Supplement of Earth Syst. Sci. Data, 14, 1193–1213, 2022
https://doi.org/10.5194/essd-14-1193-2022-supplement
© Author(s) 2022. CC BY 4.0 License.

Supplement of

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

Xing Yan et al.

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

The copyright of individual parts of the supplement might differ from the article licence.

Supplementary data

1. The parameters in Eq.(1)

The parameters in the Eq.(1) are same as those described by O'Neill et al. (2010):

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

All parameters are:

$$\begin{cases} \mathrm{a} = \left(a_{lower} + a_{upper}\right)/2 \\ a_{upper} = -0.22 \\ a_{lower} = -0.3 \end{cases}$$

$$\begin{cases} \mathrm{b^*} = \mathrm{b} + 2\alpha_{\mathrm{c}}\,\mathrm{a} \\ b = \left(b_{lower} + b_{upper}\right)/2 \\ b_{upper} = 10^{-0.2388}\lambda^{1.0275} \\ b_{lower} = 0.8 \end{cases}$$

where λ is reference wavelength (μm), in this study is 0.5 μm.

$$\begin{cases} \mathrm{c^*} = \mathrm{c} + (\mathrm{b} + \mathrm{a}\,\alpha_{\mathrm{c}})\alpha_{\mathrm{c}} - \alpha_{\mathrm{c}}' \\ c = \left(c_{lower} + c_{upper}\right)/2 \\ c_{upper} = 10^{0.2633}\lambda^{-0.4683} \\ c_{lower} = 0.63 \end{cases}$$

$$\alpha_{\mathrm{c}} = -0.15 \text{ and } \alpha_{\mathrm{c}}' = 0$$

2. α' bias error correction

This study used O'Neill et al. (2003) Appendix A1 to correct the α' bias and propagate this correction through all derived parameters:

$$\alpha'_{error} = 0.65 \times exp[-(FMF^1 - 0.78)^2 / (2 \times 0.18^2)]$$

where FMF^1 is the uncorrected estimate of FMF as shown in Eq. (2) of the main paper. Then

$$\alpha'_{corrected} = \alpha'^{1} + \alpha'_{error} \quad ,$$

$$t_{corrected} = \alpha - \alpha_c - \frac{\alpha'_{corrected} - \alpha_c'}{\alpha - \alpha_c} \quad ,$$

$$\mathrm{D}_{\mathrm{corrected}} = \sqrt{(\mathrm{t}_{\mathrm{corrected}} + \mathrm{b}^*)^2 + 4(1 - \mathrm{a})\,\mathrm{c}^*} \quad ,$$

$$\alpha_{f_{\mathrm{corrected}}} = \frac{1}{2(1 - \mathrm{a})}\left(\mathrm{t}_{\mathrm{corrected}} + \mathrm{b}^* + \mathrm{D}_{\mathrm{corrected}}\right) + \alpha_c \quad ,$$

$$FMF_{\mathrm{corrected}} = \frac{\alpha - \alpha_c}{\alpha_{f_{\mathrm{corrected}}} - \alpha_c} \quad .$$

3. Mean of extreme (MOE) modification

The error of α_f derived by SDA is (O'Neill et al., 2003):

$$\Delta\alpha_f^{\,2} = \left(\mathrm{k}_1 \frac{\partial \alpha_f}{\partial \alpha'} + \mathrm{k}_2 \frac{\partial \alpha_f}{\partial \alpha}\right)^2 \left(\frac{\Delta\tau_{\mathrm{a}}}{\tau_{\mathrm{a}}}\right)^2 + \left(\frac{\partial \alpha_f}{\partial \mathrm{a}} \Delta \mathrm{a}\right)^2 + \left(\frac{\partial \alpha_f}{\partial \mathrm{b}} \Delta \mathrm{b}\right)^2 + \left(\frac{\partial \alpha_f}{\partial \mathrm{c}} \Delta \mathrm{c}\right)^2 + \left(\frac{\partial \alpha_f}{\partial \alpha'_{\mathrm{c}}} \Delta \alpha'_{\mathrm{c}}\right)^2 + \left(\frac{\partial \alpha_f}{\partial \alpha_{\mathrm{c}}} \Delta \alpha_{\mathrm{c}}\right)^2$$

where $\mathrm{k}_1 = 10$, $\mathrm{k}_2 = -2.5$, $\Delta\tau_{\mathrm{a}}$ is the nominal root mean square error in AOD at the reference wavelength, τ_{a} is the AOD at the reference wavelength (this study is at 0.5 μm AOD), $\Delta\alpha'_{\mathrm{c}} = 0.15$, $\Delta\alpha_{\mathrm{c}} = 0.15$, and

$$\begin{cases} \Delta \mathrm{a} = (\mathrm{a}_{\mathrm{upper}} - \mathrm{a}_{\mathrm{lower}})/2 \\ \Delta \mathrm{b} = (\mathrm{b}_{\mathrm{upper}} - \mathrm{b}_{\mathrm{lower}})/2 \\ \Delta \mathrm{c} = (\mathrm{c}_{\mathrm{upper}} - \mathrm{c}_{\mathrm{lower}})/2\,. \end{cases}$$

In $\Delta\alpha_f^{\,2}$,

$$\frac{\partial \alpha_f}{\partial \alpha'} = \frac{-1}{FMF_{corrected}\ \mathrm{D}_{corrected}} ,$$

$$\frac{\partial \alpha_f}{\partial \alpha} = \frac{\mathrm{t}_{+}}{FMF_{corrected}\ \mathrm{D}_{corrected}} \quad ,$$

$$\mathrm{t}_{+} = \alpha - \alpha_c - \frac{\alpha'_{corrected} - \alpha_c'}{\alpha - \alpha_c} \quad ,$$

$$\frac{\partial \alpha_f}{\partial \mathrm{a}} = \frac{(\alpha_{f_{corrected}} - \alpha_\mathrm{c})}{(1 - \mathrm{a})} + \frac{1}{\mathrm{D}_{corrected}} \left(\alpha_\mathrm{c} (2\alpha_{f_{corrected}} - \alpha_\mathrm{c}) - \frac{\mathrm{c}^*}{(1 - \mathrm{a})} \right) \quad ,$$

$$\frac{\partial \alpha_f}{\partial \mathrm{b}} = \frac{\alpha_{f_{corrected}}}{\mathrm{D}_{corrected}} \quad ,$$

$$\frac{\partial \alpha_f}{\partial \mathrm{c}} = \frac{1}{\mathrm{D}_{corrected}} \quad ,$$

$$\frac{\partial \alpha_f}{\partial \alpha'_\mathrm{c}} = \frac{1}{\mathrm{D}_{corrected}} \left(\frac{1}{FMF_{corrected}} - 1 \right) \quad ,$$

$$\frac{\partial \alpha_f}{\partial \alpha_\mathrm{c}} = \frac{\mathrm{t}_{corrected}}{\mathrm{D}_{corrected}} \left(\frac{1}{FMF_{corrected}} - 1 \right) .$$

When we obtain the $\Delta\alpha_f$ $(=\sqrt{\Delta\alpha_f^{\ 2}})$, the SDA set the theoretical maximum of α_f is:

$\alpha_{f\,TMAX} = min(4, 10^{(0.18*log10(\lambda)+0.57)})$.

Then:

$$\alpha_{f\,MAX} = \alpha_{f_{corrected}} + \Delta\alpha_f$$

$$\alpha_{f\,Min} = \alpha_{f_{corrected}} - \Delta\alpha_f$$

If $\alpha_{f\,MAX} > \alpha_{f\,TMAX}$, $\alpha_{f\,MAX} = \alpha_{f\,TMAX}$.

If $\alpha_{f\,Min} > \alpha_{f\,TMAX}$, $\alpha_{f\,Min} = \alpha_{f\,TMAX}$.

The final output of corrected FMF (FMF_{output}) is:

if $_{f\,Min} < \alpha$ or $\alpha_c + \Delta\alpha_c > \alpha$

if $\alpha_{f\,Min} < \alpha$ and $\alpha_{f\,MAX} > \alpha$

$$b_2 = \frac{0.5 - 2/m}{2\Delta\alpha_f^{\,2}}$$

$$b_1 = \frac{1}{m \times \Delta\alpha_f} - (2\alpha - \Delta\alpha_f)b_2$$

$$b_0 = -(\alpha - \Delta\alpha_f)b_1 - (\alpha - \Delta\alpha_f)^2 \times b_2$$

$$\Omega = b_0 + b_1\alpha_{f_{corrected}} + b_2(\alpha_{f_{corrected}})^2$$

$$\alpha^1_{fcorrected} = \Omega \times \alpha_{f\,MAX} + (1-\Omega)\,\alpha$$

elseif $\alpha_{f\,MAX} \le \alpha$

$$\alpha^1_{fcorrected} = \alpha$$

end

if $\alpha < \alpha_c + abs(\Delta\alpha)$ and $\alpha > \alpha_c - abs(\Delta\alpha)$

$$c_2 = \frac{0.5 - 2/m}{2\Delta\alpha^2}$$

$$c_1 = \frac{1}{m \times abs(\Delta\alpha)} - (2\alpha_c + abs(\Delta\alpha))c_2$$

$$c_0 = -(\alpha_c - abs(\Delta\alpha))c_1 - (\alpha_c + abs(\Delta\alpha))^2 \times c_2$$

$$\Omega = c_0 + c_1\alpha + c_2\alpha^2$$

$$\alpha_c = \Omega(\alpha - abs(\Delta\alpha)) + (1-\Omega) \times (-0.15)$$

elseif $\alpha \le \alpha_c - abs(\Delta\alpha)$

$$\alpha_c = \alpha$$

end

$$FMF_{\mathrm{output}} = \frac{\alpha - \alpha_c}{\alpha^1_{fcorrected} - \alpha_c}$$

where $m = 8$ and $\Delta\alpha = \mathrm{k}_2 \dfrac{\Delta\tau_{\mathrm{a}}}{\tau_{\mathrm{a}}}$.

4. FMF frequency

To validate and study the characteristics of FMF, three levels of FMF were defined in this study (low level: FMF<0.5, medium level: 0.5<FMF<0.8, high level: FMF>0.8). The frequency for a certain level of FMF is define as:

$$F_{FMF_{bin}} = \frac{N_{FMF_{bin}}}{N_{FMF_{all}}} \times 100\%$$

Where $F_{FMF_{bin}}$ is the frequency of FMF in a certain level bin, $N_{FMF_{bin}}$ represents the total amount of FMF sample within this level bin, and $N_{FMF_{all}}$ represents the total amount of FMF sample.

Table S1. Data used for Phy-DL FMF retrieval

Name	**MOD02SSH**	**MOD09CMG**	**MOD08_D3**	**ERA5**
Data version	MODIS C6.1 L1B	MODIS C6.1 L3	MODIS C6.1 L3	reanalysis-era5-single-levels
Domain	-90~90°N, -180~180°E	-90~90°N, -180~180°E	-90~90°N, -180~180°E	-90~90°N, -180~180°E
Spatial resolution	5 km×5 km	0.05°×0.05°	1°×1°	0.25°×0.25°
Product used	TOA reflectance data: Band 1-Band 7	Surface Reflectance: Band 1-Band 7, Brightness_Temperature: Band 20 (3.360-3.840 μm) Band 21 (3.929-3.989 μm) Band 31 (10.780-11.280 μm) Band 32 (11.770-12.270 μm) Relative_Azimuth_Angle,	Aerosol_Optical_Depth_Land _Mean (at 500nm, calculated by MODIS DT-based Ångstrom exponent)	'10m_u_component_of_wind', '10m_v_component_of_wind', '2m_dewpoint_temperature', '2m_temperature', 'boundary_layer_height', 'surface_pressure',

		Solar_Zenith_Angle, View_Zenith_Angle		
Data access	https://ladsweb.modaps.eosdis.nasa.gov/search/	https://e4ftl01.cr.usgs.gov/MOLT/MOD09CMG.061/		https://climate.copernicus.eu/climate-reanalysis
Reference	http://dx.doi.org/10.5067/MODIS/MOD0SSH.061	Vermote (2015)	Platnick et al. (2015)	Hersbach et al. (2020)

Table S2. The sites from SURFRAD used for out of site validation and their locations.

Sites	Longitude	Latitude	Land type
Desert Rock (DRA)	-116.02	36.62	Barren or sparse
Fort Peck (FPK)	-105.10	48.31	Grasslands
Goodwin Creek (GWN)	-89.87	34.25	Woody savannas
Penn State (PSU)	-77.93	40.72	Mixed forests

Table S3. FMF data used for the comparison.

Name	**POLDER**	**MISR**	**MODIS**
Data version	POLDER/GRASP high precision v1.2 L3	MIL3DAEN.004	MODIS C5 MOD08
Domain	-70~69°N, -180~179°E	-89.75~89.8°N, -180~179.75°E	-90~90°N, -180~180°E
Spatial resolution	1°×1°	0.5°×0.5°	1°×1°
Product used	AODF490, AOD490	Small_Mode_Aerosol_Optical_Depth, Aerosol_Optical_Depth	Optical_Depth_Ratio_Small_Land
Data access	https://download.grasp-cloud.com/download/polder/polder-3/	https://asdc.larc.nasa.gov/data/MISR/	
Reference	Dubovik et al. (2014)	Garay et al. (2020)	Levy et al. (2007)

Table S4. The land types and corresponded value from MODIS MCD12C1 data (the International Geosphere-Biosphere Programme scheme).

Value	Land type	Value	Land type
1	Evergreen needleleaf	9	Savannas
2	Evergreen broadleaf	10	Grasslands
3	Deciduous needleleaf	11	Permanent wetlands
4	Deciduous broadleaf	12	Croplands
5	Mixed forests	13	Urban and built up
6	Closed shrubland	14	Crop natural vegetation mosaic
7	Open shrublands	15	Snow and ice
8	Woody savannas	16	Barren or sparse

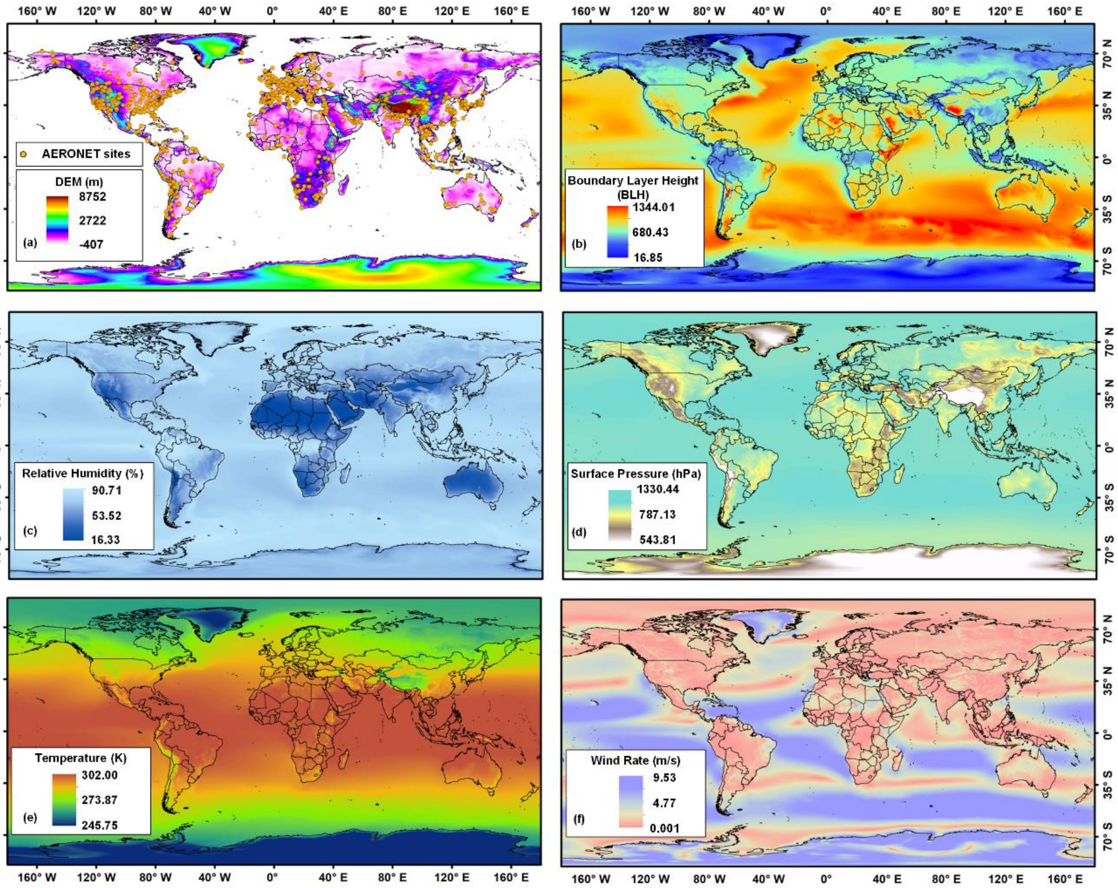

Figure S1. (a) The distribution of Global digital elevation model [DEM; base map in (a)], AERONET sites [dots in (a)], annual mean boundary layer height (BLH) in 2001-2020 (b), annual mean relative humidity (RH) in 2001-2020 (b), annual mean surface pressure in 2001-2020 (c), annual mean temperature in 2001-2020 (d), annual mean wind rate in 2001-2020 (e) used in this study.

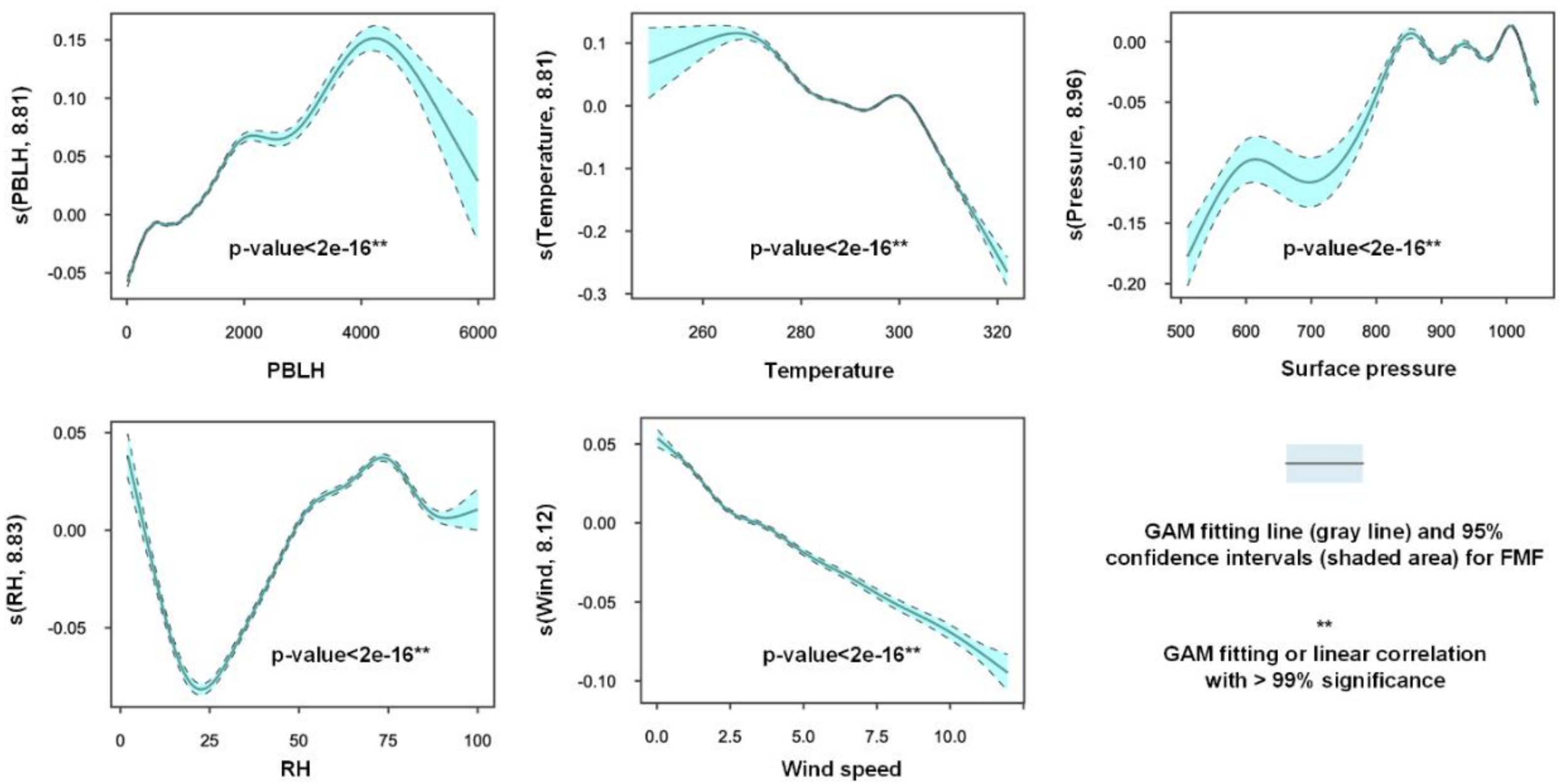

Figure S2. Generalized Additive Model (GAM) fitting plots for the meteorological variables and the FMF. Shaded areas in the GAM plots indicate 95% confidence intervals, and the y-axes shows the covariate and effective degrees of freedom of the smoothing. The asterisks (**) after each p-value indicate the 99% confidence interval of fitting.

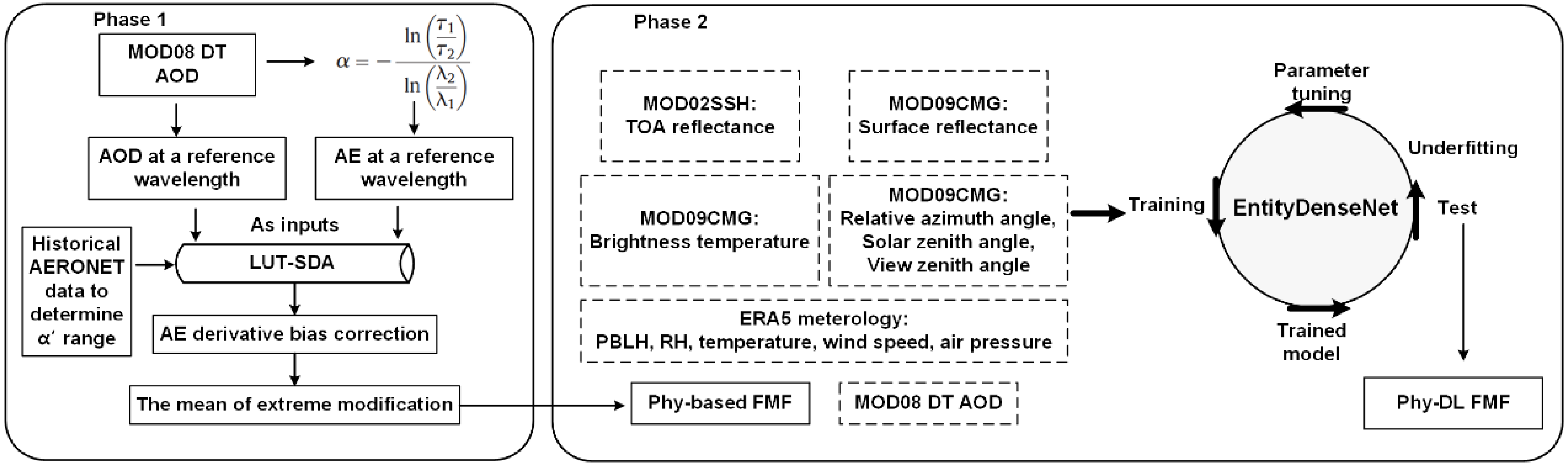

Figure S3. Schematic diagram describing the Phy-DL FMF calculation in this study.

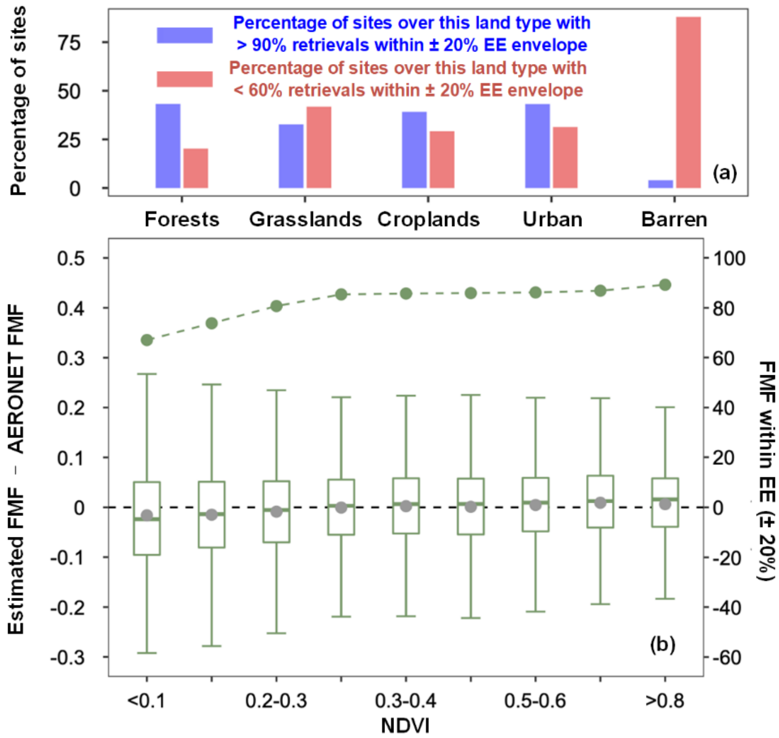

Figure S4. (a) Bar plots of the percentage of sites with > 90% of retrievals falling within the ±20% EE envelope (blue bars) and the percentage of sites with < 60% of retrievals falling within the ±20% EE envelope (red bars) for five land types. (b) Box plots of the FMF bias (estimated FMF minus AERONET FMF) as a function of NDVI. The black horizontal dashed line indicates the zero bias. The gray 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 green dots connected by the dashed curve are percentages of FMF retrievals falling within the EE envelope of ±20%.

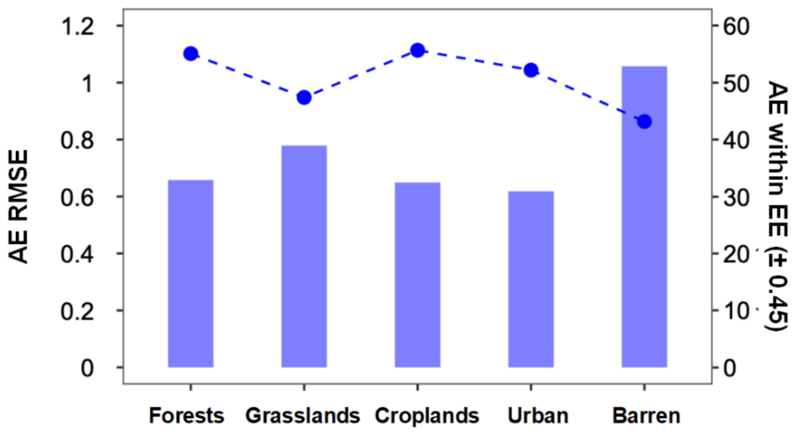

Figure S5. RMSEs (bars) and percentages of MOD08 AE falling within the EE envelope of ±0.45 (dash-dotted line) against AERONET observation for five land types. The EE envelope (±0.45) was adopted from Levy et al. (2013).

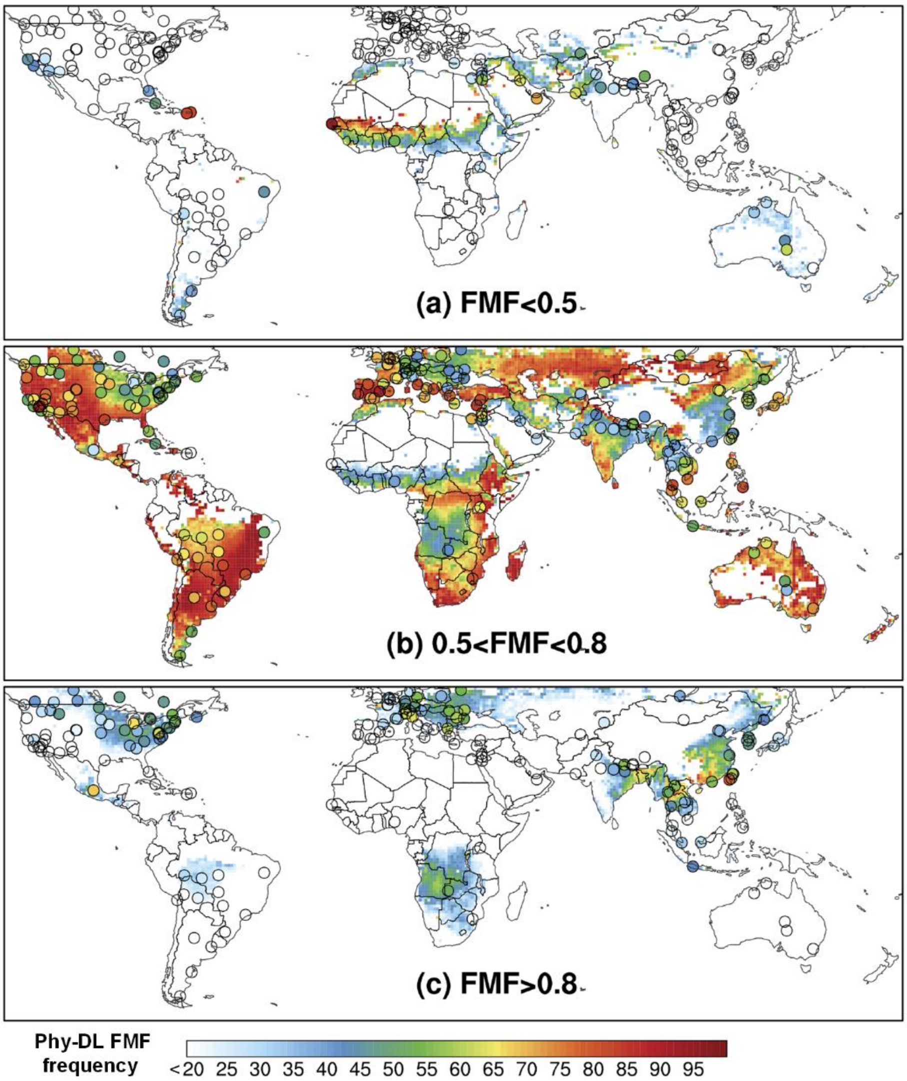

Figure S6. Frequencies of three FMF levels (low: FMF < 0.5, medium: 0.5< FMF <0.8, high: FMF > 0.8) calculated by Phy-DL (based map) and AERONET (dots) FMF during 2001 to 2020. Only pixels of Phy-DL with 120 retrievals/year and AERONET FMF covering more than 10 years were shown.

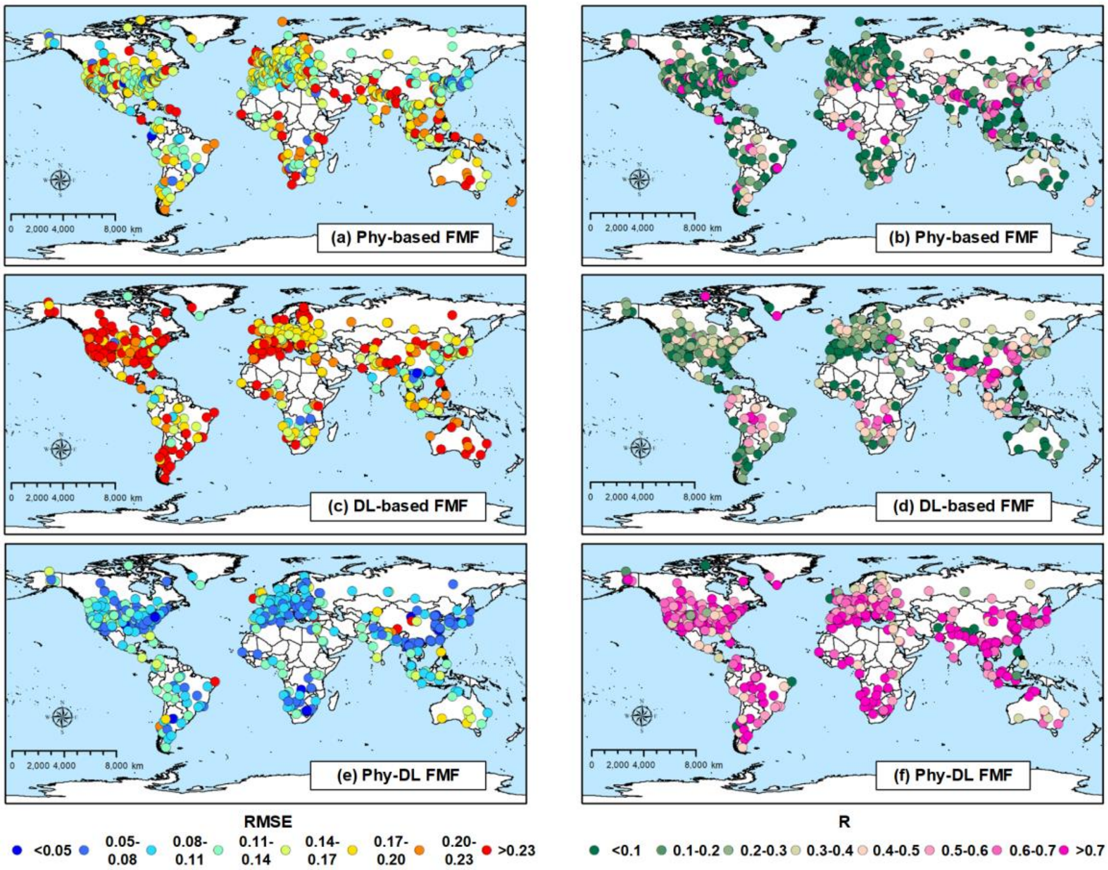

Figure S7. the validation statistics of Phy-based, DL-based and Phy-DL FMF against AERONET FMF over global AERONET sites for root mean squared error (RMSE; a, c, e) and correlation coefficient (R; b, d, f).

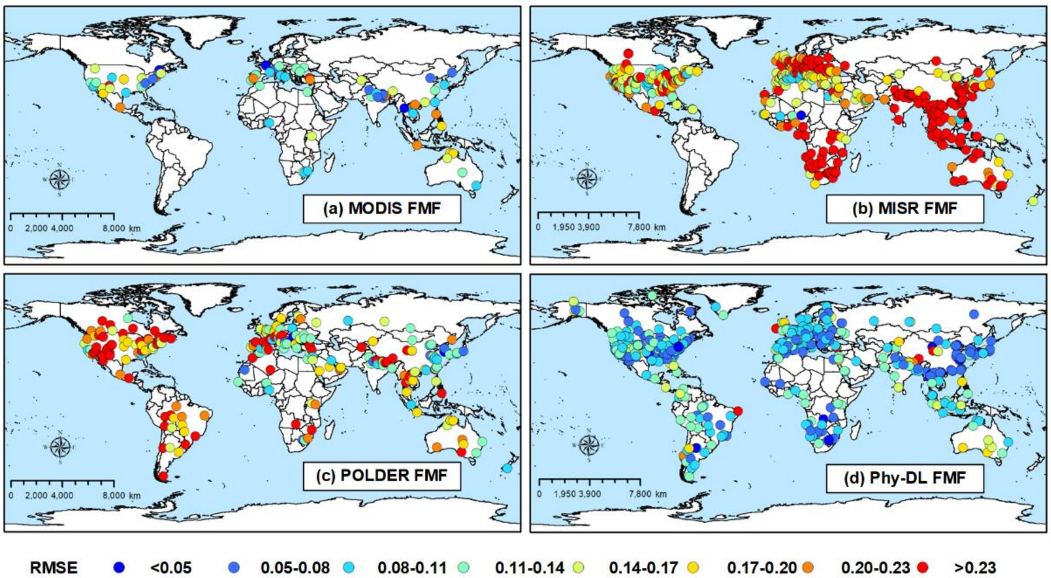

Figure S8. the validation statistics of MODIS, MISR, POLDER and Phy-DL FMF against AERONET FMF over global AERONET sites for RMSE.

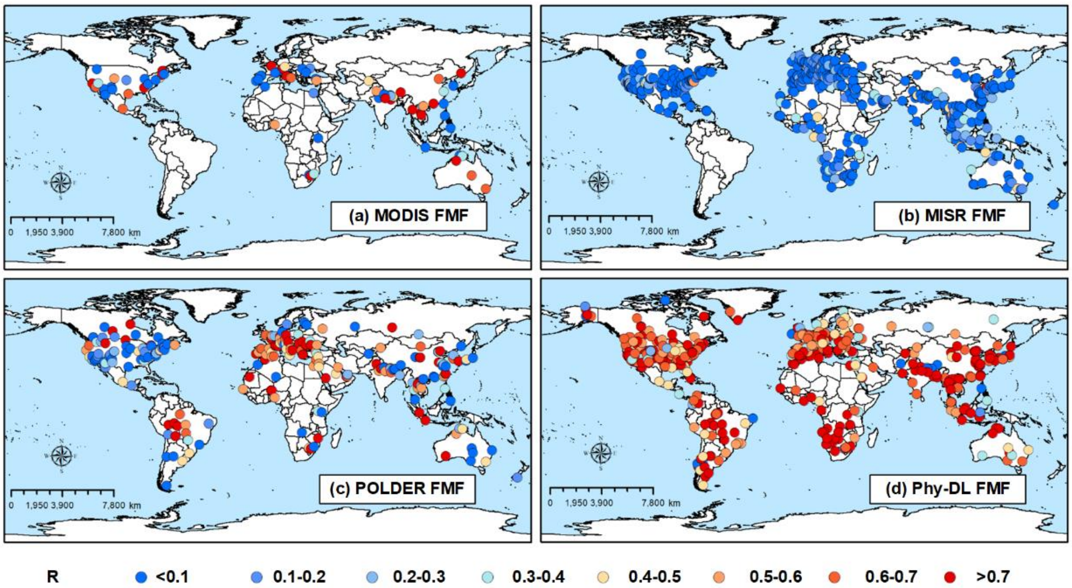

Figure S9. the validation statistics of MODIS, MISR, POLDER and Phy-DL FMF against AERONET FMF over global AERONET sites for R.

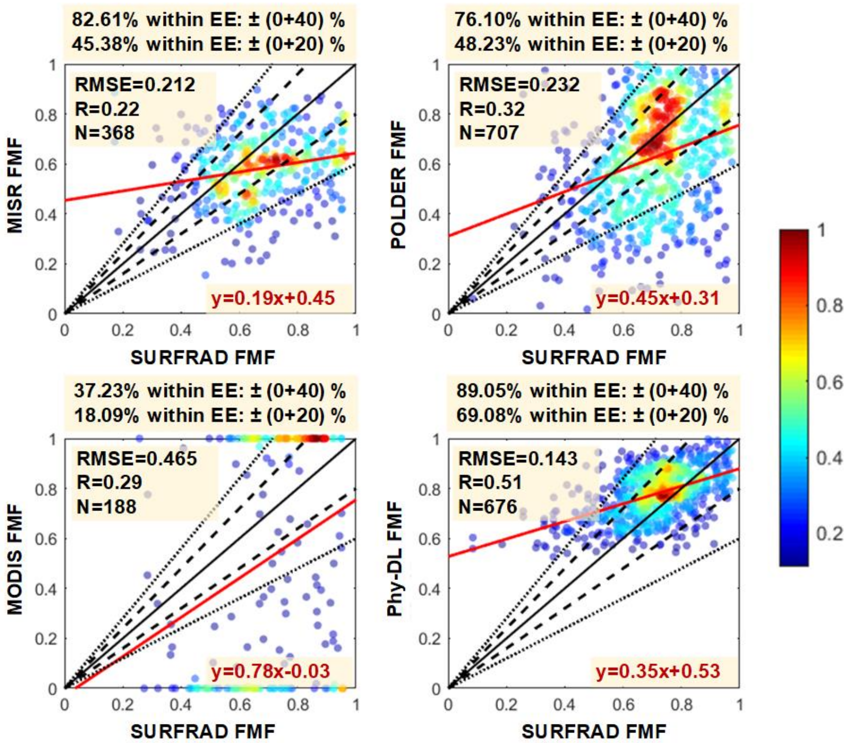

Figure S10. Evaluation of (a) MISR (550 nm), (b) POLDER (490 nm), (c) MODIS (550 nm), and (d) Phy-DL FMFs (500 nm) against SURFRAD FMFs (500 nm) from 2008 to 2013. Black and red solid lines are 1:1 reference lines and best-fit lines from linear regression, respectively. Black dashed and dotted lines represent the EE envelopes of ±20% and ±40%, respectively. The number of samples (N), root-mean-square error (RMSE), correlation coefficient (R), and linear regression relation are given in each panel.

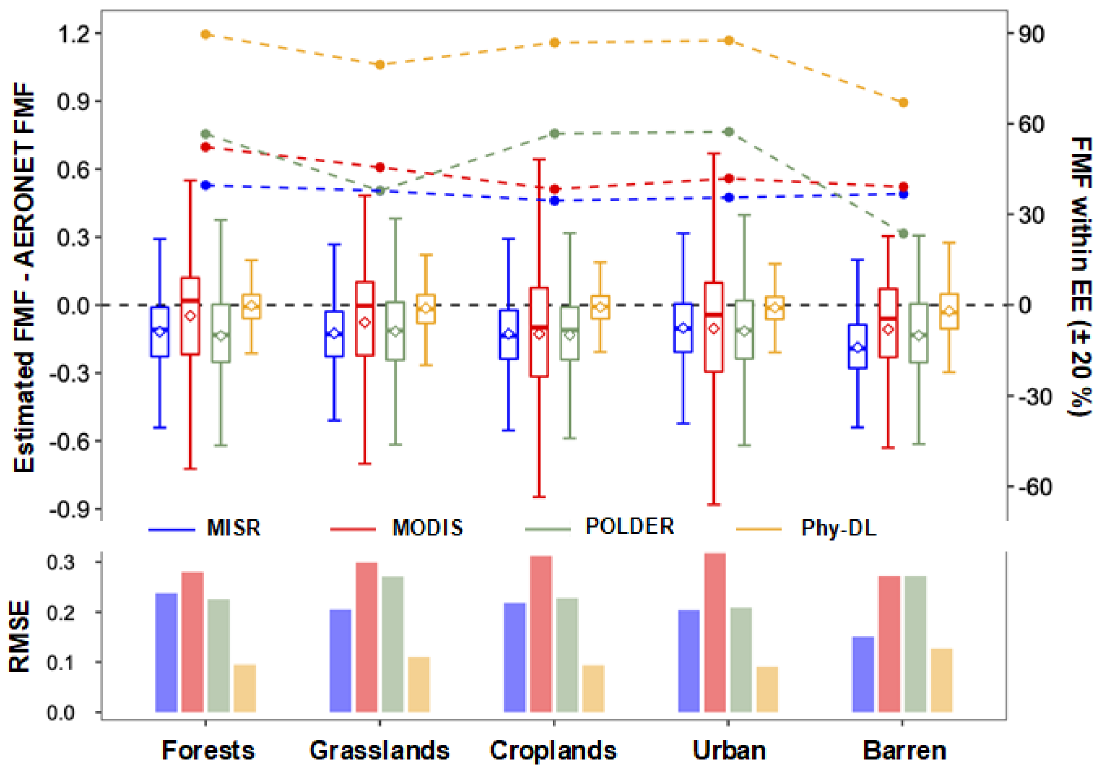

Figure S11. The MISR (blue), MODIS (red), POLDER (green) and Phy-DL FMF (orange) estimation compared with AERONET FMF (all at 500 nm, using data from 2008-2017). (a) The boxplots of bias (Estimated FMF minus AERONET FMF) and percentage of FMF estimations falls within EE of ±20% (dots and dashed lines) as the function of land types. 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 FMF bias. (b) the RMSE over each land type against AERONET FMF.

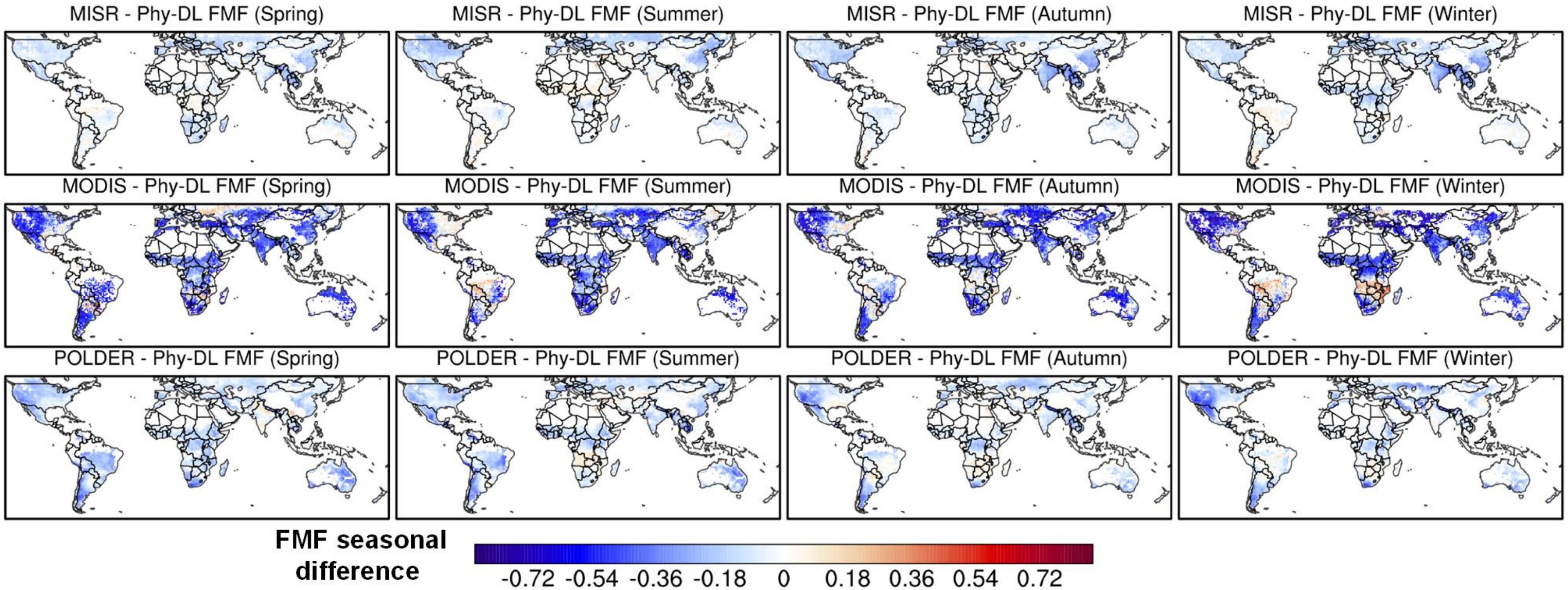

Figure S12. The seasonal mean differences of Phy-DL with MISR, MODIS and POLDER FMF during 2008-2013.

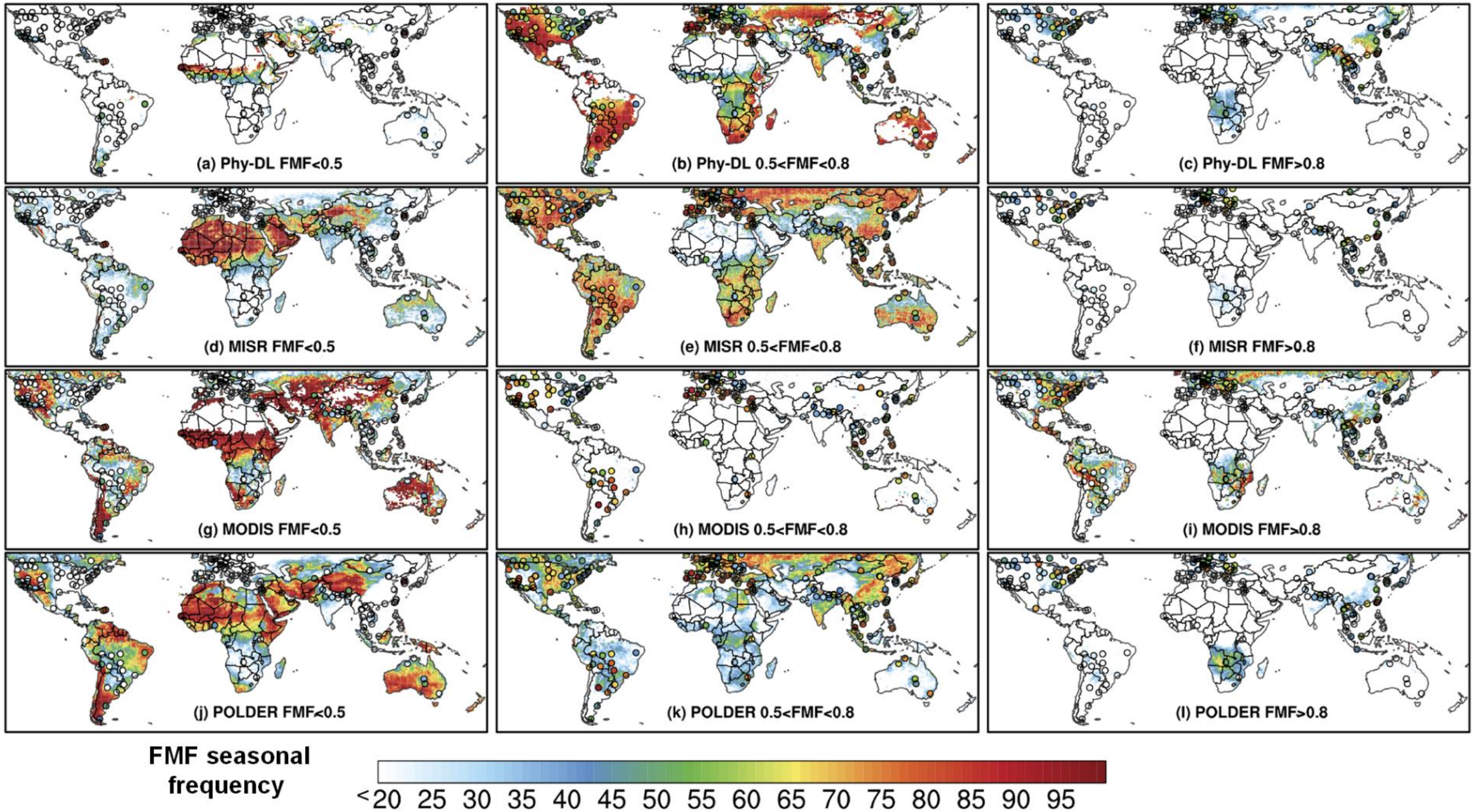

Figure S13. FMF frequency for three levels FMF (FMF<0.5, 0.5<FMF<0.8, FMF>0.8) calculated by Phy-DL, MISR, MODIS and POLDER (base maps) and AERONET (dots) during 2008-2013