Supporting Information

Substantial underestimation of fine-mode aerosol loading from wildfires and its radiative effects in current satellite-based retrievals over the United States

Xing Yan[1], Chen Zuo [1], Zhanqing Li[2*], Hans W. Chen[3], Yize Jiang[1], Qiao Wang[4], Guoqiang Wang[4], Kun Jia[1], Yinglan A[4], Ziyue Chen[1], Jiayi Chen[1]

1 State Key Laboratory of Remote Sensing Science, Faculty of Geographical 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 Department of Space, Earth and Environment, Chalmers University of Technology, Gothenburg, 41296, Sweden

4 Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China

***Corresponding author**: Zhanqing Li (zli@atmos.umd.edu)

Summary:

Number of pages: 22

Number of supporting texts: 4

Number of supporting figures: 12

Number of supporting tables: 4

Section 1

Evaluation metrics

To assess the performance of our model, the following metrics were used: the Pearson correlation coefficient (R), the root-mean-square error (RMSE), and the mean absolute error (MAE), Normalized Root Mean Square Error (NRMSE) formulated as

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

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

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}}{\max(y) - \min(y)}, \quad (1)$$

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

For the fAOD, we adopted the relative error envelope (EE) to determine if an fAOD retrieval is accurate. Initially proposed by Kaufman[EE (0.05 + 0.15×AOD)] to AOD, this metric aims to measure the interval range of all absolute and relative errors of aerosol indices[1]. Results falling within this interval range are considered to be within the allowable error range, while those falling above or below the range are typically indicative of overestimation or underestimation, respectively. Our focus is on high-precision applications where minimizing systematic errors is critical. hence, a lower fixed offset (such as 0.03) may be preferred. Additionally, considering fine aerosol optical depth (fAOD) values tend to be smaller than those of total aerosol optical depth (AOD), a proportional scaling factor of 0.1 might be more appropriate for capturing the relative variability in errors across the range of fAOD values. This choice ensures that

the EE formula adequately accounts for the inherent uncertainties in fine aerosol retrievals, particularly in scenarios where the absolute magnitude of fAOD values is relatively lower compared to AOD values.

Section 2

CLAUDIA -3 Cloud Mask Algorithm

Our study utilizes the official cloud product algorithm of GOSAT-2, known as the Cloud and Aerosol Unbiased Decision Intellectual Algorithm 3(CLAUDIA-3), developed by Yu Oishi et al[2]. This algorithm employs a Support Vector Machine (SVM) method to optimize the threshold between cloudy and clear skies, enabling automatic threshold adjustment. Comparative validation has shown that its results are comparable in accuracy to the MODIS C6.1 cloud mask algorithm and the cloud data provided by the Atmospheric Radiation Measurement Climate Database, demonstrating good consistency. CLAUDIA-3 is effective in identifying both thick and thin cloud[3]. Therefore, this study adopts the CLAUDIA-3 cloud mask method to enhance the quality of subsequent data and the reliability of the results.

Section 3

Ångström exponent interpolation

To ensure fAOD wavelength consistency in our study, we employed the Ångström exponent interpolation method to convert AERONET AOD measurements at 500nm to 550nm.The Ångström exponent interpolation method is a common approach to convert fAOD from one wavelength to another. The Ångström exponent characterizes the relationship between aerosol particle size distribution and optical properties. Its calculation is based on the formula(2):

$$\alpha = -\frac{\ln\left(fAOD_2 / fAOD_1\right)}{\ln\left(\lambda_2 / \lambda_1\right)} \quad , \quad (2)$$

which fAOD1 and fAOD2 are the fAOD values at wavelengths λ_1 and λ_2 respectively and α is Ångström exponent of fAOD. To interpolate fAOD from 500nm to 550nm, known fAOD values at 500nm and Ångström exponent of fAOD from in-situ measurement. Subsequently, utilizing this exponent and the fAOD value at the target wavelength, interpolation is performed using the formula(3):

$$fAOD_2 = fAOD_1\left(\frac{\lambda_1}{\lambda_2}\right)^{-\alpha} \quad , \quad (3)$$

which λ_1 and λ_2 represent the wavelengths at 500nm and 550nm respectively, $fAOD_1$ and $fAOD_2$ represent the fine-mode aerosol optical depth at wavelengths 500nm and 550nm respectively, and α represents the Ångström wavelength exponent.

Section 4

Model input variable

In our model inputs, we included raw satellite band data and other satellite-derived variables. As shown in Table S5, the raw data consist of ten band data from both forward and backward satellite bands, as well as two angular data points: satellite zenith angle and solar zenith angle. These angles mainly describe the geometric relationship between the satellite observations and the sun's position. The derived variables include the scattering angle (SA), and the monthly minimum and difference values for each band. The scattering angle (SA) is a key angle that measures the scattering process of sunlight in the atmosphere, reflecting the spatial scattering angle between the sunlight and the observation point. SA is often used in the atmospheric radiation transfer equation to solve for AOD and its parameter information and is commonly applied in

the physical calculations of Rayleigh scattering in atmospheric aerosols[4,5]. The monthly minimum value of each band is approximated to represent the minimum surface reflectance for the month, while the difference value in the model represents the contribution of atmospheric aerosols. Additionally, we provided categorical information such as month and season to the our model.

Table S1. Main characteristics of the GOSAT-2 spacecraft.

	Name	Parameters
Satellite parameters	Orbit type	sun synchronous
	Altitude	613 km
	Repeat cycle	6 days
	Local equator crossing time in the descending node	13:00 ± 0:15 (UTC)

Table S2. Main characteristics of the GOSAT-2/CAI-2.

Band	Spectral Range (nm)	Central Band (nm)	Spatial Resolution (m)	Tilt	Swath (km)
1	333–353	339	460	20° (Forward direction)	920
2	433–453	441	460		
3	664– 684	672	460		
4	859–879	865	460		
5	1585–1675	1630	920		
6	370–390	377	460	-20° (backward direction)	
7	540–560	546	460		
8	664–684	672	460		
9	859–879	865	460		
10	1585–1675	1630	920		

Table S3. The definition of different products for fAOD

Measurement	Parameter	Particle size definition (r_p =particle radius)	Source
MODIS	fAOD	0.1um < r_p < 0.25 um	(Kaufman et al.,[6] 2005)
MISR	fAOD	r_p < 0.35 um	(Kahn et al.[7], 2015)
VIIRS	fAOD	0.1um < r_p < 0.25 um	(Li et al.,[8] 2022)

Table S4. SURFRAD sites used for independent validation and their locations.

Site Name	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
Desert Rock, Nevada (DRA)	36.62	116.01

Table S5. Summary of our model's input variables.

Variable	Class	Name	Detailed Description
Continuous variables	Satellite Raw Variables	B1 - B5	Bands 1 to 5 of the satellite at forward viewing
		B6 - B10	Bands 6 to 10 of the satellite at backward viewing
	Satellite Derived Variables	M1 - M5	Minimum monthly value for each band at forward viewing
		M6 - M10	Minimum monthly value for each band at backward viewing
		B_M(1-5)	Raw value minus monthly minimum for each band at forward viewing
		B_M(6-10)	Raw value minus monthly minimum for each band at backward viewing
		NDVI_F	Normalized difference vegetation index at forward viewing
		NDVI_B	Normalized difference vegetation index at backward viewing
	Satellite Angle Variables	SolZ_F	Solar zenith angle at forward viewing
		SenZ_F	Satellite zenith angle at forward viewing
		SCA_F	Scattering angle at forward viewing
		SolZ_B	Solar zenith angle at forward viewing
		SenZ_B	Satellite zenith angle at forward viewing
		SCA_B	Scattering angle at backward viewing
Categorical variables		Month	
		Season	

Table S6. Hyperparameters of the model settings.

Short Name	Configuration	Detailed Description
batch_size	256	The number of data samples captured in one training step
d_model	256	The number of hidden-layer nodes in each layer
n_tf_head	4	The number of attention heads used in the multi-head attention encoder layer
n_tf_layer	2	The number of transformer encoder layers
p_tf_drop	0.1	The percentage of neurons temporarily removed from the transformer encoder layers
n_mlp_layer	1	The number of multilayer perceptrons (MLP)
p_mlp_drop	0.1	The percent of neurons temporarily removed from the MLP layers
lr	0.0005	Learning rate
max_epochs	2000	The maximum number of epochs to train for the model

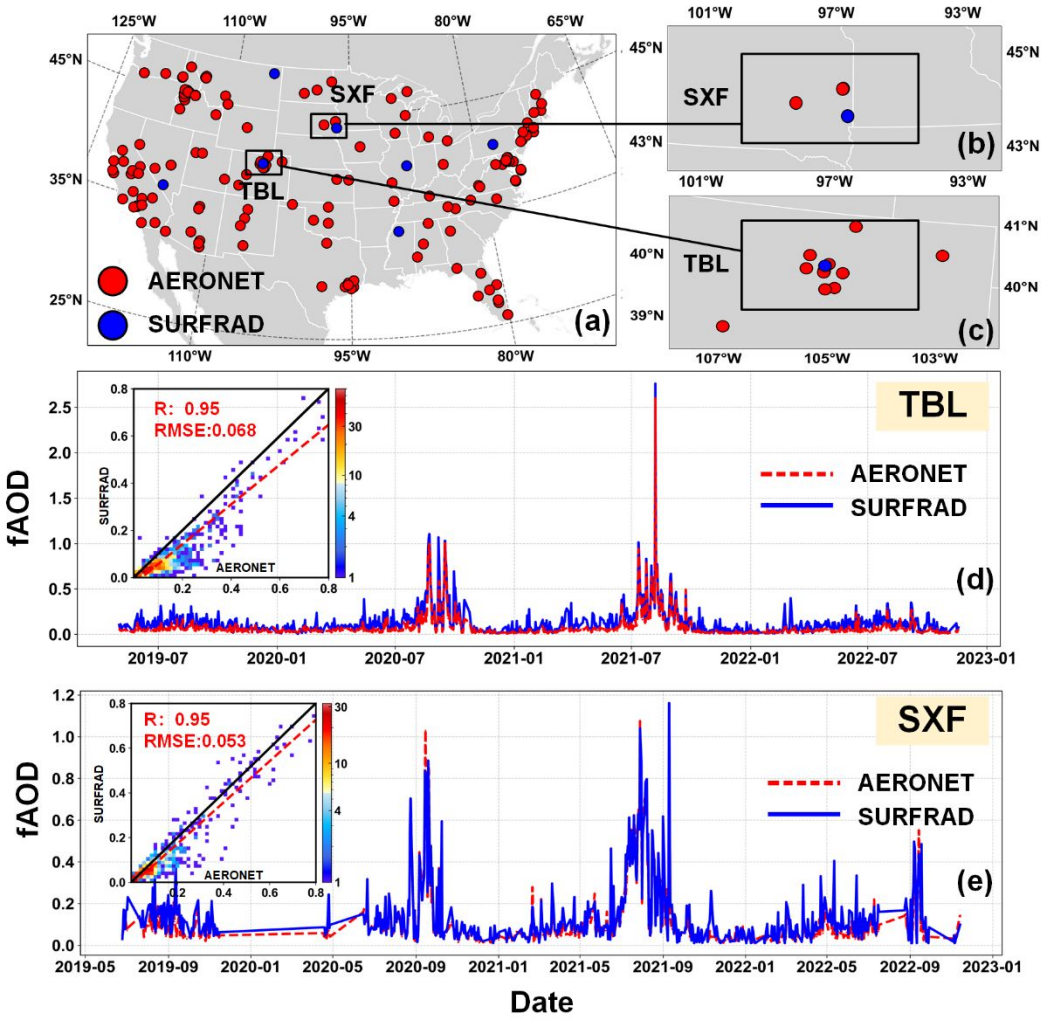

Figure S1. (a) Locations of AERONET stations (red dots) and SURFRAD stations (blue dots; see Table S3 for the names of these stations) across the contiguous United States. The black borders in the figure encircle the adjacent areas of AERONET and SURFRAD stations, respectively. (b-c) The spatial zoom-in of the selected area in Figure S1a. (d) Daily mean fAOD values of AERONET (red dotted line) and SURFRAD (blue line) within the TBL area range. Scatter density plots for both are shown in the top left corner of the figure.(e) same with figureS1c, but in SXF area.

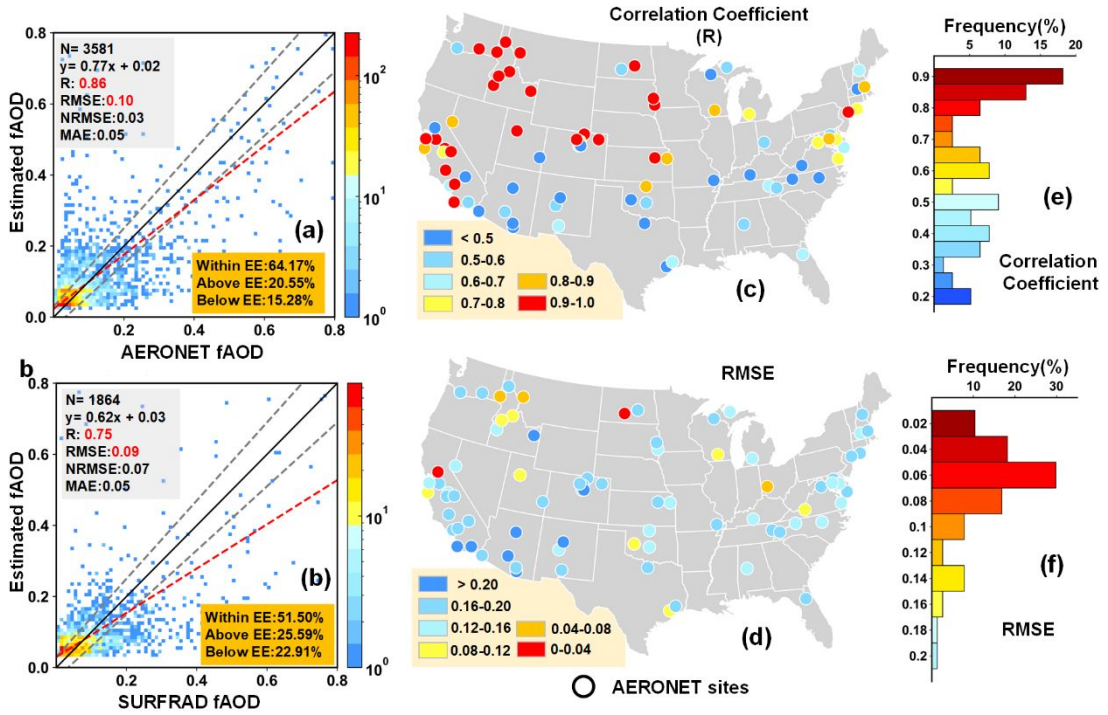

Figure S2. (a) Density scatterplots of ADLM fAOD estimates against AERONET fAOD retrievals. The red dashed line is the best-fit line from linear regression, and the black solid line is the 1:1 line. Black dashed lines outline the error envelope, EE [EE ± (0.03+10%)]. The number of samples (N), relation from linear regression, correlation coefficient (R), root-mean-square-error (RMSE), normalized RMSE (NRMSE), and mean absolute error (MAE) are given. **(b)** Same as (a) but for the independent validation against SRDRAD fAOD retrievals. **(c, e)** Spatial performance (c) and statistical characterization (e) of R for AERONET sites. **(d, f)** Spatial performance (d) and statistical distribution (e) of RMSE for AERONET sites.

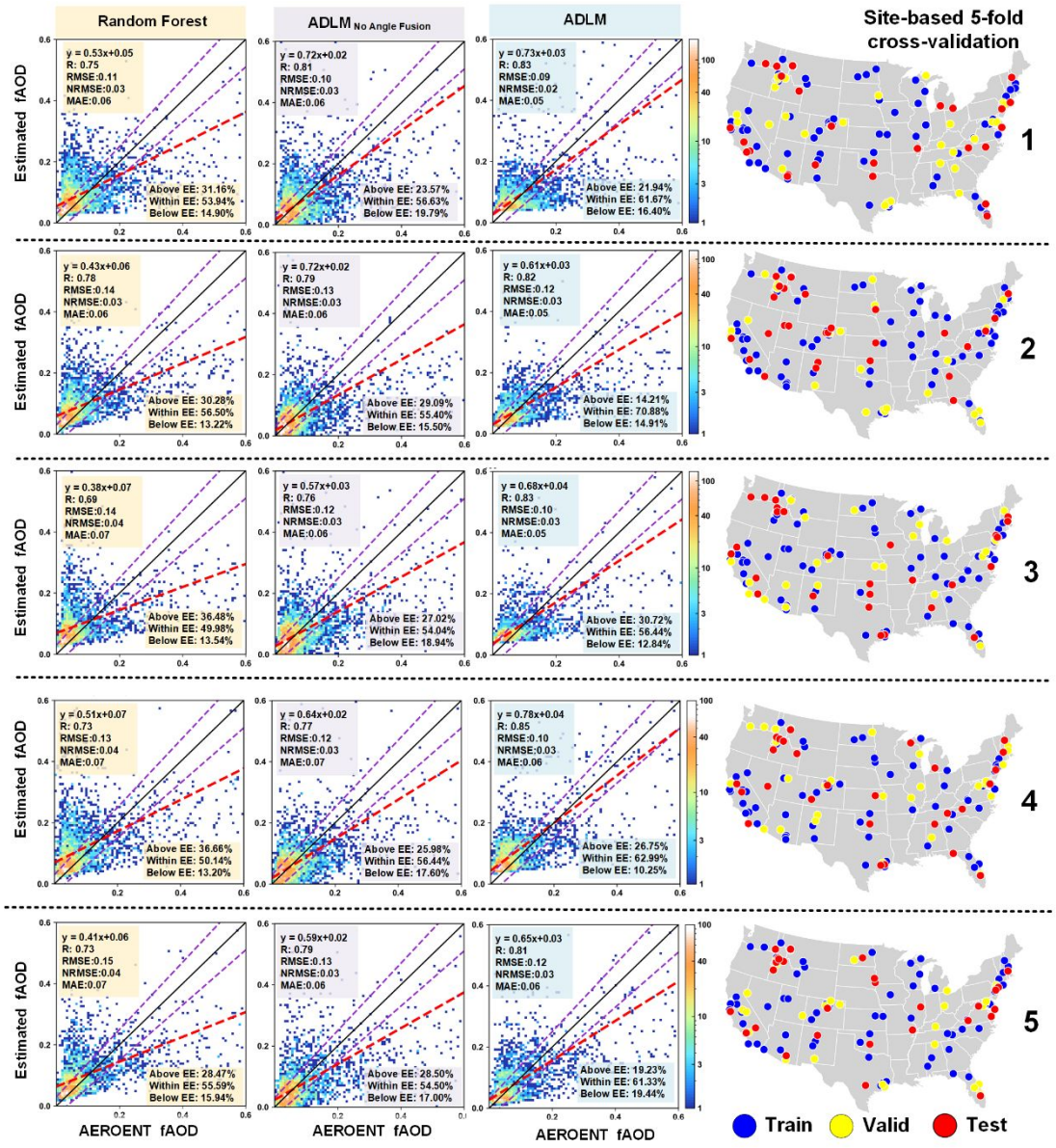

Figure S3. Site-based 5-fold cross-validation for the random forest model , $ADLM_{No\ Angle\text{-}Fusion}$ and ADLM to test their performances and abilities to generalize to unseen locations. The AERONET sites are divided into five subsets, each containing a distinct 20% of available measurement sites. These sites are not involved in the model training process and are only used for testing purposes.

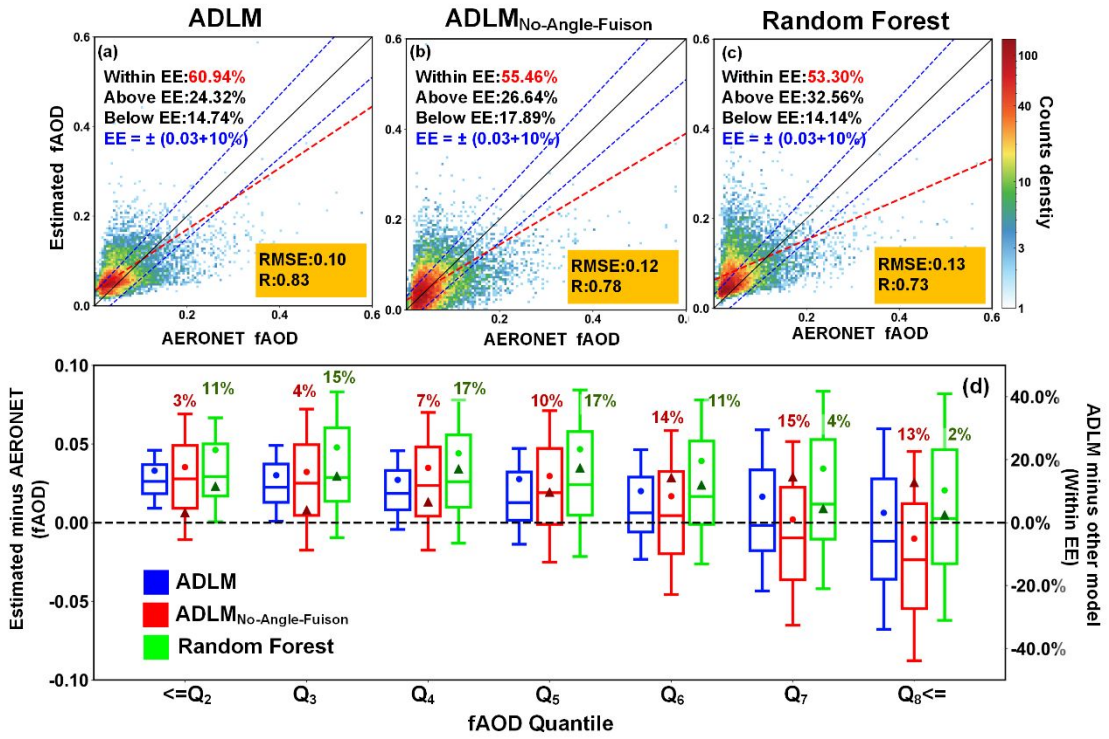

Figure S4. (a, b, c) Density scatterplots of (a) ADLM fAOD,(b)$ADLM_{no\text{-}angle\text{-}fusion}$ fAOD and (c) random forest fAOD estimates against AERONET fAODs. The red dashed line is the best-fit line from linear regression, and the black solid line is the 1:1 line. Blue dashed lines outline the error envelope, EE [EE ± (0.03+10%)]. The correlation coefficient (R) and root-mean-square-error (RMSE) are given in each panel. **(d)** Boxplots of the fAOD bias (estimated minus AERONET fAOD) of the three models (ADLM, $ADLM_{no\text{-}angle\text{-}fusion}$ and random forest) as a function of AERONET fAOD, divided into eight equal quantiles.

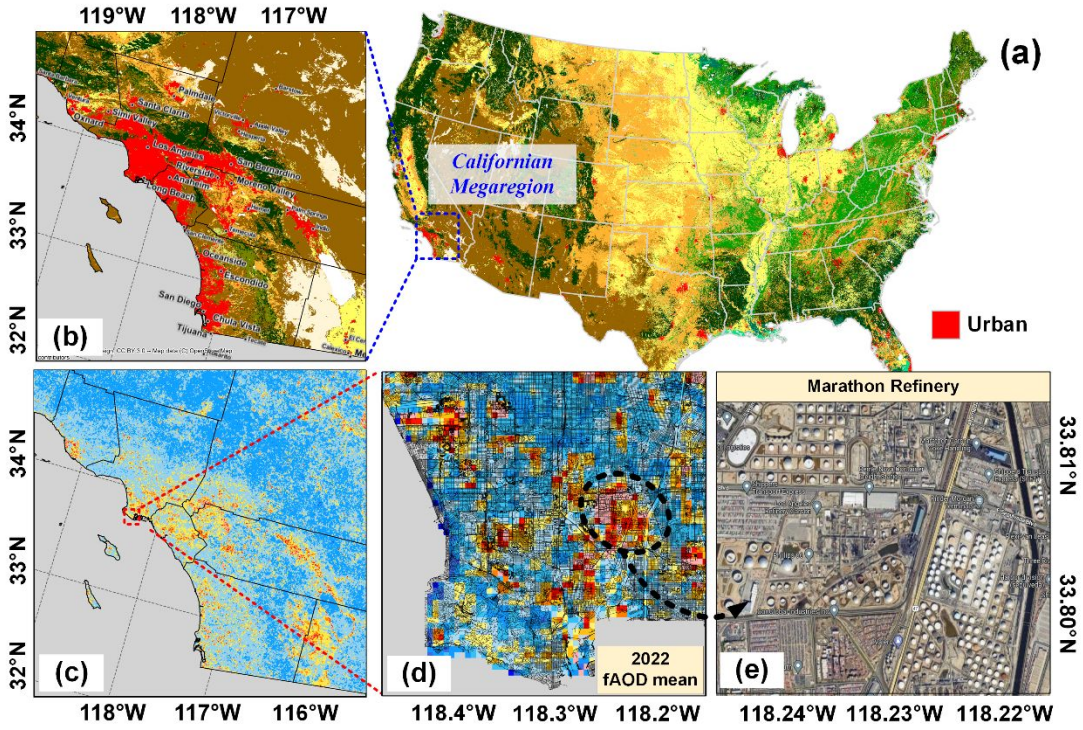

Figure S5. ADLM fAOD in an urban-scale emissions application. **(a)** Land-cover types from the ESACCI Land Cover product, with urban areas in red. **(b)** A Californian urban megaregion and **(c)** its corresponding annual fAOD mean values. **(d)** Spatial patterns of fAOD within urban areas at an ultra-high scale, superimposed with roads from the Open Street Map dataset. **(e)** Google map imagery aligned with the high fAOD area circled in (d).

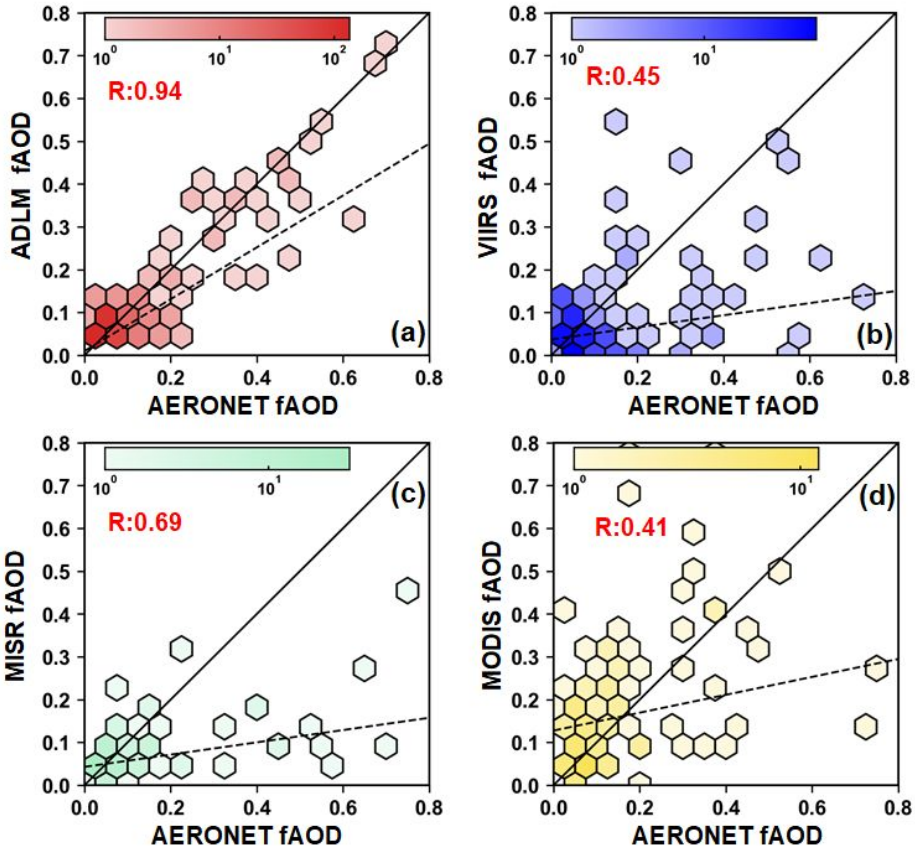

Figure S6. Density scatterplots of monthly averages for (a) ADLM fAOD, (b)VIIRS fAOD, (c)MISR fAOD, (d)MODIS fAOD estimates against AERONET fAOD over California. The black dashed line is the best-fit line from linear regression, and the black solid line is the 1:1 line.

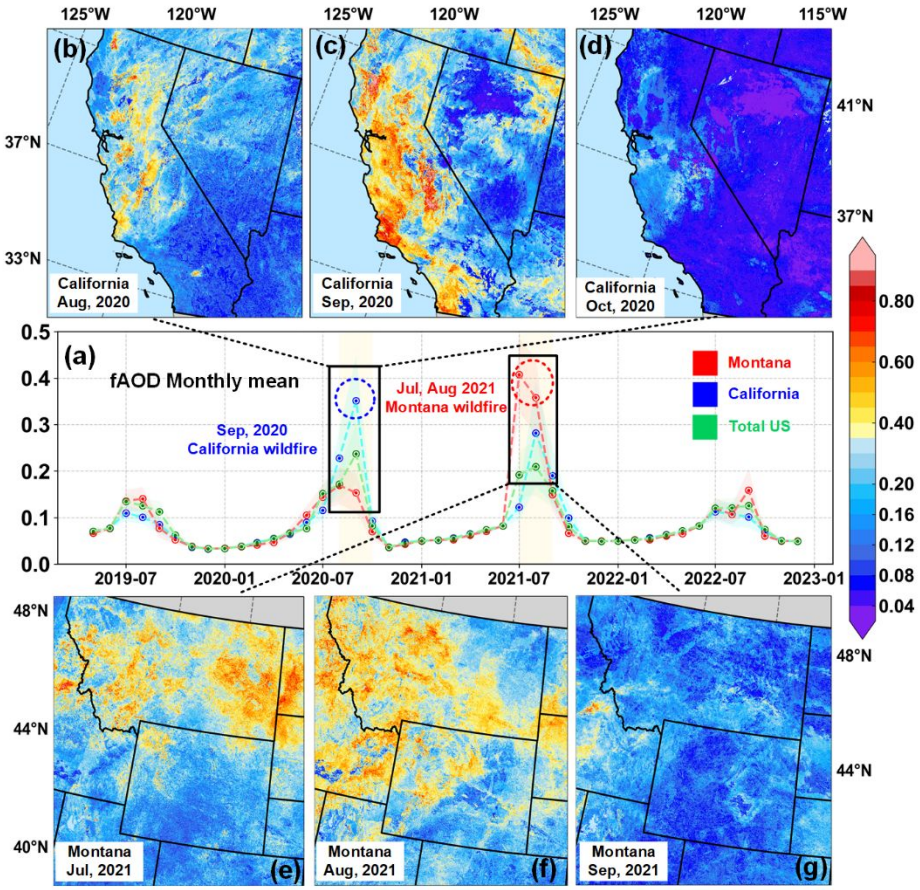

Figure S7. (a) Monthly mean fAOD estimates in California and Montana, USA from May 2019 to December 2022. Two wildfires are highlighted by black rectangles. Spatial distributions of monthly mean fAOD (b-d) in California from August to October 2020 and (e-g) in Montana from July to September 2021.

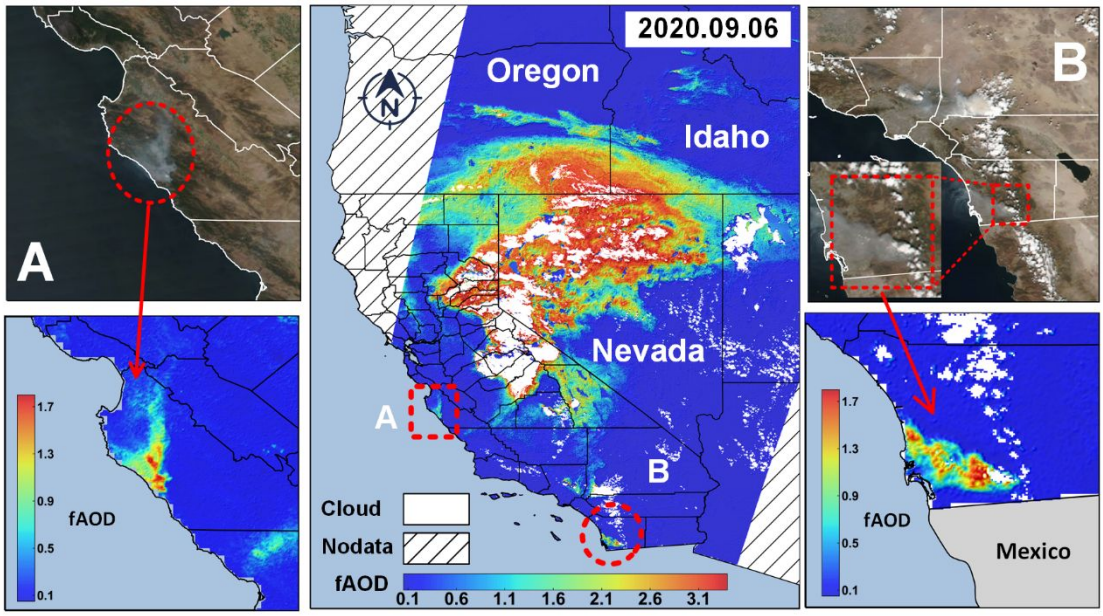

Figure S8. Spatial detail of fAOD chiefly caused by biomass combustion during wildfires on 9 September 2020.

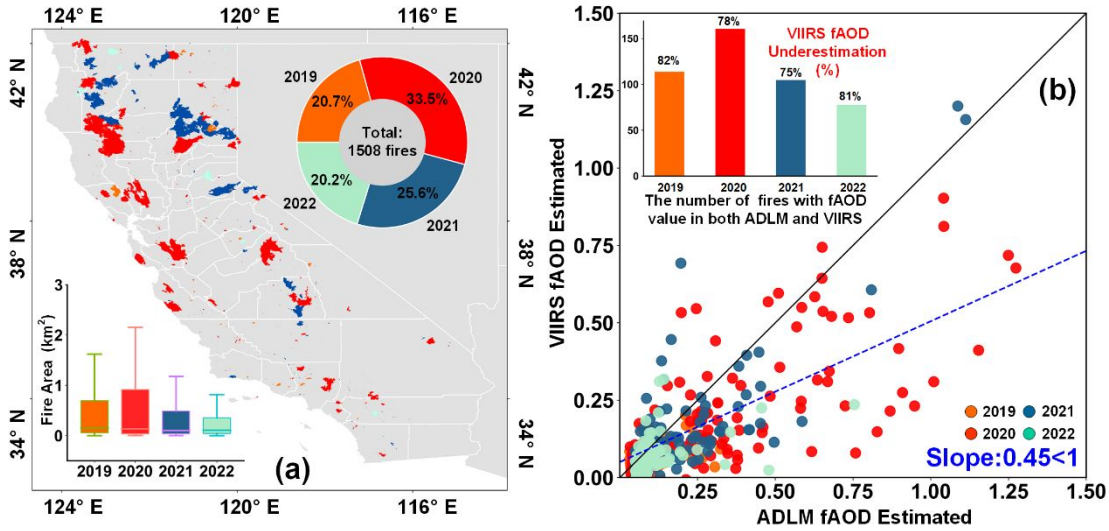

Figure S9. (a) Fire perimeter data (FRAP) in California for the years 2019 to 2022. The pie chart provides the percentage of fire occurrences by year. The boxplot inset figure shows the areal coverage of fire-affected areas from 2019 to 2022. **(b)** VIIRS fAOD as a function of the ADLM fAOD product for individuals fires in the FRAP dataset. The solid black and blue lines are the 1:1 line and the best-fit line from linear regression, respectively. Histograms show the number (and percentage) of fire events where VIIRS fAOD is underestimated compared to ADLM estimates.

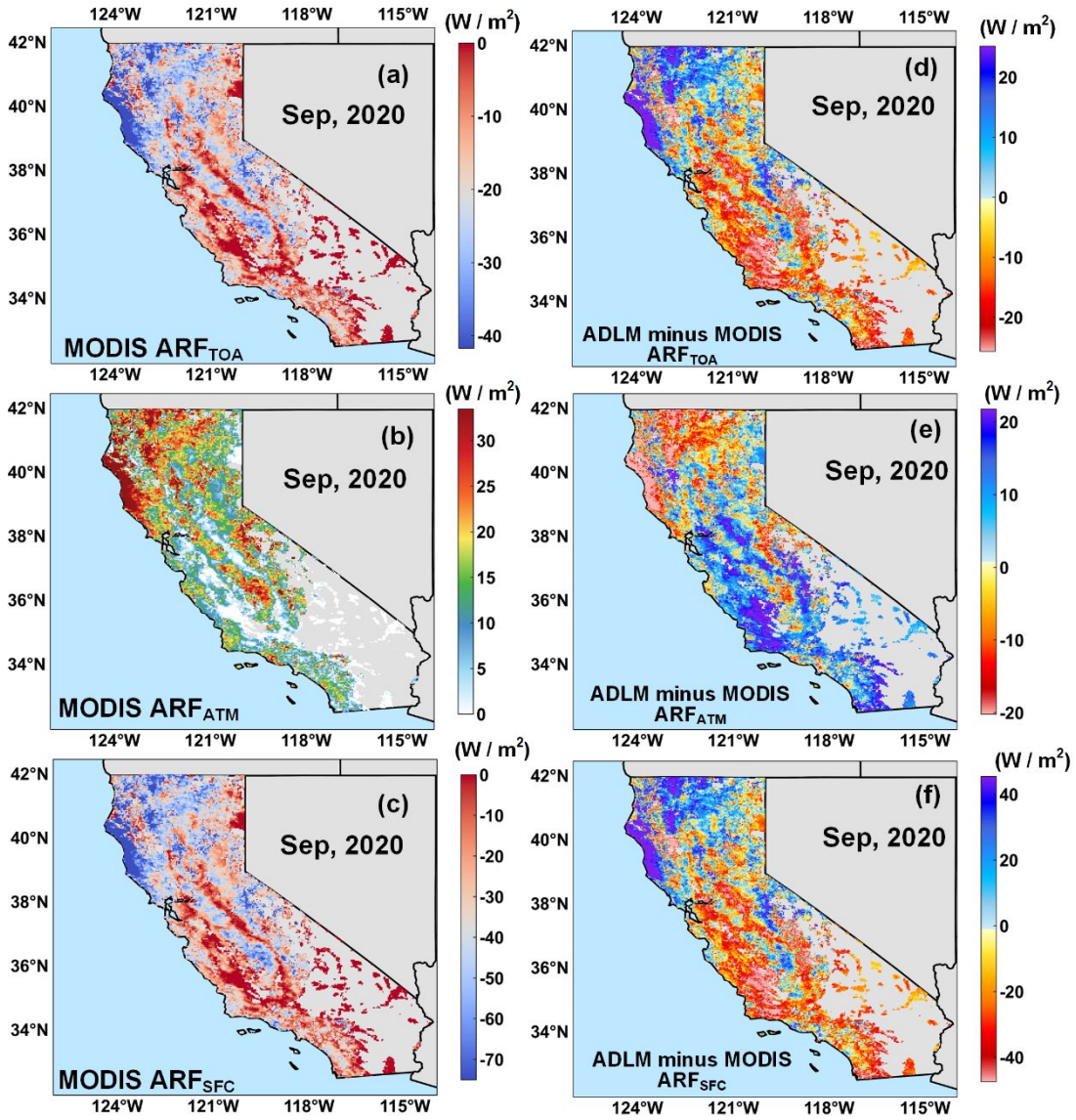

Figure S10. Mean fAOD radiative forcing from MODIS during the California wildfires in September 2020 is presented at the top of the atmosphere (a), within the atmosphere (b), and at the surface (c). The biases in fAOD radiative forcing with ADLM are depicted in d, e, and f.

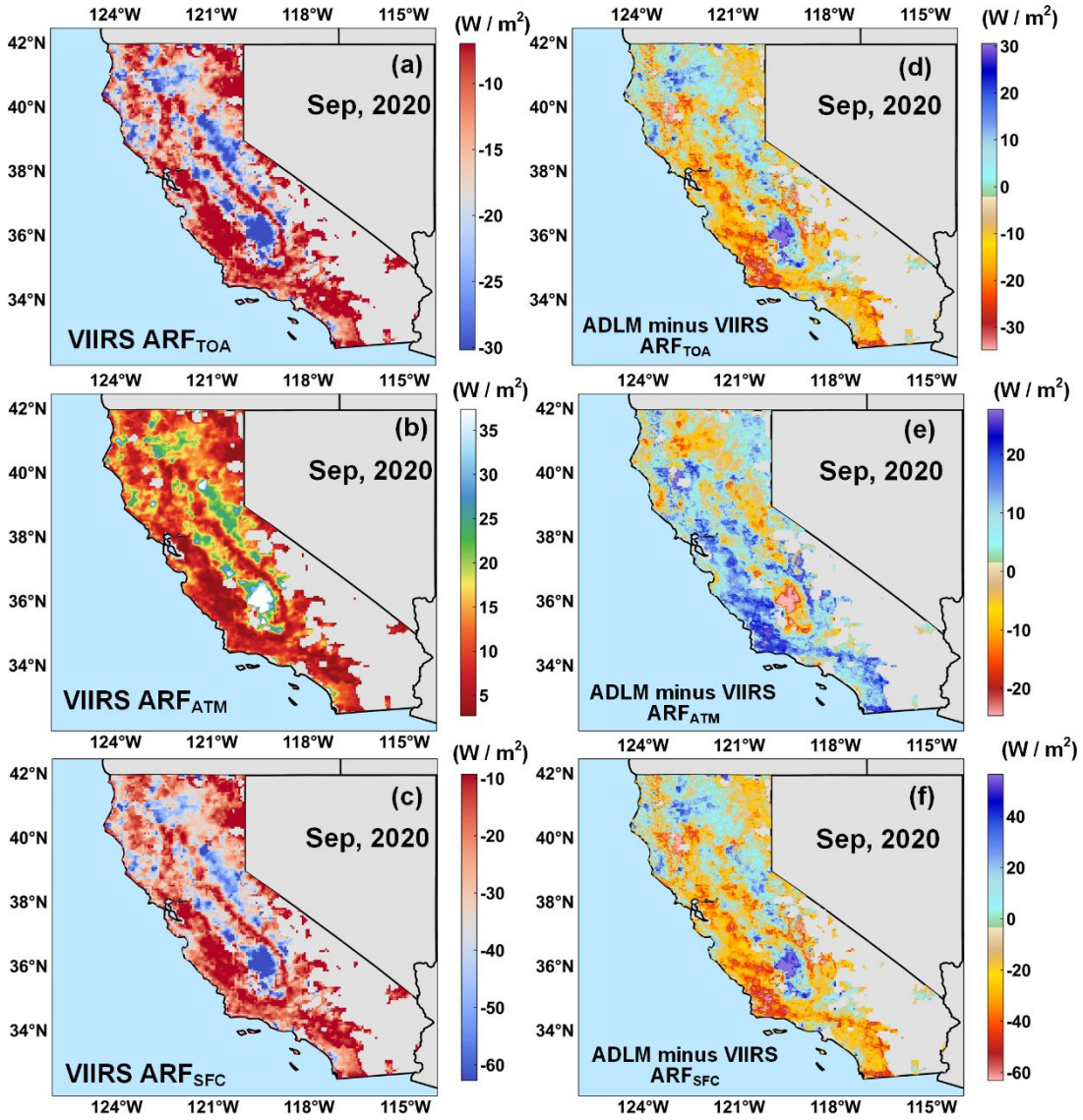

Figure S11. Same as Figure S10 but for VIIRS with ADLM comparison.

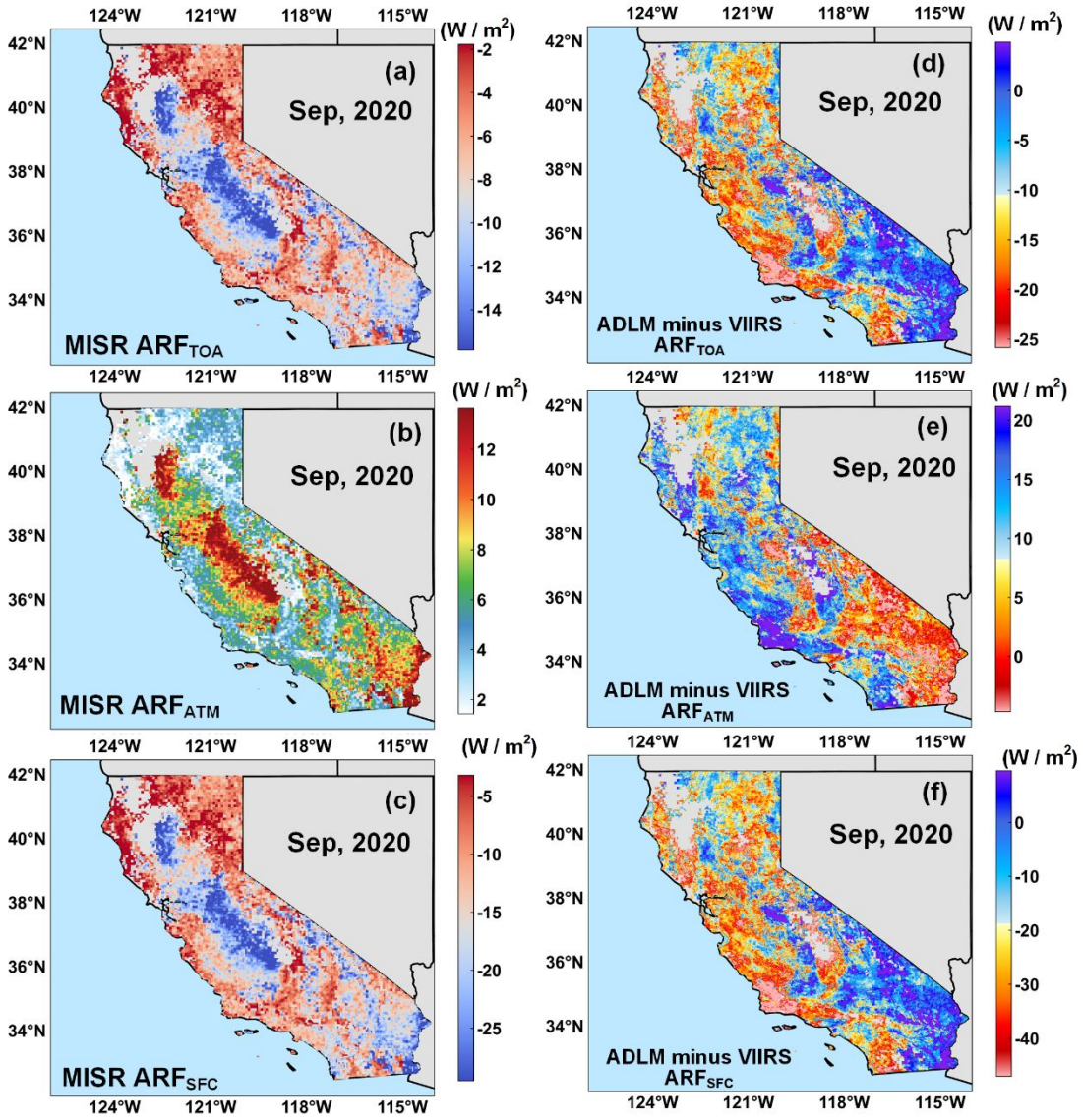

Figure S12. Same as Figure S10 but for MISR with ADLM comparison.

References

1. Kaufman, Y. J. et al. Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. *Journal of Geophysical Research, Washington, DC*. **102**, 17051-17067 (1997).
2. Ishida, H., Oishi, Y., Morita, K., Moriwaki, K. & Nakajima, T. Y. Development of a support vector machine based cloud detection method for MODIS with the adjustability to various conditions. *Remote Sens. Environ.* **205**, 390-407 (2018).
3. Oishi, Y., Ishida, H., Nakajima, T. Y., Nakamura, R. & Matsunaga, T. Preliminary verification for application of a support vector machine-based cloud detection method to GOSAT-2 CAI-2. *Atmos. Meas. Tech.* **11**, 2863-2878 (2018).
4. Arai, K., Iisasa, Y. & Liang, X. Aerosol parameter estimation with changing observation angle of ground based polarization radiometer. *Adv. Space Res.* **39**, 28-31 (2007).
5. Yan, X. et al. A minimum albedo aerosol retrieval method for the new-generation geostationary meteorological satellite Himawari-8. *Atmos. Res.* **207**, 14-27 (2018).
6. Kaufman, Y. J. et al. Dust transport and deposition observed from the Terra-Moderate Resolution Imaging Spectroradiometer (MODIS) spacecraft over the Atlantic Ocean. *Journal of Geophysical Research - Atmospheres*. **110**, D10S-D12S (2005).
7. Kahn, R. A. & Gaitley, B. J. An analysis of global aerosol type as retrieved by MISR. *Journal of geophysical research. Atmospheres*. **120**, 4248-4281 (2015).
8. Li, W. et al. Comprehensive Validation and Comparison of Three VIIRS Aerosol Products over the Ocean on a Global Scale. *Remote sensing (Basel, Switzerland)*. **14**, 2544 (2022).