Supplement

Deep learning with pre-trained framework unleashes the power of satellite-based global fine-mode aerosol retrieval

Xing Yan[a], Zhou Zang[a*], Zhanqing Li[b*], Hans W. Chen[c], Jiayi Chen[a], Yize Jiang[a], Yunhao Chen[d], Bin He[d], Chen Zuo[a],Terry Nakajima[e], Jhoon Kim[f]

[a]State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China

[b]Department of Atmospheric and Oceanic Science and ESSIC, University of Maryland, College Park, MD, 20740, USA

[c]Department of Space, Earth and Environment, Chalmers University of Technology, Gothenburg, 41296, Sweden

[d]State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China

[e]Tokyo University of Marine Science and Technology, Tokyo, 108-8477, Japan

[f]Department of Atmospheric Sciences, Yonsei University, Seoul, 03722, South Korea

***Corresponding author**: Zhou Zang (joey.zang@mail.utoronto.ca); Zhanqing Li (zli@atmos.umd.edu)

Summary: 16 pages, 10 figures, 2 tables.

Section S1 Present global fine-mode aerosol optical depth (fAOD) studies

To date, increasing efforts have been devoted to the detection of fine-mode aerosol and estimation of fAOD over land using satellite data from numerous space-borne sensors, including the Advanced Along-Track Scanning Radiometer (Sundstrom et al., 2012), the Multi-angle Imaging SpectroRadiometer (MISR) (Kahn and Gaitley, 2015), the Visible Infrared Imaging Radiometer Suite, the Moderate Resolution Imaging Spectroradiometer (MODIS) (Levy et al., 2007), and the Polarization and Directionality of the Earth's Reflectances (POLDER) (Dubovik et al., 2019) instrument.

As a polarimetric imager, POLDER can effectively use the polarized radiation at the Top Of Atmosphere (TOA) for fAOD derivation, because this information is mainly contributed by fine particles within the accumulation mode (Deuze et al., 2001). Launched in December 2004 onboard PARASOL, POLDER/PARASOL was able to provide land fAOD with global coverage every 2 days based on the GRASP algorithm (Tanré et al., 2011). However, the project was shut down in 2013 and thus had only less than 10-year data length, which seriously confines the retrievals for fAOD, especially for recent years.

While MISR takes advantage of its multi-angular observation (9 push broom cameras) to constrain the aerosol types (Diner et al., 1998), providing AOD from small mode to coarse mode (Kahn and Gaitley, 2015; Garay et al., 2020). Compared to POLDER, MISR has much longer monitoring data (2000~ to date), yet the narrow swath and long revisiting time (~16 days) make it hard to capture the fast-changing aerosol information at the local scale.

Among them, MODIS has the longest duration (over 20 years) and a short revisiting time (~1 day), ensuring daily global coverage and making it a superior candidate for global long-term fAOD monitoring. In the MODIS retrieval, the fine-mode fraction (FMF), i.e., the ratio of fAOD to AOD, extracts fAOD from AOD [fAOD = AOD × FMF] by dark target (DT) algorithm. Previous studies have applied the FMF derived from official MODIS aerosol products to calculate global-scale fAOD (Bellouin et al., 2005; Lee and Chung et al., 2013; Suman et al., 2014). However, accurate MODIS retrievals of fAOD and cAOD over land remain a challenge. This is because the retrievals of FMF are based on only four types of aerosol modes and the products are merely 11 discrete from 0 and 1 (Mielonen et al., 2011; Suman et al., 2014). In addition, frequent unrealistic cases of FMF=0 shows in spatial distribution and the evaluations with AERONET observations indicate large uncertainties in MODIS FMF (Levy et al., 2007; Yan et al., 2017, 2021; Zhang and Li., 2015). The retrievals of the components, and their ratio (FMF), are much more erroneous than the total AOD (Jethva et al., 2010; Mielonen et al., 2011; Suman et al., 2014). Evaluations using Aerosol Robotic Network (AERONET) observations indicate a large bias in the MODIS FMF (Levy et al., 2007; X. Chen et al., 2020; Yan et al., 2017, 2021; Zhang and Li, 2015).

To obtain MODIS-retrieved fAOD with better accuracy, Yan et al. (2017, 2019) proposed a lookup table (LUT) spectral deconvolution algorithm (SDA) to improve FMF retrievals over land. To improve the retrieval efficiency, Liang et al. (2021) adopted the LUT-SDA FMF to create 9-year global land fAOD products which are

better than the previous products, but still suffered from 21.4% overestimation in comparison with the AERONET-based retrievals.

Section S2 Evaluation metrics

To assess the performance of the DLFE-Satellite model, the following metrics were used: Pearson correlation coefficient (R), root-mean-square error (RMSE), and mean absolute error (MAE), formulated as

$$R = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2 \sum_{i=1}^{N}(Y_i - \overline{Y})^2}},$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2},$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|X_i - Y_i\right|.$$

For the fine-mode aerosol optical depth (fAOD), we adopted the relative error (EE) envelope of ± (0.05+15%) to determine if an fAOD retrieval is accurate, i.e., if it falls within this envelope (Bréon et al., 2011).

To compare evaluation metrics from multiple modeling results, we employed the rate of change in RMSE and MAE, formulated as

$$RateofChange = \frac{RMSE1 - RMSE2}{RMSE2} \times 100\%,$$

$$RateofChange = \frac{MAE1 - MAE2}{MAE2} \times 100\%.$$

Table S1. Data used for estimating fAOD in this study.

Type	Parameter	Spatial resolution	Temporal resolution	Source
MOD02SSH (MODIS C6.1 L1B)	TOA reflectance data: bands 1-7	5-km×5-km	1 day	https://ladsweb.modaps.eosdis.nasa.gov/
MOD09CMG (MODIS C6.1 L3)	Surface reflectance: bands 1-7 Brightness_Temperature (K) : bands 20, 21, 31, and 32 Relative_Azimuth_Angle Solar_Zenith_Angle View_Zenith_Angle	0.05°×0.05°	1 day	
MOD08_D3 (MODIS C6.1 L3)	Aerosol_Optical_Depth_Land_Mean (at 500 nm)	1°×1°	1 day	
ERA5 (reanalysis-era5-single-levels)	10m_u_component_of_wind (m/s) 10m_v_component_of_wind (m/s) 2m_dewpoint_temperature (K) 2m_temperature (K) boundary_layer_height (m) surface_pressure (Pa)	0.25°×0.25°	1 hour	https://cds.climate.copernicus.eu/

Table S2. SURFRAD sites used for independent validation and their locations.

Site	Longitude (°W)	Latitude (°N)
Bondville, Illinois (BON)	88.37	40.05
Fort Peck, Montana (FPK)	105.10	48.31
Goodwin Creek, Mississippi (GWN)	89.87	34.25
Penn. State Univ., Pennsylvania (PSU)	77.93	40.72
Sioux Falls, South Dakota (SXF)	96.62	43.73
Table Mountain, Boulder, Colorado (TBL)	105.24	40.12

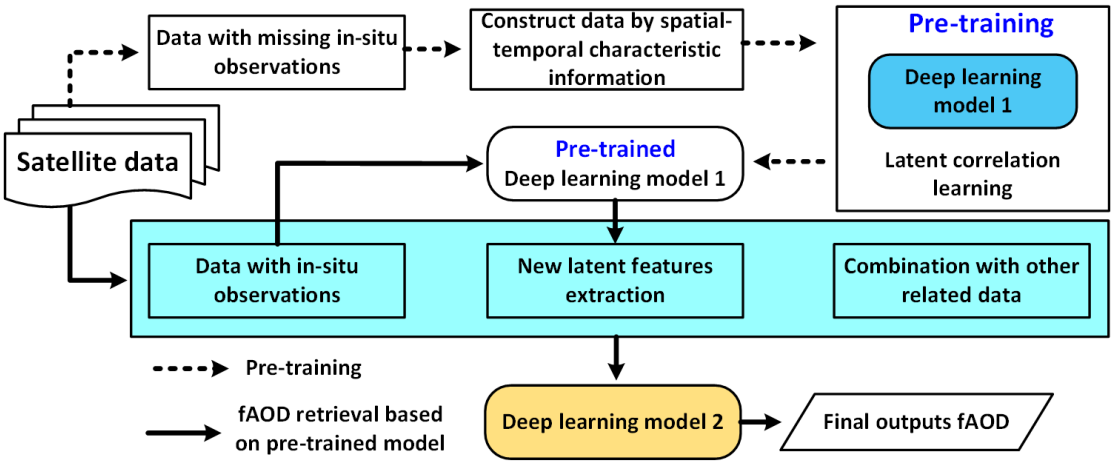

Figure S1. The schematic flowchart of DLFE-Satellite.

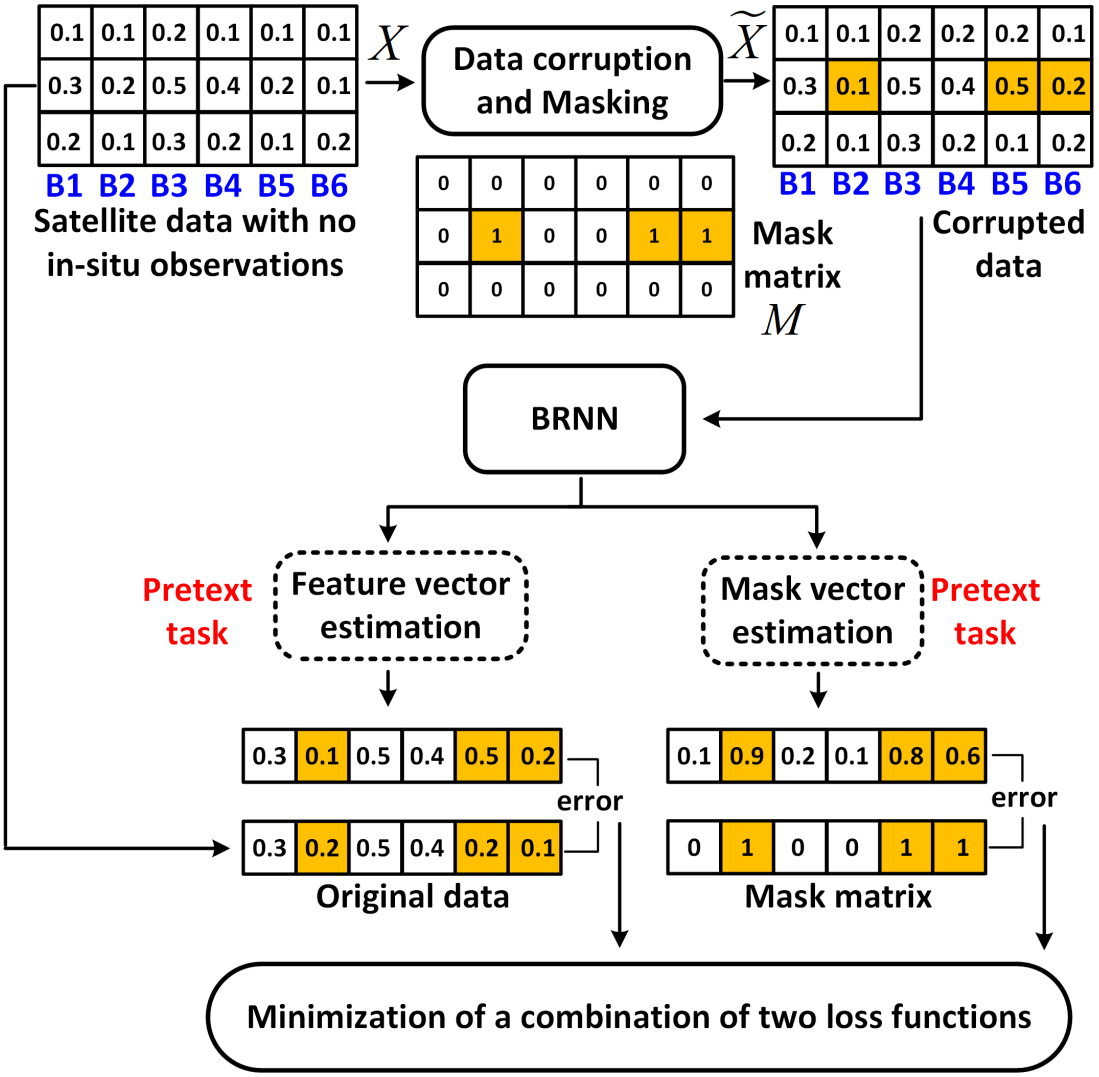

Figure S2. The self-supervised learning in pre-training step of DLFE-Satellite. For

example, using TOA of MODIS Band 1- Band 6 as the original data, the feature vector estimation and mask vector estimation try to recover the original data [0.3, 0.2, 0.5, 0.4, 0.2, 0.1] and the mask data [0, 1, 0, 0, 1, 1] from the corrupted data [0.3, 0.1, 0.5, 0.4, 0.5, 0.2]

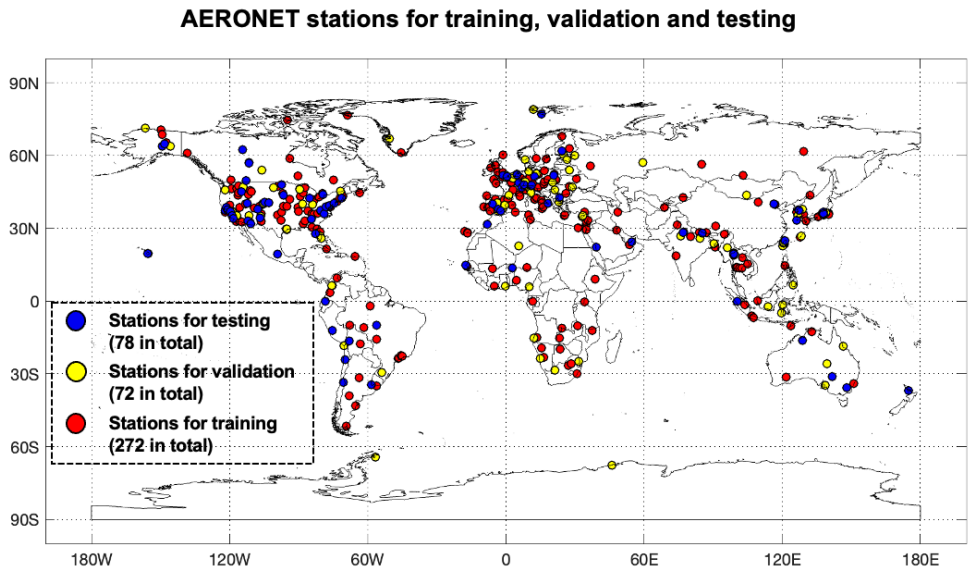

Figure S3. Site-based independent validation. The global AERONET stations were randomly partitioned into training (272 stations), validation (72 stations), and testing (78 stations) subsets.

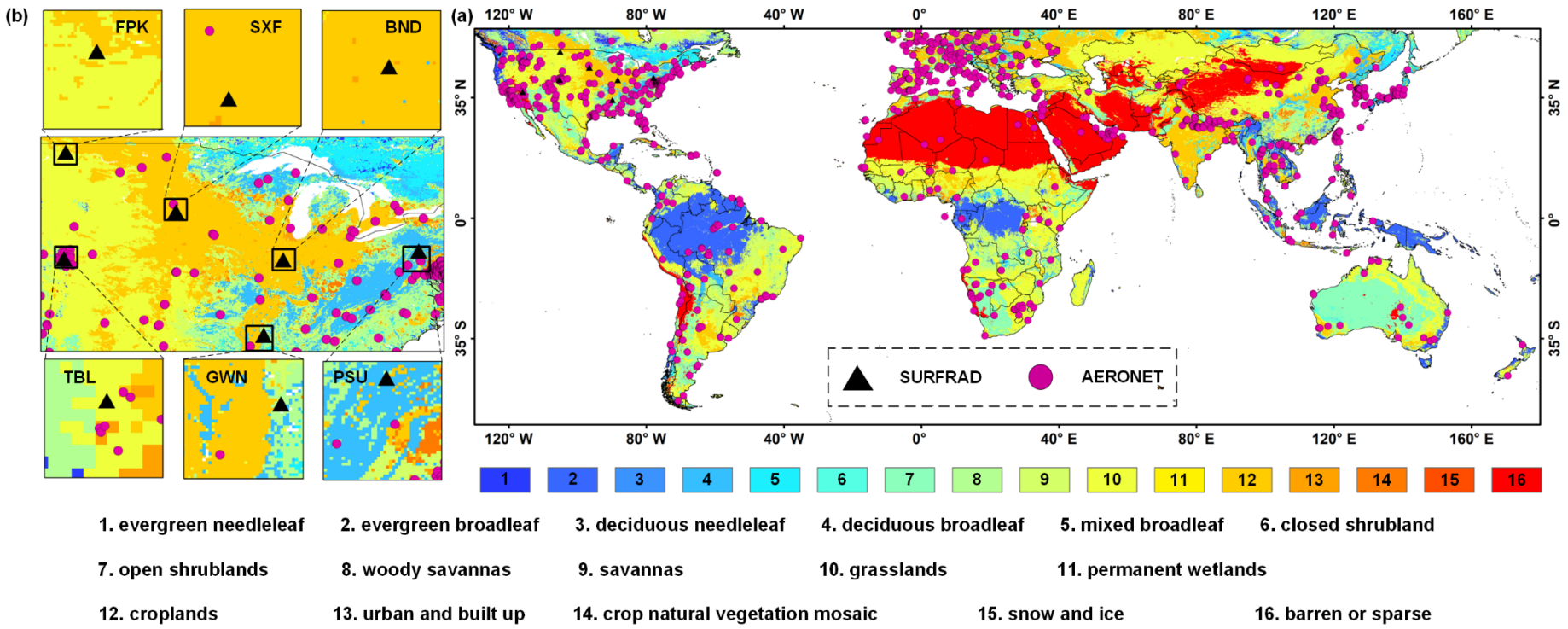

Figure S4. (**a**) AERONET (purple dots) and SURFRAD (black triangles) sites used in this study. The base map shows land cover types from the MODIS MCD12C1 product (i.e., the International Geosphere-Biosphere Programme scheme), consisting of 16 land cover types. (**b**) Zoomed-in distributions of six SURFRAD sites. Table S2 provides details about these sites.

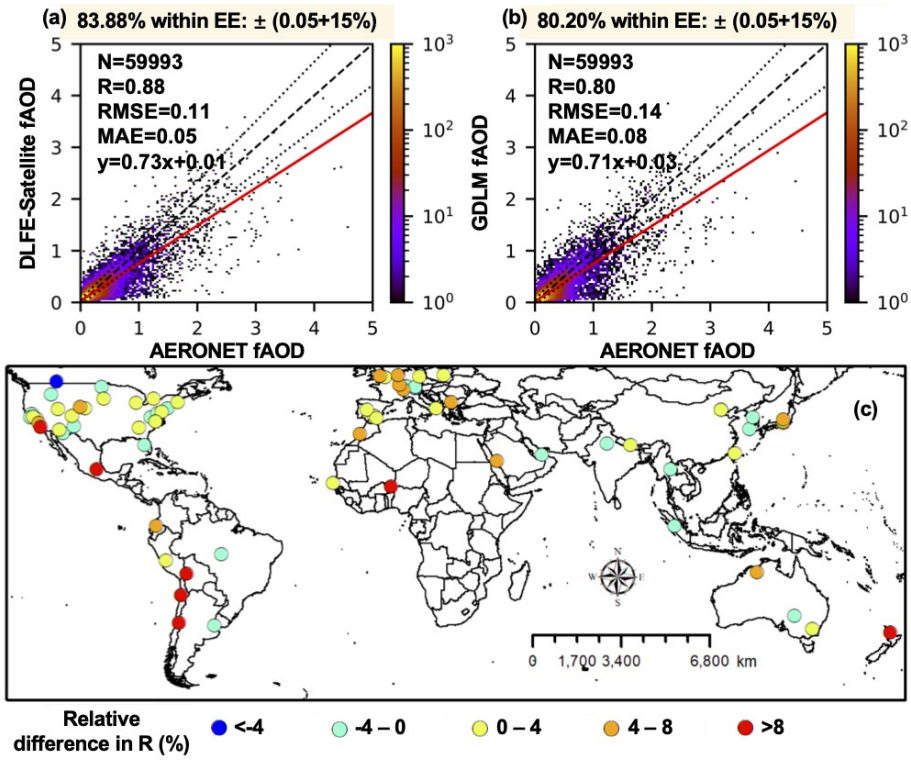

Figure S5. Application in global fAOD modeling involved the estimation of fAOD (at 0.5 μm) from 2001 to 2020 using both the GDLM and DLFE-Satellite methodologies. Figures (a) and (b) depict density scatterplots of GDLM and DLFE-Satellite fAOD approximations contrasted with AERONET fAODs from 78 testing stations. The black dashed line signifies the 1:1 relationship, the red solid line represents the linear fitting line, and two black dotted lines delineate the EE envelope of ± (0.05+15%). Figure (c) displays the relative difference of R between GDLM and DLFE-Satellite fAOD at each AERONET site [relative difference = 100 * (DLFE-Satellite R – GDLM R) / GDLM R].

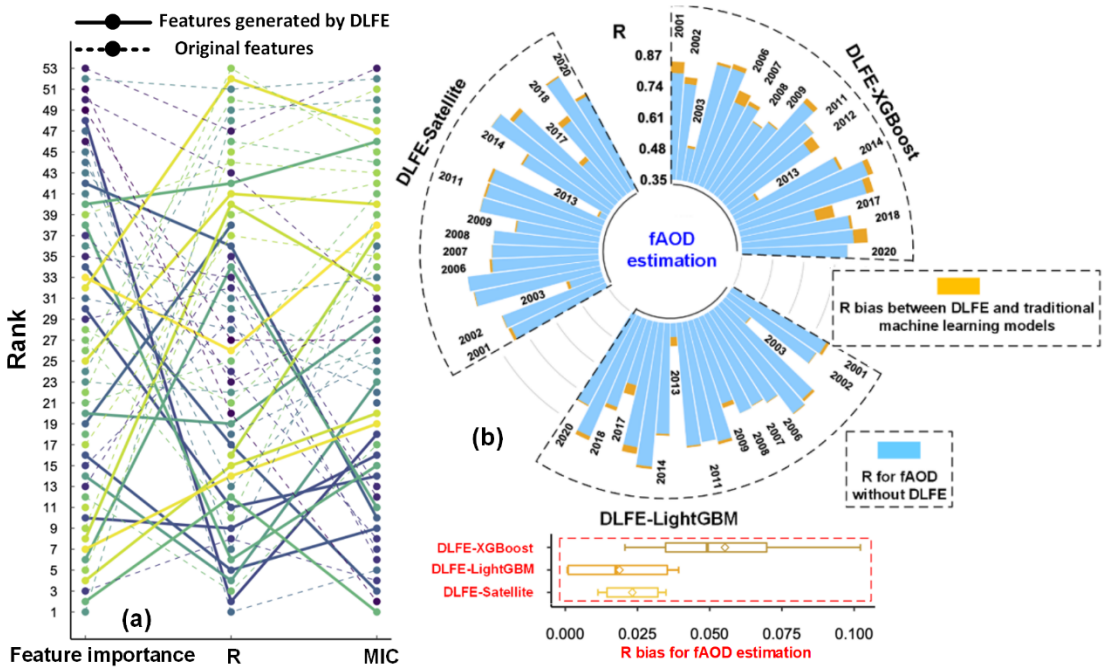

Figure S6. Exploring the correlation of DLFE-satellite generated features with fAOD and the model's generalizability across various machine models. (a) The ranking of the generated features among all features, based on their contribution as indicated by (1) the feature importance according to Random Forest, (2) the Pearson correlation value (R), and (3) the Maximal Information Coefficient (MIC) value. (b) R resulting from the fAOD estimation in each year from 2001-2020. The three sectors represent the result of using the DLFE-Satellite model with XGBoost, LightGBM, and GDLM as DLM2. In each sector, blue bars represent the R from the traditional machine learning models without DLFE-Satellite, and the orange bars represent the R bias (R bias = R estimated by the model with DLEF minus R estimated by the model without DLEF) after applying the DLFE-Satellite model with multiple machine-learning models. The distribution of R bias after applying the DLFE-Satellite to different models and tasks. In each boxplot, the diamond is the mean value of the R bias, The error bars are the 25th and 75th percentiles of the total relative difference.

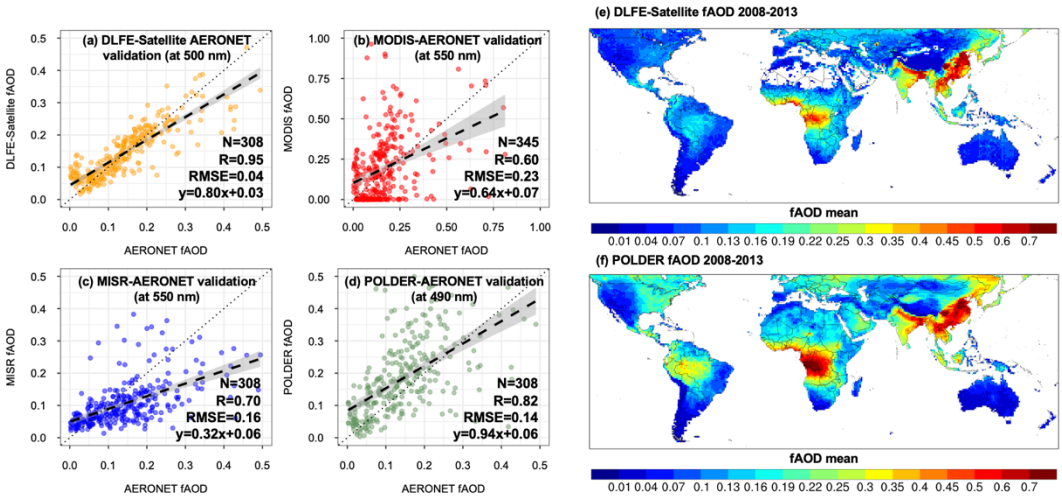

Figure S7. (a-d) Density scatterplots showcasing the monthly mean AERONET retrievals and fAODs products for DLFE-Satellite (a), MODIS (b), MISR (c), and POLDER (d). Only the matched data pairs available for the four products at the same time and location were used for comparison. The black dotted line represents the 1:1 line, and the black dashed line represents the linear fitting line. The shaded areas represent the 95% confidence intervals of the linear regression. (e-f) Global spatial distribution of fAOD by DLFE-Satellite and PODLER in 2008-2013.

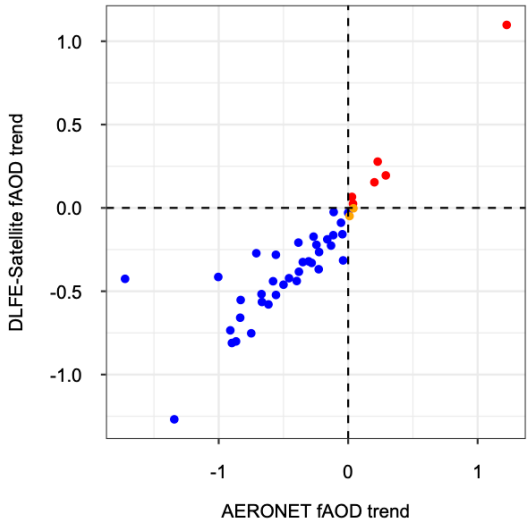

Figure S8. Scatter plot between DLFE-Satellite fAOD and AERONET fAOD trends. Each scatter represents a trend over one AERONET station during 2001-2021. The colors indicate the trend signs between DLFE-Satellite fAOD and AERONET fAOD trends (blue: both trends are negative; red: both trends are positive; orange: the trends are inconsistent). Only the matched data with >18 years of length were selected for trend comparison.

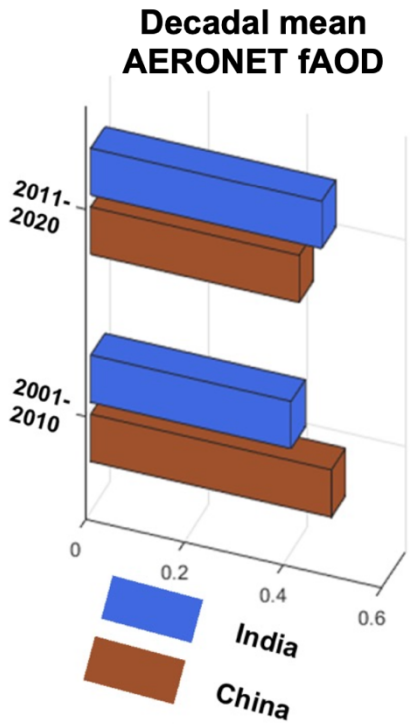

Figure S9. The decadal mean AERONET fAOD averaged for 2001-2010 and 2010-

2020 in India and China.

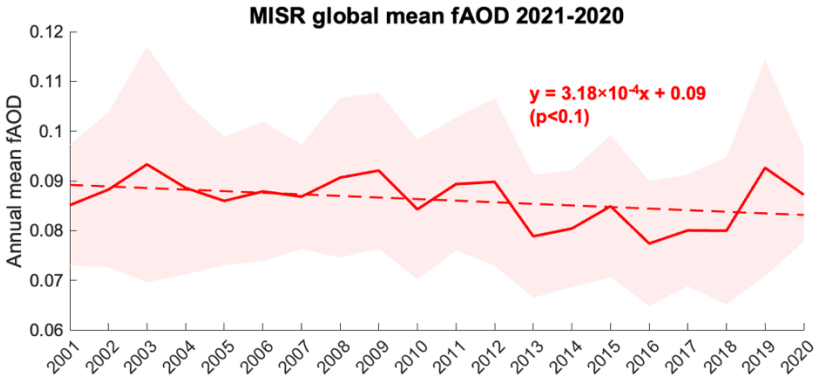

Figure S10. Global monthly mean time series of MISR fAOD from 2001 to 2020. The dashed lines are the linear fitting lines, and the shaded areas represent the monthly mean fAOD ± the monthly standard deviation of fAOD.

References

Bellouin, N.; Boucher, O.; Haywood, J.; Reddy, M. S. Global estimate of aerosol direct radiative forcing from satellite measurements. Nature 2005, 438 (7071), 1138–1141.

Bréon, F., Vermeulen, A., & Descloitres, J. An evaluation of satellite aerosol products against sunphotometer measurements, Remote Sensing of the Environment 2011, 115, 3102–311.

Chen, X.; de Leeuw, G.; Arola, A.; Liu, S.; Liu, Y.; Li, Z.; 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 2020, 249, 112006.

Deuze, J. L., Breon, F. M., Devaux, C., Goloub, P., Her- man, 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 2001, 106, 4913–4926.

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 1998, 36, 1072–1087.

Dubovik, O.; Li, Z.; Mishchenko, M. I.; Tanré, D.; Karol, Y.; Bojkov, B.; Cairns, B.; Diner, D. J.; Espinosa, W. R.; Goloub, P.; et al. Polarimetric remote sensing of

atmospheric aerosols: Instruments, methodologies, results, and perspectives. Journal of Quantitative Spectroscopy and Radiative Transfer 2019, 224, 474–511.

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 SpectroRadiometer (MISR) aerosol product, Atmospheric Measurement Techniques 2020, 13, 593–628.

Jethva, H.; Satheesh, S.; Srinivasan, J.; Levy, R. Improved retrieval of aerosol size-resolved properties from moderate resolution imaging spectroradiometer over India: Role of aerosol model and surface reflectance. Journal of Geophysical Research: Atmospheres 2010, 115 (D18), D18213.

Kahn, R. A.; Gaitley, B. J. An analysis of global aerosol type as retrieved by MISR. Journal of Geophysical Research: Atmospheres 2015, 120 (9), 4248–4281.

Lee, K. and Chung, C. E.: Observationally-constrained estimates of global fine-mode AOD, Atmospheric Chemistry and Physics 2013, 13, 2907–2921.

Levy, R. C.; Remer, L. A.; Mattoo, S.; Vermote, E. F.; 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 2007, 112 (D13), D13211.

Liang, C.; Zang, Z.; Li, Z.; Yan, X.: An Improved Global Land Anthropogenic Aerosol Product Based on Satellite Retrievals From 2008 to 2016, IEEE Geoscience and Remote Sensing Letters 2021, 18, 944– 948.

Mielonen, T.; Levy, R.; Aaltonen, V.; Komppula, M.; De Leeuw, G.; Huttunen, J.; Lihavainen, H.; Kolmonen, P.; Lehtinen, K.; Arola, A. Evaluating the assumptions of surface reflectance and aerosol type selection within the MODIS aerosol retrieval over land: the problem of dust type selection. Atmospheric Measurement Techniques 2011, 4 (2), 201–214.

Sai Suman, M.; Gadhavi, H.; Ravi Kiran, V.; Jayaraman, A.; Rao, S. Role of Coarse and Fine Mode Aerosols in MODIS AOD Retrieval: a case study over southern India. Atmospheric Measurement Techniques 2014, 7 (4), 907–917.

Sundström, A.-M.; Kolmonen, P.; Sogacheva, L.; de Leeuw, G. Aerosol retrievals over China with the AATSR dual view algorithm. Remote sensing of environment 2012, 116, 189–198.

Tanré, D., Bréon, F. M., Deuzé, J. L., Dubovik, O., Ducos, F., François, P., Goloub, P., Herman, M., Lifermann, A., and Waquet, F.: Remote sensing of aerosols by using polarized, directional and spectral measurements within the A-Train: the PARASOL mission, Atmospheric Measurement Techniques 2011, 4, 1383–1395.

Yan, X.; Li, Z.; Luo, N.; Shi, W.; Zhao, W.; Yang, X.; Liang, C.; Zhang, F.; 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 2019, 222, 90–103.

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

Yan, X.; Zang, Z.; Liang, C.; Luo, N.; Ren, R.; Cribb, M.; Li, Z.: New global aerosol fine-mode fraction data over land derived from MODIS satellite retrievals, Environmental Pollution 2021, 276, 116707.

Zhang, Y.; Li, Z. Remote sensing of atmospheric fine particulate matter (PM2. 5) mass concentration near the ground from satellite observation. Remote Sensing of Environment 2015, 160, 252–262.