Supplementary data

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. 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. 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)
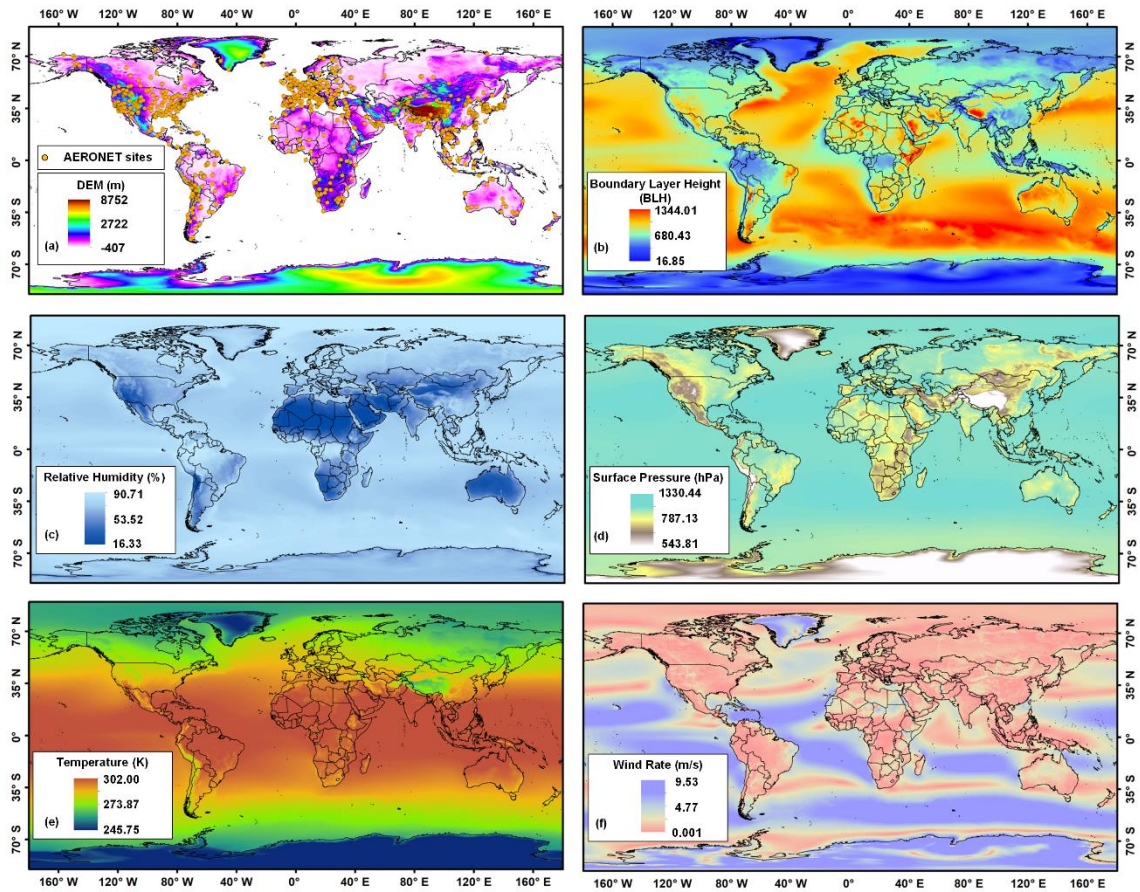
Figure S1. (a) The distribution of Global digital elevation model [DEM; base map in (a)], AERONET sites [dots in (a)], annual mean boundary layer height (BLH) in 2001-2020 (b), annual mean relative humidity (RH) in 2001-2020 (b), annual mean surface pressure in 2001-2020 (c), annual mean temperature in 2001-2020 (d), annual mean wind rate in 2001-2020 (e) used in this study.
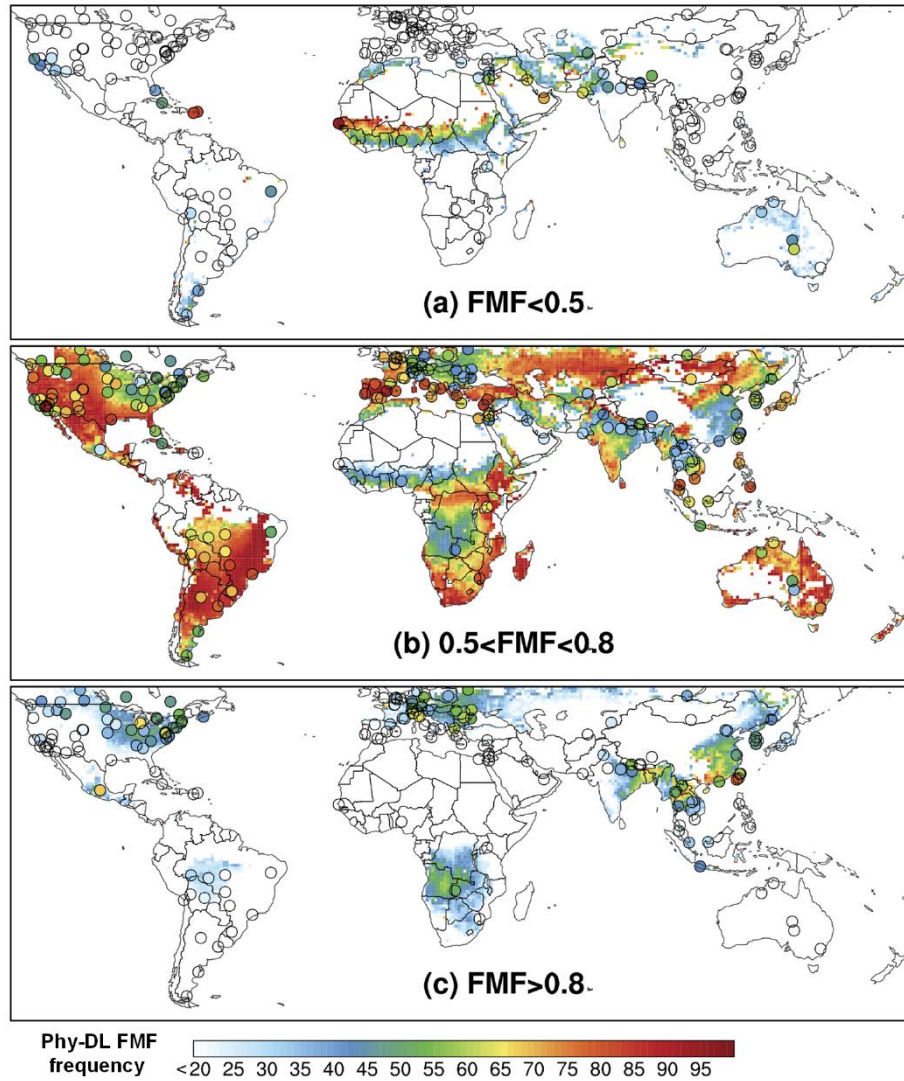
Figure S2. Frequencies of three FMF levels (low: FMF < 0.5, medium: 0.5< FMF <0.8, high: FMF > 0.8) calculated by Phy-DL (based map) and AERONET (dots) FMF during 2001 to 2020. Only pixels of Phy-DL with 120 retrievals/year and AERONET FMF covering more than 10 years were shown.
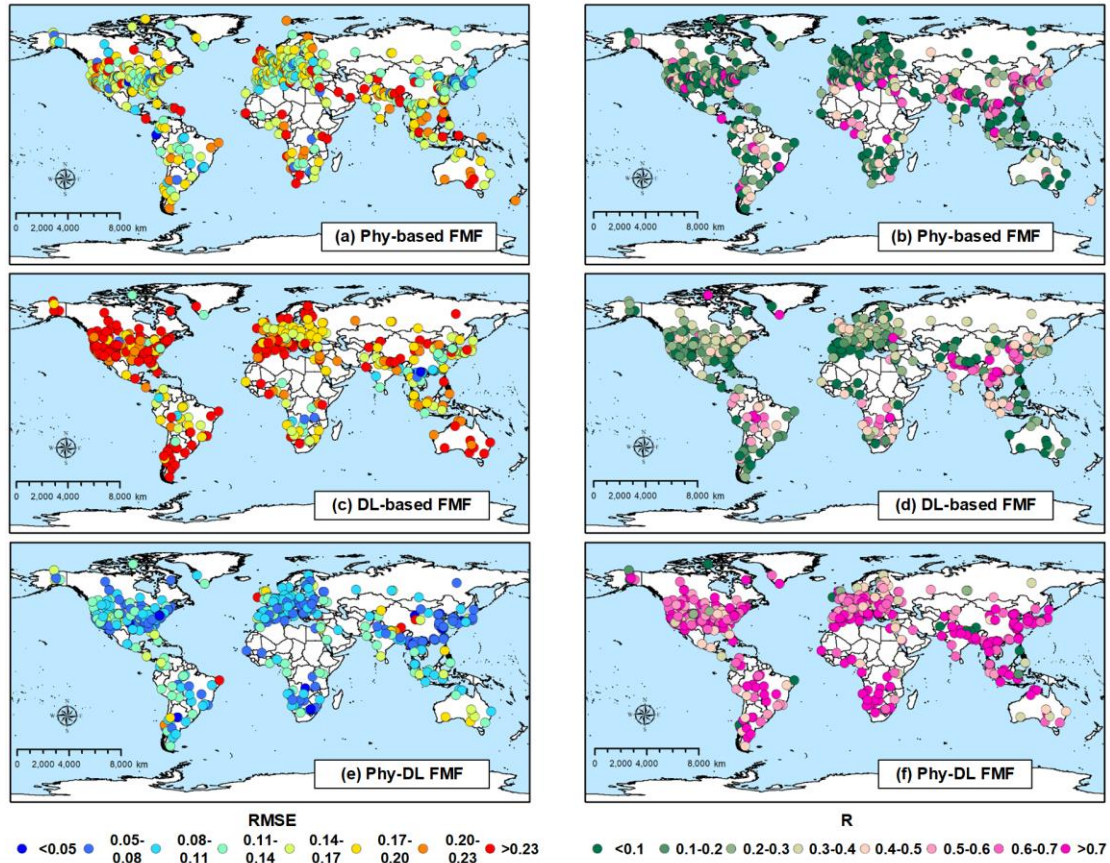
Figure S3. the validation statistics of Phy-based, DL-based and Phy-DL FMF against AERONET FMF over global AERONET sites for root mean squared error (RMSE; a, c, e) and correlation coefficient (R; b, d, f).
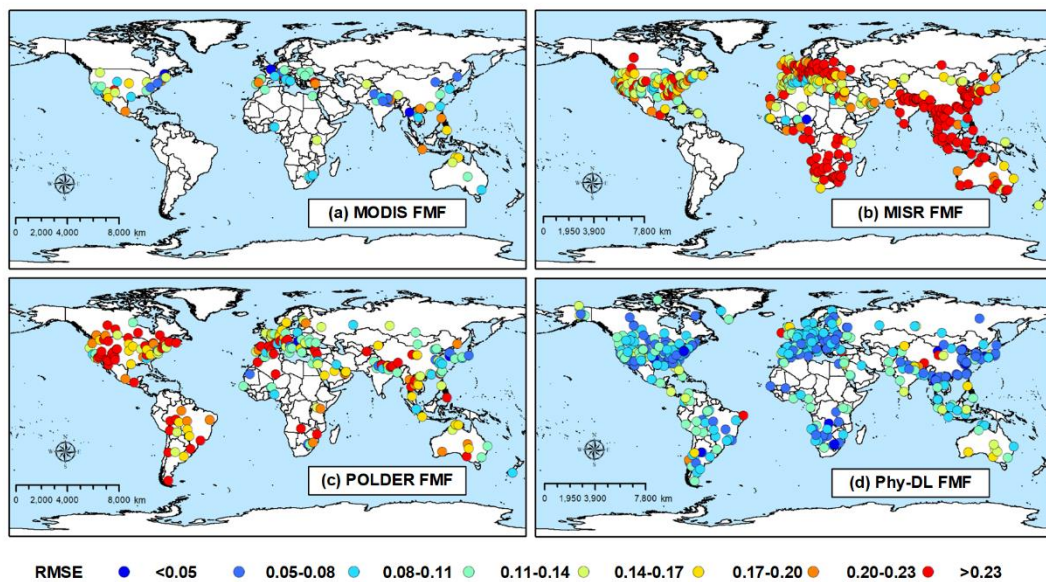

Figure S4. the validation statistics of MODIS, MISR, POLDER and Phy-DL FMF against AERONET FMF over global AERONET sites for RMSE.
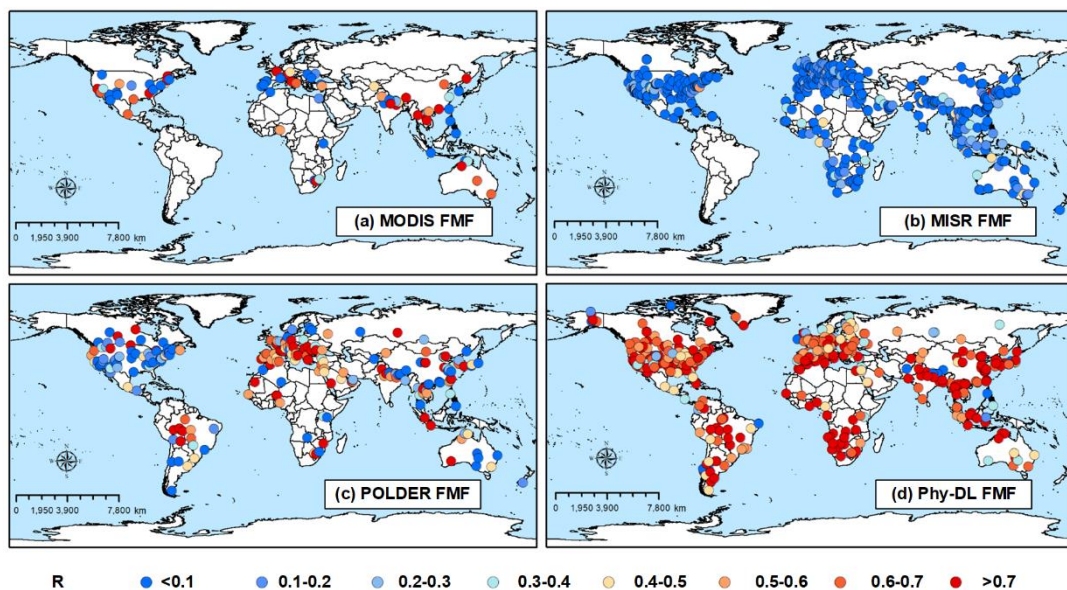
Figure S5. the validation statistics of MODIS, MISR, POLDER and Phy-DL FMF against AERONET FMF over global AERONET sites for R.
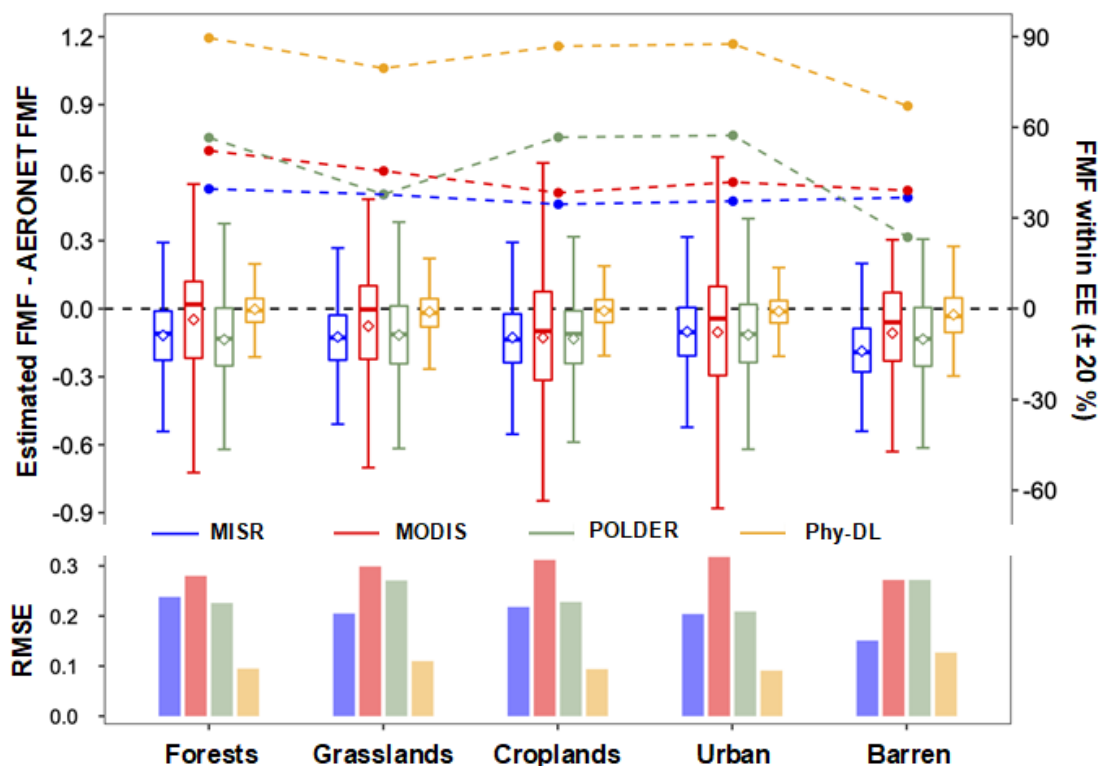

Figure S6. The MISR (blue), MODIS (red), POLDER (green) and Phy-DL FMF (orange) estimation compared with AERONET FMF (all at 500 nm, using data from 2008-2017). (a) The boxplots of bias (Estimated FMF minus AERONET FMF) and percentage of FMF estimations falls within EE of ±20% (dots and dashed lines) 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.
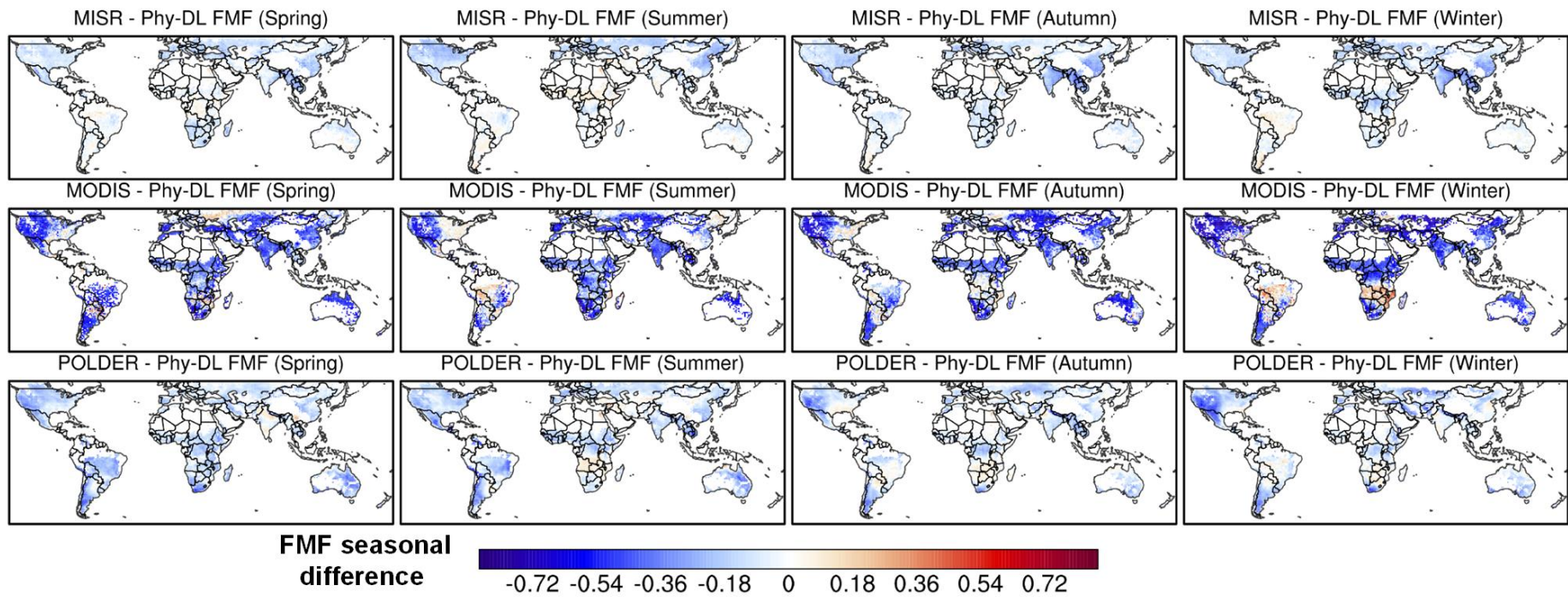
Figure S7. The seasonal mean differences of Phy-DL with MISR, MODIS and POLDER FMF during 2008-2013.

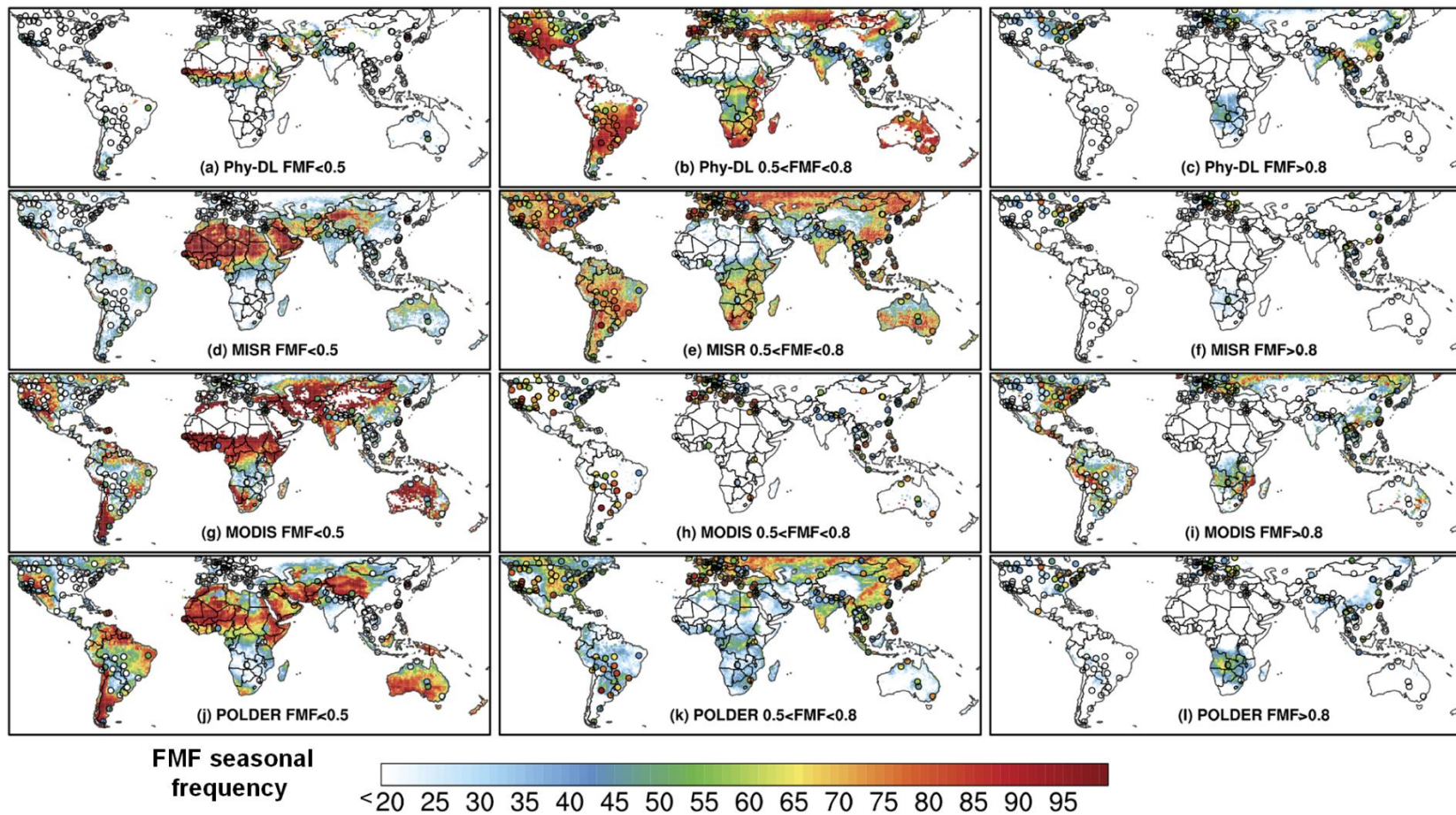**Figure S8.** FMF frequency for three levels FMF (FMF<0.5, 0.5<FMF<0.8, FMF>0.8) calculated by Phy-DL, MISR, MODIS and POLDER (base maps) and AERONET (dots) during 2008-2013