Responses to RC1:

This paper constructed a global aerosol fine mode fraction (FMF) dataset using combined physical and statistical methods. The dataset showed overall superior performance over existing satellite products. This is a good paper, providing a useful set of information to study the distribution and potentially climate impact of anthropogenic aerosols. I recommend publication of the paper after addressing a few questions. Although my recommendation is a major revision, the comments should not be difficulty to fix, but I think will increase of the credibility of the method and data.

Response: Thanks for your careful reading and comments. Your comments are valuable in improving the quality of our manuscript.

1. A 10 fold cross validation seems too simple. It would be interesting to see how the method perform on different sites or different years. Therefore, an out of site validation (e.g., predicting FMF at AERONET sites whose data are not used for training) or out of period validation (i.e., using part of the time series as training and the rest as validation) is highly recommended to increase the robustness of the results.

Response: Thank you for this valuable recommendation. We have conducted a validation based on measurements from four independent National Oceanic and Atmospheric Administration Surface Radiation Budget (SURFRAD) network sites not used for training in the deep-learning model. The SURFRAD network provides long-term, multi-band AOD observations at a temporal resolution of three minutes. As shown in **Figure R1**, the four sites (black triangles) are located across the US, covering different land types from forested land to barren land (**Tables R1 and R2**). The land types were based on MODIS MCD12C1 data from the International Geosphere-Biosphere Programme scheme.

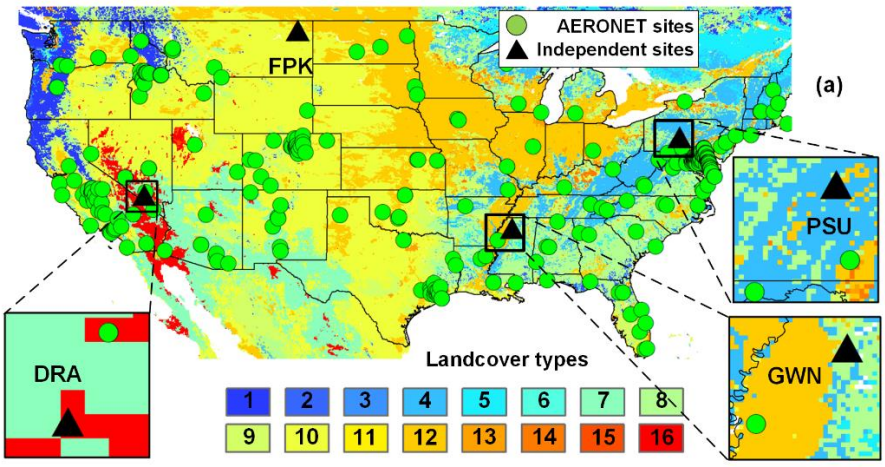

Figure R1. (a) Locations of AERONET sites (green dots) and four independent SURFRAD sites (black triangles, **Table R1**) for the independent validation of the Phy-

DL FMF algorithm. The base map shows the land types from MODIS MCD12C1 data (the International Geosphere-Biosphere Programme scheme) in 2010. **Table R2** provides details about the land-type legend.

Table R1. SURFRAD sites used for out-of-site validation and their locations and land types.

Sites	Longitude (°W)	Latitude (°N)	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 R2. Land types and corresponding values 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

We used the same AERONET method (SDA) to calculate the FMF at the SURFRAD sites. SURFRAD data were not included in the model training, so SURFRAD FMFs can be regarded as the out-of-site validation for the Phy-DL FMF algorithm. **Figure R2** shows how SURFRAD and Phy-DL FMFs compare. The correlation coefficient (R) was 0.51, and the root-mean-square error (RMSE) was 0.144, somewhat poorer performance than AERONET validation results (i.e., R=0.68 and RMSE=0.136).

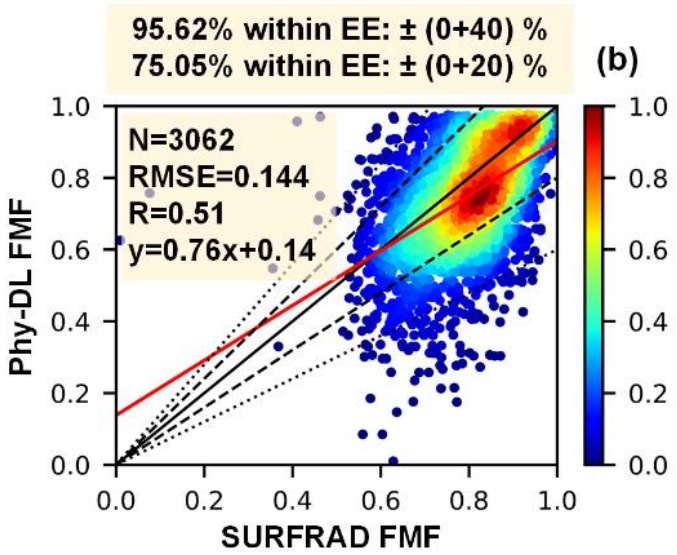

Figure R2. Phy-DL FMF at 500 nm as a function of SURFRAD FMF. The black and red solid lines are the 1:1 line and the best-fit line obtained from linear regression, respectively. The black dashed and dotted lines represent the expected error (EE) envelopes of ±20% and ±40%, respectively.

Furthermore, the Phy-DL FMF performance was validated at each SURFRAD site. **Figure R3** shows the bias of Phy-DL FMF (Phy-DL FMF minus SURFRAD FMF), retrievals falling within the ±20% EE envelope, and RMSEs at each SURFRAD site. In general, most of the sites have a mean bias and an RMSE lower than 0.1 and 0.15, respectively, with over 70% of the retrievals falling within the ±20% EE envelope. The out-of-site validation reveals that the Phy-DL FMF algorithm is reliable even in regions without AERONET sites for model training.

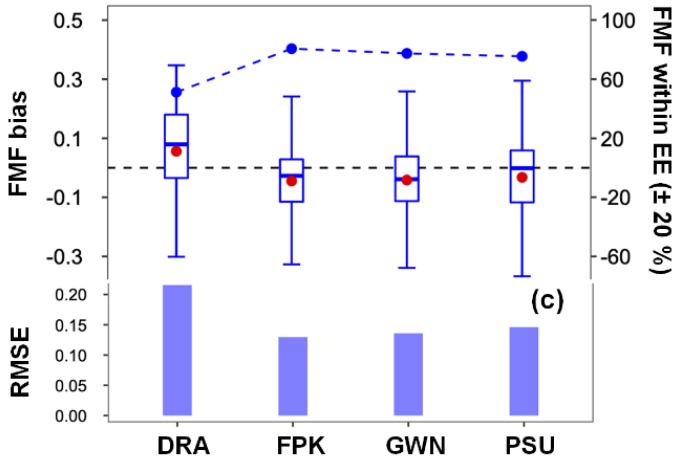

Figure R3. Boxplots of bias (Phy-DL FMF minus SURFRAD FMF), percentage of FMF estimates falling within the EE envelope of ±20% (dash-dotted line), and RMSEs at the four independent SURFRAD sites. The upper, middle, and lower lines in each box present the 75th, median, and 25th percentiles, respectively. The red point in each box represents the mean value of the FMF bias.

Changes in the manuscript: We have revised the manuscript as follows:
(1) Added a new figure (i.e., **Figure 4)** to the manuscript.

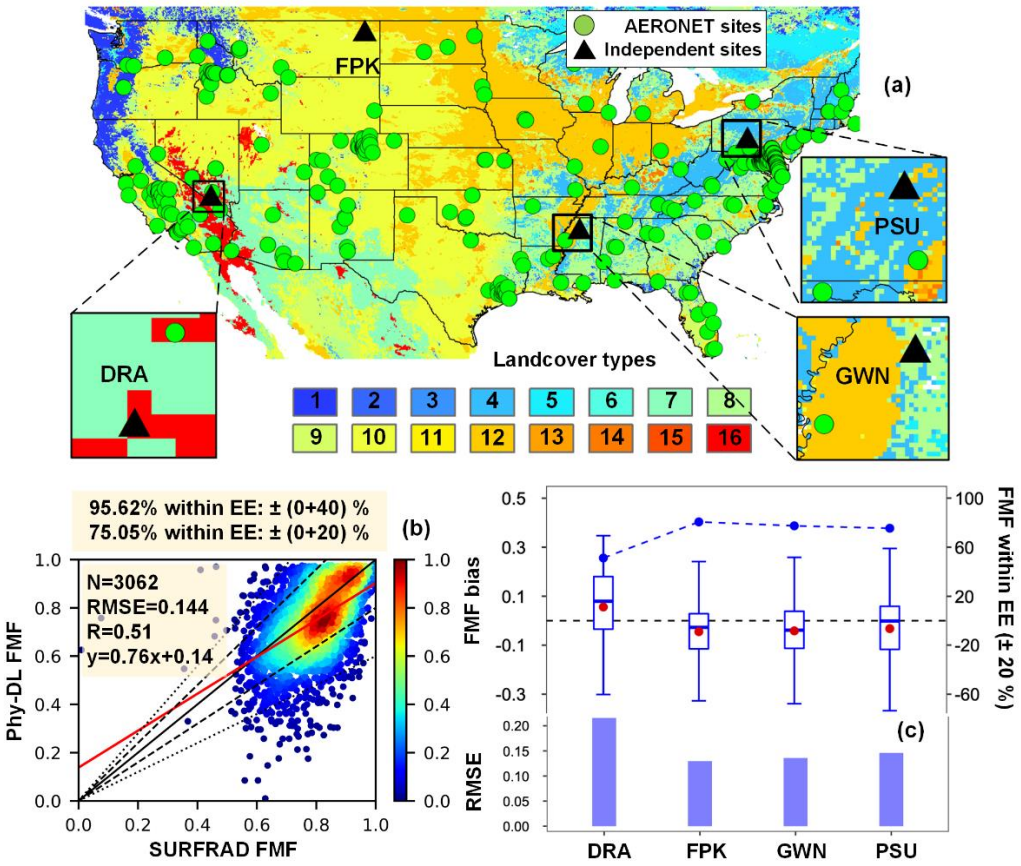

Figure 4. (a) The locations of AERONET sites (green points) and four independent SURFRAD sites (black triangles) for the independent validation of the Phy-DL FMF algorithm. The base map shows the land types from MODIS MCD12C1 data (the International Geosphere-Biosphere Programme scheme) in 2010. Table S1 provides details about the land-type legend. (b) Phy-DL FMF at 500 nm as a function of SURFRAD 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. (c) Boxplots of bias (Phy-DL FMF minus SURFRAD FMF), percentage of FMF estimates falling within the EE envelope of ±20% (dash-dotted lines), and RMSEs at the four independent SURFRAD sites. The upper, middle, and lower lines in each box present the 75th, median, and 25th percentiles, respectively. The red point in each box represents the mean value of the FMF bias.

(2) Added **Table S3** and **Table S4** to the supplementary file.

Table S3. SURFRAD sites used for out-of-site validation and their locations and land types.

Sites	Longitude (°W)	Latitude (°N)	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 S4. Land types and corresponded values 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

(3) In section 3.1 entitled “Phy-DL FMF validation”, we have added the following text:

Four sites from the National Oceanic and Atmospheric Administration’s Surface Radiation Budget (SURFRAD) network were selected for the independent validation of the Phy-DL FMF algorithm, providing long-term, multi-bands AOD observations at the temporal resolution of three minutes. As shown in Figure 4a, the four sites (black triangles) are located across the US, covering different land types from forested land to barren land. The land types were based on MODIS MCD12C1 data from the International Geosphere-Biosphere Programme scheme. We used the same AERONET method (SDA) to calculate the FMF at the SURFRAD sites. SURFRAD data were not included in the model training, so SURFRAD FMFs can be used to do the out-of-site validation for the Phy-DL FMF algorithm. Figure 4b shows how SURFRAD and Phy-DL FMFs compare. The correlation coefficient (R) was 0.51, and the root-mean-square error (RMSE) was 0.144, somewhat different than AERONET validation results (i.e., R=0.68 and RMSE=0.136).

Furthermore, the Phy-DL FMF performance was validated at each SURFRAD site. Figure 4c shows the bias of Phy-DL FMF (Phy-DL FMF minus SURFRAD FMF), percentage of retrievals falling within the ±20% EE envelope, and RMSEs at each site. In general, most of the sites have a mean bias and an RMSE lower than 0.1 and 0.15, respectively, with over 70% of the retrievals falling within the ±20% EE envelope. The out-of-site validation reveals that the Phy-DL FMF algorithm is reliable even in regions without AERONET sites for model training.

2. How are the input variables selected? Are they all necessary or there might be some redundant inputs? Since MODIS AE is used as input, I don't see how some meteorology variables, such as boundary layer height, winds, RH are physically related to FMF.

Response: Thank you for the question. One of the most important reasons for considering meteorological variables in the deep-learning model is that the physical approach (SDA) does not include these potential meteorological influences. O'Neill et al. (2008) have reported that when the temperature is low, the error of the fine-mode AOD calculated by the SDA is clearly large (**Figure R4**). Although the developers of the SDA know about this issue, the relationship between meteorological factors and the FMF is complex, difficult to describe in equations. Therefore, we incorporate meteorological variables into deep learning to model these complex relationships with the FMF.

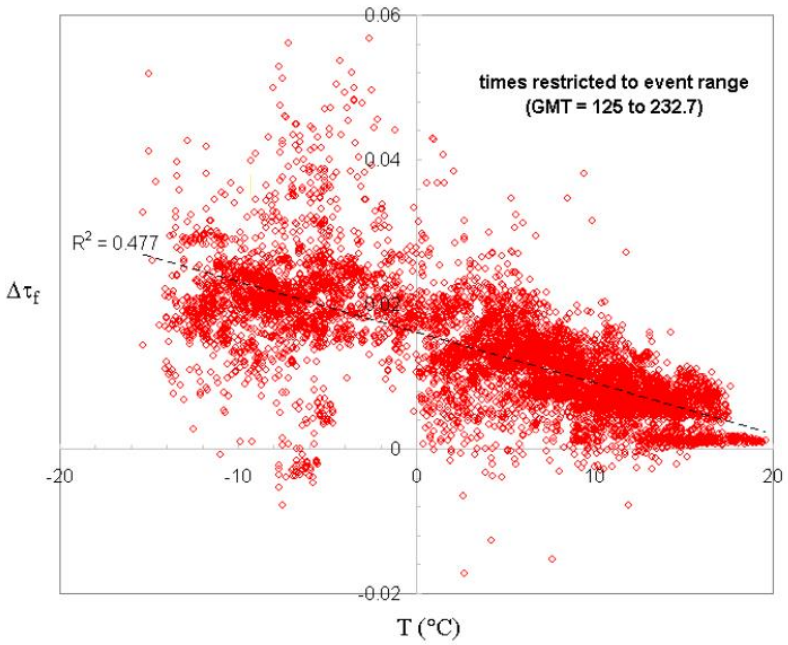

Figure R4. Variation in $\Delta\tau_f$ with detector temperature (PEARL CIMEL, May 1 to August 31, 2007). Copied from the Spectral Deconvolution Algorithm (SDA) technical memo.

We implemented the Generalized Additive Model (GAM) to fit the meteorological variables to the FMF and investigated their nonlinear relationships. **Figure R5** reveals that the meteorological variables considered, i.e., PBLH, temperature, surface pressure, RH, and wind speed, all had significant non-linear relationships with the FMF (at the 99% significance level). Both PBLH and surface pressure had similar influences on the FMF, i.e., a positive (negative) response when PBLH and surface pressure values were low (high). This is because high PBLH and surface pressure values can increase the diffusion of fine particles, decreasing the magnitude of the FMF (Tai et al., 2010). Meanwhile, the negative response of the FMF to wind speed also reflects the influence of fine-particle diffusion, as well as the contribution of dust particles strengthened by

wind speed (Luo et al., 2016). Increasing temperatures corresponded to decreasing FMFs, partly due to unfavorable diffusion conditions (Tai et al., 2010). On the other hand, more fine particles are released by heating during colder seasons than during warmer seasons (Ramachandran, 2007). RH had a strong positive influence on $PM_{2.5}$ concentrations when RH was between 25% and 75%. This reflects the secondary particle formation boosted by the increasing RH, contributing to the fine particles (Tai et al., 2010).

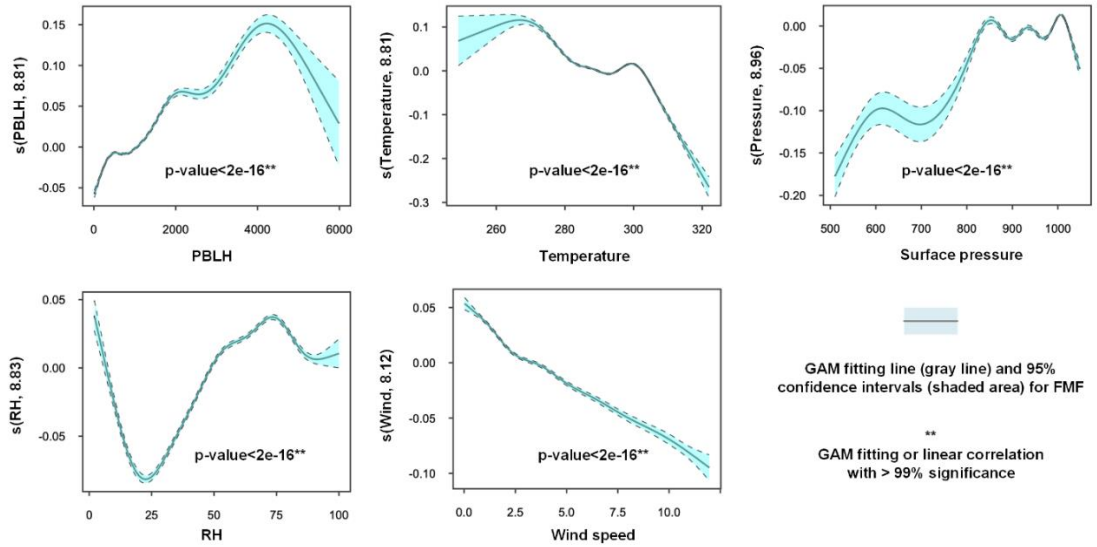

Figure R5. GAM fitting plots for the meteorological variables and the FMF. Shaded areas in the GAM plots indicate 95% confidence intervals, and the y-axes show the covariate and effective degrees of freedom of the smoothing. Asterisks (**) after each p-value indicate the 99% confidence interval of fitting.

Changes in the manuscript: We have revised the manuscript as follows:
(1) Added a new figure (i.e., **Figure S2)** in the supplementary file.

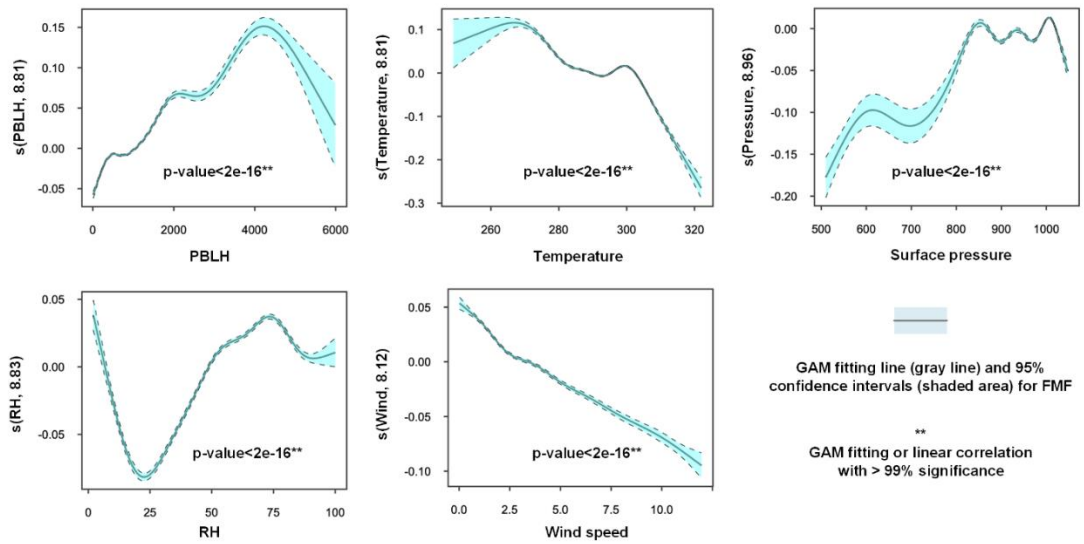

Figure S2. GAM fitting plots for the meteorological variables and the FMF. Shaded areas in the GAM plots indicate 95% confidence intervals, and the y-axes shows the

covariate and effective degrees of freedom of the smoothing. The asterisks (**) after each p-value indicate the 99% confidence interval of fitting.

(2) In section 2.3 entitled "Meteorological data", we have added the following text:

Previous studies have reported that meteorological variables are significantly correlated to fine-mode and coarse-mode aerosols. Tai et al. (2010), Liang et al. (2016), and Shen et al. (2018) all revealed that meteorological variables like temperature, RH, and wind speed explain much of the variations in $PM_{2.5}$ concentrations (> 50%). Xiang et al. (2019) and Gui et al. (2019) found a negative association between BLH and $PM_{2.5}$, and Kang et al. (2014) found that fine-mode aerosols and air pressure were correlated significantly. In this study, to investigate the correlation between meteorological variables and the FMF, we implemented the Generalized Additive Model (GAM). Figure S2 reveals that the meteorological variables considered in this study, i.e., PBLH, temperature, surface pressure, RH, and wind speed, all had significant non-linear relationships with the FMF (at the 99% significance level). Both PBLH and surface pressure had similar influences on the FMF, i.e., a positive (negative) response when PBLH and surface pressure values were low (high). This is because high PBLH and surface pressure values can increase the diffusion of fine particles, decreasing the magnitude of the FMF (Tai et al., 2010). Meanwhile, the negative response of the FMF to wind speed also reflects the influence of fine-particle diffusion, as well as the contribution of dust particles strengthened by wind speed (Luo et al., 2016). Increasing temperatures corresponded to decreasing FMFs, partly due to unfavorable diffusion conditions (Tai et al., 2010). On the other hand, more fine particles are released by heating during colder seasons than during warmer seasons (Ramachandran, 2007). RH had a strong positive influence on $PM_{2.5}$ concentrations when RH was between 25% and 75%. This reflects the secondary particle formation boosted by the increasing RH that contributed to the fine particles (Tai et al., 2010). Therefore, in this study, we used surface temperature, air pressure, PBLH, RH, and wind speed as inputs to the deep-learning model.

3. Although the proposed model produces the best FMF among different satellite products, there is still an obvious bias compared with AERONET. Also at some sites the agreement is low, such as Australia, North Africa, Mid East, etc. These all seem to be desert regions, could this be related to surface, or optical property assumption for non-spherical particles? Some more in-depth error analysis would be helpful to fully evaluate the performance of the technique.

Response: Thank you for the question. We analyzed this issue further. As shown in **Figure R6,** more than 75% of the sites located on barren land have low percentages of Phy-DL FMFs (< 60%) falling within the ±20% EE envelope. About 4% of the sites have high percentages of Phy-DL FMFs (> 90%) falling within the ±20% EE envelope.

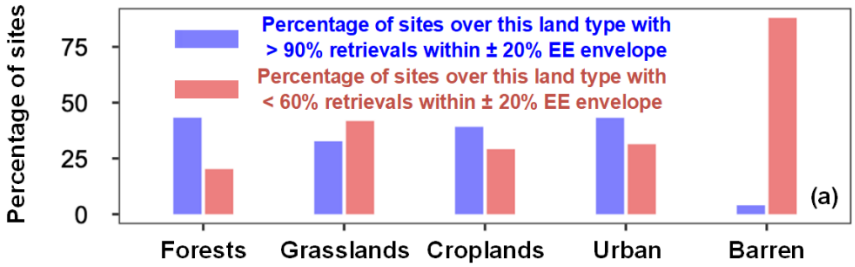

Figure R6. Bar plots of the percentage of sites with > 90% of retrievals falling within the ±20% EE envelope (blue bars) and the percentage of sites with < 60% of retrievals falling within the ±20% EE envelope (red bars) for five land types.

The barren land type is a bright surface compared to other land types where the other sites are located (**Table R1**). AODs over the bright surface used for the Phy-DL FMF retrieval were significantly overestimated, with the worst performance compared to other vegetated land-cover types (Levy et al., 2010; Petrenko and Ichoku, 2013). This suggests that the performance of the Phy-DL FMF algorithm is poor when applied to regions with barren land. **Figure R7** shows the bias of the Phy-DL FMF and the percentage of retrievals falling within the EE envelope of ±20% as a function of NDVI. As NDVI increased from < 0.1 to > 0.8, the percentage of FMF retrievals falling within the ±20% EE envelope also rose from < 70% to > 85%, and the range of bias decreased significantly. Because a higher NDVI value indicates a darker surface, **Figure R7** reveals that the Phy-DL FMF algorithm performs better over dark surfaces than over bright surfaces, resulting in a lower accuracy over the barren land type than vegetated land types.

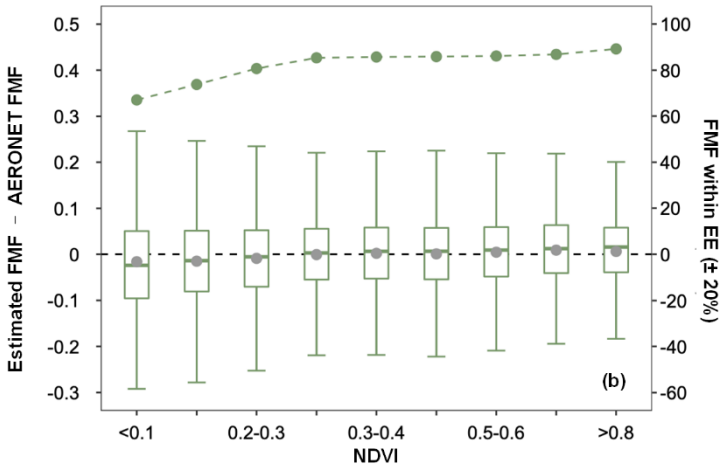

Figure R7. Box plots of the FMF bias (estimated FMF minus AERONET FMF) as a function of NDVI. The black horizontal dashed line indicates the zero bias. The gray dot in each box represents the mean value of the FMF bias. The upper, middle, and lower horizontal lines in each box show the 75th, median, and 25th percentiles, respectively. The green dots connected by the dashed curve are percentages of FMF retrievals falling within the EE envelope of ±20%.

The Ångström Exponent (AE) from the MODIS DT aerosol product is still highly

uncertain. The core of the SDA method relies on AE as input (Yan et al., 2017). The low accuracy of AE can significantly influence the performance of the Phy-DL FMF algorithm. As shown in **Figure R8**, AEs from the MODIS MOD08 product used as input to the Phy-DL FMF algorithm performed the worst over barren land, with the highest RMSE (> 1) and the lowest percentage of retrievals falling within the EE envelope of ±0.45 (< 45%). This would result in a lower performance of the Phy-DL FMF algorithm when applied to regions with barren land.

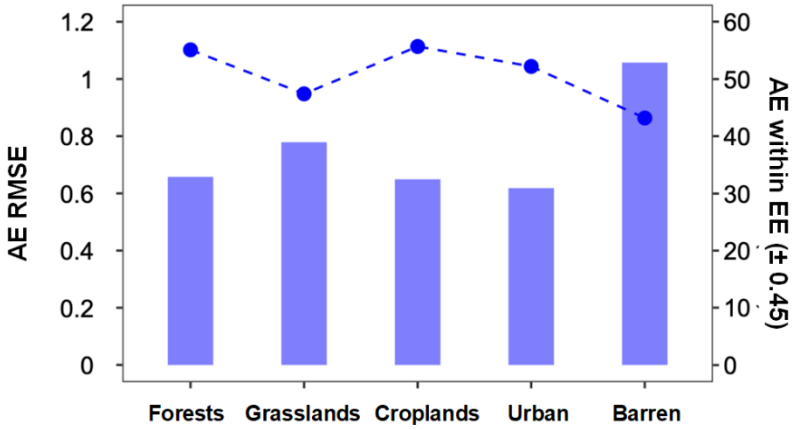

Figure R8. RMSEs (bars) and percentages of MOD08 AE falling within the EE envelope of ±0.45 (dash-dotted line) against AERONET observations for five land types. The EE envelope (±0.45) was adopted from Levy et al. (2013).

Changes in the manuscript: We have revised the manuscript as follows:
(1) Added new figures (i.e., **Figure S3** and **Figure S4)** to the supplementary file.

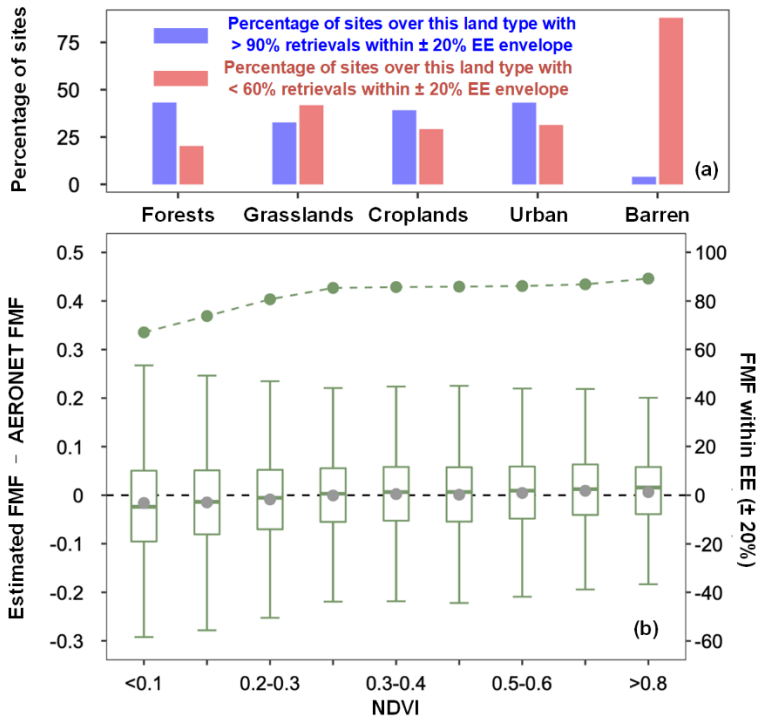

Figure S3. (a) Bar plots of the percentage of sites with > 90% of retrievals falling within the ±20% EE envelope (blue bars) and the percentage of sites with < 60% of retrievals falling within the ±20% EE envelope (red bars) for five land types. (b) Box plots of the FMF bias (estimated FMF minus AERONET 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%.

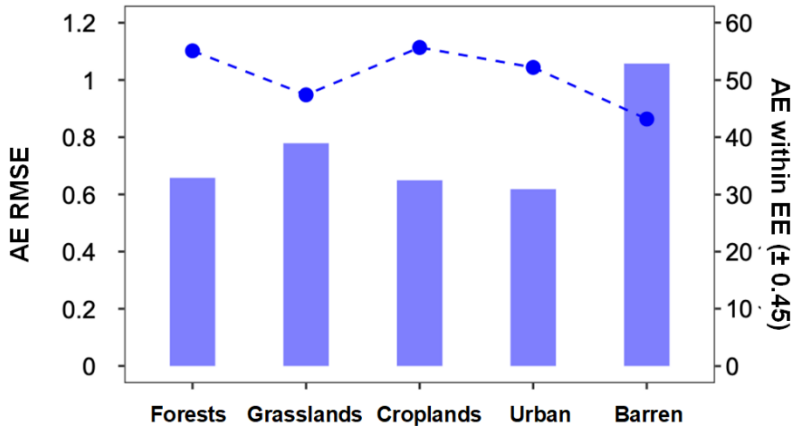

Figure S4. RMSEs (bars) and percentages of MOD08 AE falling within the EE envelope of ±0.45 (dash-dotted line) against AERONET observation for five land types. The EE envelope (±0.45) was adopted from Levy et al. (2013).

(2) In section 3.1 entitled "Phy-DL FMF validation", we have added the following text:

Figure S3a shows that more than 75% of the sites located on barren land have low percentages of Phy-DL FMFs (< 60%) falling within the EE envelope of ±20%. About 4% of the sites have high percentages of Phy-DL FMFs (> 90%) falling within the ±20% EE envelope. This suggests that the accuracy of Phy-DL FMF over barren land is much lower than over other land types. The barren land type is a bright surface compared to other land types where the other sites are located (Table S3). AODs over the bright surface used for the Phy-DL FMF retrieval were significantly overestimated, with the worst performance compared to other vegetated land-cover types (Levy et al., 2010; Petrenko and Ichoku, 2013). This suggests that the performance of the Phy-DL FMF algorithm is poor when applied to regions with barren land. Figure S3b shows the bias of the Phy-DL FMF and the percentage of retrievals falling within the EE envelope of ±20% as a function of NDVI. As NDVI increased from < 0.1 to > 0.8, the percentage of FMF retrievals falling within the ±20% EE envelope also rose from < 70% to > 85%, and the range of bias decreased significantly. The Ångström Exponent (AE) from the MODIS DT aerosol product is still highly uncertain. The core of the SDA method relies on AE as input (Yan et al., 2017). The low accuracy of AE can significantly influence the performance of the Phy-DL FMF algorithm. As shown in Figure S4, AEs from the

MODIS MOD08 product used as input to the Phy-DL FMF algorithm performed the worst over barren land, with the highest RMSE (> 1) and the lowest percentage of retrievals falling within the EE envelope of ±0.45 (< 45%). This would result in a lower performance of the Phy-DL FMF algorithm when applied to regions with barren land.

References:

Gui, K., Che, H., Wang, Y., Wang, H., Zhang, L., Zhao, H., Zheng, Y., Sun, T., and Zhang, X.: Satellite-derived $PM_{2.5}$ concentration trends over Eastern China from 1998 to 2016: relationships to emissions and meteorological parameters, Environmental Pollution, 247, 1125–1133, https://doi.org/10.1016/j.envpol.2019.01.056, 2019.

Kang, P., Feng, N., Wang, Z., Guo, Y., Wang, Z., Chen, Y., Zhan, J., Zhan, F. B., and Hong, S.: Statistical properties of aerosols and meteorological factors in Southwest China, Journal of Geophysical Research: Atmospheres, 119, 9914–9930, https://doi.org/10.1002/2014JD022083, 2014.

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, 10,399–10,420, https://doi.org/10.5194/acp-10-10399-2010, 2010.

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 Measeurement Techniques, 6, 2989–3034, https://doi.org/10.5194/amt-6-2989-2013, 2013.

Liang, X., Li, S., Zhang, S., Huang, H., and Chen, S. X.: $PM_{2.5}$ data reliability, consistency, and air quality assessment in five Chinese cities, Journal of Geophysical Research: Atmospheres, 121, 10,220–10,236, https://doi.org/10.1002/2016JD024877, 2016.

Luo, N., An, L., Nara, A., Yan, X., and Zhao, W.: GIS-based multielement source analysis of dustfall in Beijing: a study of 40 major and trace elements, Chemosphere, 152, 123–131, 2016.

O'Neill, N., Eck, T., Smirnov, A., and Holben, B. (2008). Spectral deconvolution algorithm (SDA) technical memo.

Petrenko, M. and Ichoku, C.: Coherent uncertainty analysis of aerosol measurements from multiple satellite sensors, Atmospheric Chemistry And Physics, 13, 6777–6805, https://doi.org/10.5194/acp-13-6777-2013, 2013.

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, https://doi.org/10.1029/2007jd008500, 2007.

Shen, L., Jacob, D. J., Mickley, L. J., Wang, Y., and Zhang, Q.: Insignificant effect of climate change on winter haze pollution in Beijing, Atmospheric Chemistry and Physics, 18, 17,489–17,496, https://doi.org/10.5194/acp-18-17489-2018, 2018.

Tai, A. P. K., Mickley, L. J., and Jacob, D. J.: Correlations between fine particulate matter ($PM_{2.5}$) and meteorological variables in the United States: implications for the sensitivity of $PM_{2.5}$ to climate change, Atmospheric Environment, 44, 3976–3984, https://doi.org/10.1016/j.atmosenv.2010.06.060, 2010.

Xiang, Y., Zhang, T., Liu, J., Lv, L., Dong, Y., and Chen, Z.: Atmosphere boundary layer height and its effect on air pollutants in Beijing during winter heavy pollution, Atmospheric Research, 215, 305–316, https://doi.org/10.1016/j.atmosres.2018.09.014, 2019.

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.