https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

Open Access Earth System Science Data Discussions

A global land aerosol fine-mode fraction dataset (2001–2020) retrieved from MODIS using hybrid physical and deep learning approaches

Xing Yan[1*], Zhou Zang[1], Zhanqing Li[2*], Nana Luo[3], Chen Zuo[1], Yize Jiang[1], Dan Li[1], Yushan Guo[1], Wenji Zhao[4], Wenzhong Shi[5], Maureen Cribb[2]

[1]State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing, 100875, China
[2]Department of Atmospheric and Oceanic Science and ESSIC, University of Maryland, College Park, MD, 20740, USA
[3]School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 102612, China
[4]College of Resource Environment and Tourism, Capital Normal University, Beijing, China
[5]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China

Correspondence to: Xing Yan (yanxing@bnu.edu.cn); Zhanqing Li (zli@atmos.umd.edu)

Abstract. The aerosol fine-mode fraction (FMF) is potentially valuable for discriminating natural aerosols from anthropogenic ones. However, most current satellite-based FMF products are highly unreliable. Here, we developed a new satellite-based global land daily FMF dataset (Phy-DL FMF) by synergizing the advantages of physical and deep learning methods at a 1° spatial resolution by covering the period from 2001 to 2020. The Phy-DL FMF dataset is comparable to Aerosol Robotic Network (AERONET) measurements, based on the analysis of 361,089 data samples from 1170 AERONET sites around the world. Overall, Phy-DL FMF showed a root-mean-square error of 0.136 and correlation coefficient of 0.68, and the proportion of results that fell within the ±20% expected error window was 79.15%. Phy-DL FMF showed superior performance over alternate deep learning or physical approaches (such as the spectral deconvolution algorithm presented in our previous studies), particularly for forests, grasslands, croplands, and urban and barren land types. As a long-term dataset, Phy-DL FMF is able to show an overall significant decreasing trend (at a 95% significance level) over global land areas. Based on the trend analysis of Phy-DL FMF for different countries, the upward trend in the FMFs was particularly strong over India and the western USA. Overall, this study provides a new FMF dataset for global land areas that can help improve our understanding of spatiotemporal fine- and coarse-mode aerosol changes. The datasets can be downloaded from https://doi.org/10.5281/zenodo.5105617 (Yan, 2021).

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

Open Access Earth System Science Data Discussions

35

1 Introduction

Evaluating the impact of anthropogenic aerosols on climate change and human health relies on the ability to separate the proportion of anthropogenic aerosols from the total aerosol loading (Anderson et al., 2005; Zheng et al., 2015). Although satellite remote sensing can provide global-scale data on aerosol content that are represented by the aerosol optical depth
40 (AOD), accurate monitoring of anthropogenic aerosols is still a major challenge. This is because a key parameter called the aerosol fine-mode fraction (FMF), which is used for discriminating anthropogenic aerosols from natural ones (Bellouin et al., 2005), has been regarded as highly unreliable according to satellite-based AOD retrievals, especially over land (Levy et al., 2013; Yan et al., 2017; Liang et al., 2021; Yang et al., 2020; Zang et al., 2021a).

Satellite-based FMF retrievals based on physical methods have been performed previously,; currently, five global-scale
45 FMF products exist (Figure 1) that exhibit different temporal resolutions from 1 to 16 days (Levy et al., 2007; Garay et al., 2020; Chen et al., 2020a). Of these, POLarization and Directionality of the Earth's Reflectances (POLDER) can perform multi-angle and multi-spectral polarized measurements, which provide unique advantages in the retrieval of aerosol FMF (Dubovik et al., 2011; Dubovik et al., 2019). Therefore, in recent years, several POLDER-based FMF retrieval methods have been proposed (Zhang et al., 2016; Zhang et al., 2021), such as the generalized retrieval of aerosols and surface properties (Dubovik
50 et al., 2014). However, POLDER ended its mission in 2013, whereas the Moderate Resolution Imaging Spectroradiometer (MODIS) has operated for about 20 years and continues to perform well. Currently, only the MODIS Dark Target (DT) method has been used to generate global aerosol FMF products over both land and ocean. However, the MODIS DT-derived FMF over land is highly unreliable and is not recommended for use even though it has evolved to the Collection 6.1 (C6.1) level (Levy et al., 2013; Chen et al., 2020a). To improve the accuracy of MODIS land-based FMF retrievals, improvements have
55 been made to physical approaches, such as the Look-Up-Table-based Spectral Deconvolution Algorithm (LUT-SDA) (Yan et al., 2017; Yan et al., 2019) and the Yonsei Aerosol Retrieval algorithm (Choi et al., 2016). Using the LUT-SDA model in previous research, we developed a 10-year global land FMF dataset (Yan et al., 2021b) with moderately improved retrieval accuracy (root-mean-square error, RMSE = 0.22). No multi-angle and multi-spectral polarized information, Lipponen et al. (2018) noted that MODIS-based FMF retrievals using physical methods still suffer from these major limitations.

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

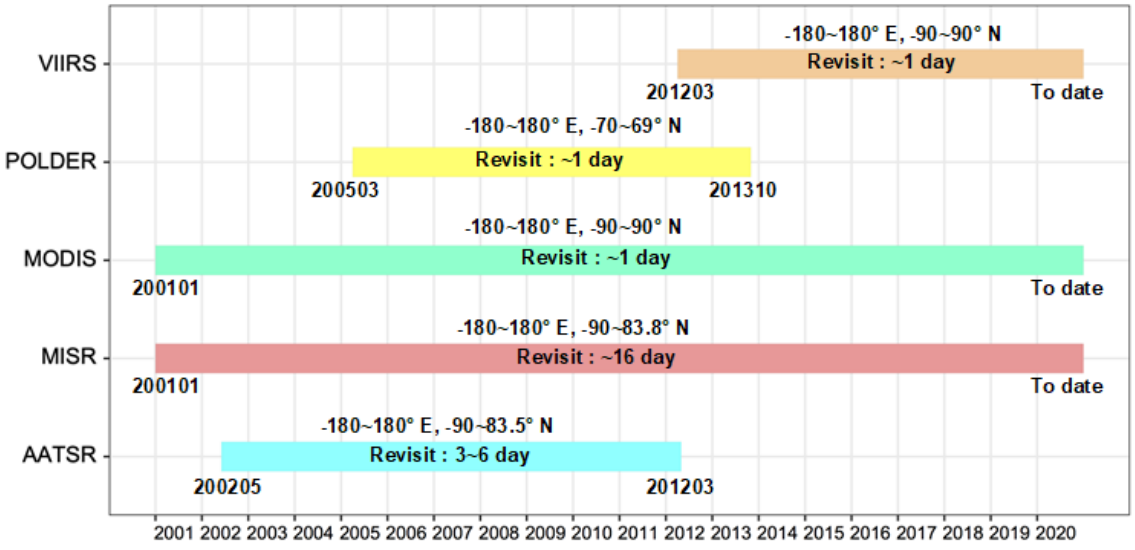


Figure 1: Overview of the time periods covered by different satellites that provide global-scale FMF products. Acronyms used in this figure: AATSR: Advanced Along-Track Scanning Radiometer; MISR: Multi-angle Imaging SpectroRadiometer; MODIS: Moderate Resolution Imaging Spectroradiometer; POLDER: POLarization and Directionality of the Earth's Reflectances; VIIRS: Visible Infrared Imaging Radiometer Suite.

In recent years, deep learning approaches have been applied to satellite satellite-based atmospheric research (Zang et al., 2021b; Yan et al., 2020a; Yuan et al., 2020; Shen et al., 2018; Ong et al., 2016), including FMF retrieval (X. Chen et al., 2020). Compared with classical machine learning methods, deep learning is more capable of approximating nonlinear relationships (Yan et al., 2021c). For example, X. Chen et al. (2020) used a convolutional neural network (CNN) to develop a deep learning model for MODIS FMF retrievals called the Neural Network based AEROsol retrieval (NNAero) method. The NNAero-derived FMF is a significant improvement over the MODIS DT-derived FMF, with the RMSE decreasing from 0.34 (DT) to 0.1567 (NNAero). However, this method has only been applied and validated over northern and eastern China, and not globally. As an important limitation, Zhang et al. (2016) noted that satellite-measured multispectral reflectance of ground-based data alone was not sufficient to retrieve FMFs with high accuracy. This limitation increases the difficulty of preparing training data for the deep learning approach. The physical method may provide a means to alleviate this deficiency, raising the question of whether combining the physical method and deep learning can improve the FMF retrieval accuracy.

To address the above issues, we synergize the advantages of the physical method and deep learning to retrieve aerosol FMFs over land on a global scale using MODIS data. We tested and validated this hybrid model using two decades of data (2001–2020) and produced a new long-term FMF dataset called Phy-DL FMF (physical-deep learning FMF). Contrary to previous studies, the proposed hybrid model considers both physical characteristics and nonlinear relationships to constrain the FMF calculation. This long-term dataset shows good promise for shedding light on the impacts of human activities on atmospheric aerosols, providing a foundation for understanding the variations in fine mode aerosols on a global scale.

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

2 Materials and methods

2.1 MODIS data

The MODIS sensor onboard Terra has provided long-term observations on a global scale every day since February 2000 (Levy et al., 2010), available at the Atmosphere Archive & Distribution System Distributed Active Archive Center. In this study, MODIS C6.1 L1B MOD02SSH data (i.e., top-of-the-atmosphere (TOA) reflectances from Band 1 to Band 7), MODIS C6.1 L3 MOD09CMG data (surface reflectances from Band 1 to Band 7), and MODIS C6.1 L3 MOD08 daily data were obtained from 2001 to 2020 for retrieving FMFs. Table S1 summarizes details about the MODIS data used in this study.

2.2 AERONET data

The AERONET is a worldwide, sun–sky photometer network providing ground-level aerosol properties, recently updated to Version 3 (Holben et al., 1998). To retrieve FMFs from AERONET solar extinction data, (O'neill et al., 2001a; O'neill et al., 2001b; O'neill et al., 2003) developed the Spectral Deconvolution Algorithm (SDA) method. The FMFs based on this inversion method (i.e., SDA FMF) have been included in the standard AERONET data offering, with an estimated uncertainty of 0.1 (O'neill et al., 2001b; O'neill et al., 2003). Since there is not enough Level 2.0 data for use as training data for modelling purposes, here, we used the Level 1.5 SDA FMF dataset generated from data from 1170 global AERONET sites covering the period of 2001 to 2020 as the ground truth for further modelling and validation (Figure S1a). These AERONET sites are spread around the world, enabling the construction of a universal model and allowing for a more thorough validation of the new FMF product.

2.3 Meteorological data

Due to the impact of meteorological factors on FMF (Yan et al., 2021a), five meteorological variables (i.e., 2-m air temperature, planetary boundary layer height, surface pressure, 10-m U/V wind components, and relative humidity) were obtained from ERA5 (Figure S1b-f). ERA5 is the fifth-generation product produced by the European Centre for Medium-Range Weather Forecasts, with hourly data available since 1950 and at a 0.25 °spatial resolution. Given the overpass time and spatial resolution of MODIS data, only monitoring-time meteorological data collected from 10:00 to 11:00 local time were used and resampled to 1 °×1 °to obtain daily averages.

2.4 Combining physical and deep learning models (Phy-DL) for retrieving FMFs

The physical model used in this research is the LUT-SDA (Yan et al., 2017). This method follows the SDA to build a LUT for the satellite-based FMF retrieval. Based on Eq. (1), a set of hypothetical values for the Ångström exponent (AE)

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

derivative (α'), the AE (α), the AE of fine-mode AOD (α_f), and AOD (at 500 nm) are input to the SDA for the FMF calculation (η), thus creating the LUT:

$$\begin{cases} \alpha_f = \frac{1}{2(1-a)} \{(\alpha - \alpha_c - \frac{\alpha' - \alpha_c'}{\alpha - \alpha_c} + b^*) + [(\alpha - \alpha_c - \frac{\alpha' - \alpha_c'}{\alpha - \alpha_c} + b^*)^2 + 4c^*(1-a)]^{1/2}\} + \alpha_c \\ \alpha_f = \frac{\alpha - \alpha_c}{\eta} + \alpha_c \end{cases} \quad (1)$$

where a, b*, c*, α_c', and α_c are fixed parameters described by O'neill (2010). In this study, the MODIS MOD08 DT-based AOD and AE are used as input to the LUT-SDA for the global land physical-model-based FMF (Phy-based FMF) retrieval (Yan et al., 2021b).

The deep learning model used in this study is called EntityDenseNet (Yan et al., 2020). The EntityDenseNet incorporates the Entity Embeddings method (Guo and Berkhahn, 2016) that can directly process spatial or time-based features, such as location, season, and month. It includes one input layer, two hidden layers, and one output layer. Each hidden layer has one fully connected layer, one rectified linear unit (ReLU) layer, one batch normalization (BN) layer, and one dropout layer. The feed-forward operation of each hidden layer can be written as

$$a^{n+1} = BN\{f[W^{n+1} D(a^n) + b^{n+1}]\} \quad (2)$$

where n is the layer number, a^n is the output vector from layer n, D() is the dropout layer for the thinning vector a^n, W^{n+1} and b^{n+1} are weights and biases, respectively, at layer n+1, $f[]$ is the ReLU activation function, and BN is the batch normalization function.

In this study, we combine Phy-based FMF into EntityDenseNet along with satellite measurements and meteorological data to reduce FMF retrieval biases. As shown by Yan et al. (2021b), the global land Phy-based FMF is still unreliable. Due to its unknown and known error sources (e.g., MODIS-derived AE) and nonlinearity in the data itself, a linear model may not be able to correct these errors. In addition, current physical retrieval methods do not use all the information provided by satellite observations for aerosol size information retrievals (Zang et al., 2021). Lipponen et al. (2018) showed that satellite TOA reflectance and geometry data can significantly improve the aerosol size data retrieval accuracy of the machine learning model. Some studies have also suggested that surface reflectance and meteorological factors can also impact the FMF retrieval accuracy (Yan et al., 2021a; X. Chen et al., 2020). Thus, besides Phy-based FMF, we input MODIS TOA reflectance data, geometry data, surface reflectance, and meteorological data into EntityDenseNet for the final Phy-DL FMF calculation (Table S1). In the deep learning model training process, 70%, 20%, and 10% of all input data are randomly separated into groups of data for training, validation, and testing, respectively. The validation data are used for the hyperparameter optimization (node numbers and dropout rate in each hidden layer) of the deep learning model. The testing data are used to evaluate the performance of the trained deep learning model. When the trained model is finally optimized by the validation and testing data, we apply this trained deep learning model to reconstruct global land FMFs for the period of 2001 to 2020.

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

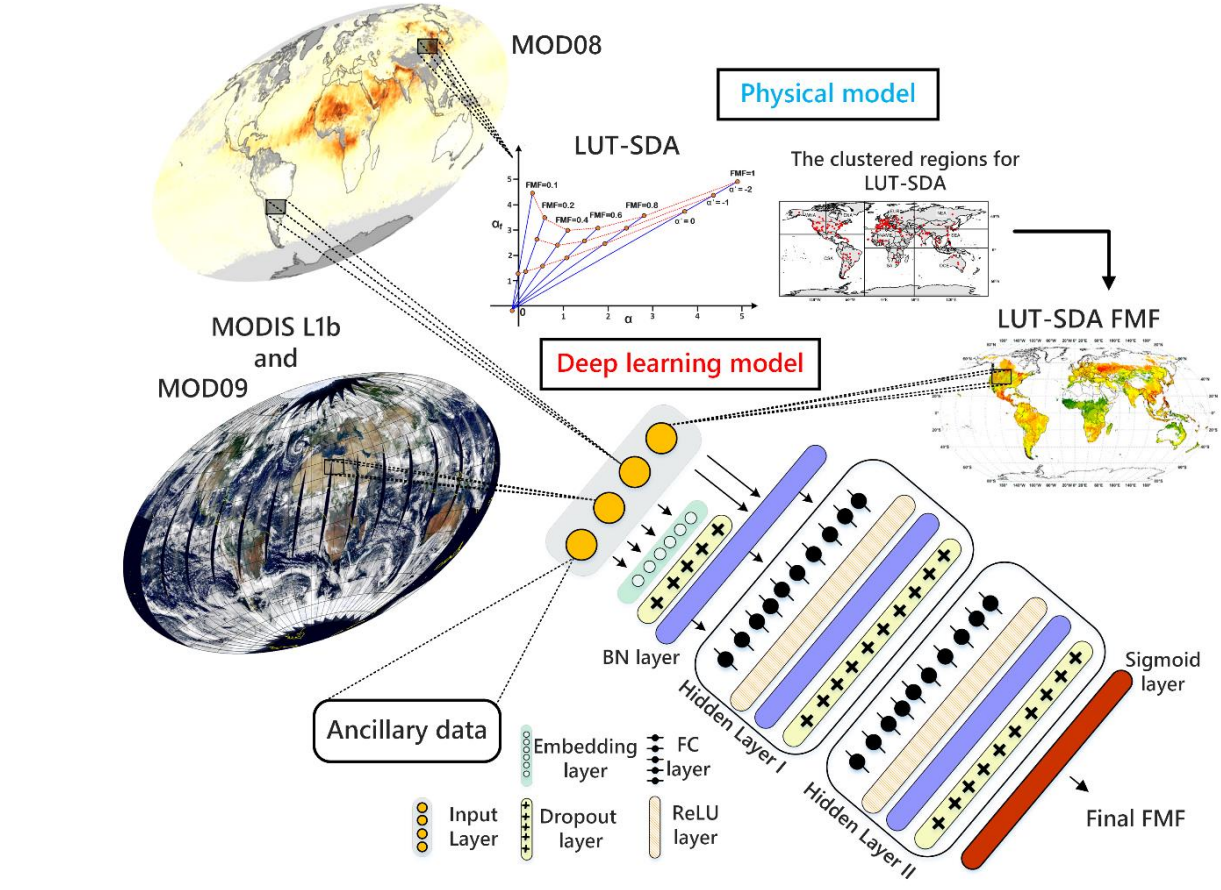


Figure 2: Technical flowchart for the production of the global land FMF product.

145

2.5 Other global FMF products for comparison

Phy-DL-derived FMFs were compared with the following FMF products from three other satellite missions (Table S2):

a. POLDER/GRASP FMF:

Launched in December 2004, POLDER-3 onboard the Polarization and Anisotropy of Reflectances for Atmospheric
150 Sciences coupled with Observations from a Lidar satellite was operational from March 2005 to October 2013, making multi-angular polarization measurements. By capitalizing on the small and fairly neutral polarized reflectances (Deuze et al., 2001), POLDER/GRASP is able to provide the fine-mode AOD (fAOD, radius < 0.35 μm) in two categories: "high-precision" and "models". Because "high-precision" fAODs perform better than "models" fAODs (Wei et al., 2020), we used monthly "high-precision" POLDER/GRASP fAODs and AODs (both at 490 nm) at a spatial resolution of 1 °for calculating FMF (at 490 nm)
155 (FMF=fAOD/AOD).

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021


Open Access Earth System Science Data Discussions

b. Multi-angle Imaging SpectroRadiometer (MISR) FMF:

The MISR instrument onboard the National Aeronautics and Space Administration Earth Observing System Terra satellite has been continuously working since 2000 (Diner et al., 1998; Kahn and Gaitley, 2015). MISR has nine push-broom cameras with different nominal viewing angles, allowing it to distinguish aerosol type, including aerosol size (Garay et al., 2020). The MISR algorithm retrieves small-mode AODs (at 550 nm) due to aerosol particles with radii less than 0.35 μm at a spatial resolution of 0.5 °. We used it to calculated FMF as the ratio of fAOD and AOD for further comparison in this study.

c. MODIS FMF:

The latest C6.1 MODIS aerosol product (Levy et al., 2013) no longer includes global-scale FMFs, so we used FMFs at 550 nm from the previous collection (C5) for comparison purposes (Levy et al., 2007). Although this MODIS FMF product is not reliable over land (Levy et al., 2010), it was used in numerous previous studies (Ramachandran, 2007; Vinoj et al., 2014) including for $PM_{2.5}$ estimations (Li et al., 2016; Zhang and Li, 2015).

3. Results

3.1 Phy-DL FMF validation

Figure 3 shows the validation of the Phy-DL FMFs against AERONET FMFs. By matching 20 years of estimated Phy-DL FMFs with AERONET FMFs (number of match-ups, N = 361,089), we first evaluated the overall performance of Phy-DL FMF (Figure 3a). The correlation coefficient (R) was 0.68, and the root-mean-square error (RMSE) was 0.136. Approximately 90% and 79% of retrievals fell within the expected error (EE) envelopes of ±40% and ±20%, respectively (these envelopes have been adopted from X. Chen et al. (2020)). Figure 3b shows the biases of the Phy-DL FMFs (estimated FMF minus AERONET FMF) as a function of the AERONET FMFs. The Phy-DL FMFs slightly underestimated the FMFs, with a negative median bias in each FMF bin. As each FMF increased, a higher percentage of retrievals fell within the ±20% EE envelope, ranging from 42.85% (when FMF < 0.3) to 91.17% (when FMF > 0.8). This indicates that the Phy-DL FMF retrieval performed better when the fine-mode aerosols dominated. Figure 3c shows the validations of Phy-DL FMFs over different AERONET sites around the world. Most sites in the eastern USA and Europe have over 70% of Phy-DL FMFs falling within the EE envelope of ±20%. Over 90% of Phy-DL FMFs fell within the ±20% EE envelope at some sites in the Amazon, southern Africa, and southeast Asia. However, at coastal AERONET sites in the Caribbean and Mediterranean regions, Australia, and South America, less than 60% of Phy-DL FMFs fell within the ±20% EE envelope. A similar result was found for some AERONET sites near deserts in southern South America, Central Asia, Northwest China, and Central Australia.

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

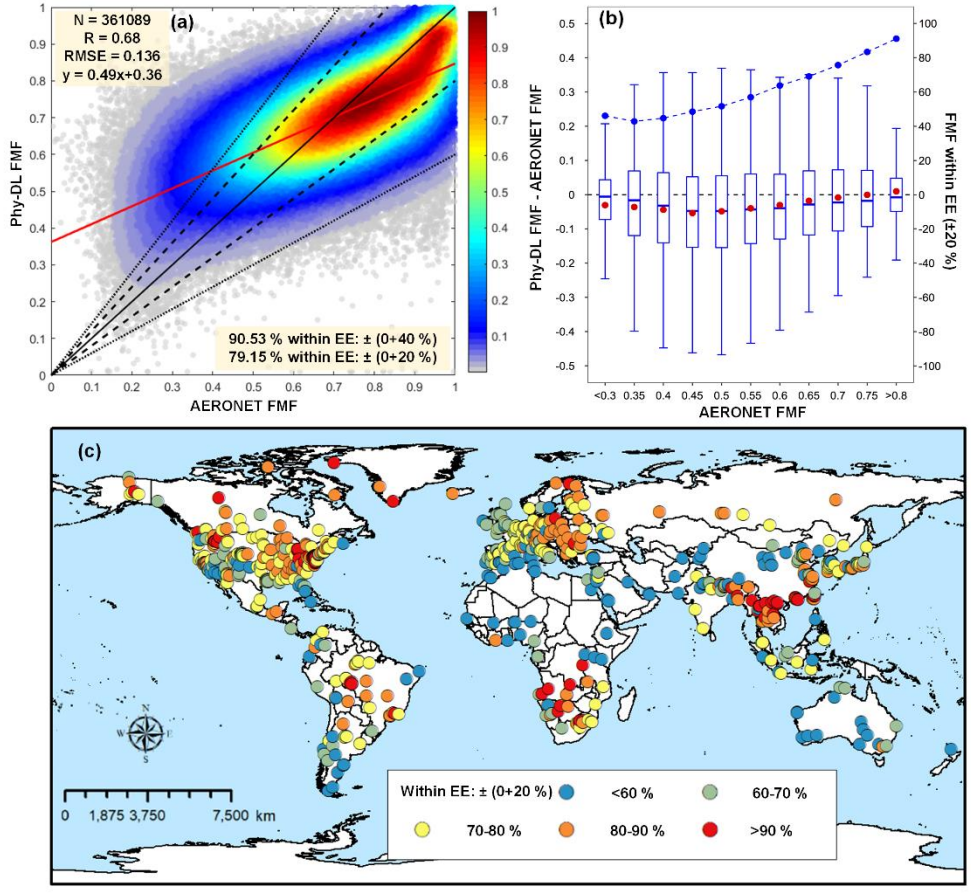


Figure 3: (a) Phy-DL FMF at 500 nm as a function of AERONET FMF. The black and red solid lines are the 1:1 line and the best-fit line obtained from linear regression, respectively. The black dashed and dotted lines represent the expected error (EE) envelopes of ±20% and ±40%, respectively. (b) Box plots of the FMF bias (estimated FMF minus AERONET FMF) as a function of AERONET FMF. The black horizontal dashed line indicates the zero bias. The red dot in each box represents the mean value of the FMF bias. The upper, middle, and lower horizontal lines in each box show the 75th, median, and 25th percentiles, respectively. The blue dots connected by the dashed curve are percentages of FMF retrievals falling within the EE envelope of ±20%. (c) Global distribution of percentages of Phy-DL FMFs falling within the EE envelope of ±20% at the AERONET sites.

Figure S2 shows the frequencies of three FMF levels (low: FMF < 0.5, medium: 0.5 < FMF < 0.8, high: FMF > 0.8, Supplementary section S1) based on Phy-DL and AERONET FMF data from 2001 to 2020. Over 60% of AERONET-derived FMFs were low over Central Asia, Central Australia, and the sub-Sahara; the AOD of these locations was dominated by coarse-mode aerosols (dust). Phy-DL-estimated FMFs were also low over Central Asia and the sub-Sahara, but slightly underestimated over Central Australia (frequency < 40%). Over 90% of both AERONET and Phy-DL FMFs were at the medium level in South America, Western Africa, Australia, Western Asia, and the Western USA. Approximately 45–55% of

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

Phy-DL and AERONET FMFs were at the medium level in the Eastern USA, Europe, and Central Africa. Over 60% of Phy-DL and AERONET FMFs were at a high level in Northern India, Southeast Asia, and Southeast China.

3.2 Global land FMF spatial distribution and trends from 2001 to 2020

Figure 4a shows the global distribution of mean Phy-DL FMFs over land from 2001 to 2020. A high proportion of fine-mode aerosols with FMFs greater than 0.77 can be seen in populated regions, including Southern China, Southeast Asia, Eastern Europe, and the Eastern USA. Low FMF values (< 0.55) were observed in Australia, Northwest China, Central Asia, the Saharan region, southern South America, and the Southeastern USA, where coarse-mode aerosols from large deserts dominate. Figure 4b shows the spatial distributions of the Phy-DL and AERONET FMF linear trends from 2001 to 2020. In general, both datasets show decreasing trends (i.e., $< -3 \times 10^{-3}$ yr^{-1}) in Northeast China, Central Asia, Europe, the Saharan region, Southern Africa, South America, Mexico, and the eastern USA. In contrast, Southeast Asia, India, Central Australia, Central Africa, and the western USA show significant increasing trends of over 3×10^{-3} yr^{-1}. The increasing FMF trend over central Australia is sporadic and could be related to an increase in fire activity (Andela et al., 2017). In South America and Africa, the long-term decrease in burning during the past two decades (Andela et al., 2017; Deeter et al., 2018) has contributed to a significant decrease in FMFs. However, the reduced biomass burning in Central Africa is also partially offset by the dramatic growth in anthropogenic emissions (Zheng et al., 2019), leading to a slight increasing trend in FMF (3×10^{-3} to 7×10^{-3} yr^{-1}). The decreasing FMF trends in Europe and the eastern USA are driven by reduced anthropogenic emissions from transportation sources (Crippa et al., 2016; Jiang et al., 2018). The decreasing FMF trend in Northeast China is likely more associated with a decrease in industrial and residential emissions due to the implementation of clean air policies (Van Der Werf et al., 2017; Yang et al., 2018; Zheng et al., 2019). In the western USA, the dramatically increasing FMF trend is likely partly attributed to the increase in smoke from wildfires (Parks and Abatzoglou, 2020; O'dell et al., 2019). In India, the significant increase in FMF likely reflects an increase in vehicular anthropogenic emissions and crop residue burning (Jethva et al., 2019; Manoj et al., 2019). Figure 4c shows the time series of the global monthly mean Phy-DL and AERONET FMFs from 2001 to 2020. Both time series show similar annual cycles and decreasing trends (i.e., negative slopes). However, only the Phy-DL-estimated FMF decreasing trend was significant (-1.9×10^{-3} yr^{-1} at 95% significance level). This is because the Phy-DL dataset has greater spatial coverage than that of the point-scale AERONET dataset.

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

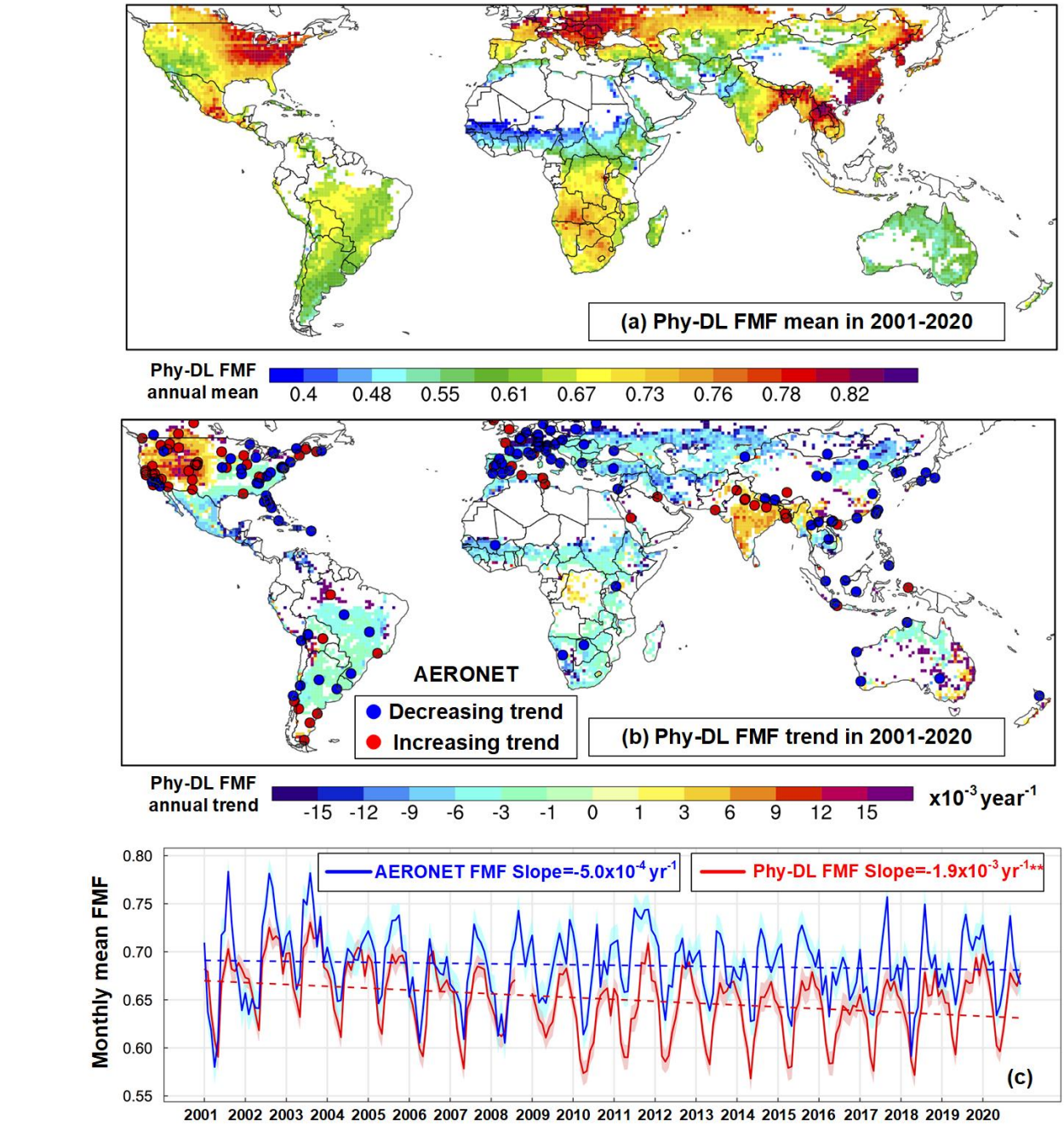


Figure 4: (a) Global distribution of Phy-DL FMF mean values over the 2001–2020 period. Only those pixels with over 120 retrievals yr^{-1} were considered. (b) Global distribution of Phy-DL FMF linear trends from 2001 to 2020. Only those pixels with trends at the 95% significance level were considered. The red and blue dots represent AERONET stations with increasing and decreasing linear trends, respectively, at the 95% significance level. (c) Global monthly mean Phy-DL FMF (red line) and AERONET FMF (blue line). The shaded

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

areas around each line represent the monthly mean FMF value ± 0.1 × the monthly standard deviation. The double-asterisks "**" indicate that the linear trend was at the 95% significance level.

Figure 5 shows the global distributions of seasonal Phy-DL-estimated FMFs from 2001 to 2020. In Central Africa, spring had the lowest FMF, especially in northern Central Africa, due to the transportation of dust (Huebert et al., 2003). Meanwhile, FMFs in summer and autumn were higher than those during winter. This is partly attributed to the high temperature and humidity conditions conducive to the formation of fine-mode aerosols (Tan et al., 2015). In addition, dense vegetation produces a larger amount of submicron aerosols through biomass emissions (Carslaw et al., 2010), which may have also had an effect on increasing FMF values.

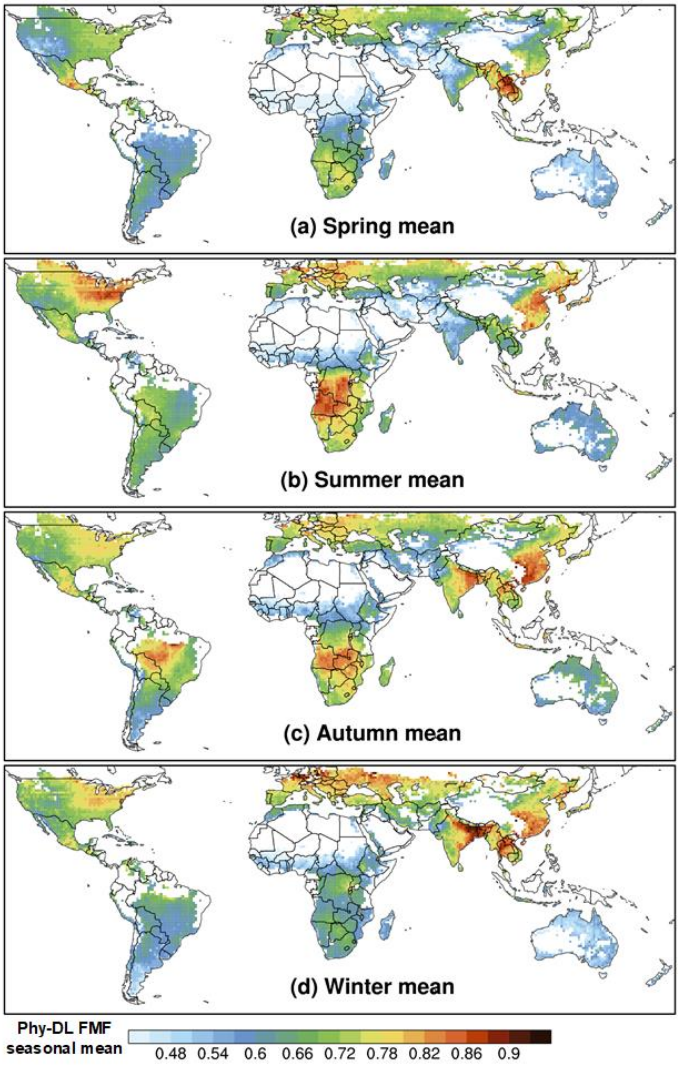


Figure 5: Phy-DL-estimated FMF seasonal mean values from 2001 to 2020. The seasons are defined as spring (March, April, and May), summer (June, July, and August), autumn (September, October, and November), and winter (December, January, and February) for both the

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021


Open Access Earth System Science Data Discussions

Northern Hemisphere and Southern Hemisphere. Only those pixels with 120 retrievals yr^{-1} were considered when calculating the mean values.

In India, FMFs are noticeably higher in autumn and winter, especially in Northern India (i.e., the Indo-Gangetic Plain), where the FMF was greater than 0.87. During spring and summer, FMFs were usually less than 0.63 over India. Mhawish et al. (2021) also reported the same seasonal pattern. This is likely related to spring in India being the pre-monsoon season when dust particles from nearby deserts are frequently transported to the country (Gautam et al., 2009). During that season, the dominant coarse-mode aerosols decrease from west to east over the Indo-Gangetic Plain (Kalapureddy and Devara, 2008), thereby leading to lower FMFs, particularly in the western Indo-Gangetic Plain. In the post-monsoon seasons (autumn and winter), the higher FMF is attributed to the low boundary layer and non-convective atmosphere, which induces haze and stagnant conditions. Frequent biomass burning events also occur during these seasons (Ramachandran, 2007), which contributes to higher FMFs.

In Central Africa and the Amazon, FMFs in summer and autumn were higher than those in spring and winter (seasons here correspond to the seasons in the Northern Hemisphere). This coincides with local biomass burning, which mainly occurs from early summer to the middle of autumn (Generoso et al., 2003; Perez-Ramirez et al., 2017). Accordingly, fine-mode aerosols including black carbon and organic carbon can contribute to higher FMFs in summer and autumn. Although Australia had low FMFs (< 0.6) in all seasons, some sporadic pixels in autumn had FMFs near 0.7, which may also be related to the frequent wildfires in autumn (Shi et al., 2021; Liu et al., 2021).

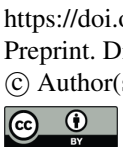
In the eastern USA, FMFs are the highest in summer; however, in the western USA, FMFs are highest in winter. Across the entire USA, FMFs were lowest in spring. In the eastern USA, it is thought that accelerated photochemical reactions and stagnant conditions in summer produce the highest amount of ammonium sulfate in all seasons (Tai et al., 2010). Moreover, ammonium nitrate is the main component of fine-mode particles in the western USA whose content peaks in winter (Hand et al., 2012). This explains why FMF maxima occur in different seasons on both sides of the country.

In Eastern China, summer and autumn had higher FMFs (> 0.8) than those in spring and winter (< 0.78). This is probably because warm seasons with relatively high humidity and temperature can enhance the generation of secondary fine particles by gas-to-particle conversions (Tan et al., 2015). In addition, springtime dust transportation in Northeastern China results in increasing coarse dust particles, thereby affecting the FMF (Huebert et al., 2003). In contrast, Southeastern Asia had exceedingly higher FMFs in winter and spring (> 0.86) than those in summer and autumn (< 0.8), owing to the intense biomass burning from January to April (Yin et al., 2019).

3.3 Comparison between Phy-DL, DL-based, and Phy-based FMFs

To analyze the differences in FMFs obtained by different methods, FMFs generated by the Phy-DL method, deep learning (DL) method (meaning no Phy-based FMF as input), and Phy-based method (i.e., the LUT-SDA) from 2008 to 2017 were compared using AERONET FMFs as the ground truth. Figure 6a shows the three FMF estimations that were averaged

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021



Open Access Earth System Science Data Discussions

into 20 bins with AERONET measurements AOD >0.2 based on the method in Levy et al. (2007). Compared with Phy-based FMF, DL-based FMF has better estimation for low FMF (<0.6), showing the overall improvement in R from 0.51 to 0.60. However, there is still significant underestimation for DL-based FMF when AERONET FMF is greater than 0.6. Phy-DL FMF ameliorated both the underestimation for high FMF values and overestimation for low FMF values, with the highest R (0.81) among three FMFs. The regression equation of Phy-DL FMF also improved tremendously, with smaller intercept and slope closer to 1.

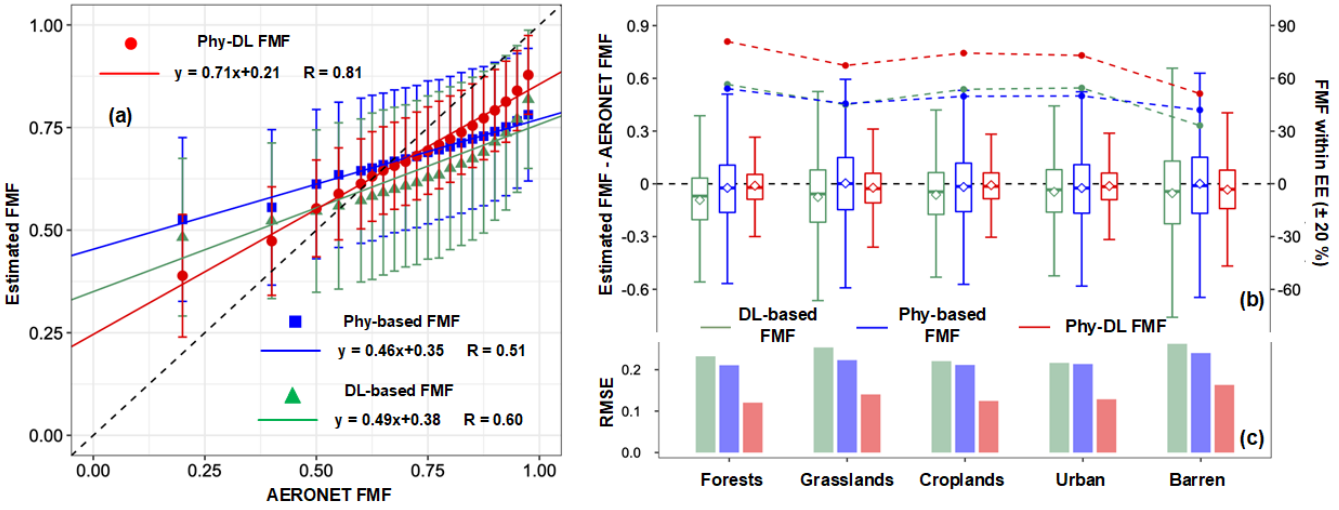


Figure 6: Phy-DL (red), Phy-based (blue), and DL-based FMF (green) estimation compared with AERONET FMFs for AOD > 0.2 (at 500 nm, using data from 2008 to 2017). (a) The dots and the error bars indicate the means and standard deviations of the FMF estimates in 20 bins of AERONET FMF. The solid blue and red lines are the best-fit lines from linear regression. The black dashed line represents the 1:1 line. Linear regression relations and correlation coefficients (R) are given. (b) Boxplots of bias (estimated FMF minus AERONET FMF) and percentage of FMF estimates falling within the EE envelope of ±20 % (dotted, dashed lines) as a function of land type. The upper, middle, and lower lines in each box presents the 75th, median, and 25th percentiles, respectively. The diamond in each box represents the mean value of the FMF bias. (c) The RMSE for each land type against that of the AERONET FMFs.

Figures 6b and 6c compare the accuracy between Phy-DL, Phy-based, and DL-based FMFs over five land types (forests, grasslands, croplands, urban, and barren). The five land types were selected based on MODIS MCD12C1 data from the International Geosphere-Biosphere Programme scheme. Figure 6b shows that Phy-DL FMF had the lowest bias with mean values close to 0, smallest range of bias, and highest FMF retrievals (within ±20% EE) over all land types. Although DL-based FMF had a slightly smaller range of bias and higher FMF retrievals (within ±20% EE) than those of Phy-based FMF over forests, croplands, and urban land types, DL-based FMF still had the largest mean bias and showed the worst performance over barren land types. In addition, the DL-based FMF had the highest RMSE among all the FMFs for all land types. Figure 6b shows that Phy-DL, Phy-based, and DL-based FMFs all had the best performance over forests, with RMSE values of 0.120,

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

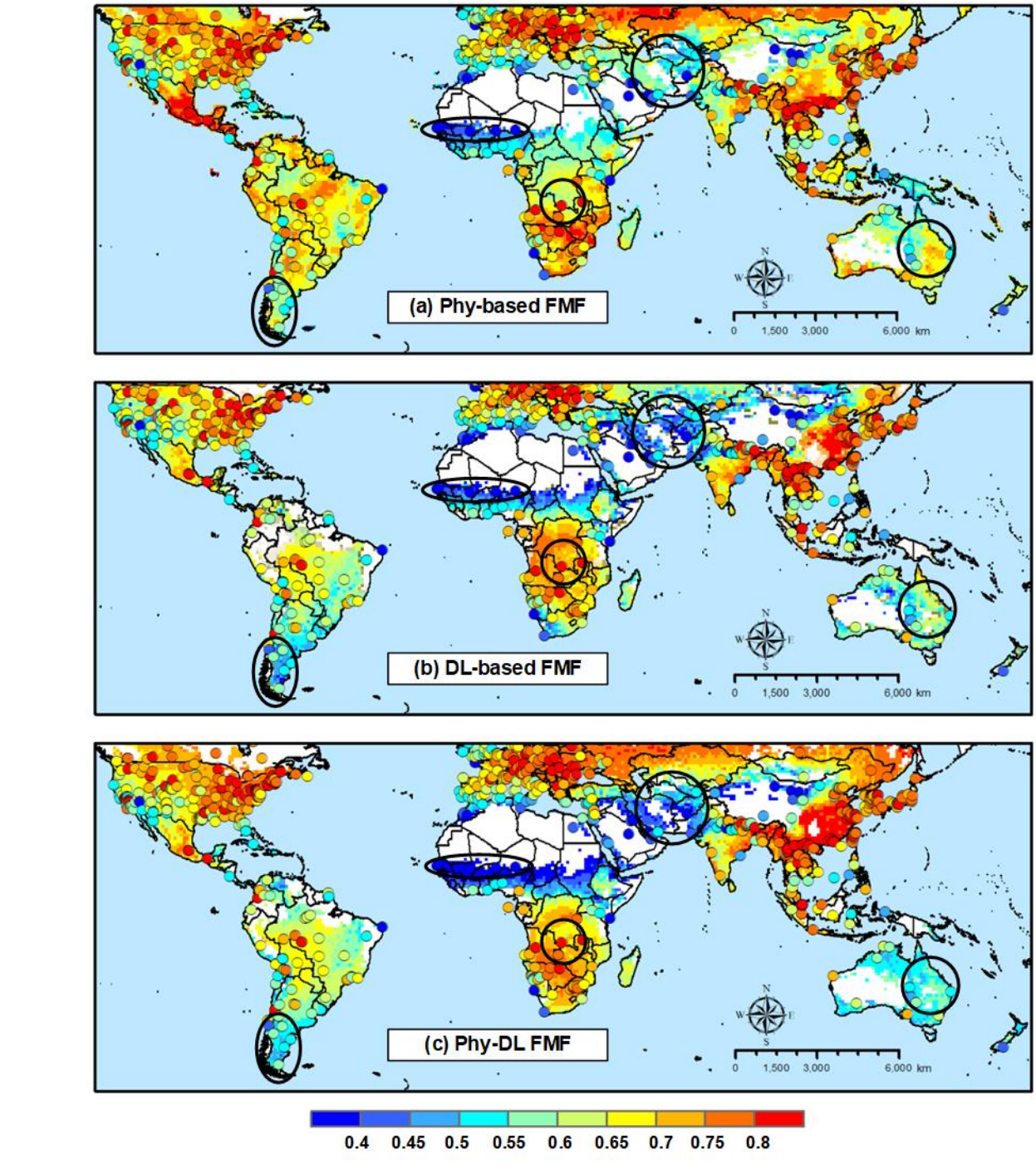

Open Access Earth System Science Data Discussions

0.211, and 0.223, respectively (Figure 6c). Likewise, all performed the worst over barren land, showing a significant negative bias, with less than 50% of the FMF retrievals falling within the ±20% EE envelope. Overall, Phy-DL-estimated FMFs showed a significant improvement over the Phy-based and DL-based FMFs, especially over forests, croplands, and urban land types, where the RMSEs and biases were noticeably reduced.

For further evaluation, Phy-based, DL-based and Phy-DL FMF were validated against AERONET FMFs over AERONET sites to show their spatial performance (Figure S3). DL-based FMF have generally the highest RMSE, with 93.2% sites having RMSE greater than 0.11, compared to 81.0% sites for Phy-based FMF and only 34.8% sites for Phy-DL FMF. Especially, in Australia, India, southern South America, Mediterranean region and North America, DL-based FMF has RMSE dominantly exceeds 0.23 but the RMSE of Phy-based FMF ranges from 0.11~0.23 and Phy-DL FMF is lower than 0.17. Phy-DL FMF performed well in eastern Asia, southern Africa, Europe and eastern US, with RMSE typically lower than 0.11. In contrast, Phy-based FMF in these regions has RMSE greater than 0.11, and DL-based FMF even has large amount of site with RMSE over 0.23. Regarding to R, 69% sites of Phy-DL FMF have R over 0.6, but only 21% sites of Phy-based FMF and 11% sites of DL-based FMF reach R over 0.6. According to Figure S3b, d and f, although DL-based FMF has fewer sites with R less than 0.1 in Europe and North America than Phy-based FMF, there are limited sites for both FMFs in eastern China, India, southeastern Asia, Saharan region and eastern US having high R (>0.6). However, most of sites for Phy-DL FMF achieves this high R.

Figure 7 compares the annual mean FMF from 2008-2017 based on Phy-based, DL-based and Phy-DL FMF estimations. In general, high FMFs (> 0.7) were well captured by both estimation methods over Eastern China, Southeast Asia, Europe, Southern Africa, the eastern USA, and Mexico. However, compared to DL-based and Phy-DL FMF, Phy-based FMF tends to underestimate the hotspots of FMF, such as eastern China and central Africa. While in some regions with comparatively low FMF (<0.55), the estimations also show large differences. For example, in northeast Australia and southern South America, Phy-DL and AERONET FMFs agreed well with values less than 0.55, but Phy-based FMFs were clearly overestimated by ~0.1. In addition, in regions dominated by coarse mode aerosol such as Saharan region and central Asia, only Phy-DL FMF captured this low FMF (<0.45), while Phy-based FMF shows overestimation by ~0.1. DL-based FMF also captured the low FMF in central Asia, yet overestimated the FMF in Saharan region. In central Africa, FMF value is relatively high (>0.7) according to AERONET. Phy-DL and DL-based FMF captured this high value yet Phy-based FMF is greatly underestimated, with FMF less than 0.7. In Australia, only Phy-DL FMF well agreed with AERONET FMF of values less than 0.6, while both DL-based and Phy-based FMF show severely overestimations with FMF values reach over 0.65.

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021



330

Figure 7: Annual mean (a) Phy-DL FMF estimates, (b) Phy-based FMF estimates, and (c) DL-based FMF estimates. The colored dots in (a),(b), and (c) show annual mean AERONET FMFs. Areas outlined in black show regions with noticeably large differences in the FMF estimates. Only those pixels with over 120 retrievals yr^{-1} were considered. Data from 2008 to 2017 were averaged.

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

Open Access Earth System Science Data Discussions

335 **3.4 Comparison with other satellite-based FMF products**

Figure 8a–d shows the performance of MISR, POLDER, MODIS, and Phy-DL FMFs against AERONET FMFs. Because these three satellite FMF products cover different yet overlapping time ranges, we only compare retrievals made from 2008 to 2013 when all products were available. The Phy-DL FMF performed the best, with R and RMSE values of 0.78 and 0.100, respectively. In addition, 96.31% (84.74%) of Phy-DL FMFs fell within the EE envelope of ±40% (±20%), an improvement 340 over other FMF products. The next best-performing FMFs were from POLDER and MISR. POLDER shows R and RMSE values of 0.48 and 0.233, respectively, and 76.05% (46.99%) of the retrievals falling within the EE envelope of ±40% (±20%), while MISR FMF has R and RMSE of 0.42 and 0.204, and 85.01% (45.85%) retrievals between the EE envelop of ±40% (±20%). Both POLDER and MISR FMFs were underestimated compared to AERONET FMFs, especially when the AERONET FMF was greater than 0.6. In contrast, MODIS FMFs were overestimated compared to AERONET FMFs, 345 especially for AERONET FMFs greater than 0.6, where MODIS FMFs reached values near 1. The overall performance of MODIS FMFs was also the worst, with R and RMSE values of 0.37 and 0.282, respectively, and 68.88% (44.48%) of the retrievals falling within the EE envelope of ±40% (±20%). Figure 8e shows the probability density functions (PDFs) of the FMF biases (estimated FMF minus AERONET FMF). The Phy-DL PDF reveals that most of the biases were close to zero, suggesting the robustness of the Phy-DL method. MISR and POLDER PDFs show underestimations, with most of the biases 350 near -0.2 and -0.1, respectively. The MODIS PDF shows overestimations with biases concentrated near 0.05. Overall, compared with AERONET FMFs, of the four FMF products, the Phy-DL-estimated FMFs agreed the best. Figure S4 shows the global distributions of RMSE from validations of MISR, POLDER, MODIS, and Phy-DL FMFs against AERONET FMFs at the AERONET sites. Concerning MISR FMFs, 47.9% of the sites had RMSEs higher than 0.23, and 5.3% of the sites had RMSEs lower than 0.11, showing the worst performance. Concerning POLDER FMFs, 29.7% of the sites had RMSEs higher 355 than 0.23, mainly in the USA, the Amazon, Southern Africa, Western Europe, and Southeast Asia. MODIS FMFs performed well in Eastern China, India, Europe, and the eastern USA, with 40.0% of the sites having RMSEs lower than 0.11. Concerning Phy-DL FMFs, 65.2% of the sites had RMSEs lower than 0.11. In addition, the number of match-ups of Phy-DL-estimated and AERONET FMFs was the highest (N = 566), indicating a higher data coverage compared with the other FMF products. In terms of R (Figure S5), at 82.2% of the AERONET sites, R for MISR FMFs was less than 0.2 (Figure S5a before S5b). At 360 33.8% of the AERONET sites, mainly in Eastern China, India, and Australia, R for MODIS FMFs was greater than 0.5, but at most sites in the USA and Europe, R was less than 0.2 (Figure S5a). At 39.7% of the AERONET sites, R for POLDER FMFs was greater than 0.5 in Europe, the Amazon, and Eastern China, but at most sites in the USA, India, and Australia, R was less than 0.2 (Figure S5c). The R for Phy-DL FMFs was greater than 0.5 at 79.0% of the AERONET sites, agreeing better with AERONET FMFs than did POLDER and MODIS FMFs in the USA, Africa, Southeast Asia, and Europe (Figure S5d).

365

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

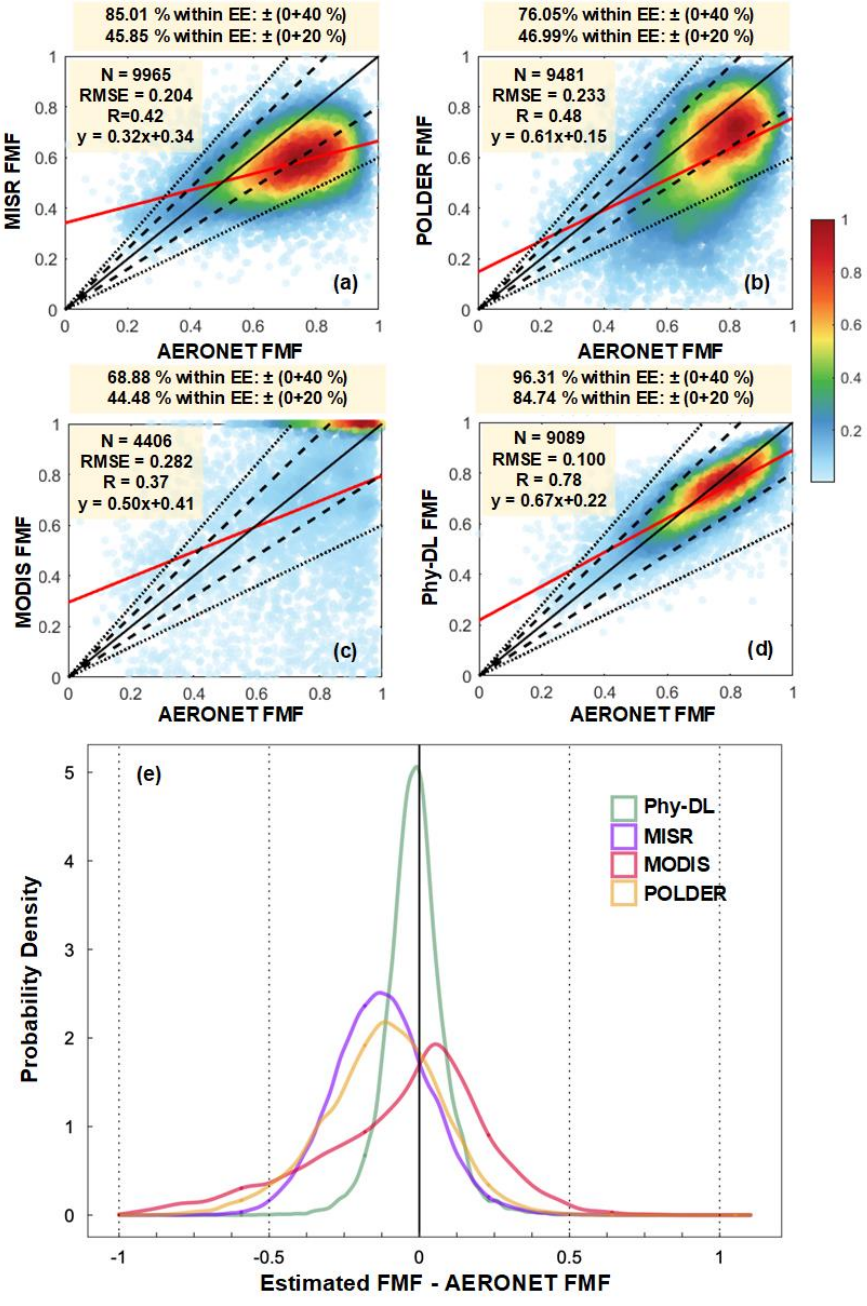


Figure 8: Evaluation of (a) MISR (550 nm), (b) POLDER (490 nm), (c) MODIS (550 nm), and (d) Phy-DL FMFs (500 nm) against AERONET FMFs (500 nm) from 2008 to 2013. Black and red solid lines are 1:1 reference lines and best-fit lines from linear regression, respectively. Black dashed and dotted lines represent the EE envelopes of ±20% and ±40%, respectively. The number of samples (N), RMSE, correlation coefficient (R), and linear regression relation are given in each panel. (e) Probability density functions of the FMF bias (estimated FMF minus AERONET FMF) for MISR (blue), POLDER (orange), MODIS (red), and Phy-DL (green) FMFs.

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021


Open Access Earth System Science Data Discussions

Figure 9 compares the spatial distributions of annual mean MISR-, MODIS-, POLDER-, and Phy-DL-estimated FMFs from 2008 to 2013. In general, Phy-DL FMFs were higher than the satellite-based FMFs over areas of known biomass burning and urban areas, including the eastern USA, the Amazon, Southern Africa, Eastern China, and Australia. Phy-DL and AERONET FMFs in Eastern China reached over 0.7, while POLDER, MISR, and MODIS FMFs were significantly underestimated (~0.6–0.7, ~0.5–0.6, and generally < 0.4, respectively). In the western USA, Phy-DL and AERONET FMFs were higher than 0.6, but MODIS FNFs were < 0.4, and MISR and POLDER FMFs were < 0.6. In Central Africa, POLDER, Phy-DL, and AERONET FMFs were similar (> 0.7), but MISR FMFs ranged from 0.6 to 0.7, and MODIS FMFs exceeded 0.8. In Australia and the Amazon, Phy-DL and MISR FMFs agreed well with AERONET FMFs (0.5–0.6 for Australia and ~0.6–0.7 for the Amazon), but POLDER and MODIS FMFs (< 0.4) were significantly underestimated compared with AERONET FMFs. Figure S6 shows the bias, the percentage of FMF retrievals falling within the EE envelope of ±20%, and the RMSEs of MISR, MODIS, POLDER, and Phy-DL FMFs over five land types (forests, grasslands, croplands, urban, and barren), using data from 2008 to 2013. Over all land types considered and compared with the satellite-based retrievals, Phy-DL FMFs had biases nearest to 0, a higher percentage of FMFs falling within the EE envelope (> 67%), and the lowest RMSE (< 0.127). Both POLDER and MISR FMFs had significant negative biases of -0.2 and -0.1, respectively, over all land types. MODIS FMFs had significant positive biases over forests and grasslands and negative biases over croplands, urban areas, and barren areas. Over forests, grasslands, croplands, and urban areas, MODIS FMFs had the largest RMSE (> 0.280), and MISR FMFs had the lowest percentage of FMFs falling within the EE envelope (< 40%). Over barren land and of all FMF products, POLDER FMFs were the poorest (23.68% of the FMFs falling within the EE envelope, and RMSE = 0.326).

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

Open Access Earth System Science Data Discussions

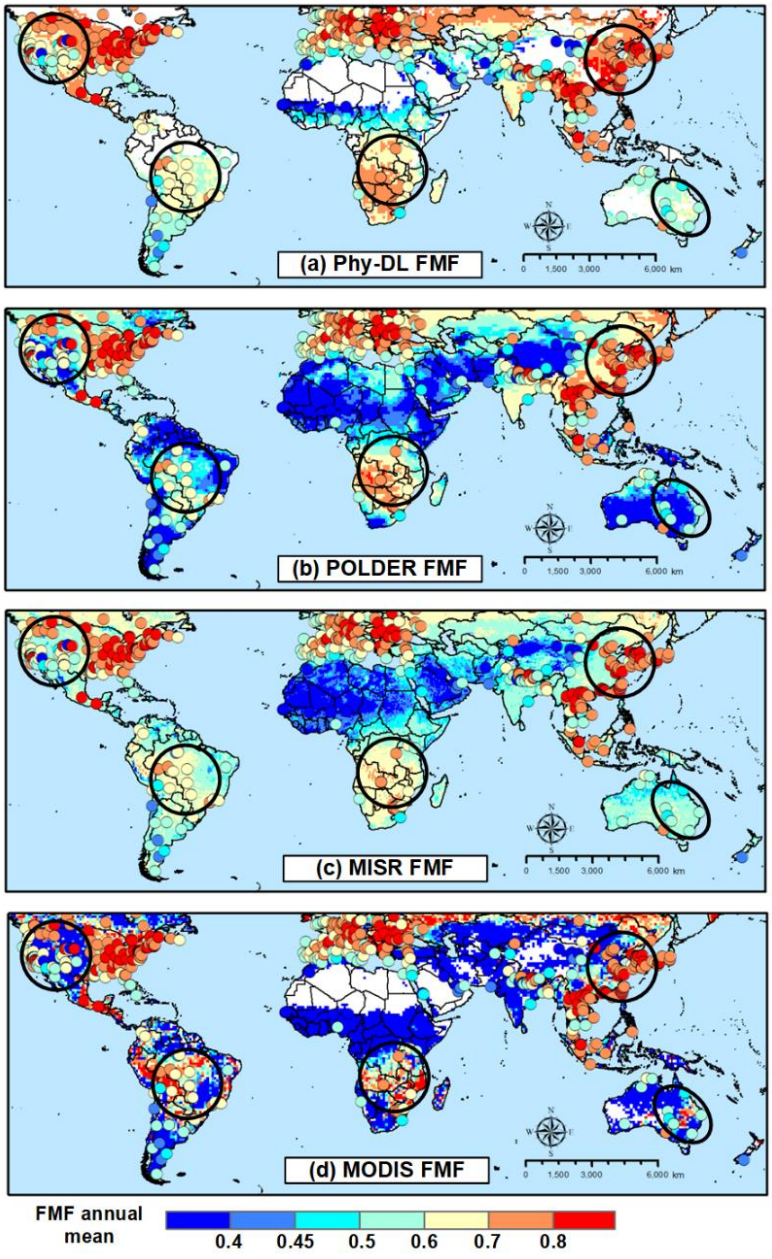


Figure 9: Annual mean FMFs based on (a) Phy-DL, (b) POLDER, (c) MISR, and (d) MODIS. The colored dots show annual mean AERONET FMFs. Areas outlined in black circles show regions with noticeably large differences in the FMF estimates. Only those pixels with over 120 retrievals yr^{-1} were considered in the Phy-DL estimation. Data from 2008 to 2013 were averaged.

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

Next, we conducted a comprehensive comparison of these satellite-based FMF products over Central Africa. Regarding FMF annual mean values (Figure 10a–d), the POLDER and Phy-DL FMFs agreed the best with AERONET FMFs, which captured the high values in the middle part of Central Africa (> 0.76) and the low values along the coasts (< 0.7). Although MISR FMFs also captured the low FMFs over coastal regions, FMFs were underestimated in the interior (< 0.7). However, MODIS FMFs were significantly overestimated along the western coast (> 0.85) and underestimated in the southeastern part of Central Africa (< 0.4). Linear trends were also calculated for all the FMF products (Figure 10e–h). Note that only the linear trends significant at the 95% level were examined. AERONET showed a significant increasing trend in the northern part of Central Africa (+0.01 yr^{-1}) and a decreasing trend in the southern region (-0.01 yr^{-1}). Of all the FMF products, Phy-DL FMF trends agreed best with AERONET FMF trends. POLDER and MODIS FMF trends were greatly enhanced in the southern region (+0.05 yr^{-1}), while MISR FMF trends did not reflect the AERONET FMF trends well. Overall, Figure 10 illustrates that in Central Africa, compared with the three satellite-based FMF products, Phy-DL FMFs are more accurate and reliable.

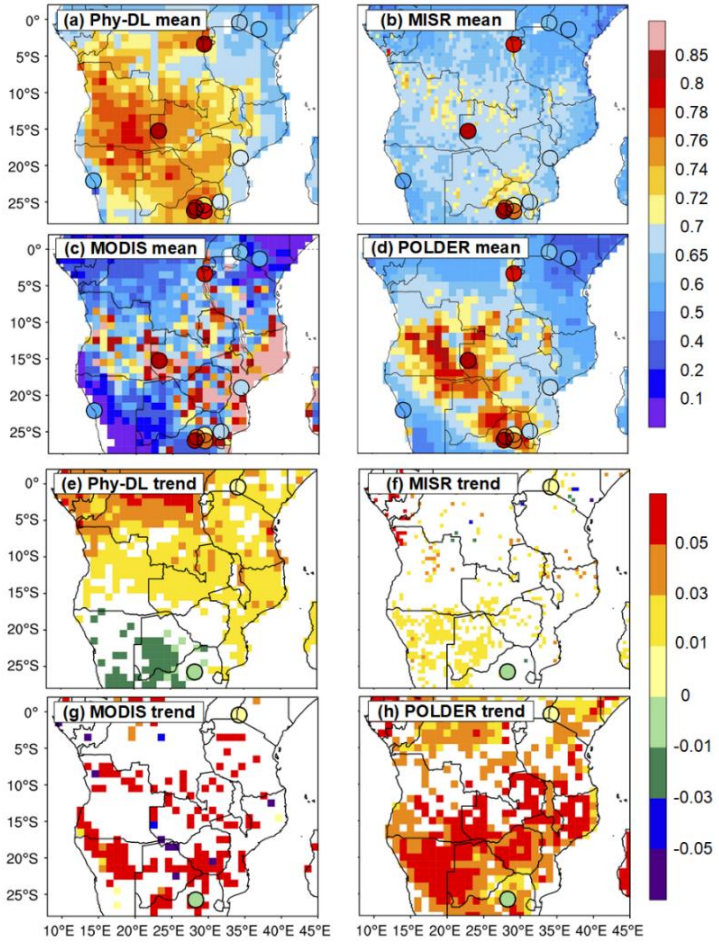


Figure 10: (a–d) Spatial distributions of annual mean FMF (averaged from 2008 to 2013) over Central Africa based on Phy-DL, MISR, MODIS, and POLDER. The colored dots show annual mean AERONET FMFs. (e–h) Spatial distributions of the FMF linear trend from

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

2008 to 2013 over Central Africa based on Phy-DL, MISR, MODIS, and POLDER. The colored dots show linear trends at the AERONET sites. Only pixels and dots with linear trends at the 95% significance level are shown.

To compare seasonal differences between the evaluated methods, Figure 11 compares the seasonal mean MISR-, MODIS-, POLDER-, and Phy-DL-estimated FMFs from 2008 to 2013, and Figure S7 shows their differences (i.e., satellite estimates minus Phy-DL estimates). In all seasons, Phy-DL FMFs were generally higher than MISR FMFs over urban areas and regions where biomass burning was prevalent, such as the USA, Eastern China, and India. During fine-mode-particle-dominated seasons (FMF > 0.8), such as summertime for the eastern USA and wintertime for Eastern China and India, differences between Phy-DL and MISR FMFs reached < -0.18. MODIS FMFs (< 0.2) were much lower than Phy-DL FMFs (> 0.6) in sub-Saharan Africa, India, China, Australia, and the western USA in all four seasons, with differences < -0.5. Conversely, during winters in the Amazon and Central Africa, MODIS FMFs (> 0.74) were slightly higher than Phy-DL FMFs (~0.66); the differences of POLDER FMFs (~0.2) were globally lower than Phy-DL FMFs in all four seasons. In the eastern USA during autumn and winter, POLDER FMFs were < 0.2 and Phy-DL FMFs were > 0.6, resulting in large differences (< -0.4). Figure S8 shows MISR-, MODIS-, POLDER-, and Phy-DL-estimated FMF frequencies at three levels (low: FMF < 0.5, medium: 0.5 < FMF < 0.8, high: FMF > 0.8) from 2008 to 2013. In the low-level category, MODIS and POLDER FMFs were more frequent than AERONET FMFs (50% and 20%, respectively), especially over the Amazon and western USA. The frequencies of MISR, Phy-DL, and AERONET FMFs in this category were in good agreement. In the medium-level category, high frequencies of Phy-DL and AERONET FMFs occurred over Australia and the Amazon (> 80%), and low frequencies of Phy-DL and AERONET FMFs occurred in sub-Saharan Africa, Central Africa, and Eastern China (< 30%). The MISR slightly overestimated the frequency of medium-level FMFs in Central Africa and underestimated it in Northern Australia and the Amazon. The frequencies of medium-level POLDER FMFs were underestimated over the Amazon and western USA and overestimated over Southeast China. MODIS was unable to capture medium-level FMFs globally, with frequencies of < 20%. High-level FMFs mainly appeared over areas experiencing biomass burning and urban regions, with frequencies commonly < 50%. The frequencies of MODIS, Phy-DL, and AERONET FMFs in the high-level category over Central Africa, Southern China, and the eastern USA agreed well. However, the frequencies of high-level MODIS FMFs were overestimated over the Amazon and underestimated over Northern India. The frequencies of high-level POLDER FMFs were captured well over Central Africa, but significantly underestimated over Northern India, Southern China, and the eastern USA. Moreover, MODIS was unable to capture high-level FMFs globally with frequencies of < 20%.

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

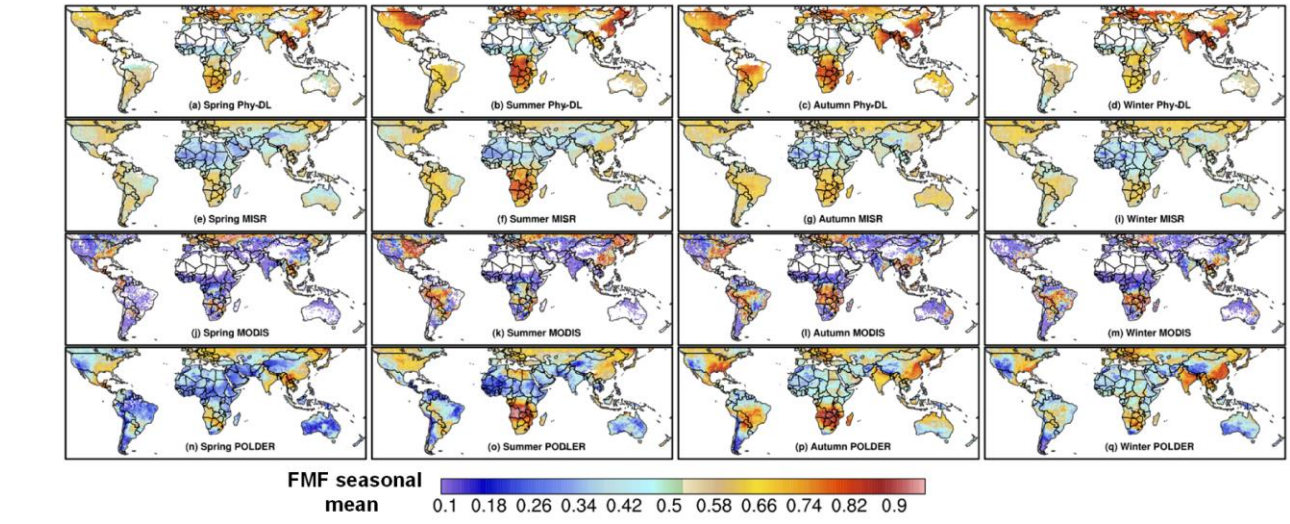


Figure 11: Seasonal mean FMF, averaged from 2008 to 2013, based on (from top to bottom) Phy-DL, MISR, MODIS, and POLDER. Columns from left to right are for spring, summer, autumn, and winter.

4. Data availability

The global land FMF dataset (2001–2020) developed in this study, Phy-DL FMF, is available at https://doi.org/10.5281/zenodo.5105617. The FMF data are in the Geotiff format on a daily scale.

5. Conclusion

This study developed an improved long-term global aerosol FMF (at 500 nm) dataset (2001–2020) over land by combining physical and deep learning approaches called Phy-DL FMF. This new dataset was extensively evaluated against AERONET FMF retrievals, revealing its higher accuracy (RMSE = 0.136, based on 361,089 validation samples; 79.15% of the data fell within the ±20 % EE envelope) and generally good agreement with AERONET FMFs with respect to mean values, trends, and frequencies. Compared with physical–based (calculated using LUT-SDA, i.e., Phy-based FMF) and deep learning-based (DL-based FMF) FMF results, the accuracy of Phy-DL FMFs was substantially improved over five land types (forests, grasslands, croplands, urban area, and barren land), thereby showing less underestimations for high FMF values and less overestimations for low FMF values. Geographically, Phy-DL FMFs captured the low FMFs well over the Saharan region, Central Asia, Australia, and southern South America, while Phy-based FMFs showed significant overestimations. Phy-DL FMFs were also compared with three satellite-based official global FMF products (MISR, POLDER, and MODIS DT-based

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

FMFs). Phy-DL FMFs showed a significant improvement in terms of the accuracy and spatial distribution of trends. In Central Africa, Eastern China, Australia, the Amazon, and the western USA, Phy-DL FMFs agreed well with AERONET FMFs, while the other three satellite-based FMFs showed significant underestimations. In particular, in Southern Africa, the accuracy of the annual average was substantially improved, and the linear trends of Phy-DL FMFs corresponded better with AERONET FMFs. The Phy-DL FMF dataset also captured the seasonality and frequencies of FMFs well, thereby showing better agreement with AERONET FMFs.

By examining Phy-DL FMFs from 2001 to 2020, we found a significant decreasing trend of -1.9×10^{-3} yr^{-1} at the significance level of 95%, which was not revealed by AERONET point-scale measurements. However, both Phy-DL and AERONET FMFs showed significant increasing trends in FMF over the western USA and India ($> +3 \times 10^{-3}$ yr^{-1}). The new dataset captured high-level FMFs (> 0.80) over Southern China, South Asia, Eastern Europe, and the eastern USA. FMFs were consistently < 0.3 in Northwest China, the Saharan region, and southern South America, indicating coarse-particle desert emissions. The promising results generated from the Phy-DL FMF dataset suggest that this newly developed dataset may be useful for investigating the impact of fine- and coarse-mode aerosols on the atmospheric environment and climate, thus deepening our understanding of global aerosol size variations.

Competing interests.

The authors declare that they have no conflict of interest.

Acknowledgements.

The authors gratefully acknowledge the MODIS, MISR, POLDER and AERONET teams for their effort in making the data available.

Financial support.

This work was supported by the National Natural Science Foundation of China (42030606, 41801329 and 91837204), the National Key Research and Development Plan of China (2017YFC1501702), the Open Fund of State Key Laboratory of Remote Sensing Science (OFSLRSS201915) and the Fundamental Research Funds for the Central Universities.

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

References

Andela, N., Morton, D. C., Giglio, L., Chen, Y., van der Werf, G. R., Kasibhatla, P. S., DeFries, R. S., Collatz, G. J., Hantson, S., Kloster, S., Bachelet, D., Forrest, M., Lasslop, G., Li, F., Mangeon, S., Melton, J. R., Yue, C., and Randerson, J. T.: A human-driven decline in global burned area, Science, 356, 1356-1361, 10.1126/science.aal4108, 2017.

Anderson, T. L., Wu, Y. H., Chu, D. A., Schmid, B., Redemann, J., and Dubovik, O.: Testing the MODIS satellite retrieval of aerosol fine-mode fraction, Journal of Geophysical Research-Atmospheres, 110, 10.1029/2005jd005978, 2005.

Bellouin, N., Boucher, O., Haywood, J., and Reddy, M. S.: Global estimate of aerosol direct radiative forcing from satellite measurements, Nature, 438, 1138-1141, 10.1038/nature04348, 2005.

Carslaw, K. S., Boucher, O., Spracklen, D. V., Mann, G. W., Rae, J. G. L., Woodward, S., and Kulmala, M.: A review of natural aerosol interactions and feedbacks within the Earth system, Atmospheric Chemistry and Physics, 10, 1701-1737, 10.5194/acp-10-1701-2010, 2010.

Chen, C., Dubovik, O., Fuertes, D., Litvinov, P., Lapyonok, T., Lopatin, A., Ducos, F., Derimian, Y., Herman, M., Tanre, D., Remer, L. A., Lyapustin, A., Sayer, A. M., Levy, R. C., Hsu, N. C., Descloitres, J., Li, L., Torres, B., Karol, Y., Herrera, M., Herreras, M., Aspetsberger, M., Wanzenboeck, M., Bindreiter, L., Marth, D., Hangler, A., and Federspiel, C.: Validation of GRASP algorithm product from POLDER/PARASOL data and assessment of multi-angular polarimetry potential for aerosol monitoring, Earth System Science Data, 12, 3573-3620, 10.5194/essd-12-3573-2020, 2020a.

Chen, X., de Leeuw, G., Arola, A., Liu, S., Liu, Y., Li, Z., and Zhang, K.: Joint retrieval of the aerosol fine mode fraction and optical depth using MODIS spectral reflectance over northern and eastern China: Artificial neural network method, Remote Sensing of Environment, 249, 10.1016/j.rse.2020.112006, 2020b.

Choi, M., Kim, J., Lee, J., Kim, M., Park, Y.-J., Jeong, U., Kim, W., Hong, H., Holben, B., Eck, T. F., Song, C. H., Lim, J.-H., and Song, C.-K.: GOCI Yonsei Aerosol Retrieval (YAER) algorithm and validation during the DRAGON-NE Asia 2012 campaign, Atmospheric Measurement Techniques, 9, 1377-1398, 10.5194/amt-9-1377-2016, 2016.

Crippa, M., Janssens-Maenhout, G., Dentener, F., Guizzardi, D., Sindelarova, K., Muntean, M., Van Dingenen, R., and Granier, C.: Forty years of improvements in European air quality: regional policy-industry interactions with global impacts, Atmospheric Chemistry and Physics, 16, 3825-3841, 10.5194/acp-16-3825-2016, 2016.

Deeter, M. N., Martinez-Alonso, S., Andreae, M. O., and Schlager, H.: Satellite-Based Analysis of CO Seasonal and Interannual Variability Over the Amazon Basin, Journal of Geophysical Research-Atmospheres, 123, 5641-5656, 10.1029/2018jd028425, 2018.

Deuze, J. L., Breon, F. M., Devaux, C., Goloub, P., Herman, M., Lafrance, B., Maignan, F., Marchand, A., Nadal, F., Perry, G., and Tanre, D.: Remote sensing of aerosols over land surfaces from POLDER-ADEOS-1 polarized measurements, Journal of Geophysical Research-Atmospheres, 106, 4913-4926, 10.1029/2000jd900364, 2001.

Diner, D. J., Beckert, J. C., Reilly, T. H., Bruegge, C. J., Conel, J. E., Kahn, R. A., Martonchik, J. V., Ackerman, T. P., Davies, R., Gerstl, S. A. W., Gordon, H. R., Muller, J. P., Myneni, R. B., Sellers, P. J., Pinty, B., and Verstraete, M. M.: Multi-angle Imaging SpectroRadiometer (MISR) - Instrument description and experiment overview, Ieee Transactions on Geoscience and Remote Sensing, 36, 1072-1087, 10.1109/36.700992, 1998.

Dubovik, O., Herman, M., Holdak, A., Lapyonok, T., Tanré, D., Deuzé, J. L., Ducos, F., Sinyuk, A., and Lopatin, A.: Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations, Atmos. Meas. Tech., 4, 975–1018, 10.5194/amt-4-975-2011, 2011.

Dubovik, O., Lapyonok, T., Litvinov, P., Herman, M., Fuertes, D., Ducos, F., Lopatin, A., Chaikovsky, A., Torres, B., Derimian, Y., Huang, X., Aspetsberger, M., and Federspiel, C.: GRASP: a versatile algorithm for characterizing the atmosphere, SPIE Newsroom, 25, 2.1201408, https://doi.org/10.1117/2.1201408.005558, 2014.

Dubovik, O., Li, Z., Mishchenko, M. I., Tanré, D., Karol, Y., Bojkov, B., Cairns, B., Diner, D. J., Espinosa, W. R., Goloub, P., Gu, X., Hasekamp, O., Hong, J., Hou, W., Knobelspiesse, K. D., Landgraf, J., Li, L., Litvinov, P., Liu, Y., Lopatin, A., Marbach, T., Maring, H., Martins, V., Meijer, Y., Milinevsky, G., Mukai, S., Parol, F., Qiao, Y., Remer, L., Rietjens, J., Sano, I., Stammes, P., Stamnes, S., Sun, X., Tabary, P., Travis, L. D., Waquet, F., Xu, F., Yan, C., and Yin, D.: Polarimetric remote sensing of atmospheric aerosols: Instruments, methodologies, results, and perspectives, Journal of Quantitative Spectroscopy and Radiative Transfer, 224, 474-511, /10.1016/j.jqsrt.2018.11.024, 2019.

Garay, M. J., Witek, M. L., Kahn, R. A., Seidel, F. C., Limbacher, J. A., Bull, M. A., Diner, D. J., Hansen, E. G., Kalashnikova, O. V., Lee, H., Nastan, A. M., and Yu, Y.: Introducing the 4.4 km spatial resolution Multi-Angle Imaging

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

SpectroRadiometer (MISR) aerosol product, Atmospheric Measurement Techniques, 13, 593-628, 10.5194/amt-13-593-2020, 2020.

Gautam, R., Liu, Z., Singh, R. P., and Hsu, N. C.: Two contrasting dust-dominant periods over India observed from MODIS and CALIPSO data, Geophysical Research Letters, 36, 10.1029/2008gl036967, 2009.

Generoso, S., Breon, F. M., Balkanski, Y., Boucher, O., and Schulz, M.: Improving the seasonal cycle and interannual variations of biomass burning aerosol sources, Atmospheric Chemistry and Physics, 3, 1211-1222, 10.5194/acp-3-1211-2003, 2003.

Guo, C. and Berkhahn, F.: Entity Embeddings of Categorical Variables, 2016.

Hand, J. L., Schichtel, B. A., Pitchford, M., Malm, W. C., and Frank, N. H.: Seasonal composition of remote and urban fine particulate matter in the United States, Journal of Geophysical Research-Atmospheres, 117, 10.1029/2011jd017122, 2012.

Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET - A federated instrument network and data archive for aerosol characterization, Remote Sensing of Environment, 66, 1-16, 10.1016/s0034-4257(98)00031-5, 1998.

Huebert, B. J., Bates, T., Russell, P. B., Shi, G. Y., Kim, Y. J., Kawamura, K., Carmichael, G., and Nakajima, T.: An overview of ACE-Asia: Strategies for quantifying the relationships between Asian aerosols and their climatic impacts, Journal of Geophysical Research-Atmospheres, 108, 10.1029/2003jd003550, 2003.

Jethva, H., Torres, O., Field, R. D., Lyapustin, A., Gautam, R., and Kayetha, V.: Connecting Crop Productivity, Residue Fires, and Air Quality over Northern India, Scientific Reports, 9, 10.1038/s41598-019-52799-x, 2019.

Jiang, Z., McDonald, B. C., Worden, H., Worden, J. R., Miyazaki, K., Qu, Z., Henze, D. K., Jones, D. B. A., Arellano, A. F., Fischer, E. V., Zhu, L., and Boersma, K. F.: Unexpected slowdown of US pollutant emission reduction in the past decade, Proceedings of the National Academy of Sciences of the United States of America, 115, 5099-5104, 10.1073/pnas.1801191115, 2018.

Huebert, B.J., Bates, T., Russell, P.B., et al., 2003. An overview of ACE-Asia: strategies for quantifying the relationships between Asian aerosols and their climatic impacts. J. Geophys. Res.: Atmosphere 108 (D23).

Kahn, R. A. and Gaitley, B. J.: An analysis of global aerosol type as retrieved by MISR, Journal of Geophysical Research-Atmospheres, 120, 4248-4281, 10.1002/2015jd023322, 2015.

Kalapureddy, M. C. R. and Devara, P. C. S.: Characterization of aerosols over oceanic regions around India during pre-monsoon 2006, Atmospheric Environment, 42, 6816-6827, 10.1016/j.atmosenv.2008.05.022, 2008.

Levy, R. C., Remer, L. A., Mattoo, S., Vermote, E. F., and Kaufman, Y. J.: Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance, Journal of Geophysical Research-Atmospheres, 112, 10.1029/2006jd007811, 2007.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques, 6, 2989-3034, 10.5194/amt-6-2989-2013, 2013.

Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., and Eck, T. F.: Global evaluation of the Collection 5 MODIS dark-target aerosol products over land, Atmospheric Chemistry and Physics, 10, 10399-10420, 10.5194/acp-10-10399-2010, 2010.

Li, Z., Zhang, Y., Shao, J., Li, B., Hong, J., Liu, D., Li, D., Wei, P., Li, W., Li, L., Zhang, F., Guo, J., Deng, Q., Wang, B., Cui, C., Zhang, W., Wang, Z., Lv, Y., Xu, H., Chen, X., Li, L., and Qie, L.: Remote sensing of atmospheric particulate mass of dry PM2.5 near the ground: Method validation using ground-based measurements, Remote Sensing of Environment, 173, 59-68, 10.1016/j.rse.2015.11.019, 2016.

Liang, C., Zang, Z., Li, Z., and Yan, X.: An Improved Global Land Anthropogenic Aerosol Product Based on Satellite Retrievals From 2008 to 2016, Ieee Geoscience and Remote Sensing Letters, 18, 944-948, 10.1109/lgrs.2020.2991730, 2021.

Lipponen, A., Mielonen, T., Pitkanen, M. R. A., Levy, R. C., Sawyer, V. R., Romakkaniemi, S., Kolehmainen, V., and Arola, A.: Bayesian aerosol retrieval algorithm for MODIS AOD retrieval over land, Atmospheric Measurement Techniques, 11, 1529-1547, 10.5194/amt-11-1529-2018, 2018.

Liu, T., Mickley, L. J., and McCarty, J. L.: Global search for temporal shifts in fire activity: potential human influence on southwest Russia and north Australia fire seasons, Environmental Research Letters, 16, 10.1088/1748-9326/abe328, 2021.

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

Manoj, M. R., Satheesh, S. K., Moorthy, K. K., Gogoi, M. M., and Babu, S. S.: Decreasing Trend in Black Carbon Aerosols Over the Indian Region, Geophysical Research Letters, 46, 2903-2910, 10.1029/2018gl081666, 2019.

Mhawish, A., Sorek-Hamer, M., Chatfield, R., Banerjee, T., Bilal, M., Kumar, M., Sarangi, C., Franklin, M., Chau, K., Garay, M., and Kalashnikova, O.: Aerosol characteristics from earth observation systems: A comprehensive investigation over South Asia (2000?2019), Remote Sensing of Environment, 259, 10.1016/j.rse.2021.112410, 2021.

O'Dell, K., Ford, B., Fischer, E. V., and Pierce, J. R.: Contribution of Wildland-Fire Smoke to US PM2.5 and Its Influence on Recent Trends, Environmental Science & Technology, 53, 1797-1804, 10.1021/acs.est.8b05430, 2019.

O'Neill, N. T.: Comment on "Classification of aerosol properties derived from AERONET direct sun data" by Gobbi et al. (2007), Atmospheric Chemistry and Physics, 10, 10017-10019, 10.5194/acp-10-10017-2010, 2010.

O'Neill, N. T., Dubovik, O., and Eck, T. F.: Modified angstrom ngstrom exponent for the characterization of submicrometer aerosols, Applied Optics, 40, 2368-2375, 10.1364/ao.40.002368, 2001a.

O'Neill, N. T., Eck, T. F., Smirnov, A., Holben, B. N., and Thulasiraman, S.: Spectral discrimination of coarse and fine mode optical depth, Journal of Geophysical Research-Atmospheres, 108, 10.1029/2002jd002975, 2003.

O'Neill, N. T., Eck, T. F., Holben, B. N., Smirnov, A., Dubovik, O., and Royer, A.: Bimodal size distribution influences on the variation of Angstrom derivatives in spectral and optical depth space, Journal of Geophysical Research-Atmospheres, 106, 9787-9806, 10.1029/2000jd900245, 2001b.

Ong BT, Sugiura K, Zettsu K. Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM2.5, Neural Comput Appl., 27:1553-1566. doi: 10.1007/s00521-015-1955-3, 2015.

Parks, S. A. and Abatzoglou, J. T.: Warmer and Drier Fire Seasons Contribute to Increases in Area Burned at High Severity in Western US Forests From 1985 to 2017, Geophysical Research Letters, 47, 10.1029/2020gl089858, 2020.

Perez-Ramirez, D., Andrade-Flores, M., Eck, T. F., Stein, A. F., O'Neill, N. T., Lyamani, H., Gasso, S., Whiteman, D. N., Veselovskii, I., Velarde, F., and Alados-Arboledas, L.: Multi year aerosol characterization in the tropical Andes and in adjacent Amazonia using AERONET measurements, Atmospheric Environment, 166, 412-432, 10.1016/j.atmosenv.2017.07.037, 2017.

Ramachandran, S.: Aerosol optical depth and fine mode fraction variations deduced from Moderate Resolution Imaging Spectroradiometer (MODIS) over four urban areas in India, Journal of Geophysical Research-Atmospheres, 112, 10.1029/2007jd008500, 2007.

Shen, H., Li, T., Yuan, Q., & Zhang, L.: Estimating regional ground-level PM2.5 directly from satellite top-of-atmosphere reflectance using deep belief networks, Journal of Geophysical Research: Atmospheres, 123, 13875-13886, /10.1029/2018JD028759, 2018.

Shi, G., Yan, H., Zhang, W., Dodson, J., Heijnis, H., and Burrows, M.: Rapid warming has resulted in more wildfires in northeastern Australia, Science of the Total Environment, 771, 10.1016/j.scitotenv.2020.144888, 2021.

Tai, A. P. K., Mickley, L. J., and Jacob, D. J.: Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change, Atmospheric Environment, 44, 3976-3984, 10.1016/j.atmosenv.2010.06.060, 2010.

Tan, C., Zhao, T., Xu, X., Liu, J., Zhang, L., and Tang, L.: Climatic analysis of satellite aerosol data on variations of submicron aerosols over East China, Atmospheric Environment, 123, 392-398, 10.1016/j.atmosenv.2015.03.054, 2015.

van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J., and Kasibhatla, P. S.: Global fire emissions estimates during 1997-2016, Earth System Science Data, 9, 697-720, 10.5194/essd-9-697-2017, 2017.

Vinoj, V., Rasch, P. J., Wang, H., Yoon, J.-H., Ma, P.-L., Landu, K., and Singh, B.: Short-term modulation of Indian summer monsoon rainfall by West Asian dust, Nature Geoscience, 7, 308-313, 10.1038/ngeo2107, 2014.

Wei, Y., Li, Z., Zhang, Y., Chen, C., Dubovik, O., Zhang, Y., Xu, H., Li, K., Chen, J., Wang, H., Ge, B., and Fan, C.: Validation of POLDER GRASP aerosol optical retrieval over China using SONET observations, Journal of Quantitative Spectroscopy & Radiative Transfer, 246, 10.1016/j.jqsrt.2020.106931, 2020.

Yan, X.: Physical and deep learning retrieved fine mode fraction (Phy-DL FMF) [dataset], http://doi.org/10.5281/zenodo.5105617, 2021.

Yan, X., Zang, Z., Zhao, C., and Husi, L.: Understanding global changes in fine-mode aerosols during 2008-2017 using statistical methods and deep learning approach, Environment International, 149, 10.1016/j.envint.2021.106392, 2021a.

https://doi.org/10.5194/essd-2021-326
Preprint. Discussion started: 30 September 2021

Yan, X., Zang, Z., Luo, N., Jiang, Y., and Li, Z.: New interpretable deep learning model to monitor real-time PM2.5 concentrations from satellite data, Environment International, 144, 10.1016/j.envint.2020.106060, 2020.

Yan, X., Li, Z., Shi, W., Luo, N., Wu, T., and Zhao, W.: An improved algorithm for retrieving the fine-mode fraction of aerosol optical thickness, part 1: Algorithm development, Remote Sensing of Environment, 192, 87-97, 10.1016/j.rse.2017.02.005, 2017.

Yan, X., Zang, Z., Liang, C., Luo, N., Ren, R., Cribb, M., and Li, Z.: New global aerosol fine-mode fraction data over land derived from MODIS satellite retrievals, Environmental Pollution, 276, 10.1016/j.envpol.2021.116707, 2021b.

Yan, X., Zang, Z., Jiang, Y., Shi, W., Guo, Y., Li, D., Zhao, C., and Husi, L.: A Spatial-Temporal Interpretable Deep Learning Model for improving interpretability and predictive accuracy of satellite-based PM2.5, Environmental Pollution, 273, 10.1016/j.envpol.2021.116459, 2021c.

Yan, X., Li, Z., Luo, N., Shi, W., Zhao, W., Yang, X., Liang, C., Zhang, F., and Cribb, M.: An improved algorithm for retrieving the fine-mode fraction of aerosol optical thickness. Part 2: Application and validation in Asia, Remote Sensing of Environment, 222, 90-103, 10.1016/j.rse.2018.12.012, 2019.

Yang, X., Jiang, L., Zhao, W., Xiong, Q., Zhao, W., & Yan, X.: Comparison of ground-based PM2. 5 and PM10 concentrations in China, India, and the US. International journal of environmental research and public health, 15(7), 1382, /10.3390/ijerph15071382, 2018.

Yang, X., Zhao, C., Luo, N., Zhao, W., Shi, W., & Yan, X.: Evaluation and Comparison of Himawari-8 L2 V1. 0, V2. 1 and MODIS C6. 1 aerosol products over Asia and the oceania regions, Atmospheric Environment, 220, 117068, /10.1016/j.atmosenv.2019.117068, 2020.

Yin, S., Wang, X., Zhang, X., Guo, M., Miura, M., and Xiao, Y.: Influence of biomass burning on local air pollution in mainland Southeast Asia from 2001 to 2016, Environmental Pollution, 254, 112949, /10.1016/j.envpol.2019.07.117, 2019.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., & Zhang, L.: Deep learning in environmental remote sensing: Achievements and challenges, Remote Sensing of Environment, 241, /10.1016/j.rse.2020.111716, 2020.

Zang, Z., Guo, Y., Jiang, Y., Chen, Z., Li, D., Shi, W. and Yan, X.: Tree-Based Ensemble Deep Learning Model for Spatiotemporal Surface Ozone (O3) Prediction and Interpretation, International Journal of Applied Earth Observation and Geoinformation, 103, 102516, /10.1016/j.jag.2021.102516, 2021a.

Zang, Z., Li, D., Guo, Y., Shi, W., and Yan, X.: Superior PM2.5 Estimation by Integrating Aerosol Fine Mode Data from the Himawari-8 Satellite in Deep and Classical Machine Learning Models, Remote Sensing, 13, 10.3390/rs13142779, 2021b.

Zhang, Y. and Li, Z.: Remote sensing of atmospheric fine particulate matter (PM2.5) mass concentration near the ground from satellite observation, Remote Sensing of Environment, 160, 252-262, 10.1016/j.rse.2015.02.005, 2015.

Zhang, Y., Li, Z., Liu, Z., Wang, Y., Qie, L., Xie, Y., Hou, W., and Leng, L.: Retrieval of aerosol fine-mode fraction over China from satellite multiangle polarized observations: validation and comparison, Atmospheric Measurement Techniques, 14, 1655-1672, 10.5194/amt-14-1655-2021, 2021.

Zhang, Y., Li, Z., Qie, L., Zhang, Y., Liu, Z., Chen, X., Hou, W., Li, K., Li, D., and Xu, H.: Retrieval of Aerosol Fine-Mode Fraction from Intensity and Polarization Measurements by PARASOL over East Asia, Remote Sensing, 8, 10.3390/rs8050417, 2016.

Zheng, B., Chevallier, F., Yin, Y., Ciais, P., Fortems-Cheiney, A., Deeter, M. N., Parker, R. J., Wang, Y., Worden, H. M., and Zhao, Y.: Global atmospheric carbon monoxide budget 2000-2017 inferred from multi-species atmospheric inversions, Earth System Science Data, 11, 1411-1436, 10.5194/essd-11-1411-2019, 2019.

Zheng, X., Zhao, W., Yan, X., Shu, T., Xiong, Q., & Chen, F.: Pollution characteristics and health risk assessment of airborne heavy metals collected from Beijing bus stations, International journal of environmental research and public health, 12(8), 9658-9671, /10.3390/ijerph120809658, 2015.