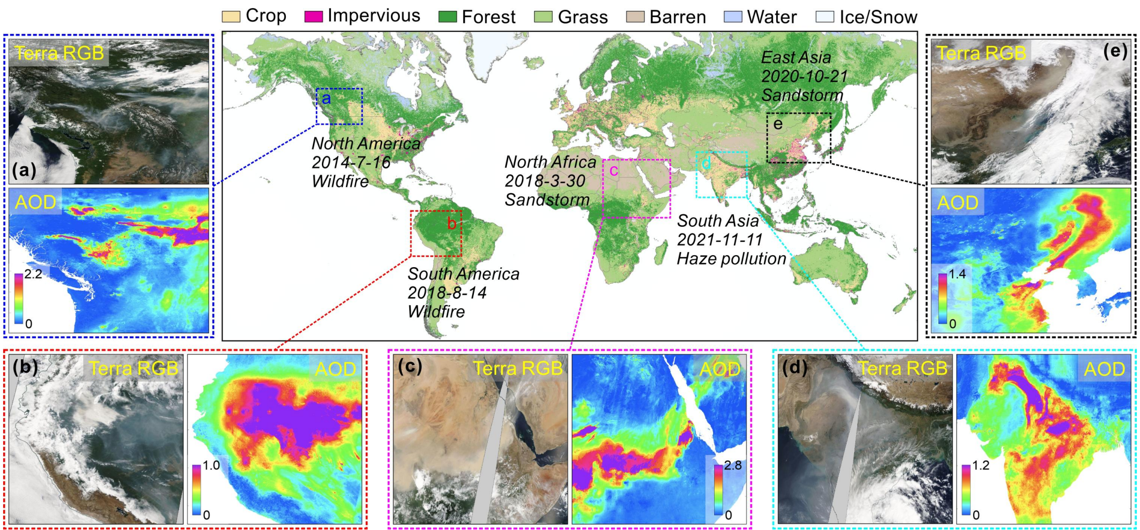

Figure 9. An illustration of AOD responses to ~~wild fire~~wildfires, ~~sand storm~~sandstorms, and haze pollution episodes across the globe, as characterized by gap-free AOD in the LGHAP v2 dataset. The ~~g~~Global map in the middle panel shows ~~a~~the spatial distribution of ~~the~~ major land cover types in 2020.

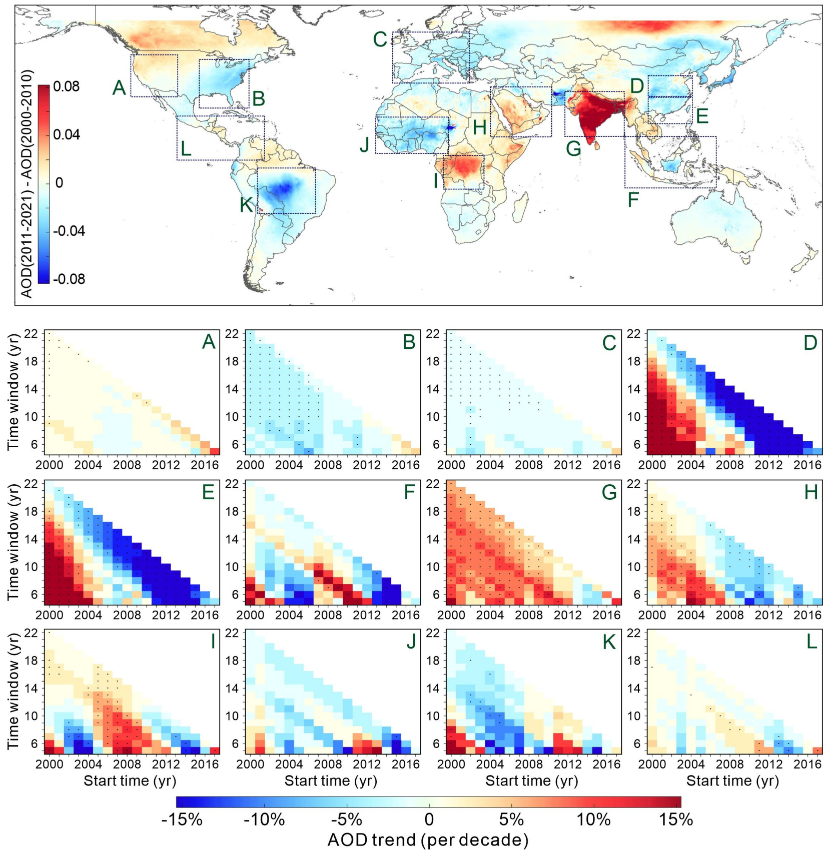

Figure 10. AOD trends over twelve regions of interest ~~across the globe~~worldwide from 2000 to 2021 estimated from gap-free AODs in the LGHAP v2 dataset. The top panel shows ~~a~~ the spatial distribution of global AOD deviations between the first and second decade in the 2000s. Twelve diagrams in the bottom panel show the linear trend of mean AOD over the outlined region of interest at different starting times with varying time window sizes.

Global AOD variation trends were carefully examined by~~By~~ taking advantage of the LGHAP v2 AOD dataset~~, global AOD variation trends were carefully examined~~. Figure~~.~~ 10a presents the AOD deviations between the AOD averages during the first and the second decade in 2000s across the globe. As shown, substantial AOD increases in the twenty-first~~21st~~ century ~~present~~ primarily present over India ~~(G)~~ and central Africa ~~(I)~~, with remarkable AOD decreases observed in the middle of South America. In North America, AOD increases were mainly observed in Canada and the western United States~~US (A)~~ whereas AOD decreases were found in the eastern United States~~US (B)~~. Additionally,~~Also,~~ in reference~~referring~~ to temporally ~~varied~~ varying AOD trends in regions A and B, ~~we may observe~~ evident AOD increasing trends ~~have been~~were observed in the United States~~US~~ from~~since~~ 2012 onwards, while ~~the~~ significant decreasing trends in the eastern United States~~US~~ were ~~even totally~~entirely reversed after 2015. This effect could be partially ~~linked~~ attributed to more frequent and intensive wildfire emissions in north America ~~in~~ during the second decade of the 2000s ~~in north America~~ (Burke et al., 2023; Wei et al., 2021b). A s~~S~~imilar effect was also observed in Europe ~~(C)~~, with an apparent slowdown in the AOD decreasing trend after 2010.

I~~Apparent i~~nverse effects were also observed in China but with totally different temporal transition patterns. As shown, statistically significant AOD increasing trends were observed in eastern ~~(D)~~ and southern ~~(E)~~ China in the first decade, ~~whereas increasing trends started to slow down since 2007~~with a slowdown starting around 2007, ~~and~~ followed by a sudden reversion~~e~~ to decreasing trends ~~was observed~~ after 2010. ~~More importantly, T~~this was also the most significant AOD decreasing trend ~~in~~ during the 2010s around the world. This~~ese~~ observational evidence~~s~~ confirms~~affirm~~ the ~~great~~ success of clean air action~~s~~ in improving air quality in China during ~~the past~~recent decades (Bai et al., 2022a; Liang et al., 2020; Zhang et al., 2019). A s~~S~~imilar temporal variation pattern was also observed in the Middle East ~~(H)~~ but with relatively weak trends. In contrast, India

~~(G)~~ was a~~the~~ hotspot area showing an increasing trend in AOD throughout the 2000s, despite a short period of increasing hiatus from~~during~~ 2013~~–~~ to 2015.

~~In this study, g~~Global gap-free $PM_{2.5}$ concentrations were derived ~~on the basis of~~based on gap-filled AOD grids by taking advantage of a novel SCAGAT ~~deep- learning~~ model~~, which was specifically developed to fulfil~~fulfill~~ global~~ ~~global-scale PM2.5 concentration mapping~~. ~~Differing from~~Unlike many other data-driven model~~ing practices~~s, the spatial representativeness ~~of data-driven models~~ was accounted for ~~by~~ in the SCAGAT model, providing a unique solution to model $PM_{2.5}$ concentrations over regions even without $PM_{2.5}$ monitoring sites. ~~The availability of d~~Daily gap-free $PM_{2.5}$ concentration grids ~~also~~ favors the assessment of the pandemic's influence ~~impacts~~ on regional air quality. Figures~~s.~~ 11a and 11b~~, in the middle panel,~~ present ~~a~~the spatial distribution of $PM_{2.5}$ concentrations before and during the COVID-19 pandemic, respectively. Neglecting long-term variation trends in $PM_{2.5}$ concentrations, the substantial $PM_{2.5}$ decreases in ~~the~~ middle and eastern China, as well as in central Europe, clearly indicate the positive effect of ~~pandemic~~ pandemic-related mobility restrictions on air quality improvement~~,~~ (by comparing $PM_{2.5}$ concentration in 2019 and 2020 during the synchronous period). In contrast, $PM_{2.5}$ reductions were relatively small in the United States~~US~~ due to the lack of mobility restriction measures, with apparent $PM_{2.5}$ reductions observed mainly in regions like Chicago. Overall, ~~the availability of~~ the LGHAP v2 dataset enables us to better investigate global aerosol variations and ~~to~~ ~~assess~~ assess $PM_{2.5}$~~-5~~-related health exposure risks~~ via exposure assessment.~~.

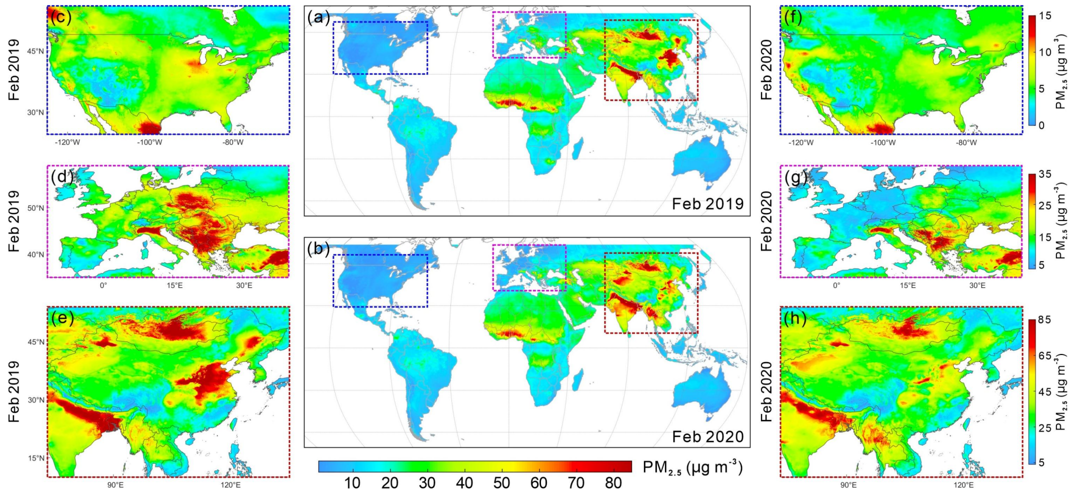

Figure 11. Influence~~mpacts~~ of the COVID-19 pandemic on $PM_{2.5}$ concentrations in United States, Europe, and China. ~~The~~ $PM_{2.5}$ concentrations from LGHAP v2 were averaged over ~~athe~~ synchronous periods in 2019 and 2020 for inter-comparison.

6. Data Availability

The LGHAP v2 dataset provides global gap-free AOD and $PM_{2.5}$ concentration grids from 2000 to 2021 with a daily 1-km resolution. To facilitate the data sharing, each daily map was saved ~~as one~~ a separate~~single~~ NetCDF file, and the data in each individual month w~~as~~ere then archived as ~~a~~ one zip file. ~~Because of~~~~Due to the data storage limitations, one year of data in one year were archived as one single dataset.~~ Table 4 ~~provides~~ summarizes the permanent digital object identifiers for data in each ~~individual dataset~~calenda year from 2000 to 2021. All these datasets were publicly available at the LGHAP community link via https://zenodo.org/communities/ecnu_lghap (Bai et al., 2023a). The d~~D~~ata user guide and visualization codes (Python, MATLAB, R, and IDL) were also provided to guide the users in~~to~~ retrieving~~e~~ data from the NetCDF files, which can be accessed~~ible~~ at https://doi.org/10.5281/zenodo.10216396.

Table 4. List of data links for AOD and $PM_{2.5}$ concentration grids in the LGHAP v2 dataset for each individual year.

Year	LGHAP v2 AOD grids	LGHAP v2 $PM_{2.5}$ grids

2000	https://doi.org/10.5281/zenodo.8281206	https://doi.org/10.5281/zenodo.8307595
2001	https://doi.org/10.5281/zenodo.8281216	https://doi.org/10.5281/zenodo.8307597
2002	https://doi.org/10.5281/zenodo.8281218	https://doi.org/10.5281/zenodo.8307599
2003	https://doi.org/10.5281/zenodo.8281222	https://doi.org/10.5281/zenodo.8307601
2004	https://doi.org/10.5281/zenodo.8281226	https://doi.org/10.5281/zenodo.8307605
2005	https://doi.org/10.5281/zenodo.8281228	https://doi.org/10.5281/zenodo.8307607
2006	https://doi.org/10.5281/zenodo.8287125	https://doi.org/10.5281/zenodo.8308225
2007	https://doi.org/10.5281/zenodo.8287129	https://doi.org/10.5281/zenodo.8308227
2008	https://doi.org/10.5281/zenodo.8287133	https://doi.org/10.5281/zenodo.8308231
2009	https://doi.org/10.5281/zenodo.8287995	https://doi.org/10.5281/zenodo.8308233
2010	https://doi.org/10.5281/zenodo.8288389	https://doi.org/10.5281/zenodo.8308237
2011	https://doi.org/10.5281/zenodo.8288395	https://doi.org/10.5281/zenodo.8310586
2012	https://doi.org/10.5281/zenodo.8288397	https://doi.org/10.5281/zenodo.8310590
2013	https://doi.org/10.5281/zenodo.8287207	https://doi.org/10.5281/zenodo.8310702
2014	https://doi.org/10.5281/zenodo.8288387	https://doi.org/10.5281/zenodo.8310704
2015	https://doi.org/10.5281/zenodo.8289613	https://doi.org/10.5281/zenodo.8310706
2016	https://doi.org/10.5281/zenodo.8289615	https://doi.org/10.5281/zenodo.8310708
2017	https://doi.org/10.5281/zenodo.8294100	https://doi.org/10.5281/zenodo.8310711
2018	https://doi.org/10.5281/zenodo.8301364	https://doi.org/10.5281/zenodo.8313603
2019	https://doi.org/10.5281/zenodo.8301367	https://doi.org/10.5281/zenodo.8313611
2020	https://doi.org/10.5281/zenodo.8301375	https://doi.org/10.5281/zenodo.8313613
2021	https://doi.org/10.5281/zenodo.8301379	https://doi.org/10.5281/zenodo.8313615

7. Conclusion

In this study, the LGHAP v2 dataset, a heritage of the LGHAP~~-~~, ~~which provides long-term gap-free AOD and PM concentration grids with a daily 1-km resolution in China,~~ was generated to provide global gap-free AOD and $PM_{2.5}$ concentration grids with a daily 1-km resolution ~~with the same resolution~~from 2000 to 2021~~(as of 2000daily and 1km) across the globe),~~ , by ~~taking advantage of~~leveraging an improved big ~~earth~~Earth data analytics approach. The g~~G~~round validation results ~~demonstrate~~ confirm high accuracies of these two gap-free products, with AOD having ~~a correlation~~n an R of 0.85 and an RMSE of 0.14 compared to the AERONET AOD observations~~,~~, which are slightly worse than the original MCD19A2 product (R = 0.88 and RMSE = 0.11). Similarly, ~~The sSite-based validation results also indicate that the~~ $PM_{2.5}$ concentration estimates derived from gap-free AOD via the SCAGAT method show an~~a good~~ agreement with the withheld~~held-out~~ ground-based $PM_{2.5}$ measurements, ~~with~~ achieving an R of 0.91 and an RMSE of 9.57 μg m^{-3}~~;, and Furthermore,~~, while the data accuracy was ~~further~~ improved to an R of 0.95 and an RMSE of 5.7 μg m^{-3} with the fusion of ground-measured $PM_{2.5}$ ~~measurements~~concentrations. ~~To our knowledge, this is the first two-decade-longtwenty-year global gap-free AOD and PM2.5 concentration dataset with such a high resolution.~~

~~The dData gaps in satellite-based AOD images were filled using a similar~~ Several new algorithmic enhancement modules were incorporated to the big data analytics ~~approach~~ framework to ~~what wasthat~~ as ~~developedused~~improve both the computing speed and the reconstruction accuracy ~~to for generatinge the LGHAP dataset in China, albeitbut with several new algorithmic improvements~~. The ablation experiments ~~well~~ demonstrated the effectiveness and advantages of ~~applying incorporating~~ the newly implemented attention mechanism to weigh~~t~~ each slice of soft data in the AOD tensor ~~during the tensor completion procedure~~. ~~Also,Additionally, u~~Updating prior information in the target image after each tensor reconstruction iteration ~~not only~~ help~~s~~ed mitigate the ~~probability~~ risk of error propagation from numerical aerosol diagnostics to the final reconstructed field ~~andbut also , while also~~ and improv~~e~~ing~~es~~ the convergence speed of tensor completion. ~~Moreover~~Overall, this study provides a ~~good~~ compelling illustration of big ~~earth~~Earth data analytics to generate high-quality remote sensing datasets by synergistically integrating and assimilating multimodal data from diverse sources via ~~machine~~ machine-learning techniques. ~~The last but not leFastinally~~Additionally, this big data analytics approach ~~can~~ could be also used for ~~be also used to fulfilfulfill~~

near-term gap-free AOD mapping by ~~leveraging~~simply replacing~~ by simply replacing aerosol reanalysis with~~ numerical AOD reanalysis with forecasting fields~~s~~ (e.g., CAMS ~~AOD~~ forecasts).

This study also provides new insights on how to deal with the scal~~ing effect~~e problem when ~~establishing~~ developing ~~large~~ large-scale ~~$PM_{2.5}$~~environmental variable (e.g. $PM_{2.5}$ concentration) ~~prediction~~ mapping models. Instead of~~Rather than creating~~ constructing a global model ~~by gathering~~with all paired data samples ~~into one a single training set~~, site-specific $PM_{2.5}$ prediction models were first~~ly~~ established using a random forest model~~, and. Follow that~~, and a graph attention network was ~~then~~ then ~~applied~~ developed to establish an ensemble learning~~spatial interpolation~~ model~~,~~ to integrat~~ing~~e multiple ~~on the basis of~~ $PM_{2.5}$ estimates derived from site-specific random forest models trained over sites with similar scene features as the target grid. By ~~fully taking advantage of~~accounting for the scene similarity ~~of~~between ~~distant data sample~~geographic regions, the proposed deep-learning m~~ethod~~odel ~~effectively~~attempted to~~BecauseSince there is no need to establish regional estimation models, thissuch a philosophy not only improves the modeling accuracy andbut also~~ ~~solve~~addresss the ~~scaling~~ scale problem in large-~~-~~scale $PM_{2.5}$ modeling practices.

The LGHAP v2 dataset is publicly accessible using~~from~~ the aforementioned links~~ given above~~. The ~~Given the merit of the~~ gap-free and high-resolution~~ merit, this~~ dataset can ~~be used to deepen our understanding of~~be used as a reliable data source for assessing ~~aerosol~~ aerosol-~~climatic~~ climate ~~effects~~interactions, as well as $PM_{2.5}$ exposure risks and related health outcomes ~~at the global scale~~around the world. ~~Also,Additionally, the r~~Researchers are also encouraged to use this dataset to ~~better~~ evaluate the status and trends of urban aerosol pollutions across the globe to support the assessment of ~~sustainable~~ Sustainable ~~development~~ Development ~~goals~~ Goals~~ related to urban air quality across the globe~~.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant No. 42171309), the International Research Center of Big Data for Sustainable Development Goals (Grant No. CBAS2022GSP07), the Foreign Technical Cooperation and Scientific Research Program (Grant No. E3KZ0301), ~~and~~ the Director's Fund of Key Laboratory of Geographic Information Science (Ministry of Education), and East China Normal University (Grant No. KLGIS2023C01). The authors would like to express gratitude to relevant organizations and data archive services for generating and sharing~~their great efforts in providing~~ essential datasets ~~sources~~ used in this study.~~ to support the generation of global LGHAP v2 dataset.~~

References

Bai, K., and Li, K.: LGHAP: Long-term Gap-free High-resolution Air Pollutants concentration dataset, Zenodo [dataset], https://zenodo.org/communities/ecnu_lghap, 2023a.

Bai, K., Chang, N.-B., and Chen, C.-F.: Spectral Information Adaptation and Synthesis Scheme for Merging Cross-Mission Ocean Color Reflectance Observations from MODIS and VIIRS, IEEE Transactions on Geoscience and Remote Sensing, 54, 311–329, https://doi.org/10.1109/TGRS.2015.2456906, 2016a.

Bai, K., Chang, N.-B., Yu, H., and Gao, W.: Statistical bias correction for creating coherent total ozone record from OMI and OMPS observations, Remote Sensing of Environment, 182, 150–168, https://doi.org/10.1016/j.rse.2016.05.007, 2016b.

Bai, K., Li, K., Chang, N.-B., and Gao, W.: Advancing the prediction accuracy of satellite-based PM2.5 concentration mapping: A perspective of data mining through in situ PM2.5 measurements, Environmental Pollution, 254, https://doi.org/10.1016/j.envpol.2019.113047, 2019.

Bai, K., Li, K., Guo, J., and Chang, N.-B.: Multiscale and multisource data fusion for full-coverage PM2.5 concentration mapping: Can spatial pattern recognition come with modeling accuracy? ISPRS Journal of Photogrammetry and Remote Sensing, 184, 31–44, https://doi.org/10.1016/j.isprsjprs.2021.12.002, 2022b.

Bai, K., Li, K., Guo, J., Yang, Y., and Chang, N.-B.: Filling the gaps of in situ hourly PM2.5 concentration data with the aid of empirical orthogonal function analysis constrained by diurnal cycles, Atmospheric Measurement Techniques, 13, 1213–1226, https://doi.org/10.5194/amt-13-1213-2020, 2020.

Bai, K., Li, K., Ma, M., Li, K., Li, Z., Guo, J., Chang, N.-B., Tan, Z., and Han, D.: LGHAP: the Long-term Gap-free High-resolution Air Pollutant concentration dataset, derived via tensor-flow-based multimodal data fusion, Earth System Science Data, 14, 907–927, https://doi.org/10.5194/essd-14-907-2022, 2022a.

Bai, K., Li, K., Sun, Y., Wu, L., Zhang, Y., Chang, N.-B., and Li, Z.: Global synthesis of two decades of research on improving PM2.5 estimation models from remote sensing and data science perspectives, Earth-Science Reviews, 241, 104461, https://doi.org/10.1016/j.earscirev.2023.104461, 2023b.

Beckers, J. M. and Rixen, M.: EOF calculations and data filling from incomplete oceanographic datasets, Journal Of Atmospheric And

Oceanic Technology, 20, 1839–1856, https://doi.org/10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2, 2003.

Bi, J., Belle, J. H., Wang, Y., Lyapustin, A. I., Wildani, A., and Liu, Y.: Impacts of snow and cloud covers on satellite-derived PM2.5 levels, Remote Sensing of Environment, 221, 665–674, https://doi.org/10.1016/j.rse.2018.12.002, 2019.

Buchard, V., Randles, C. A., da Silva, A. M., Darmenov, A., Colarco, P. R., Govindaraju, R., Ferrare, R., Hair, J., Beyersdorf, A. J., Ziemba, L. D., and Yu, H.: The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part II: Evaluation and Case Studies, Journal of Climate, 30, 6851–6872, https://doi.org/10.1175/JCLI-D-16-0613.1, 2017.

Burke, M., Childs, M. L., de la Cuesta, B., Qiu, M., Li, J., Gould, C. F., Heft-Neal, S., and Wara, M.: The contribution of wildfire to PM2.5 trends in the USA, Nature, 622, 761–766, https://doi.org/10.1038/s41586-023-06522-6, 2023.

Che, H., Zhang, X.-Y., Xia, X., Goloub, P., Holben, B., Zhao, H., Wang, Y., Zhang, X.-C., Wang, H., Blarel, L., Damiri, B., Zhang, R., Deng, X., Ma, Y., Wang, T., Geng, F., Qi, B., Zhu, J., Yu, J., Chen, Q., and Shi, G.: Ground-based aerosol climatology of China: aerosol optical depths from the China Aerosol Remote Sensing Network (CARSNET) 2002−2013, Atmospheric Chemistry And Physics, 15, 7619–7652, https://doi.org/10.5194/acp-15-7619-2015, 2015.

Chen, X., Ding, J., Liu, J., Wang, J., Ge, X., Wang, R., and Zuo, H.: Validation and comparison of high-resolution MAIAC aerosol products over Central Asia, Atmospheric Environment, 251, 118273, https://doi.org/10.1016/j.atmosenv.2021.118273, 2021.

Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., and Lyapustin, A. I.: 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, 169−209, https://doi.org/10.5194/amt-12-169-2019, 2019.

Guo, B., Wang, Z., Pei, L., Zhu, X., Chen, Q., Wu, H., Zhang, W., and Zhang, D.: Reconstructing MODIS aerosol optical depth and exploring dynamic and influential factors of AOD via random forest at the global scale, Atmospheric Environment, 315, 120159, https://doi.org/10.1016/j.atmosenv.2023.120159, 2023.

Guo, J., Deng, M., Lee, S. S., Wang, F., Li, Z., Zhai, P., Liu, H., Lv, W., Yao, W., and Li, X.: Delaying precipitation and lightning by air pollution over the Pearl River Delta. Part I: Observational analyses, Journal of Geophysical Research: Atmospheres, 121, 6472–6488, https://doi.org/10.1002/2015JD023257, 2016.

Guo, J., Su, T., Chen, D., Wang, J., Li, Z., Lv, Y., Guo, X., Liu, H., Cribb, M., and Zhai, P.: Declining summertime local-scale precipitation frequency over China and the United States, 1981–2012: The disparate roles of aerosols. Geophysical Research Letters, 46, 13281–13289. https://doi.org/10.1029/2019GL085442, 2019.

He, Q., Wang, W., Song, Y., Zhang, M., and Huang, B.: Spatiotemporal high-resolution imputation modeling of aerosol optical depth for investigating its full-coverage variation in China from 2003 to 2020, Atmospheric Research, 281, 106481, https://doi.org/10.1016/j.atmosres.2022.106481, 2023.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Huang, X., Song, Y., Yang, J., Wang, W., Ren, H., Dong, M., Feng, Y., Yin, H., and Li, J.: Toward accurate mapping of 30-m time-series global impervious surface area (GISA), International Journal of Applied Earth Observation and Geoinformation, 109, 102787, https://doi.org/10.1016/j.jag.2022.102787, 2022.

Jiang, J., Liu, J., Jiao, D., Zha, Y., and Cao, S.: Evaluation of MODIS DT, DB, and MAIAC Aerosol Products over Different Land Cover Types in the Yangtze River Delta of China, Remote Sensing (Basel), 15, 275, https://doi.org/10.3390/rs15010275, 2023.

Johnson, J. M., Khoshgoftaar, T. M.: Survey on deep learning with class imbalance, Journal of Big Data, 6, 27. https://doi.org/10.1186/s40537-019-0192-5, 2019.

Li, K., Bai, K., Jiao, P., Sun, Y., Shao, L., Li, X., Liu, C., Ma, M., Qiu, S., Zheng, Z., Han, D., Li, R., Li, Z., Guo, J., Chang, N.: SCAGAT: A scene-aware ensemble learning graph attention network for global PM2.5 pollution mapping, in preparation.

Li, K., Bai, K., Li, Z., Guo, J., and Chang, N.-B.: Synergistic data fusion of multimodal AOD and air quality data for near real-time full coverage air pollution assessment, Journal of Environmental Management, 302, 114121, https://doi.org/10.1016/j.jenvman.2021.114121, 2022b.

Li, K., Bai, K., Ma, M., Guo, J., Li, Z., Wang, G., and Chang, N.-B.: Spatially gap free analysis of aerosol type grids in China: First retrieval via satellite remote sensing and big data analytics, ISPRS Journal of Photogrammetry and Remote Sensing, 193, 45–59, https://doi.org/10.1016/j.isprsjprs.2022.09.001, 2022a.

Li, L., Franklin, M., Girguis, M., Lurmann, F., Wu, J., Pavlovic, N., Breton, C., Gilliland, F., and Habre, R.: Spatiotemporal imputation of MAIAC AOD using deep learning with downscaling, Remote Sensing of Environment, 237, 111584, https://doi.org/10.1016/j.rse.2019.111584, 2020.

Li, Z. Q., Xu, H., Li, K. T., Li, D. H., Xie, Y. S., Li, L., Zhang, Y., Gu, X. F., Zhao, W., Tian, Q. J., Deng, R. R., Su, X. L., Huang, B., Qiao, Y. L., Cui, W. Y., Hu, Y., Gong, C. L., Wang, Y. Q., Wang, X. F., Wang, J. P., Du, W. B., Pan, Z. Q., Li, Z. Z., and Bu, D.: Comprehensive study of optical, physical, chemical, and radiative properties of total columnar atmospheric aerosols over China: An overview of sun–Sky radiometer observation network (SONET) measurements, Bulletin of the American Meteorological Society, 99, 739−755, https://doi.org/10.1175/BAMS-D-17-0133.1, 2018.

Li, Z., Guo, J., Ding, A., Liao, H., Liu, J., Sun, Y., Wang, T., Xue, H., Zhang, H., and Zhu, B.: Aerosol and boundary-layer interactions and impact on air quality, National Science Review, 4, 810–833, https://doi.org/10.1093/nsr/nwx117, 2017.

Li, Z., Wang, Y., Guo, J., Zhao, C., Cribb, M. C., Dong, X., Fan, J., Gong, D., Huang, J., Jiang, M., Jiang, Y., Lee, S. S., Li, H., Li, J., Liu, J., Qian, Y., Rosenfeld, D., Shan, S., Sun, Y., Wang, H., Xin, J., Yan, X., Yang, X., Yang, X., Zhang, F., and Zheng, Y.: East Asian Study of Tropospheric Aerosols and their Impact on Regional Clouds, Precipitation, and Climate (EAST-AIRCPC), Journal of Geophysical Research: Atmospheres, 124, 13026–13054, https://doi.org/10.1029/2019JD030758, 2019.

Liang, F., Xiao, Q., Huang, K., Yang, X., Liu, F., Li, J., Lu, X., Liu, Y., and Gu, D.: The 17-y spatiotemporal trend of PM2.5 and its mortality

burden in China, Proceedings of the National Academy of Sciences, 117, 25601–25608, https://doi.org/10.1073/pnas.1919641117, 2020.
Liu, J., Ren, C., Huang, X., Nie, W., Wang, J., Sun, P., Chi, X., and Ding, A.: Increased Aerosol Extinction Efficiency Hinders Visibility Improvement in Eastern China, Geophysical Research Letters, 47, https://doi.org/10.1029/2020GL090167, 2020.
Liu, N., Zou, B., Feng, H., Wang, W., Tang, Y., and Liang, Y.: Evaluation and comparison of multiangle implementation of the atmospheric correction algorithm, Dark Target, and Deep Blue aerosol products over China, Atmospheric Chemistry and Physics, 19, 8243–8268, https://doi.org/10.5194/acp-19-8243-2019, 2019.
Liu, X. and Wang, M.: Filling the gaps of missing data in the merged VIIRS SNPP/NOAA-20 ocean color product using the DINEOF method, Remote Sensing (Basel), 11, https://doi.org/10.3390/rs11020178, 2019.
Lyapustin, A., Wang, Y., Korkin, S., and Huang, D.: MODIS Collection 6 MAIAC algorithm, Atmospheric Measurement Techniques, 11, 5741–5765, https://doi.org/10.5194/amt-11-5741-2018, 2018.
Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., Levy, R., and Reid, J. S.: Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm, Journal of Geophysical Research Atmospheres, 116, 1–15, https://doi.org/10.1029/2010JD014986, 2011.
Ma, Z., Liu, Y., Zhao, Q., Liu, M., Zhou, Y., and Bi, J.: Satellite-derived high resolution PM2.5 concentrations in Yangtze River Delta Region of China using improved linear mixed effects model, Atmospheric Environment, 133, 156–164, https://doi.org/10.1016/j.atmosenv.2016.03.040, 2016.
Martins, V. S., Lyapustin, A., Carvalho, L. A. S., Barbosa, C. C. F., and Novo, E. M. L. M.: Validation of high-resolution MAIAC aerosol product over South America, Journal of Geophysical Research: Atmospheres, 122, 7537–7559, https://doi.org/10.1002/2016JD026301, 2017.
Mhawish, A., Banerjee, T., Sorek-Hamer, M., Lyapustin, A., Broday, D. M., and Chatfield, R.: Comparison and evaluation of MODIS Multi-angle Implementation of Atmospheric Correction (MAIAC) aerosol product over South Asia, Remote Sensing of Environment, 224, 12–28, https://doi.org/10.1016/j.rse.2019.01.033, 2019.
Qin, W., Fang, H., Wang, L., Wei, J., Zhang, M., Su, X., Bilal, M., and Liang, X.: MODIS high-resolution MAIAC aerosol product: Global validation and analysis, Atmospheric Environment, 264, 118684, https://doi.org/10.1016/j.atmosenv.2021.118684, 2021.
Randles, C. A., da Silva, A. M., Buchard, V., Colarco, P. R., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J., Shinozuka, Y., and Flynn, C. J.: The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part I: System Description and Data Assimilation Evaluation, Journal of Climate, 30, 6823–6850, https://doi.org/10.1175/JCLI-D-16-0609.1, 2017.
Shannon, C. E.: A Mathematical Theory of Communication, Bell System Technical Journal, 27, 379–423, 1948.
Shi, H., Zhang, Y., Chen, Y., Ji, S., Dong, Y.: Resampling algorithms based on sample concatenation for imbalance learning, Knowledge-Based Systems, 245, 108592. https://doi.org/10.1016/j.knosys.2022.108592, 2022.
Sogacheva, L., Popp, T., Sayer, A. M., Dubovik, O., Garay, M. J., Heckel, A., Christina Hsu, N., Jethva, H., Kahn, R. A., Kolmonen, P., Kosmale, M., De Leeuw, G., Levy, R. C., Litvinov, P., Lyapustin, A., North, P., Torres, O., and Arola, A.: Merging regional and global aerosol optical depth records from major available satellite products, Atmospheric Chemistry and Physics, 20, 2031–2056, https://doi.org/10.5194/acp-20-2031-2020, 2020.
Tang, Q., Bo, Y., and Zhu, Y.: Spatiotemporal fusion of multiple-satellite aerosol optical depth (AOD) products using Bayesian maximum entropy method, Journal of Geophysical Research: Atmospheres, 121, 4034–4048, https://doi.org/10.1002/2015JD024571, 2016.
Up in the aerosol, Nature Geoscience, 15, 157, https://doi.org/10.1038/s41561-022-00915-4, 2022.
Wang, Y. W. and Yang, Y. H.: China's dimming and brightening: Evidence, causes and hydrological implications, Annales Geophysicae, 32, 41–55, https://doi.org/10.5194/ANGEO-32-41-2014, 2014.
Wang, Y., Yuan, Q., Zhou, S., and Zhang, L.: Global spatiotemporal completion of daily high-resolution TCCO from TROPOMI over land using a swath-based local ensemble learning method, ISPRS Journal of Photogrammetry and Remote Sensing, 194, 167–180, https://doi.org/10.1016/j.isprsjprs.2022.10.012, 2022.
Wei, J., Li, Z., Lyapustin, A., Sun, L., Peng, Y., Xue, W., Su, T., and Cribb, M.: Reconstructing 1-km-resolution high-quality PM2.5 data records from 2000 to 2018 in China: spatiotemporal variations and policy implications, Remote Sensing of Environment, 252, 112136, https://doi.org/10.1016/j.rse.2020.112136, 2021a.
Wei, X., Bai, K., Chang, N.-B., and Gao, W.: Multisource hierarchical data fusion for high-resolution AOD mapping in a forest fire event, International Journal of Applied Earth Observation and Geoinformation, 102, 102366, https://doi.org/10.1016/j.jag.2021.102366, 2021b.
Wei, X., Chang, N.-B., Bai, K., and Gao, W.: Satellite remote sensing of aerosol optical depth: advances, challenges, and perspectives, Critical Reviews in Environmental Science and Technology, 50, 1640–1725, https://doi.org/10.1080/10643389.2019.1665944, 2020.
WHO: Ambient air pollution, 2022.
Wild, M., Wacker, S., Yang, S., and Sanchez-Lorenzo, A.: Evidence for Clear-Sky Dimming and Brightening in Central Europe, Geophysical Research Letters, 48, https://doi.org/10.1029/2020GL092216, 2021.
Xiao, Q., Geng, G., Cheng, J., Liang, F., Li, R., Meng, X., Xue, T., Huang, X., Kan, H., Zhang, Q., and He, K.: Evaluation of gap-filling approaches in satellite-based daily PM2.5 prediction models, Atmospheric Environment, 244, 117921, https://doi.org/10.1016/j.atmosenv.2020.117921, 2021.
Xiao, Q., Wang, Y., Chang, H. H., Meng, X., Geng, G., Lyapustin, A., and Liu, Y.: Full-coverage high-resolution daily PM2.5 estimation using MAIAC AOD in the Yangtze River Delta of China, Remote Sensing of Environment, 199, 437–446, https://doi.org/10.1016/j.rse.2017.07.023, 2017.
Xu, H., Guang, J., Xue, Y., de Leeuw, G., Che, Y. H., Guo, J., He, X. W., and Wang, T. K.: A consistent aerosol optical depth (AOD) dataset over mainland China by integration of several AOD products, Atmospheric Environment, 114, 48–56, https://doi.org/10.1016/j.atmosenv.2015.05.023, 2015.
Yang, X., Zhao, C., Zhou, L., Wang, Y., and Liu, X.: Distinct impact of different types of aerosols on surface solar radiation in China, Journal of Geophysical Research: Atmospheres, 121, 6459–6471, https://doi.org/10.1002/2016JD024938, 2016.
Yang, Y., Ren, L., Li, H., Wang, H., Wang, P., Chen, L., Yue, X., and Liao, H.: Fast Climate Responses to Aerosol Emission Reductions During the COVID-19 Pandemic, Geophysical Research Letters, 47, https://doi.org/10.1029/2020GL089788, 2020.
Zhang, Q., Zheng, Y., Tong, D., Shao, M., Wang, S., Zhang, Y., Xu, X., Wang, J., He, H., Liu, W., Ding, Y., Lei, Y., Li, J., Wang, Z., Zhang,

X., Wang, Y., Cheng, J., Liu, Y., Shi, Q., Yan, L., Geng, G., Hong, C., Li, M., Liu, F., Zheng, B., Cao, J., Ding, A., Gao, J., Fu, Q., Huo, J., Liu, B., Liu, Z., Yang, F., He, K., and Hao, J.: Drivers of improved PM2.5 air quality in China from 2013 to 2017, Proceedings of the National Academy of Sciences of the United States of America, 116, 24463–24469, https://doi.org/10.1073/pnas.1907956116, 2019.

Zhang, T., Zhou, Y., Zhao, K., Zhu, Z., Asrar, G. R., and Zhao, X.: Gap-filling MODIS daily aerosol optical depth products by developing a spatiotemporal fitting algorithm, Giscience & Remote Sensing, 59, 762–781, https://doi.org/10.1080/15481603.2022.2060596, 2022.

Zhao, C., Yang, Y., Fan, H., Huang, J., Fu, Y., Zhang, X., Kang, S., Cong, Z., Letu, H., and Menenti, M.: Aerosol characteristics and impacts on weather and climate over the Tibetan Plateau, National Science Review, 7, 492–495, https://doi.org/10.1093/nsr/nwz184, 2020.