Supplementary materials

Estimating daily full-coverage and high-accuracy 5-km ambient particulate matters across China: considering their precursors and chemical compositions

Yuan Wang[1], Qiangqiang Yuan[1,4,5], Tongwen Li[2], Siyu Tan[1], Liangpei Zhang[3,5]

[1]School of Geodesy and Geomatics, Wuhan University, Wuhan, Hubei, 430079, China.

[2]School of Geospatial Engineering and Science, Sun Yat-sen University, Zhuhai, Guangdong, 519082, China.

[3]The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei, 430079, China.

[4]The Key Laboratory of Geospace Environment and Geodesy, Ministry of Education, Wuhan University, Wuhan, Hubei, 430079, China.

[5]The Collaborative Innovation Center for Geospatial Technology, Wuhan, Hubei, 430079, China.

Correspondence to: Qiangqiang Yuan (yqiang86@gmail.com)

Figures

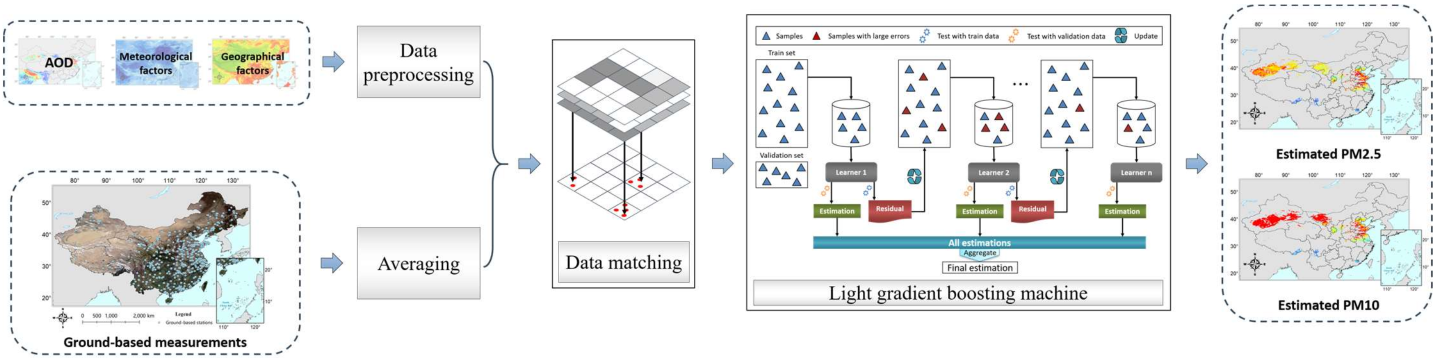

Figure S1. The flowchart of the baseline (AOD-based) in our study. The models for the estimation of $PM_{2.5}$ and PM_{10} are separately trained.

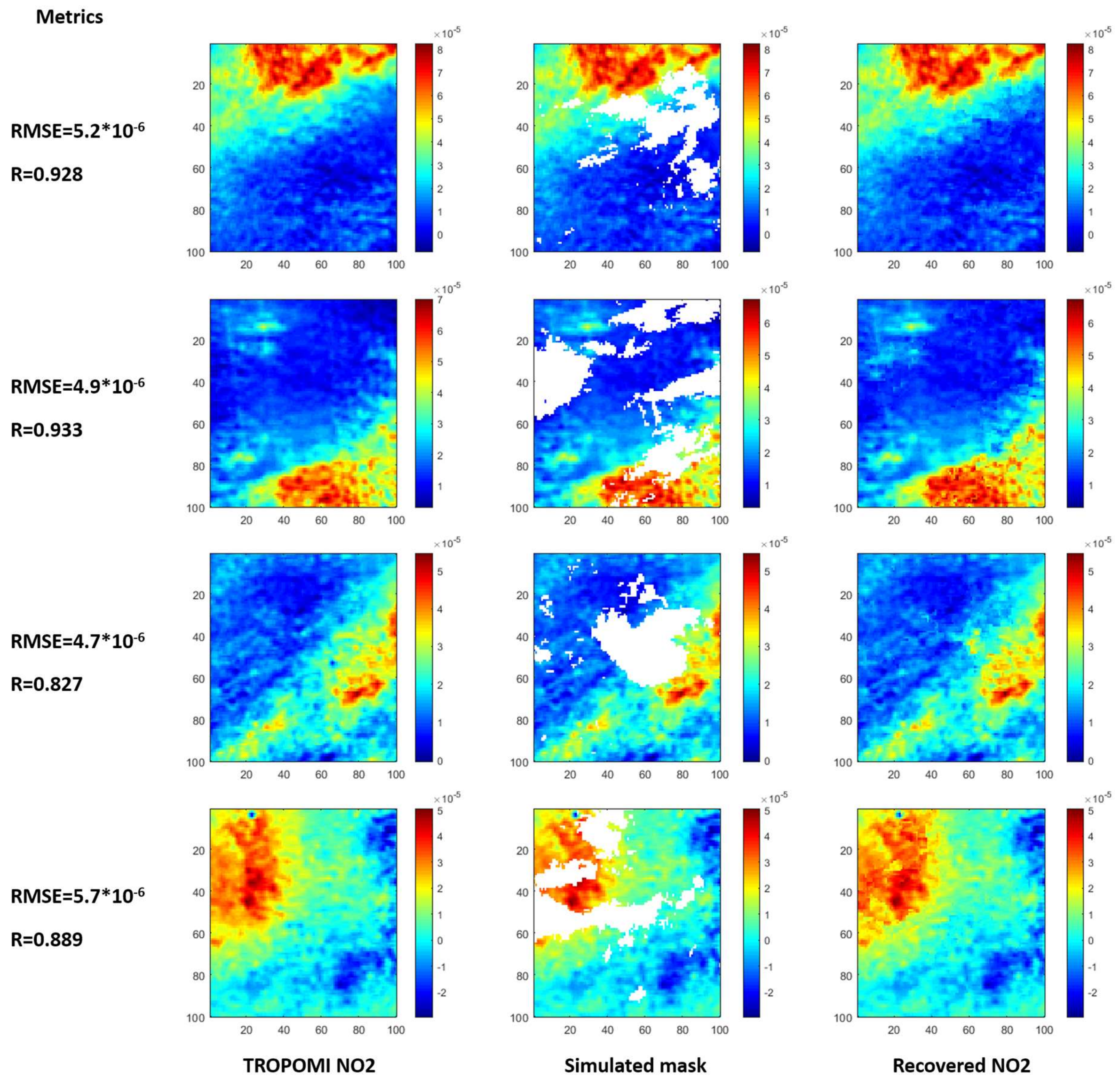

Figure S2. The examples of the simulated experiments for recovering the TROPOMI NO_2 product. The simulated masks are acquired from the shapes of the real missing data in the TROPOMI NO_2 product. The color bar illustrates the values of the tropospheric NO_2 column and the color of white represents the invalid values. Unit: mol/m^2.

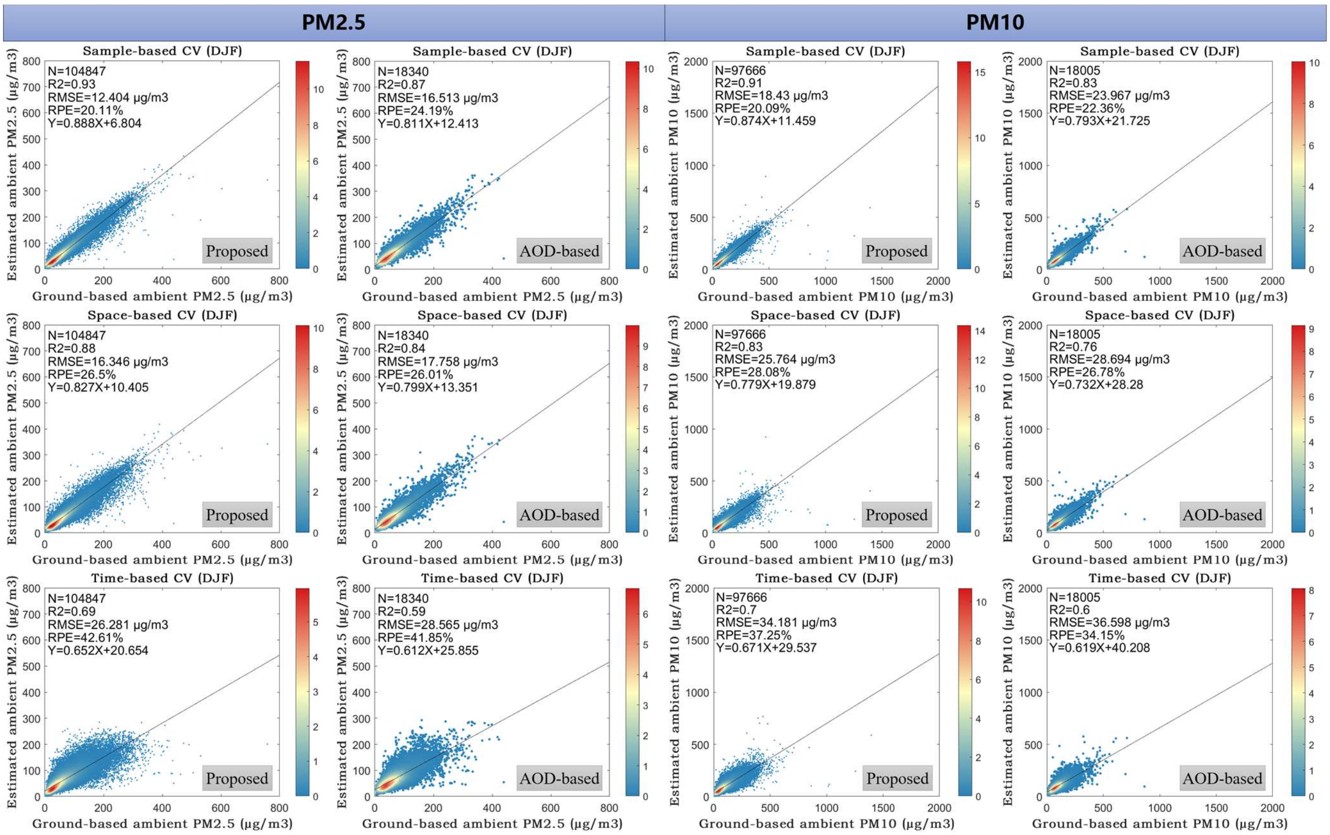

Figure S3. The density scatter plots of the validation results for DJF in 2019 over the study area. The black solid line signifies the fitted line and the color bar denotes the density of samples. Y: estimated ambient concentrations of $PM_{2.5}$ and PM_{10}; X: ground-based ambient concentrations of $PM_{2.5}$ and PM_{10}.

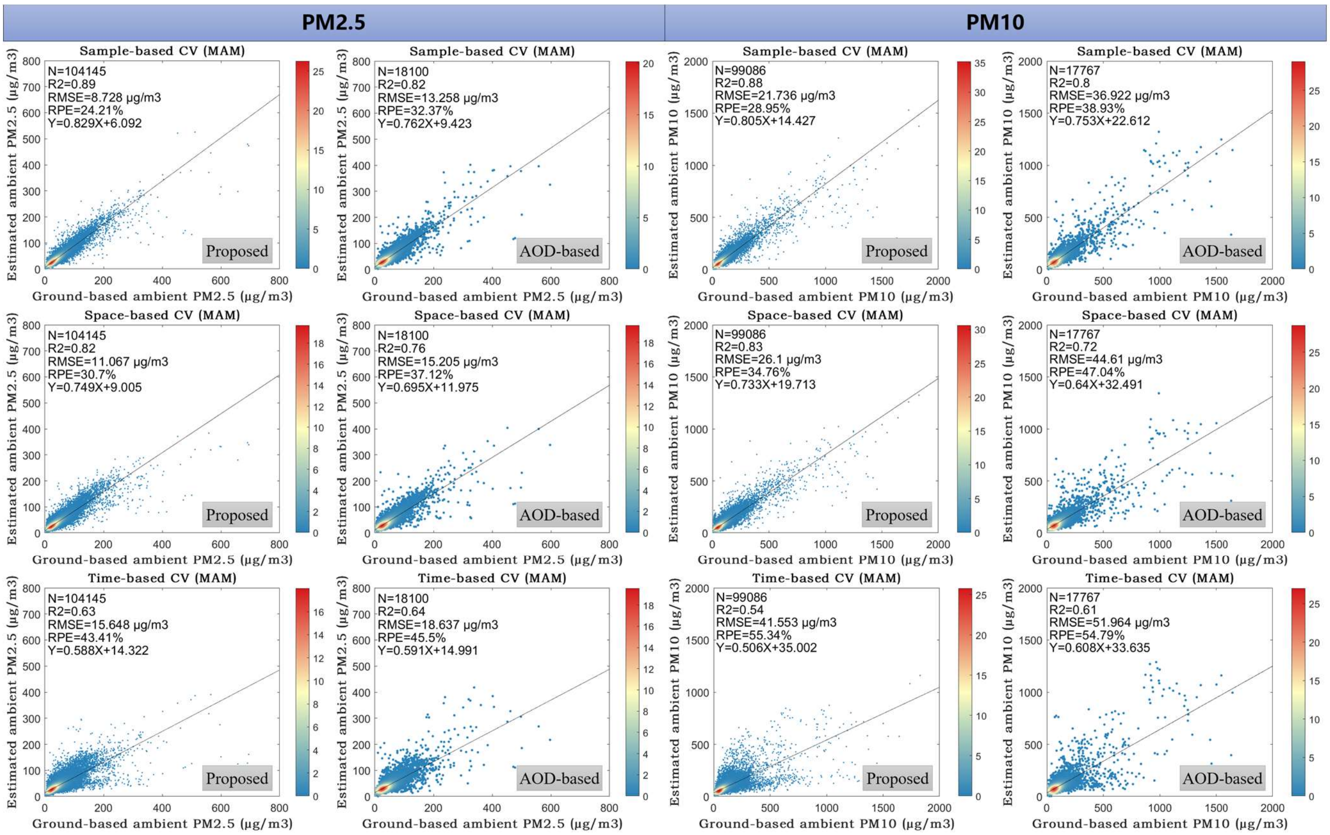

Figure S4. The density scatter plots of the validation results for MAM in 2019 over the study area. The black solid line signifies the fitted line and the color bar denotes the density of samples. Y: estimated ambient concentrations of $PM_{2.5}$ and PM_{10}; X: ground-based ambient concentrations of $PM_{2.5}$ and PM_{10}.

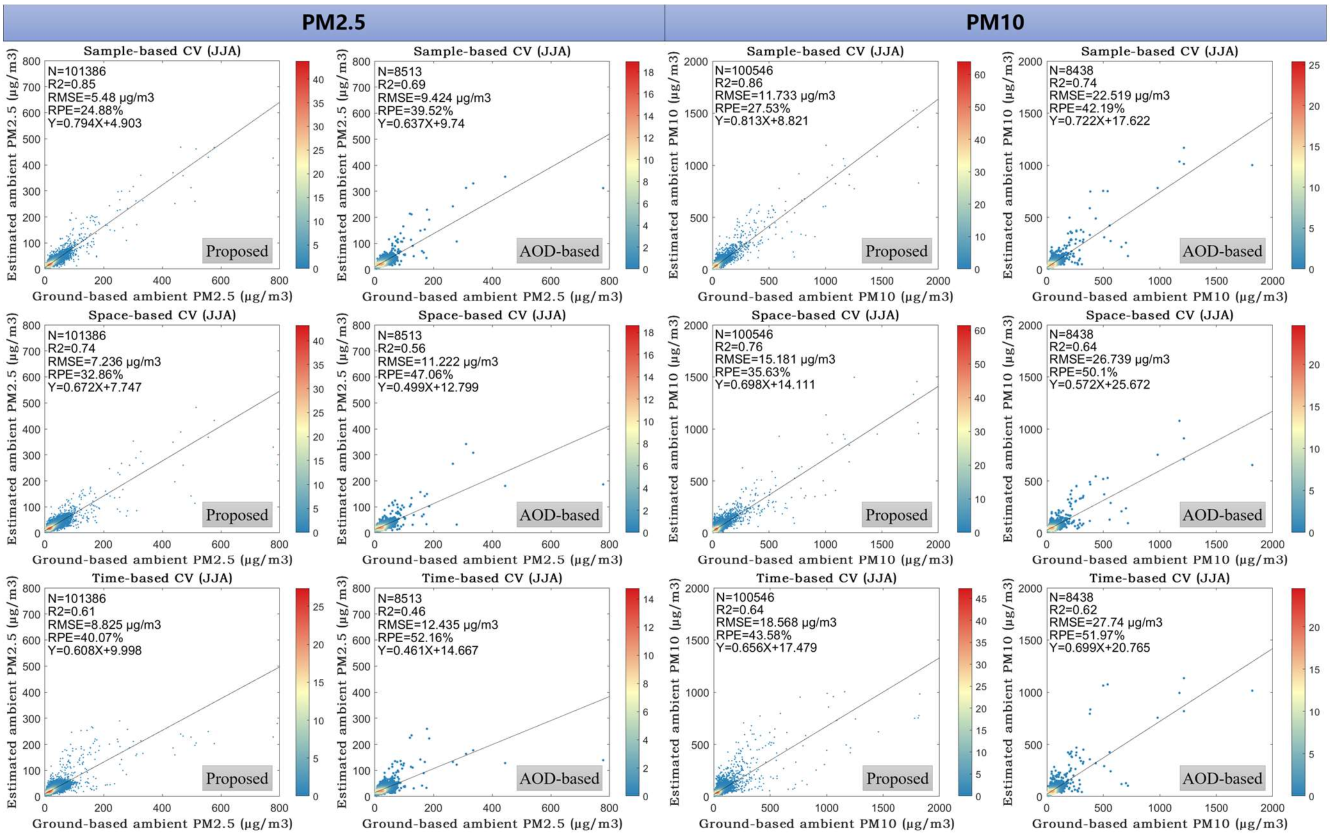

Figure S5. The density scatter plots of the validation results for JJA in 2019 over the study area. The black solid line signifies the fitted line and the color bar denotes the density of samples. Y: estimated ambient concentrations of $PM_{2.5}$ and PM_{10}; X: ground-based ambient concentrations of $PM_{2.5}$ and PM_{10}.

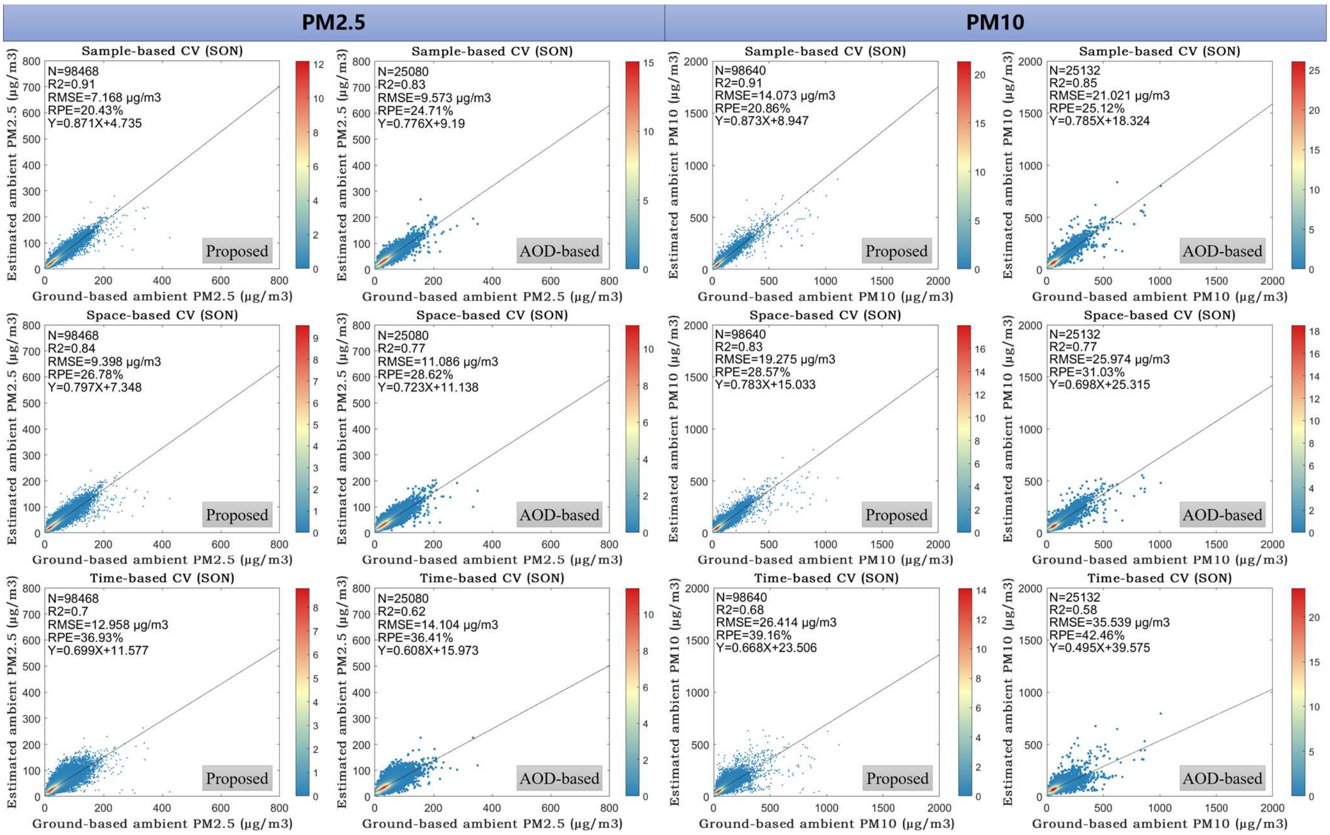

Figure S6. The density scatter plots of the validation results for SON in 2019 over the study area. The black solid line signifies the fitted line and the color bar denotes the density of samples. Y: estimated ambient concentrations of $PM_{2.5}$ and PM_{10}; X: ground-based ambient concentrations of $PM_{2.5}$ and PM_{10}.

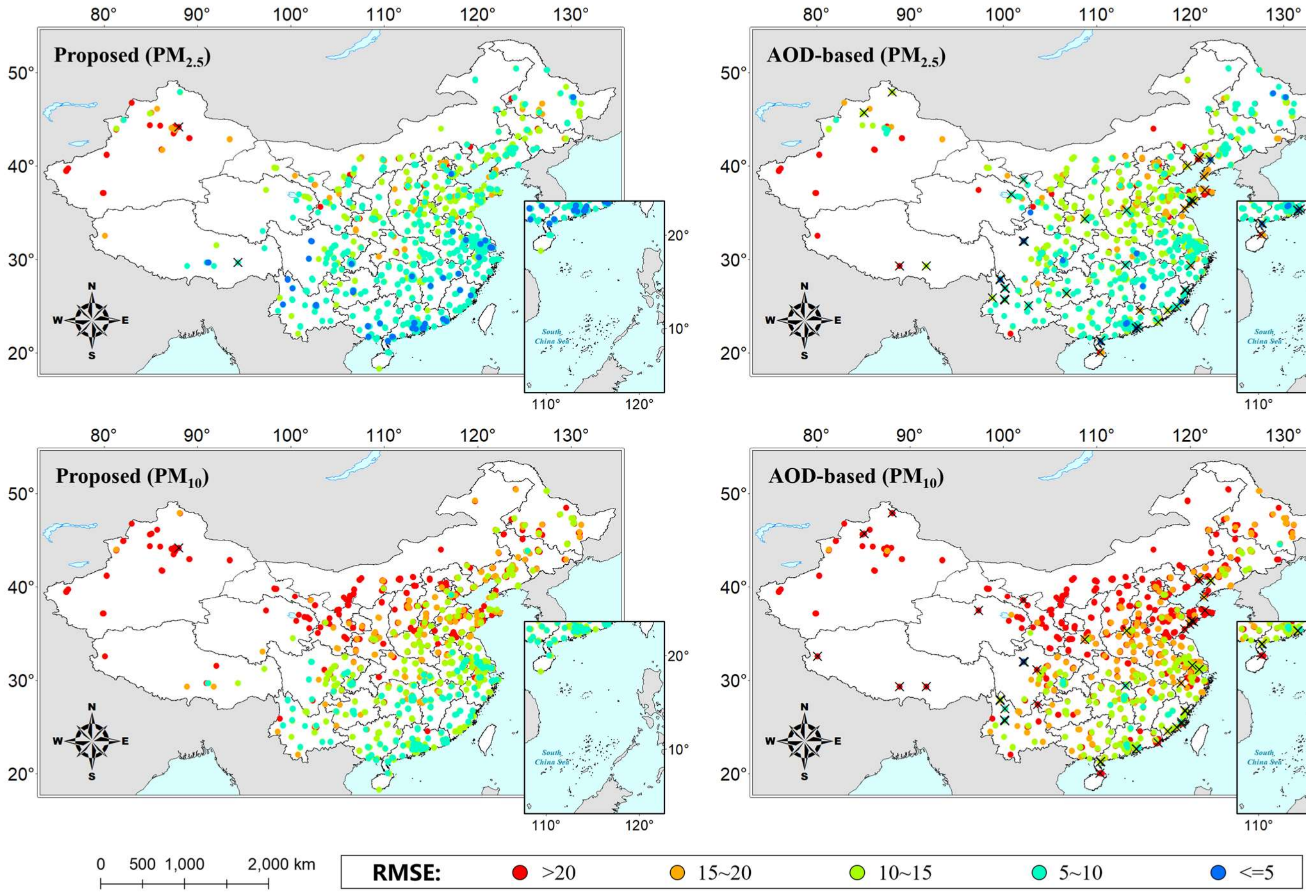

Figure S7. The spatial distribution of RMSEs for the space-based CV at each matched grid over China. The black crosses denote that the significance levels (p) of the metrics are not less than 0.01 at these matched grids. Unit: μg/m^3.

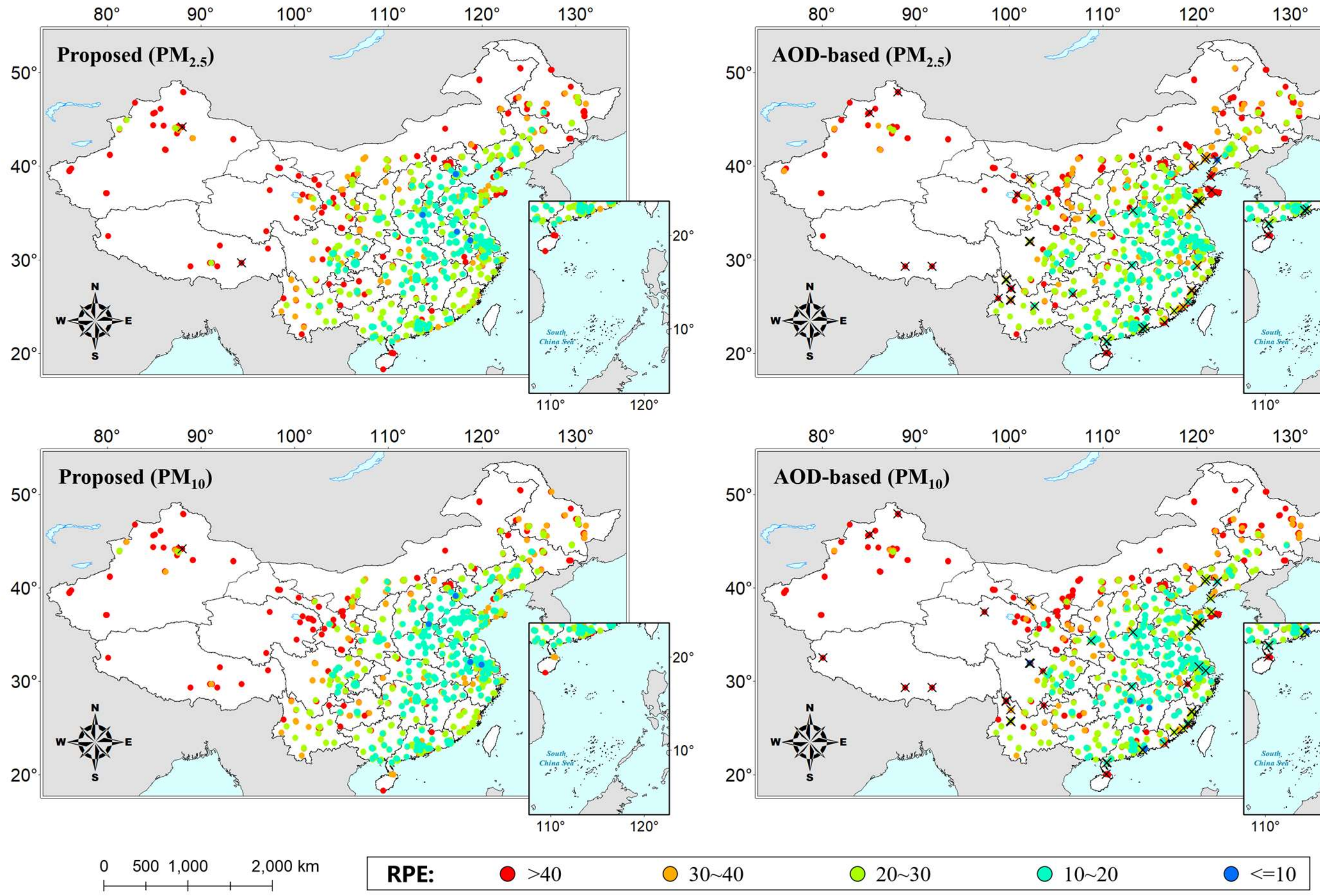

Figure S8. The spatial distribution of RPEs for the space-based CV at each matched grid over China. The black crosses denote that the significance levels (p) of the metrics are not less than 0.01 at these matched grids. Unit: %.

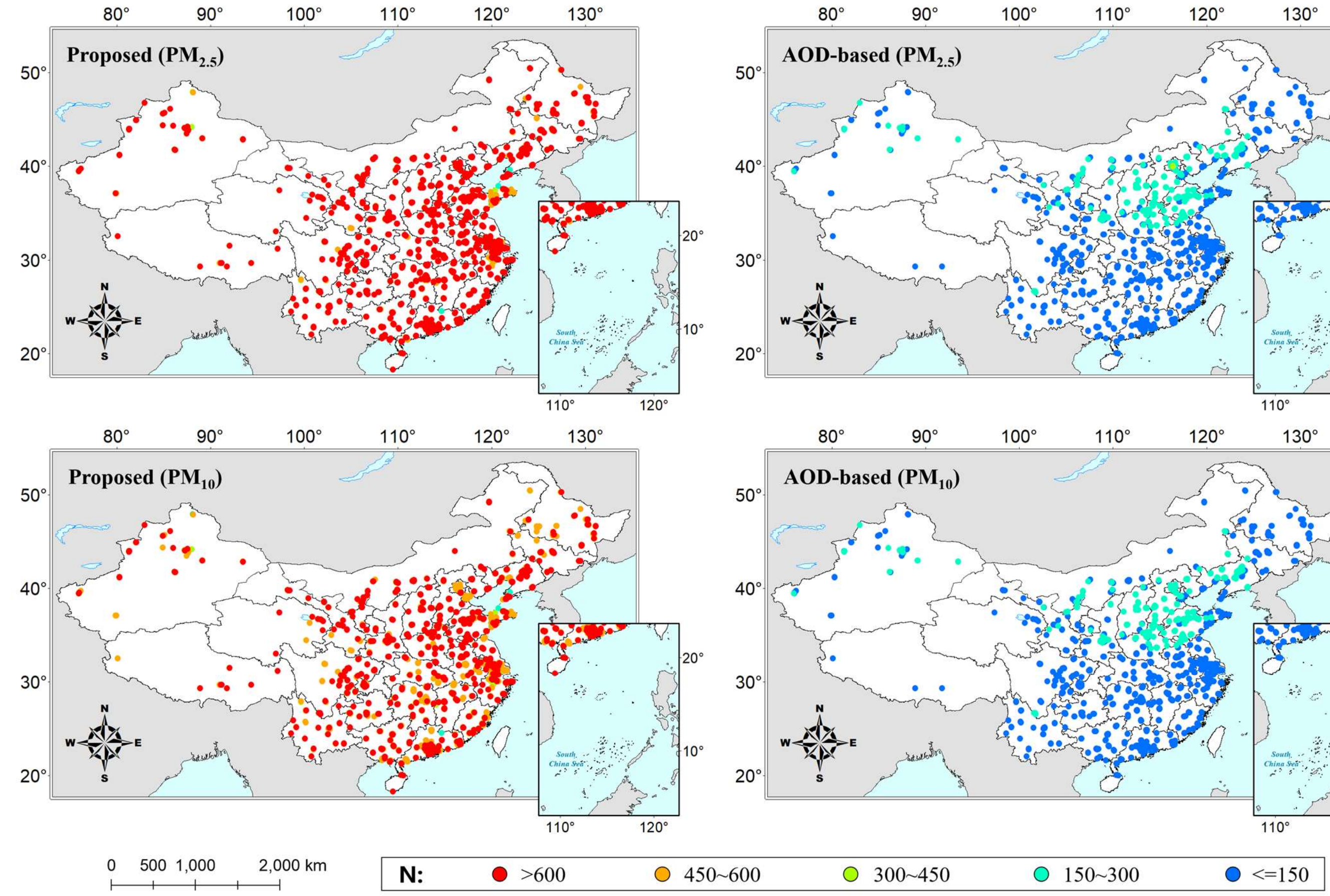

Figure S9. The spatial distribution of the sample numbers for the space-based CV at each matched grid over China.

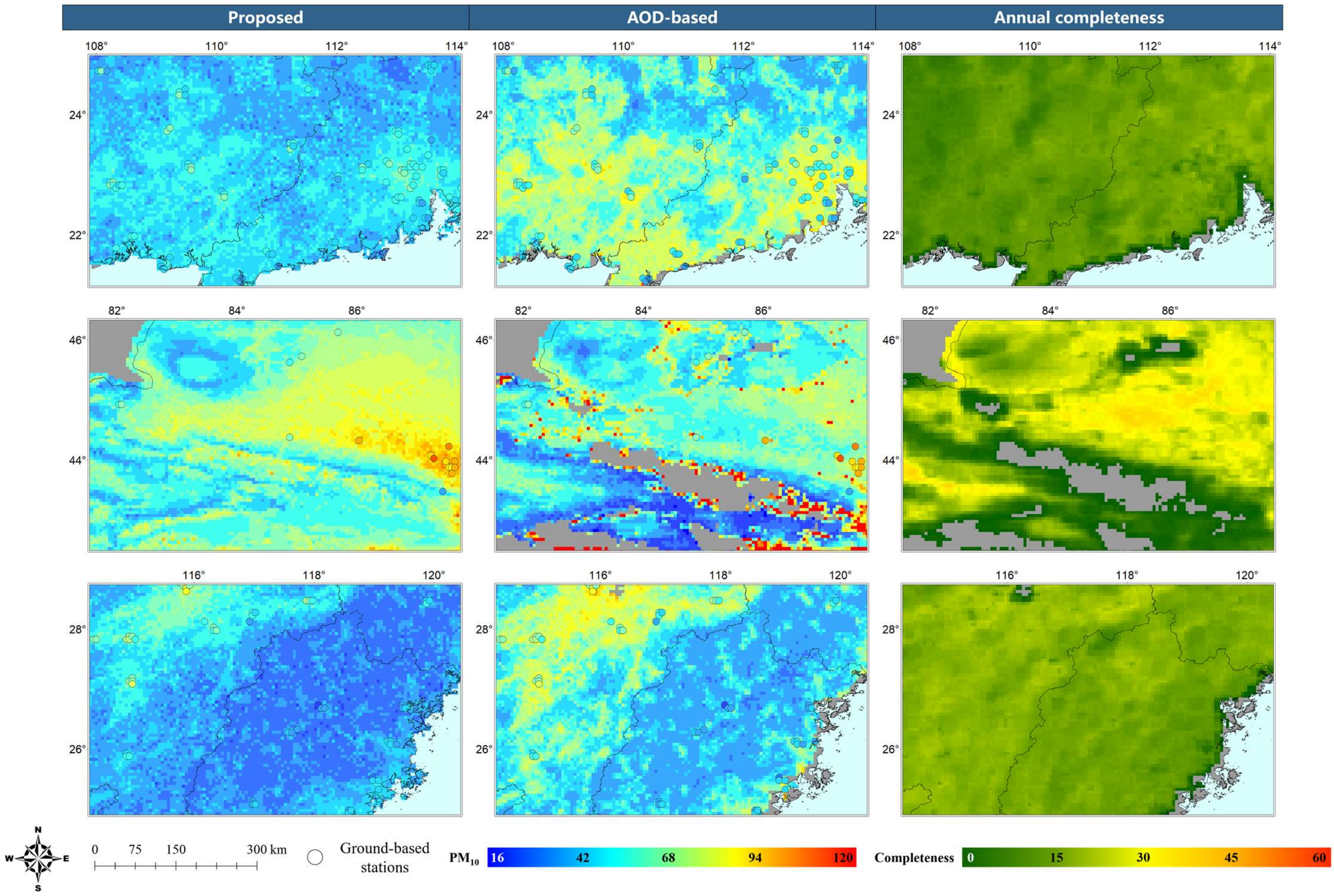

Figure S10. The annual estimated ambient concentrations of PM_{10} for the proposed and AOD-based over local regions in 2019. The left color bar represents the values of the estimated results and ground truth-values. The right color bar denotes the completeness of VIIRS DB AOD. Units: μg/m^3 for $PM_{2.5}$ and % for completeness.

Tables

Table S1. Detailed information about the datasets used in the proposed approach. TR: temporal resolution; OSR: original spatial resolution; RM: resampling method; MD: missing data; MDRM: missing data recovery method; NNI: nearest neighbor interpolation; CI: cubic interpolation; AA: area-weighted aggregation; EA: exemplar-based algorithm; IDW: inverse distance weighted; T: true; F: false; GPW: gridded population of the world; SP: study period (2018.06.01 to 2020.03.31); x-h: x-hour; x-d: x-day. x represents any number.

Type	Source	Name	Identifier	TR	OSR	RM	MD	MDRM	Period
Main variates	TROPOMI	sulfurdioxide_total_vertical_column_1km	S5P_ L2__SO2	Daily	7*3.5 km^2 5.5*3.5 km^2	NNI	T	EA	SP
		nitrogendioxide_tropospheric_column	S5P_ L2__NO2						
Auxiliary variates		ozone_total_vertical_column	S5P_ L2__O3						
Main variates	GEOS-FP	Black Carbon Column Mass Density	tavg3_2d_aer_Nx	3-h	0.25°*0.3125°	CI	F	-	
		Organic Carbon Column Mass Density							
		Nitrate Column Mass Density							
		SO4 Column Mass Density							
		Dust Column Mass Density							
		Ammonium Column Mass Density							
		Sea Salt Column Mass Density							
Auxiliary variates		10-meter Specific Humidity	tavg1_2d_slv_Nx	Hourly					
		10-meter Air Temperature							
		10-meter Eastward Wind							
		10-meter Northward Wind							
		Total Precipitable Water Vapor							
		Pbltop Pressure							
		Surface Pressure							
		Planetary Boundary Layer Height	tavg1_2d_flx_Nx						
		Air Density at Surface							
		Surface Velocity Scale							
		Evaporation from Turbulence							
	MODIS	1_km_16_days_NDVI	MYD13A2	16-d	1*1 km^2	AA	T	IDW	
		The fraction of forest	MCD12Q1 (see Table S6)	Yearly	500*500 m^2		F	-	2018
		The fraction of savanna							
		The fraction of grassland							
		The fraction of cropland							
		The fraction of urban							
		The fraction of aridland							
	OpenStreetMap	Road density	-		0.01°*0.01°				2019
	GPW	Population density	gpw_v4		~1*1 km^2				2020

Table S2. Detailed information about the datasets used in the baseline (AOD-based). TR: temporal resolution; OSR: original spatial resolution; RM: resampling method; MD: missing data; MDRM: missing data recovery method; NNI: nearest neighbor interpolation; CI: cubic interpolation; AA: area-weighted aggregation; EA: exemplar-based algorithm; IDW: inverse distance weighted; T: true; F: false; GPW: gridded population of the world; SP: study period (2018.06.01 to 2020.03.31); x-h: x-hour; x-d: x-day. x represents any number.

Type	Source	Name	Identifier	TR	OSR	RM	MD	MDRM	Period
Main variate	VIIRS	Deep_Blue_Aerosol_Optical_Depth_550_Land_Best_Estimate	AERDB_L2_VIIRS_SNPP	Daily	6*6 km^2	NNI	T	-	SP
Auxiliary variates	GEOS-FP	10-meter Specific Humidity	tavg1_2d_slv_Nx	Hourly	0.25°*0.3125°	CI	F	-	
		10-meter Air Temperature							
		10-meter Eastward Wind							
		10-meter Northward Wind							
		Total Precipitable Water Vapor							
		Pbltop Pressure							
		Surface Pressure							
		Planetary Boundary Layer Height	tavg1_2d_flx_Nx						
		Air Density at Surface							
		Surface Velocity Scale							
		Evaporation from Turbulence							
	MODIS	1_km_16_days_NDVI	MYD13A2	16-d	1*1 km^2	AA	T	IDW	
		The fraction of forest	MCD12Q1 (see Table S6)	Yearly	500*500 m^2		F	-	2018
		The fraction of savanna							
		The fraction of grassland							
		The fraction of cropland							
		The fraction of urban							
		The fraction of aridland							
	OpenStreetMap	Road density	-		0.01°*0.01°				2019
	GPW	Population density	gpw_v4		~1*1 km^2				2020

Table S3. Detailed information about the records used in our study.

Type	Full name	Relationship	Abbreviation
Sulfate-related	sulfurdioxide_total_vertical_column_1km	Precursor	SO2_T
	SO4 Column Mass Density	Chemical composition	SO4CMASS
Nitrate-related	nitrogendioxide_tropospheric_column	Precursor	NO2_T
	Nitrate Column Mass Density	Chemical composition	NICMASS
Carbon-related	Black Carbon Column Mass Density	Chemical composition	BCCMASS
	Organic Carbon Column Mass Density	Precursor	OCCMASS
Dust-related	Dust Column Mass Density	Chemical composition	DUCMASS
Ammonium-related	Ammonium Column Mass Density	Chemical composition	NH4MASS
Sea salt-related	Sea Salt Column Mass Density	Chemical composition	SSCMASS
O_3	ozone_total_vertical_column	-	O3_T
AOD	Deep_Blue_Aerosol_Optical_Depth_550_Land_Best_Estimate		AOD_V
Meteorological factors	10-meter Specific Humidity		PBLTOP
	10-meter Air Temperature		PS
	10-meter Eastward Wind		QV10M
	10-meter Northward Wind		T10M
	Total Precipitable Water Vapor		U10M
	Planetary Boundary Layer Height		V10M
	Air Density at Surface		TQV
	Surface Velocity Scale		PBLH
	Evaporation from Turbulence		RHOA
	Pbltop Pressure		USTAR
	Surface Pressure		EVAP
Geographical factors	1_km_16_days_NDVI		NDVI
	The fraction of forest		F_f
	The fraction of savanna		F_s
	The fraction of grassland		F_g
	The fraction of cropland		F_c
	The fraction of urban		F_u
	The fraction of aridland		F_a
	Road density		R_d
	Population density		P_d

Table S4. The parameters of LGBM for the proposed approach and baseline (AOD-based). The setting of other parameters is default.

Parameters	Proposed	AOD-based
learning_rate	0.05	
num_leaves	1200	
max_depth	-1	
lambda_l1	5	
lambda_l2	5	
subsample	0.75	
colsample_bytree	0.75	
num_boost_rounds	early_stopping_rounds: 10, maxima: 5000	

Table S5. Detailed information about the previous related works over China. SACV: sample-based CV; SPCV: space-based CV; TICV: time-based CV; SR: spatial resolution; TR: temporal resolution; FC: full-coverage; T: true; F: false; MF: the factors which lead to the missing values in the estimated results.

Type	Reference	Metric	SACV	SPCV	TICV	SR	TR	Study period	FC	MF
$PM_{2.5}$	Proposed	R^2	0.93	0.88	0.73	5-km	Daily	2019※	T	None
		RMSE	8.87 μg/m³	11.56 μg/m³	17.3 μg/m³					
		RPE	22.8%	29.8%	44.5%					
	Wei et al., 2019	R^2	0.85	0.83	0.63	1-km	Daily	2016	F	Cloud, snow/ice
		RMSE	15.57 μg/m³	16.63 μg/m³	24.83 μg/m³					
		RPE	-	-	-					
	He et al., 2018	R^2	0.8	-	-	3-km	Daily	2015	F	Cloud, snow/ice, bright surface
		RMSE	18 μg/m³							
		RPE	-							
	Yao et al., 2019	R^2	-	0.6	-	6-km	Daily	2014	F	Cloud, snow/ice, bright surface
		RMSE		21.76 μg/m³						
		RPE		-						
	Li et al., 2020	R^2	0.8	0.79	-	10-km	Daily	2015	F	Cloud, snow/ice
		RMSE	17.38 μg/m³	17.81 μg/m³						
		RPE	31.5%	32.29%						
	Jiang et al., 2020	R^2	0.85	0.74	-	1-km	Daily*	2018.03.01-2019.02.28	T	None
		RMSE	11.02 μg/m³	14.65 μg/m³						
		RPE	-	-						
	Kong et al., 2020	R^2	-	0.86	-	15-km	Daily*	2013–2018	T	None
		RMSE		15.1 μg/m³						
		RPE		-						
PM_{10}	Proposed	R^2	0.91	0.84	0.67	5-km	Daily	2019※	T	None
		RMSE	16.92 μg/m³	22.03 μg/m³	31.33 μg/m³					
		RPE	24.5%	31.9%	45.4%					
	Chen et al., 2018	R^2	-	0.78	-	10-km	Daily	2005–2016	F	Cloud, snow/ice
		RMSE		31.54 μg/m³						
		RPE		-						
	Kong et al., 2020	R^2	-	0.81	-	15-km	Daily*	2013–2018	T	None
		RMSE		28.8 μg/m³						
		RPE		-						

Note:

1. The symbols of * represent that the works could provide the estimated results at various temporal resolutions, while the metrics listed in the table are computed from the daily estimation.

2. ※: Only the metrics computed from the estimated results through the proposed approach for a whole year (2019) are listed in the table to be fairly compared to previous works. The study period of this paper is from June 1, 2018 to March 31, 2020.

Table S6. Detailed information about the land cover map used in our study.

Source	Identifier	Scheme	Specific type	Major type
MODIS	MCD12Q1	International Geosphere-Biosphere Programme (IGBP)	Evergreen Needleleaf Forest	Forest
			Evergreen Broadleaf Forest	
			Deciduous Needleleaf Forest	
			Deciduous Broadleaf Forest	
			Mixed Forests	
			Woody Savannas	Savanna
			Savannas	
			Grasslands	Grassland
			Croplands	Cropland
			Cropland - Natural Vegetation Mosaic	
			Urban Areas	Urban
			Barren or Sparsely Vegetated	Aridland

Eq. s1:

$$ARD = \frac{\left|ER_P - ER_A\right|}{ER_p} \times 100\%$$

where *ARD* represents the absolute relative difference. ER_P and ER_A signify the annual estimated results through the proposed and AOD-based, respectively.

Reference

Chen, G., Wang, Y., Li, S., Cao, W., Ren, H., Knibbs, L. D., and Guo, Y.: Spatiotemporal patterns of PM10 concentrations over China during 2005–2016: A satellite-based estimation using the random forests approach, Environmental Pollution., 242, 605-613, 2018.

He, Q. and Huang, B.: Satellite-based mapping of daily high-resolution ground PM2.5 in China via space-time regression modeling, Remote Sensing of Environment., 206, 72-83, 2018.

Jiang, T., Chen, B., Nie, Z., Zhehao, R., Xu, B., and Tang, S.: Estimation of hourly full-coverage PM2. 5 concentrations at 1-km resolution in China using a two-stage random forest model, Atmospheric Research., 105146, 2020.

Kong, L., Tang, X., Zhu, J., Wang, Z., Li, J., Wu, H., and Liu, B.: A Six-year long (2013–2018) High-resolution Air Quality Reanalysis Dataset over China base on the assimilation of surface observations from CNEMC, Earth System Science Data Discussions., 1-44, 2020.

Li, T., Shen, H., Yuan, Q., and Zhang, L.: Geographically and temporally weighted neural networks for satellite-based mapping of ground-level PM2.5, ISPRS Journal of Photogrammetry and Remote Sensing., 167, 178-188, 2020.

Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., and Cribb, M.: Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach, Remote Sensing of Environment., 231, 111221, 2019.

Yao, F., Wu, J., Li, W., and Peng, J.: A spatially structured adaptive two-stage model for retrieving ground-level PM2.5 concentrations from VIIRS AOD in China, ISPRS Journal of Photogrammetry and Remote Sensing., 151, 263-276, 2019.