Supplementary materials to “Pan-European, High-Resolution, Daily Estimates of Total, Fine-Mode and Coarse-Mode Aerosol Optical Depth based on Quantile Machine Learning”

Zhao-Yue Chen[1,4], Raul Méndez[1], Hervé Petetin[2], Aleksander Lacima[2], Carlos Pérez García-Pando[2,3] and Joan Ballester[1]

1ISGLOBAL, Barcelona, Spain
2Barcelona Supercomputing Center, Barcelona, Spain
3ICREA, Catalan Institution for Research and Advanced Studies, Barcelona,Spain
4Universitat Pompeu Fabra (UPF), Barcelona, Spain

Contents

Contents 1
1. Supplementary Figure 2
2. Supplementary Table 15
3. Preliminary analysis for satellite AOD 21
   3.1 Study design 21
   3.2 Preliminary results 21

1. Supplementary Figure

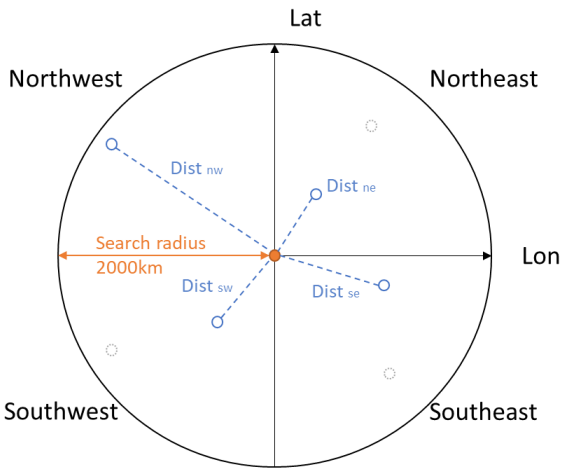

Figure S1. the sketch map of spatial calibration variables ($Dist_{direction}$) calculation. Within 4000km search radius, Orange point is the target/predicted point; Four blue points are the nearest sites in different directions and grey points

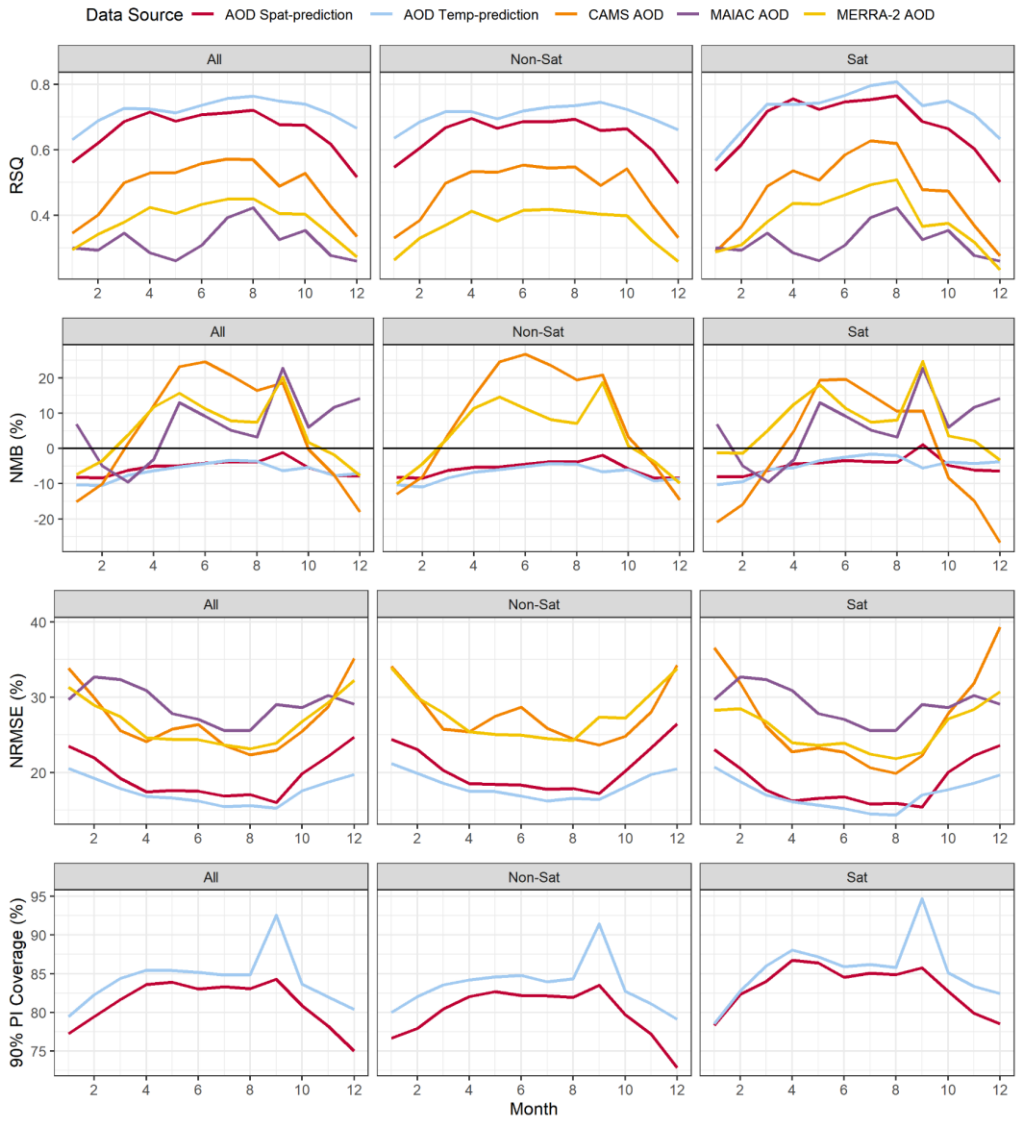

Figure S2. the monthly performance (Rsquare, NMB, NRMSE, 90% PI (predictive intervals) coverage) comparison between different data sources: QML spatial out-of-sample prediction (AOD Spat-Prediction), temporal out-of-sample prediction (AOD Temp-Prediction), CAMS AOD, MAIAC (Satellite) AOD, MERRA-2 AOD.

Supplementary Figure

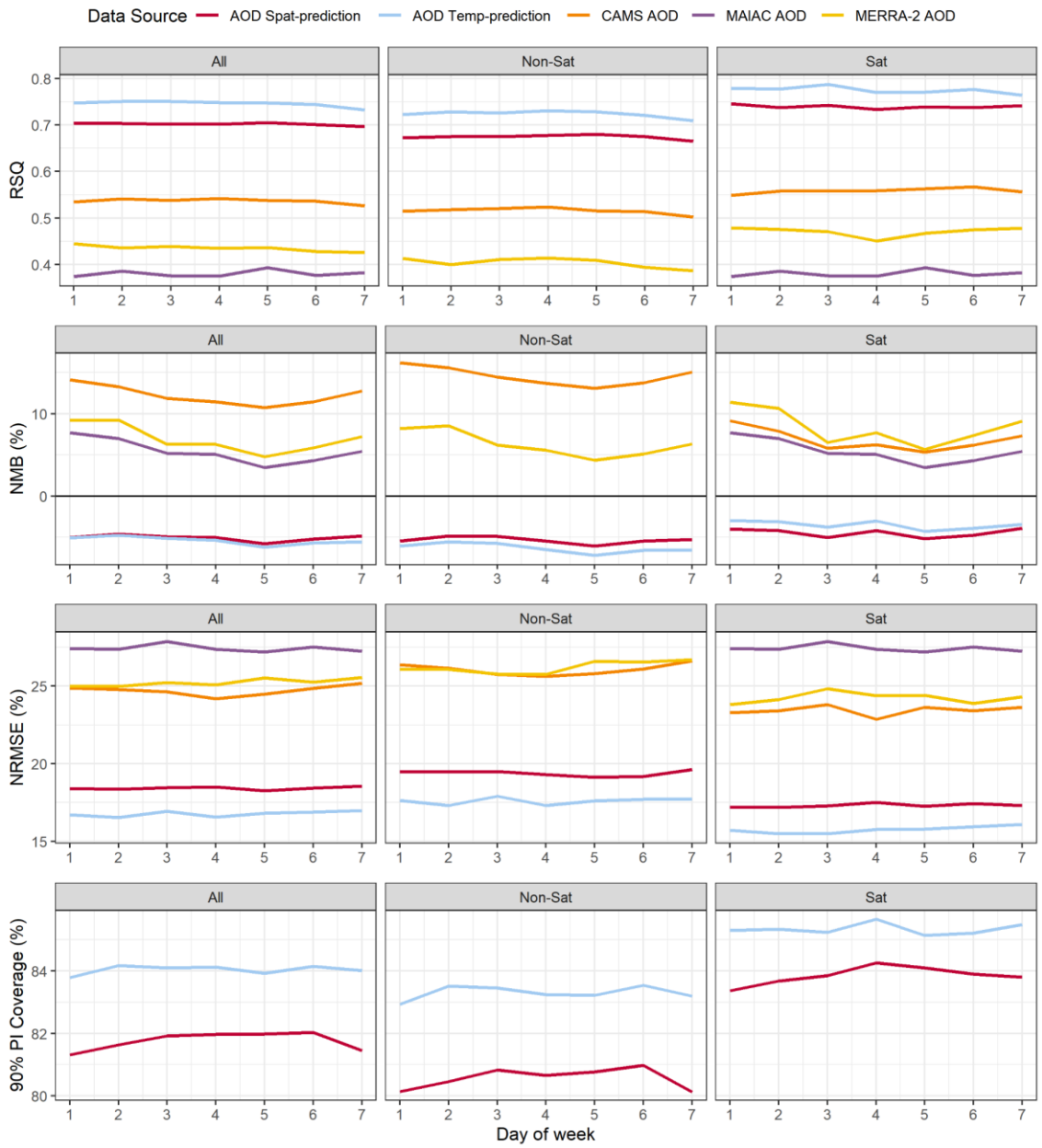

Figure S3. the weekly performance (Rsquare, NMB, NRMSE, 90% PI (predictive intervals) coverage) comparison between different data sources: QML spatial out-of-sample prediction (AOD Spat-Prediction), temporal out-of-sample prediction (AOD Temp-Prediction), CAMS AOD, MAIAC (Satellite) AOD, MERRA-2 AOD.

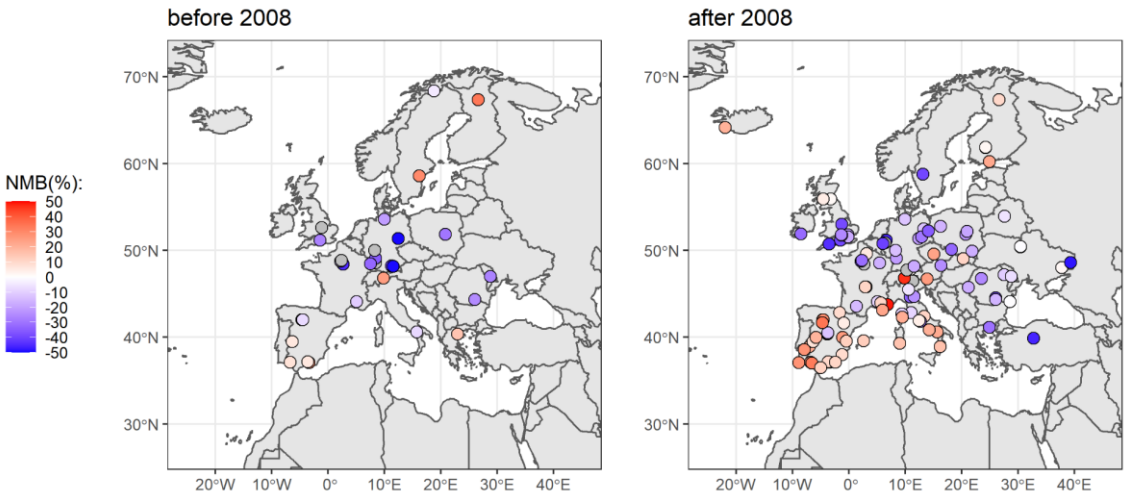

Figure S4. the spatial distribution maps of normalize mean bias (NMB) for the Phy-DL Satellite fAOD (PDL MAIAC fAOD) before 2008 and after 2008

Supplementary Figure

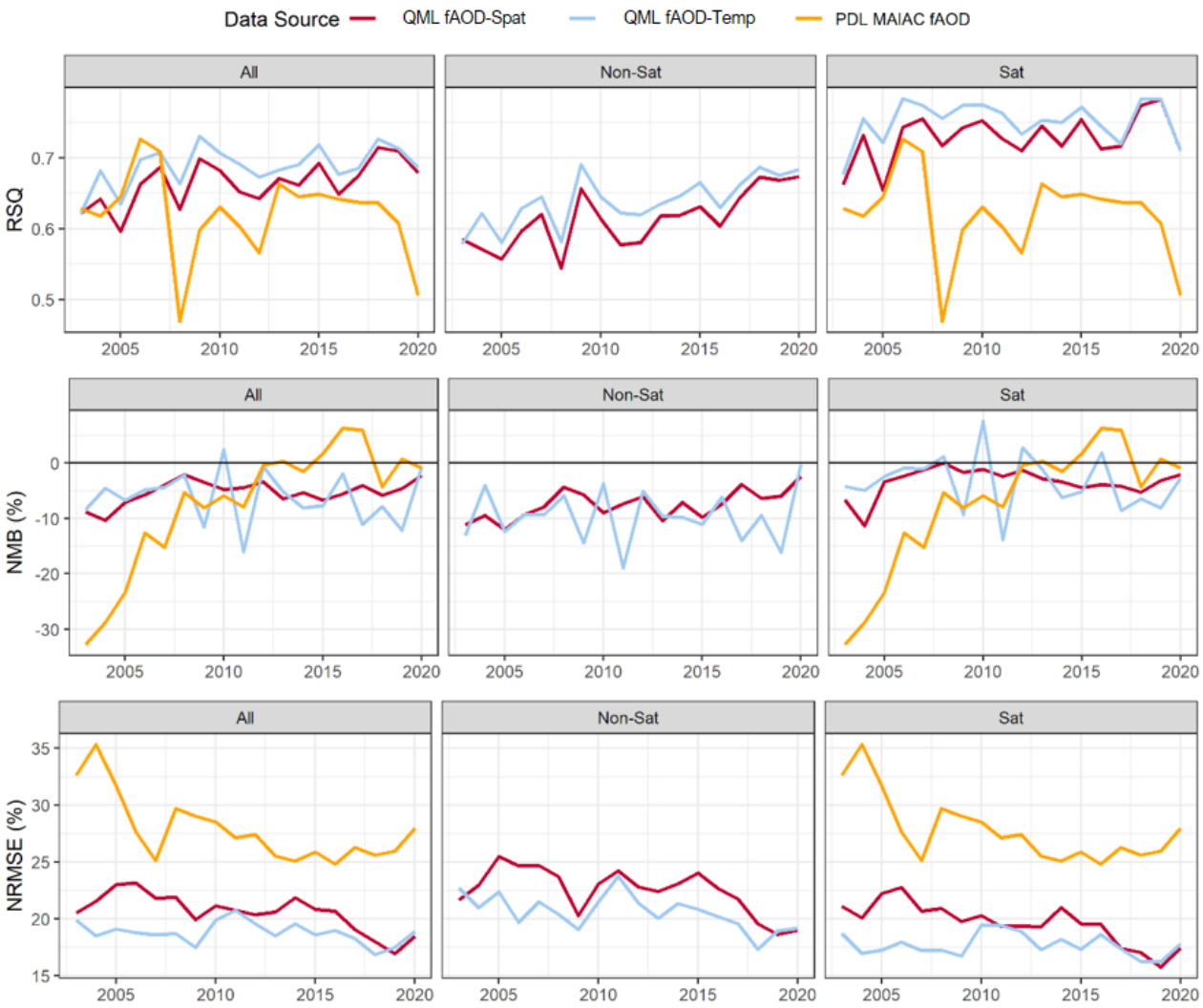

Figure S5. the annual performance (R-squared, NMB, NRMSE) comparison between different data sources: QML spatial out-of-sample prediction (fAOD Spat-Prediction), temporal out-of-sample prediction (fAOD Temp-Prediction), Phy-DL Satellite fAOD(PDL MAIAC fAOD).

Supplementary Figure

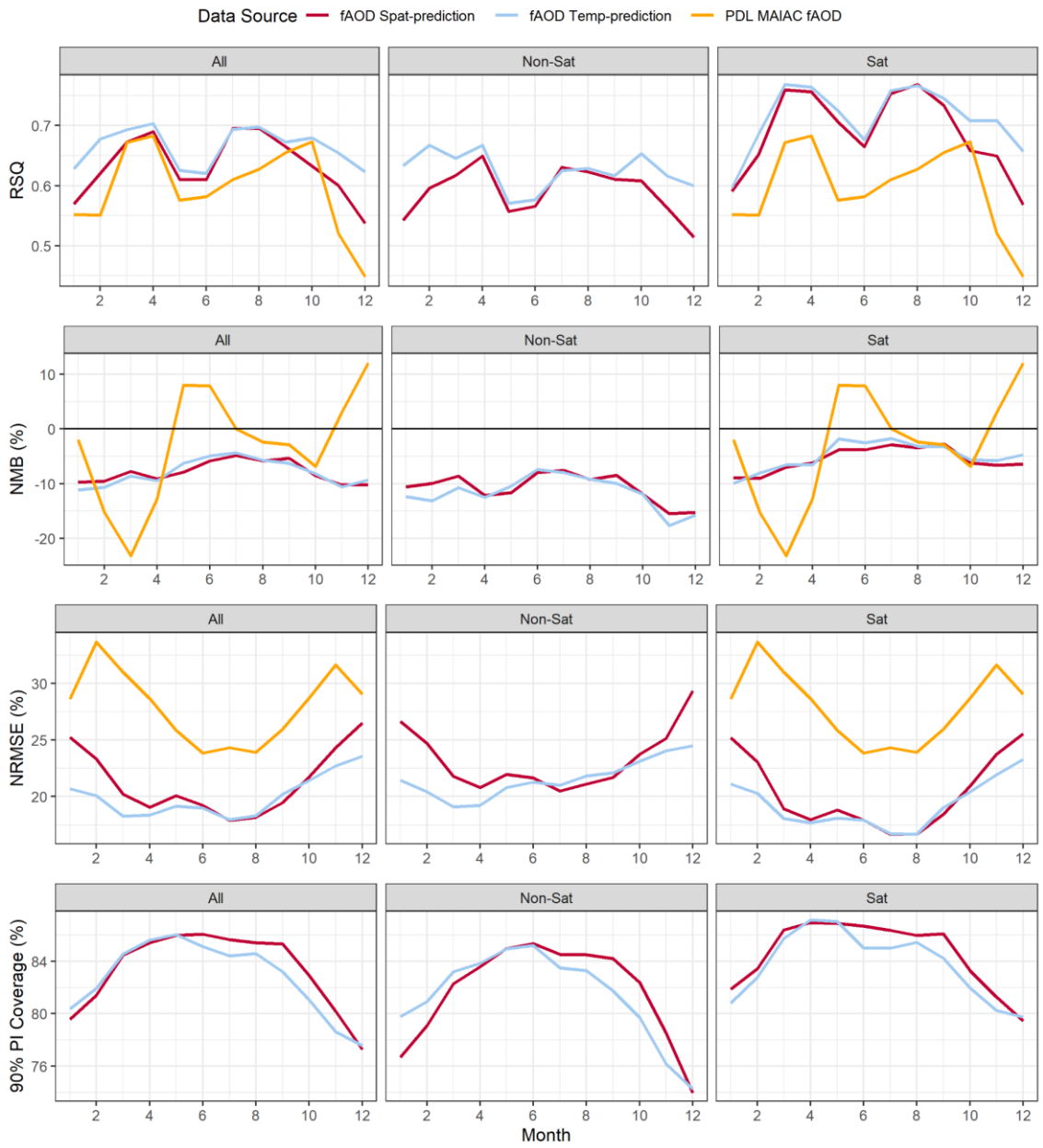

Figure S6. the monthly performance (Rsquare, NMB, NRMSE, 90% PI (predictive intervals) coverage) comparison between different data sources: QML spatial out-of-sample prediction (fAOD Spat-Prediction), temporal out-of-sample prediction (fAOD Temp-Prediction), Phy-DL Satellite fAOD(PDL MAIAC fAOD).

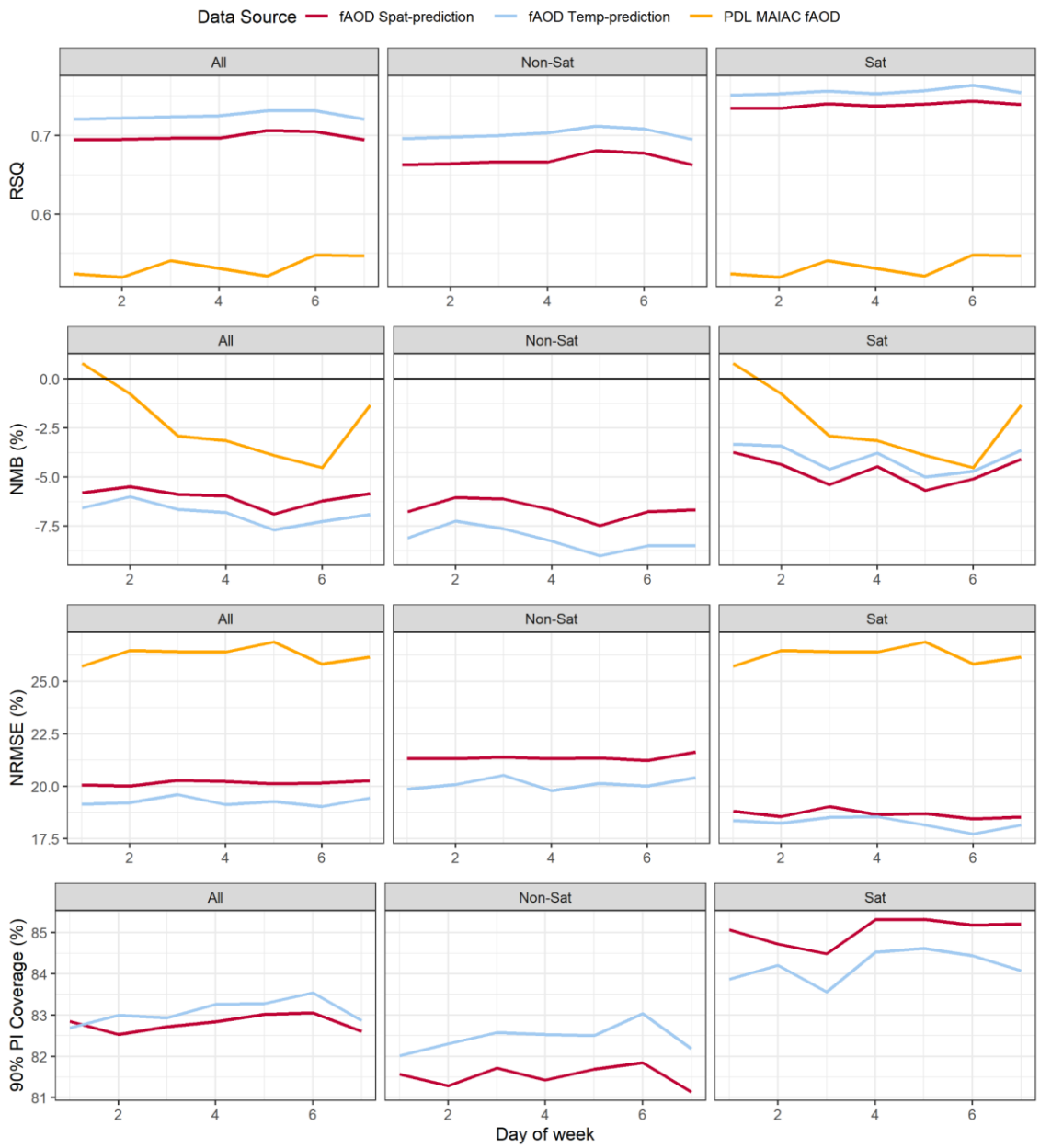

Figure S7. the weekly performance (Rsquare, NMB, NRMSE, 90% PI (predictive intervals) coverage) comparison between different data sources: QML spatial out-of-sample prediction (fAOD Spat-Prediction), temporal out-of-sample prediction (fAOD Temp-Prediction), Phy-DL Satellite fAOD(PDL MAIAC fAOD).

Supplementary Figure

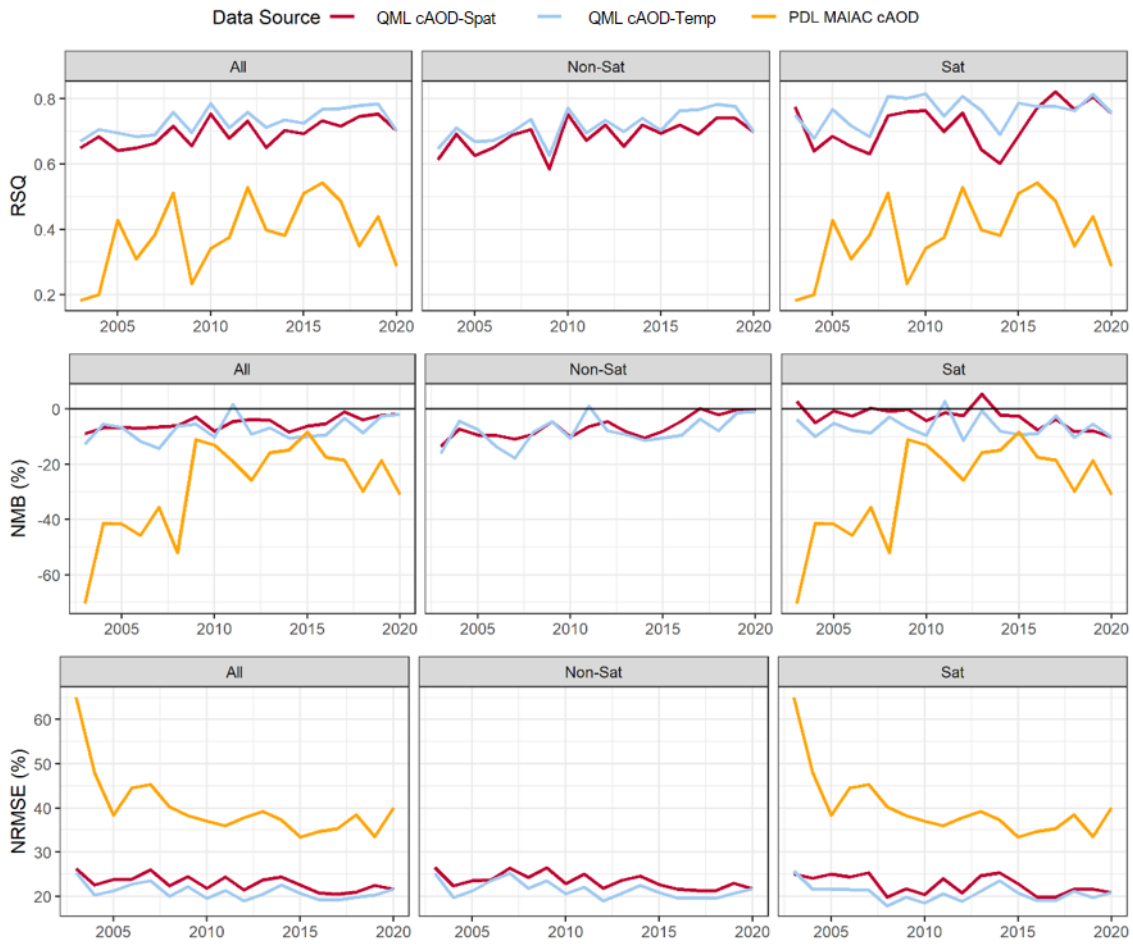

Figure s8. the performance (R-squared, NMB, NRMSE) comparison between different data sources over years: our spatial out-of-sample prediction (cAOD Spat-Prediction), temporal out-of-sample prediction (cAOD Temp-Prediction) and Phy-DL Satellite cAOD (PDL MAIAC cAOD).

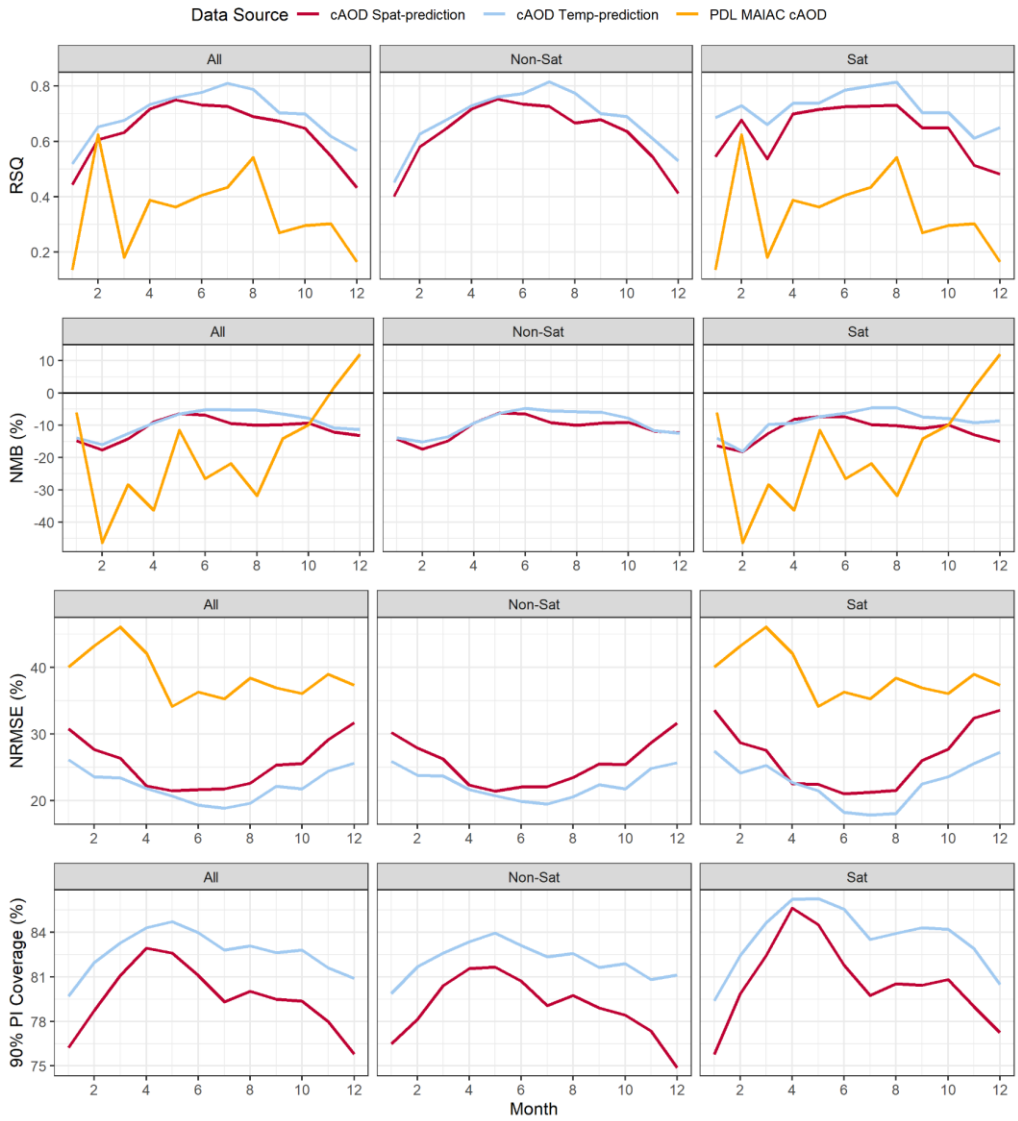

Figure S9. the monthly performance (Rsquare, NMB, NRMSE, 90% PI (predictive intervals) coverage) comparison between different data sources: QML spatial out-of-sample prediction (cAOD Spat-Prediction), temporal out-of-sample prediction (cAOD Temp-Prediction), Phy-DL Satellite cAOD(PDL MAIAC cAOD).

Supplementary Figure

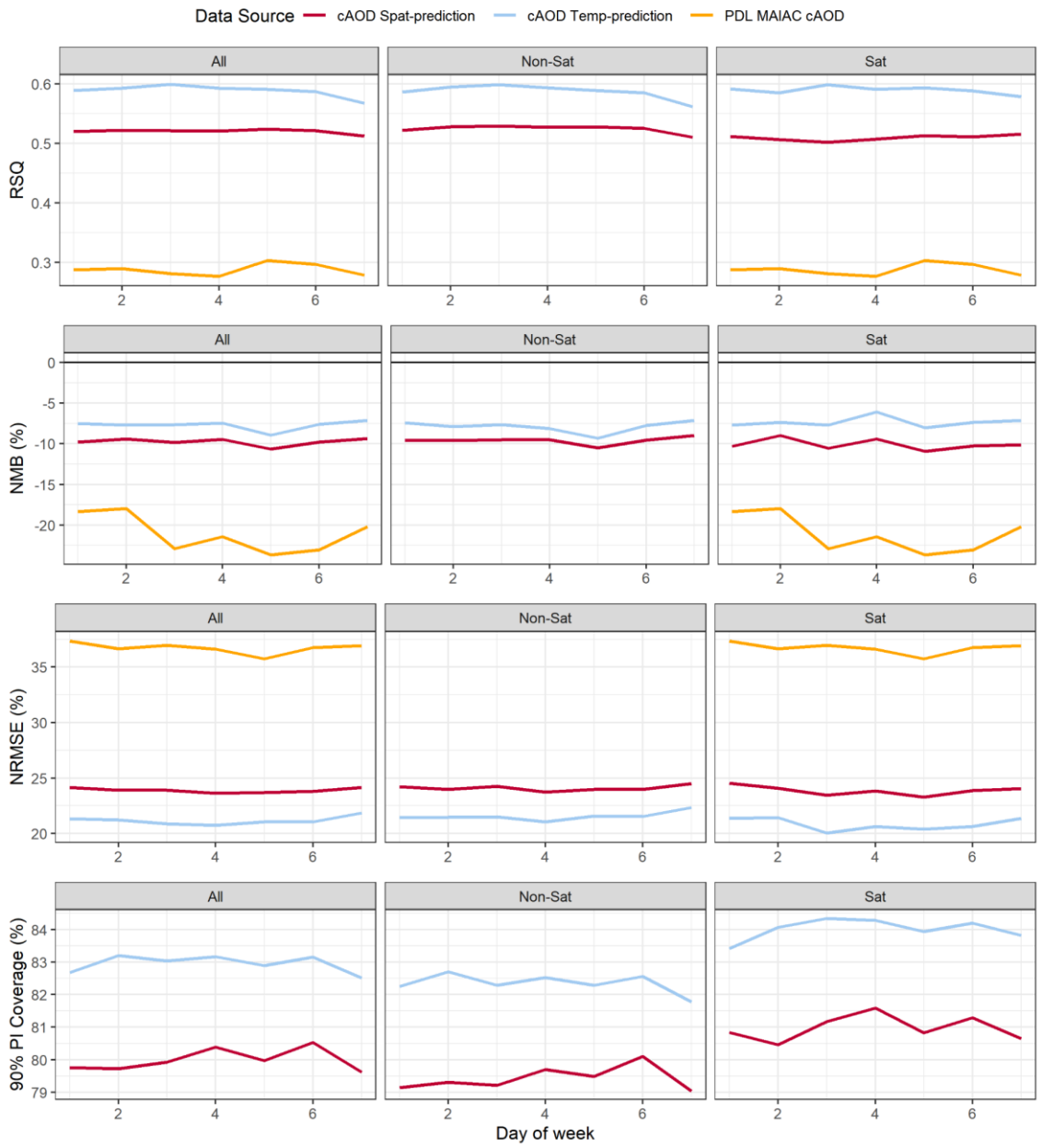

Figure S10. the weekly performance (Rsquare, NMB, NRMSE, 90% PI (predictive intervals) coverage) comparison between different data sources: QML spatial out-of-sample prediction (cAOD Spat-Prediction), temporal out-of-sample prediction (cAOD Temp-Prediction), Phy-DL Satellite cAOD(PDL MAIAC cAOD).

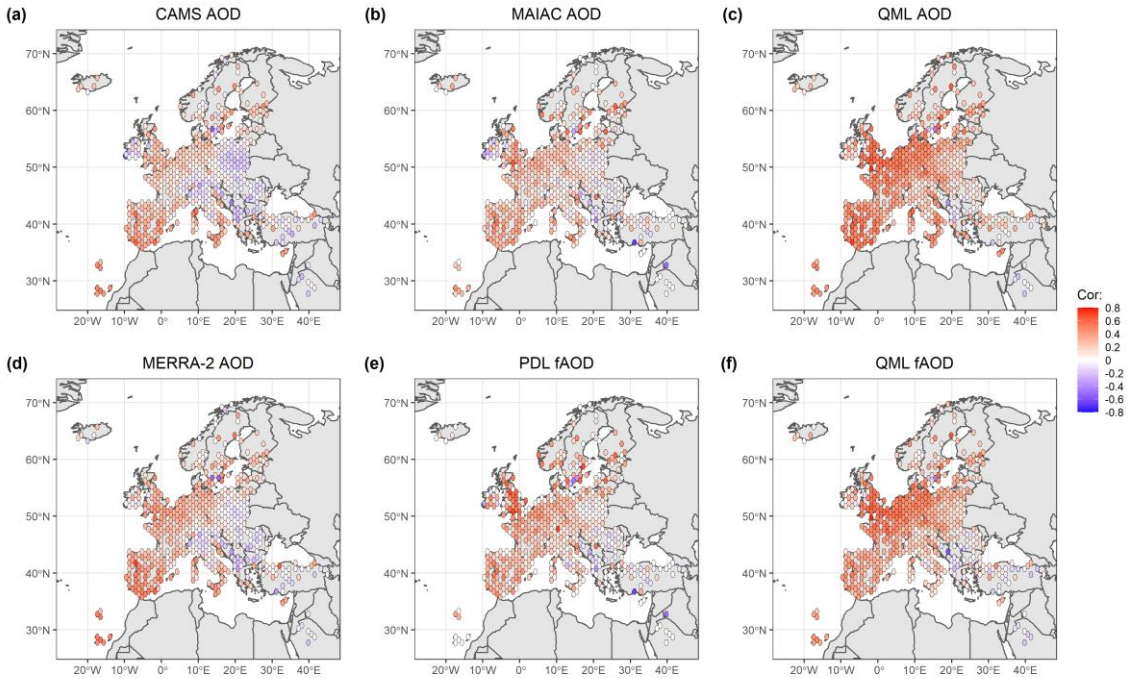

Figure S11. the correlation maps between PM10 and 6 different AOD product 2003-2020: CAMS AOD, MERRA-2 AOD. Satellite AOD, Phy-DL Satellite fAOD (PDL MAIAC fAOD), QML total AOD predictions (QML AOD), and QML fine-mode AOD predictions (QML fAOD)

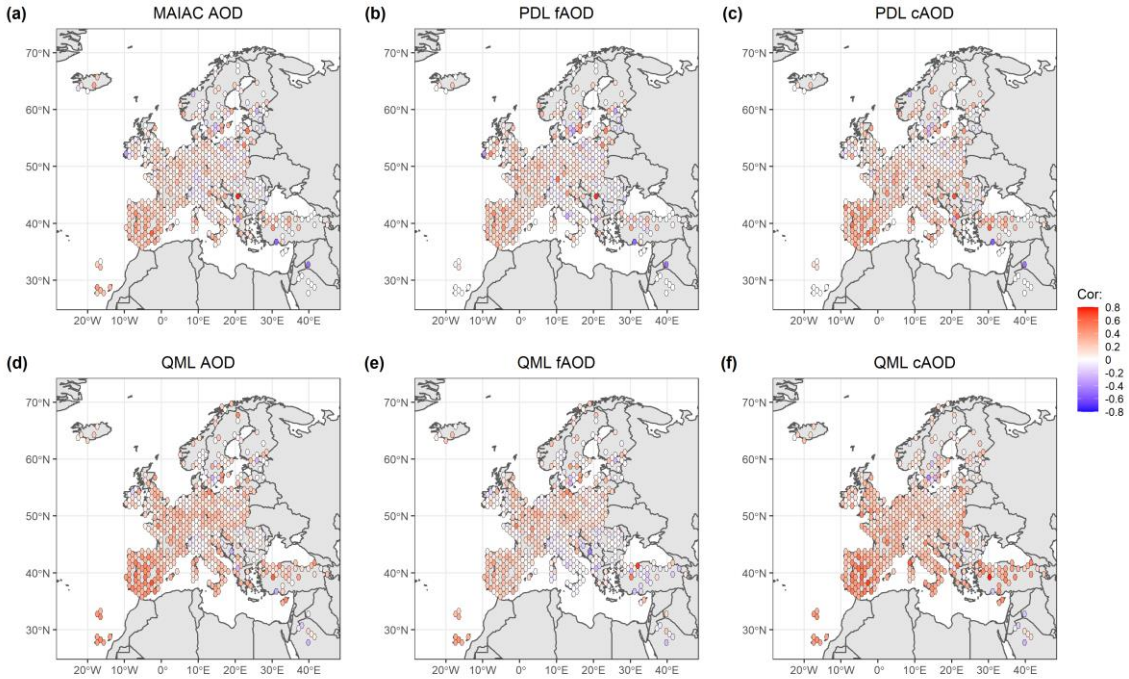

Figure S12. the correlation maps between Coarse PM and 6 different AOD product 2003-2020: Satellite AOD, Phy-DL Satellite fAOD (PDL MAIAC fAOD), Phy-DL Satellite cAOD (PDL MAIAC cAOD), QML total AOD (QML AOD), fine-mode AOD predictions (QML fAOD) and coarse-mode AOD predictions (QML cAOD).

Supplementary Figure

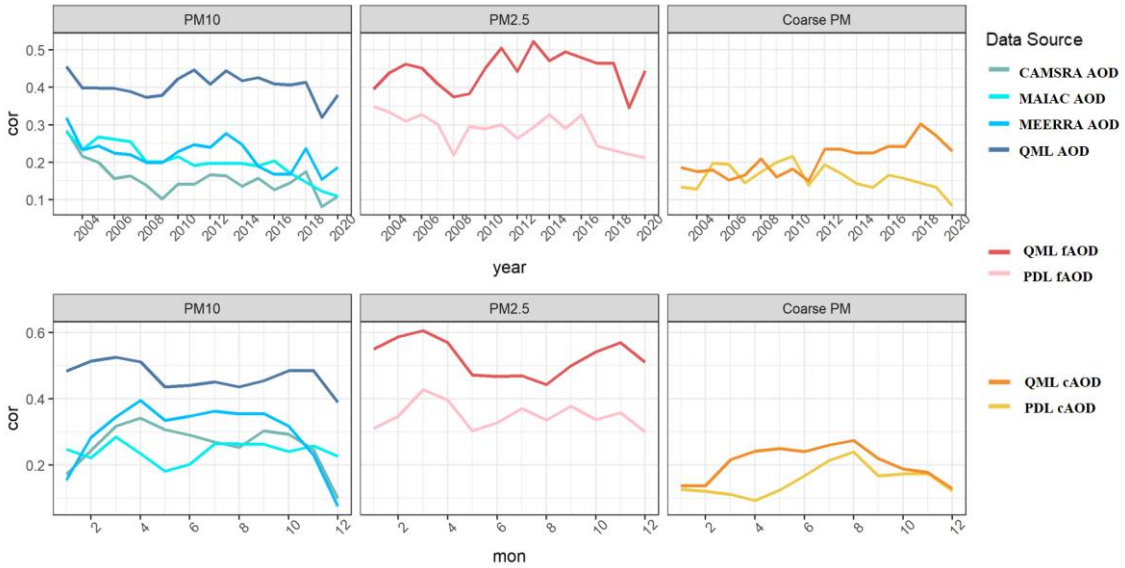

Figure s13. the spearman correlation between PM10 and total AOD products (CAMS, MERRA-2, Satellite AOD and QML AOD); PM2.5 and Fine-mode AOD products (PDL fAOD and QML fAOD); Coarse PM and coarse-mode AOD products (PDL cAOD and QML cAOD).

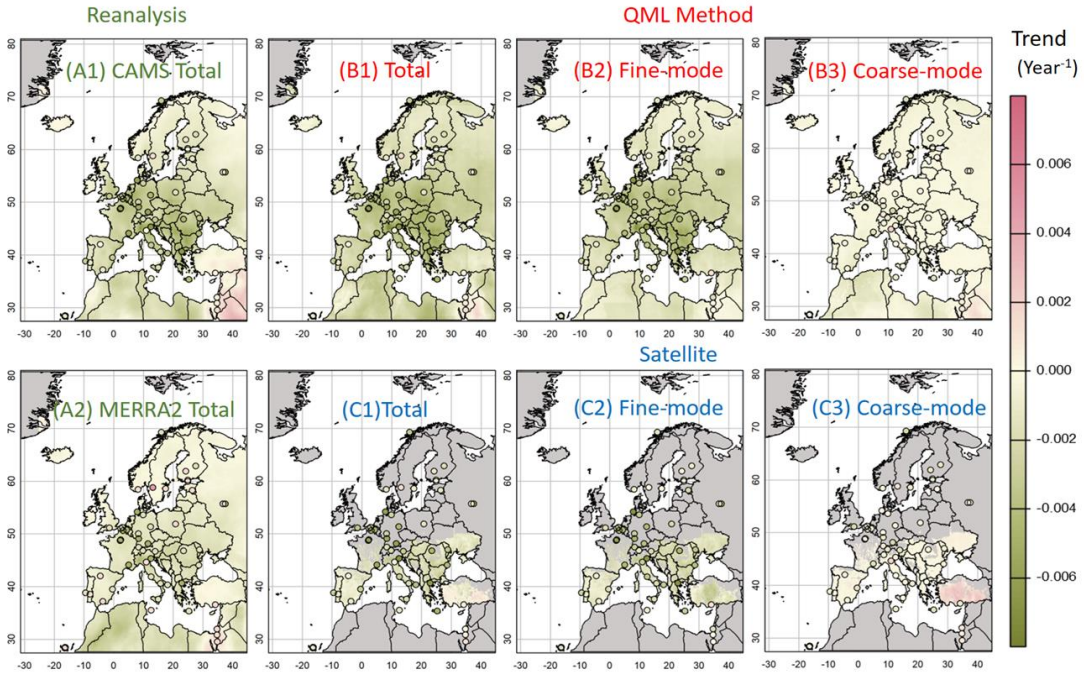

Figure s14. the maps of 18-year trend from 2003 to 2020: Reanalysis AOD (CAMS and MERRA-2) (A1-A2), QML product (AOD, fAOD and cAOD) (B1-B3), Satellite AOD (MAIAC AOD, PDL fAOD and PDL cAOD) (C1-C3) and the corresponding 46 ground-level sites, with more than ten years and at least 50 daily observation for each year.

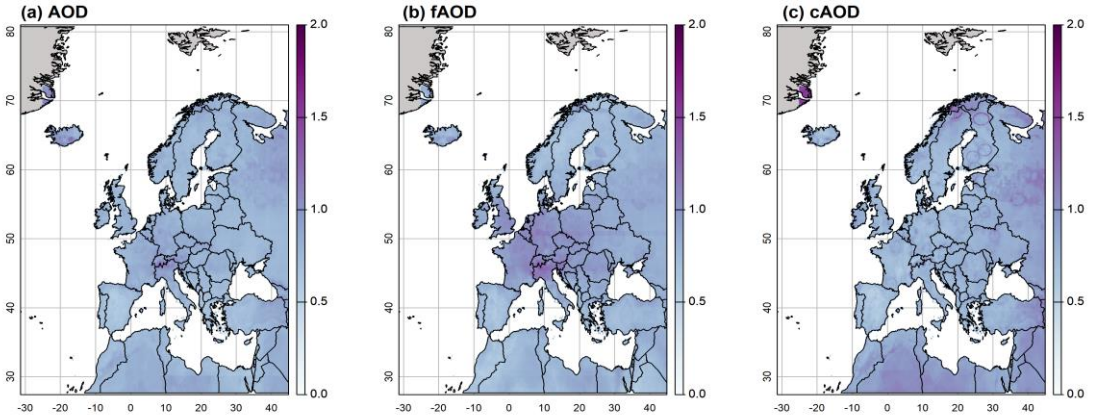

Figure s15. the averages of relative 90%-predictive-intervals width (RPIW= (90% PI upper-90% PI lower)/estimates) among tAOD, fAOD and cAOD products in the Pan European domain from 2003-2020. Higher RPIW means higher likelihood to obtain the uncertainty predictions.

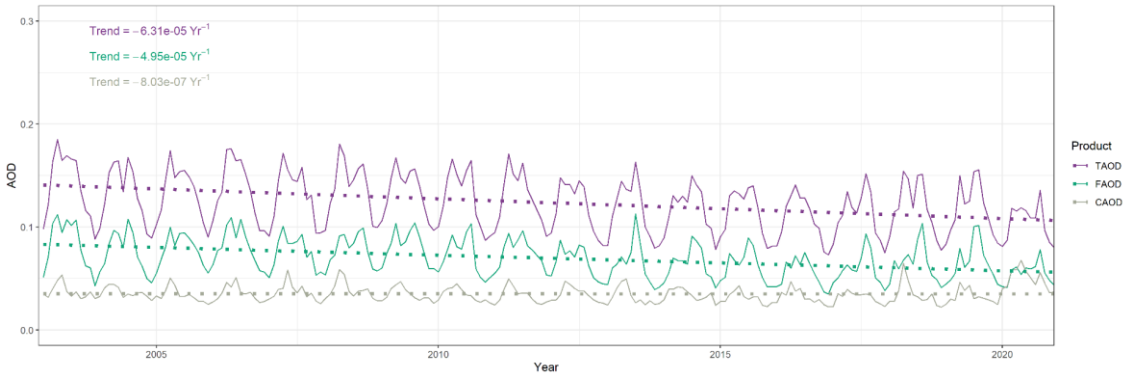

Figure S16. the time series of monthly mean tAOD, fAOD and cAOD products in the Pan European domain from 2003-2020.

2. Supplementary Table

Table S1. The comparison of model settings and results in spatial and temporal cross-validation

tAOD			
	Final[a]	Spatial[a]	Temporal[a]
Variables[b]	u10, v10, rh, t2m, asn, bld, blh, mcc, lai_hv, lai_lv,lcc, msdwswrf,sp,tco3,aluvd,aluvp, Year, DOW, DOY, Lat,Lon, NE,SE,SW,NW, cams_bcaod550, cams_duaod550, cams_omaod550, cams_ssaod550, cams_suaod550, merra_aod		
Out-of-sample R-squared[c]	0.72	0.68-0.74	0.65-0.79
Max depth	18	18(4);16(1)	18(3);17(2);16(1)
Subsample	0.8	0.8(4);0.7(1)	0.8(4);0.9(2)
Learning rate	0.05	0.05	0.05
Colsample by tree	0.8	0.8(4);0.7(1)	0.8(4);0.7(2)
fAOD			
	Final	Spatial	Temporal
Variables	u10, v10, d2m, t2m, bld, blh, mcc,tcc,slt, msdwswrf,sp,tco3,tp,aluvd,aluvp, Year, DOW, DOY, Lat,Lon, NE,SE,SW,NW, cams_bcaod550, cams_duaod550, cams_omaod550, cams_ssaod550, cams_suaod550, merra_aod		
Out-of-sample R-squared	0.67	0.65-0.73	0.61-0.75
Max depth	17	17(3);18(2)	17(4);16(2)
Subsample	0.8	0.8(4);0.7(1)	0.8(4);0.9(2)
Learning rate	0.05	0.05	0.05
Colsample by tree	0.8	0.8(4);0.7(1)	0.8(4);0.7(2)
cAOD			
	Final	Spatial	Temporal
Variables	u10, v10, d2m, t2m, bld, blh, mcc, lai_hv, lai_lv,lcc, msdwswrf,sp,tco3,aluvd,aluvp, Year, DOW, DOY, Lat,Lon, NE,SE,SW,NW, cams_bcaod550, cams_duaod550, cams_omaod550, cams_ssaod550, cams_suaod550, merra_aod		
Out-of-sample R-squared	0.68	0.68-0.74	0.62-0.78
Max depth	15	15(4);16(1)	16(3);15(2);14(1)
Subsample	0.8	0.8(4);0.7(1)	0.8(5);0.7(1)
Learning rate	0.05	0.05	0.05
Colsample by tree	0.8	0.8(4);0.7(1)	0.8(5);0.7(1)

[a] Final: the final trained models; Spatial: five models generated by different spatial subsamples, Temporal: Six models generated by different temporal subsamples. And the numbers in brackets for each parameter means how many models selected this parameter settings by their own parameter validation.

[b] the names of variables shown Table s2

[c] The performance in test datasets (not involving in the model development and parameter optimization) for final models. The performance in the spatial/temporal test fold of sub-sample model.

[d] Max depth: higher mean higher complexity of trees; Subsample: Subsample ratio of the training instances for growing each tree, to control overfit problem; Learning Rate:Step size shrinkage used in update to prevents overfitting; Colsample by tree: hSubsample ratio of inputs used for growing each tree.

Table S2. The list of variables used in this study

SHORT NAME	SOURCE	LONG NAME	UNIT
U10	ERA5	10m u component of wind	m s**-1
V10	ERA5	10m v component of wind	m s**-1
RH	ERA5	relatively humidity	(0 - 100)
D2M	ERA5	2m dewpoint temperature	K
T2M	ERA5	2m temperature	K
BLD	ERA5	boundary layer dissipation	J m**-2
BLH	ERA5	boundary layer height	m
HCC	ERA5	high cloud cover	(0 - 1)
LAI_HV	ERA5	leaf area index high vegetation	m**2 m**-2
LAI_LV	ERA5	leaf area index low vegetation	m**2 m**-2
TCC	ERA5	total cloud cover	(0 - 1)
LCC	ERA5	low cloud cover	(0 - 1)
SLT	ERA5	soil type	1-7, higher is finer soil with stronger ability contains water
MSDWSWRF	ERA5	mean surface downward short wave radiation flux	W m**-2
ASN	ERA5	snow albedo	(0 - 1)
MCC	ERA5	medium cloud cover	(0 - 1)
SP	ERA5	surface pressure	Pa
TCO3	ERA5	total column ozone	J m**-2
TP	ERA5	total precipitation	m
ALUVD	ERA5	uv visible albedo for diffuse radiation	(0 - 1)
ALUVP	ERA5	uv visible albedo for direct radiation	(0 - 1)
YEAR	Time	Year	\
DOW	Time	day of week	\
DOY	Time	day of year	\
LAT	Spatial	latitude	\
LON	Spatial	longtitude	\
NE	Minimum directional distance	minimum distance to nearest sites in North-east direction	m
SE	Minimum directional distance	minimum distance to nearest sites in South-east direction	m
SW	Minimum directional distance	minimum distance to nearest sites in South-west direction	m
NW	Minimum directional distance	minimum distance to nearest sites in North-west direction	m
CAMS_BCAOD550	CAMSRA	black carbon aerosol optical depth 550nm	\
CAMS_DUAOD550	CAMSRA	dust aerosol optical depth 550nm	\
CAMS_OMAOD550	CAMSRA	organic matter aerosol optical depth 550nm	\
CAMS_SSAOD550	CAMSRA	sea salt aerosol optical depth 550nm	\
CAMS_SUAOD550	CAMSRA	sulphate aerosol optical depth 550nm	\
MERRA_AOD	MERRA-2	MERRA2 aerosol optical depth 550nm	\

Table S3. The performance comparison of additional techniques in parameter validation datasets (20% sites), with R-squared (Rsq), Root Mean Square Error (RMSE) and their standard deviation(SD).

	tAOD		**fAOD**		**cAOD**	
All validation sites						
	Rsq(SD)	RMSE(SD)	Rsq(SD)	RMSE(SD)	Rsq(SD)	RMSE(SD)
Origin (O)	0.64(0.20)	0.05(0.07)	0.63(0.25)	0.03(0.05)	0.61(0.35)	0.013(0.02)
M1	0.64(0.17)	0.04(0.05)	0.64(0.23)	0.02(0.04)	0.62(0.33)	0.012(0.03)
M2	0.66(0.10)	0.04(0.03)	0.65(0.13)	0.02(0.01)	0.63(0.17)	0.009(0.01)
M3	0.71(0.12)	0.03(0.04)	0.68(0.15)	0.02(0.03)	0.67(0.20)	0.007(0.02)
The validation sites far away from nearest sites [a]						
	Rsq(SD)	RMSE(SD)	Rsq(SD)	RMSE(SD)	Rsq(SD)	RMSE(SD)
Origin (O)	0.42(0.25)	0.07(0.10)	0.44(0.28)	0.04(0.05)	0.36(0.28)	0.02(0.02)
M1	0.48(0.24)	0.06(0.04)	0.47(0.26)	0.03(0.04)	0.41(0.24)	0.015(0.02)
M2	0.49(0.23)	0.05(0.03)	0.48(0.22)	0.03(0.02)	0.46(0.21)	0.013(0.02)
M3	0.58(0.24)	0.04(0.04)	0.56(0.21)	0.02(0.03)	0.52(0.22)	0.011(0.02)

[a] The distance from these sites to their nearest sites is larger than 463km (the average distance from 257 AERONET sites to their nearest sites)

[b] M1= O+ weighted loss function; M2= M1+ data augment; M3= M2 + minimum directional distance

Table S3. The main evaluation criteria used in this paper

Criteria	Formula	Function	Range
R-squared	$SS_{regression}/SS_{total}$	Proportion of variance explained by models	0 to 1, higher is better
Normalized Mean Bias (NMB %)	$\frac{\sum_{k=1}^{n}(Pred_k - Obs_k)}{\sum_{k=1}^{n}(Obs_k)}$	Overestimates/ Underestimates evaluation from normalized average forecast error	-Inf to Inf, closer to 0 is better
Normalized Root Mean Square Error (NRMSE %)	$\frac{\sqrt{\frac{\sum_{k=1}^{n}(Pred_k - Obs_k)^2}{N}}}{Avg(Obs)}$	Bias evaluation, give more weight to the largest errors.	0 to Inf, lower is better
Predictive Intervals (PI) Coverage (%)	$N_{within\_intervals}/N$	Stability evaluation for PI	0 to the PI levels it set (e.g., 90% for 90% PI), higher is better
Within 20% expected error envelopes (Within 20% EE %)	N_{within_EE}/N EE:0.05 ±20 % observation for AOD; 0.025 ±20 % observation for fAOD and cAOD	Evaluate how many sample´s bias controlled within 20% error.	0 to 100, higher is better
Robustness for R-squared	$\frac{Var(R^2)}{Avg(R^2)}$	level of dispersion for cross-validation R-squared	0 to Inf, lower is better
Trend inconsistency (TI)	$\left\lvert ln\left(\frac{T_{Pred}}{T_{Obs}}\right)\right\rvert$	Trend consistency index for Predictions and observation	0 to Inf, closer to 0 is better. Inf represent the opposite trends

Table S4. The coverage and width of 90% predictive intervals (PI) for tAOD model in 10 bins

tAOD Range	N	N of Coverage (%)	N of over-estimations$^{\alpha}$ (%)	N of under-estimations$^{\beta}$ (%)	Interval width (avg)	Interval width (sd)
Spatial						
[0.00478,0.0424]	26476	14636(55.28)	11457(43.27)	550(2.08)	0.068	0.045
(0.0424,0.0584]	26493	22060(83.27)	4153(15.68)	277(1.05)	0.081	0.045
(0.0584,0.0731]	26457	23360(88.29)	2328(8.8)	768(2.9)	0.092	0.048
(0.0731,0.089]	26475	23688(89.47)	1536(5.8)	1251(4.73)	0.103	0.050
(0.089,0.107]	26476	23386(88.33)	1547(5.84)	1543(5.83)	0.116	0.056
(0.107,0.131]	26475	23406(88.41)	1022(3.86)	2046(7.73)	0.132	0.058
(0.131,0.162]	26475	23281(87.94)	854(3.23)	2338(8.83)	0.149	0.062
(0.162,0.205]	26475	22980(86.8)	501(1.89)	2992(11.3)	0.173	0.069
(0.205,0.279]	26475	22383(84.54)	208(0.79)	3883(14.67)	0.207	0.080
(0.279,1.99]	26476	18103(68.38)	45(0.17)	8327(31.45)	0.293	0.141
Temporal						
[0.00478,0.0424]	26476	17611(66.52)	8171(30.86)	724(2.73)	0.056	0.040
(0.0424,0.0584]	26493	22496(84.91)	3457(13.05)	539(2.03)	0.076	0.044
(0.0584,0.0731]	26457	23616(89.26)	2031(7.68)	808(3.05)	0.090	0.047
(0.0731,0.089]	26475	23953(90.47)	1380(5.21)	1142(4.31)	0.102	0.051
(0.089,0.107]	26476	23800(89.89)	984(3.72)	1691(6.39)	0.115	0.055
(0.107,0.131]	26475	23903(90.29)	805(3.04)	1766(6.67)	0.132	0.061
(0.131,0.162]	26475	23721(89.6)	571(2.16)	2183(8.25)	0.149	0.064
(0.162,0.205]	26475	23527(88.86)	398(1.5)	2548(9.62)	0.173	0.072
(0.205,0.279]	26475	22727(85.84)	218(0.82)	3528(13.33)	0.209	0.085
(0.279,1.99]	26476	20244(76.46)	44(0.17)	6187(23.37)	0.313	0.155

over-estimations$^{\alpha}$: the lower bounds of PI higher than the real tAOD, under-estimations$^{\beta}$: the upper bounds of PI lower than the real tAOD

Table S5. The coverage and width of 90% predictive intervals (PI) for fAOD model in 10 bins

fAOD Range	N	N of Coverage (%)	N of over-estimations$^{\alpha}$ (%)	N of under-estimations$^{\beta}$ (%)	Interval width (avg)	Interval width (sd)
Spatial						
[6.36e-05,0.0228]	26475	16491(62.29)	9493(35.86)	515(1.95)	0.045	0.030
(0.0228,0.0333]	26475	21423(80.92)	4508(17.03)	546(2.06)	0.060	0.037
(0.0333,0.0433]	26475	23125(87.35)	2516(9.5)	834(3.15)	0.067	0.040
(0.0433,0.0538]	26524	23221(87.55)	2105(7.94)	1195(4.51)	0.078	0.043
(0.0538,0.0662]	26475	23734(89.65)	1342(5.07)	1399(5.28)	0.087	0.046
(0.0662,0.0821]	26479	23781(89.81)	919(3.47)	1779(6.72)	0.102	0.050
(0.0821,0.103]	26420	23666(89.58)	594(2.25)	2160(8.18)	0.117	0.054
(0.103,0.134]	26475	23636(89.28)	316(1.19)	2522(9.53)	0.135	0.060
(0.134,0.19]	26475	22533(85.11)	135(0.51)	3806(14.38)	0.163	0.069
(0.19,1.97]	26475	18979(71.69)	17(0.06)	7477(28.24)	0.245	0.120
Temporal						
[6.36e-05,0.0228]	26475	16701(63.08)	8973(33.89)	860(3.25)	0.037	0.029
(0.0228,0.0333]	26475	21683(81.9)	4025(15.2)	767(2.9)	0.054	0.037
(0.0333,0.0433]	26475	23320(88.08)	2228(8.42)	927(3.5)	0.064	0.040
(0.0433,0.0538]	26524	23470(88.49)	1457(5.49)	1596(6.02)	0.075	0.045
(0.0538,0.0662]	26475	23773(89.79)	1013(3.83)	1688(6.38)	0.085	0.048
(0.0662,0.0821]	26479	23683(89.44)	738(2.79)	2058(7.77)	0.101	0.053
(0.0821,0.103]	26420	23870(90.35)	459(1.74)	2091(7.91)	0.116	0.058
(0.103,0.134]	26475	23523(88.85)	249(0.94)	2702(10.21)	0.135	0.063
(0.134,0.19]	26475	22602(85.37)	112(0.42)	3760(14.2)	0.167	0.075
(0.19,1.97]	26475	19896(75.15)	8(0.03)	6569(24.81)	0.260	0.134

over-estimations$^{\alpha}$: the lower bounds of PI higher than the real fAOD, under-estimations$^{\beta}$: the upper bounds of PI lower than the real fAOD

Table S6. The coverage and width of 90% predictive intervals (PI) for cAOD model in 10 bins

fAOD Range	N	N of Coverage (%)	N of over-estimations$^{\alpha}$ (%)	N of under-estimations$^{\beta}$ (%)	Interval width (avg)	Interval width (sd)
Spatial						
[5.06e-09,0.00649]	26660	4358(16.35)	21502(80.65)	29(0.11)	0.034	0.020
(0.00649,0.0106]	26439	16172(61.17)	9472(35.83)	199(0.75)	0.038	0.021
(0.0106,0.0155]	26381	22649(85.85)	2411(9.14)	119(0.45)	0.039	0.022
(0.0155,0.0211]	26494	23978(90.5)	1065(4.02)	237(0.89)	0.042	0.021
(0.0211,0.0267]	26493	23002(86.82)	1108(4.18)	794(3)	0.045	0.025
(0.0267,0.0331]	26544	22460(84.61)	702(2.64)	1883(7.09)	0.048	0.025
(0.0331,0.0425]	26443	20763(78.52)	666(2.52)	2934(11.1)	0.050	0.027
(0.0425,0.0595]	26493	18940(71.49)	810(3.06)	5051(19.07)	0.060	0.035
(0.0595,0.104]	26703	17382(65.09)	505(1.89)	6784(25.41)	0.088	0.048
(0.104,1.87]	26284	17083(64.99)	1013(3.85)	7338(27.92)	0.173	0.113
Temporal						
[5.06e-09,0.00649]	26660	8678(32.55)	16778(62.93)	450(1.69)	0.035	0.022
(0.00649,0.0106]	26439	16831(63.66)	8869(33.55)	139(0.53)	0.039	0.023
(0.0106,0.0155]	26381	23438(88.84)	1528(5.79)	211(0.8)	0.039	0.022
(0.0155,0.0211]	26494	24300(91.72)	674(2.54)	306(1.15)	0.044	0.023
(0.0211,0.0267]	26493	23458(88.54)	1036(3.91)	410(1.55)	0.046	0.025
(0.0267,0.0331]	26544	23471(88.42)	654(2.46)	920(3.47)	0.049	0.024
(0.0331,0.0425]	26443	22110(83.61)	349(1.32)	1905(7.2)	0.052	0.027
(0.0425,0.0595]	26493	20512(77.42)	465(1.76)	3825(14.44)	0.063	0.032
(0.0595,0.104]	26703	19514(73.08)	352(1.32)	4806(18)	0.090	0.043
(0.104,1.87]	26284	19872(75.6)	116(0.44)	5447(20.72)	0.177	0.113

over-estimations$^{\alpha}$: the lower bounds of PI higher than the real cAOD, under-estimations$^{\beta}$: the upper bounds of PI lower than the real cAOD

Table S7. The spearman correlation between different-size particulate matter (PM10, PM2.5, Coarse PM (from 2.5 to 10 micrometers)) and different AOD products.

	PM10	PM25	Coarse PM
QML AOD	0.414*	0.392*	0.205*
MAIAC AOD	0.227*	0.166*	0.165*
MERRA AOD	0.234*	0.183*	0.158*
CAMS AOD	0.170*	0.096*	0.157*
QML fAOD	0.375*	0.446*	0.099*
PDL fAOD	0.285*	0.285*	0.135*
QML cAOD	0.181*	0.021	0.258*
PDL cAOD	0.119*	0.011	0.163*

*with significant Spatial R

3. Preliminary analysis for satellite AOD

3.1 Study design

In the preliminary study, we designed three different routes for the comparison: Route A: without Satellite AOD; Route B: Satellite AOD Included; Route C: gap-filled Satellite AOD included are developed to compare and to determine the optimal model structure (Figure B1). The Route A is to train the Non-sat model without additional satellite data, well described in the manuscript.

The Route B is to answer whether the satellite AOD can bring more benefits to the models, under its relatively high spatial resolution. Due to the high missing rate of satellite data, this route needs to split the dataset into two scenarios: “Sat scenario” and “Non-Sat scenario”. For the sat scenario, we built the satellite calibration model to dig out the potential relationship between satellite MAIAC AOD and AERONET AOD. For the rest scenario, it remains the same model structure as Route A.

As for the route C, they try to answer whether we can maximize the information usage of satellite AOD observation. There are two stages in route C: First stage is designed to fill the missing gaps of MAIAC AOD, though modelling with reanalysis and other data. The route C assumed that splitting scenarios (like route B) may cause some discontinues surface while merging two scenarios´ data, and this discontinuous distribution may further lead to some bias in the future application. Second stage is to calibrate the missing-filled MAIAC AOD.

3.2 Preliminary results

Given in the high missing rate of Satellite AOD (75.79%) in our research domain during 2003-2020, three models route (Route A: without Satellite AOD; Route B: Satellite AOD Included; Route C: missing-filled Satellite AOD included) are developed to evaluate the contribution from satellite AOD, and to determine the optimal model structure. Figure s15 shows the radar plot for three different model routes under Six dimensions: Spatial and Temporal Accuracy (Spatial and Temporal CV R^2); Spatial and Temporal Robustness (the variation coefficients of Spatial CV R^2 among different locations and Temporal CV R^2 among different years); simplicity (the numbers of models required); Correlation with PM2.5. Notedly, for simplicity, there is only one model for Route A, while there are two for Route B (Route A model+ Sat Model) and Route C (gap-filled model+ Sat Model).

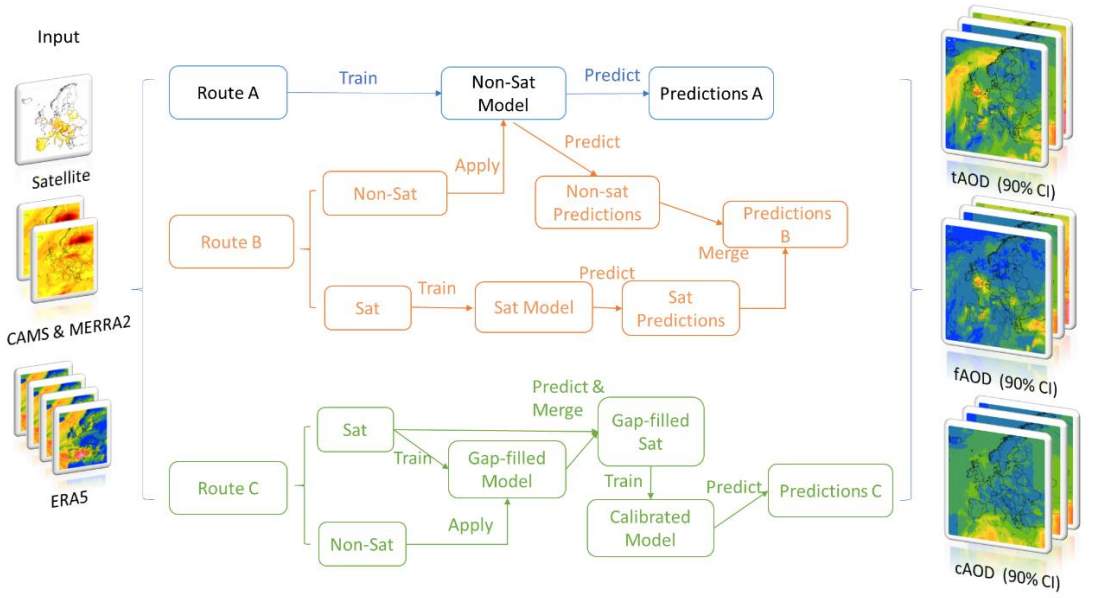

Figure s15. the workflows for three different model routes

For the total AOD (Fig s16(a)), the Spatial and temporal performance among three different model routes are similar (Temporal CV $R^2_{\text{Route A VS B VS C}}$=0.715 VS 0.718 Vs 0.718; Spatial CV $R^2_{\text{Route A VS B VS C}}$=0.691 VS 0.696 VS 0.695), and Satellite AOD only bring small improvement (0.015 in Spatial CV R^2). There is no difference between different satellite AOD applications: Building separate models for satellite AOD (Route B) and using gap-filled satellite AOD (Route C). In another word, the gap-filled product based on association between satellite AOD and meteorological factors cannot provide extra information for estimating AERONet AOD data.

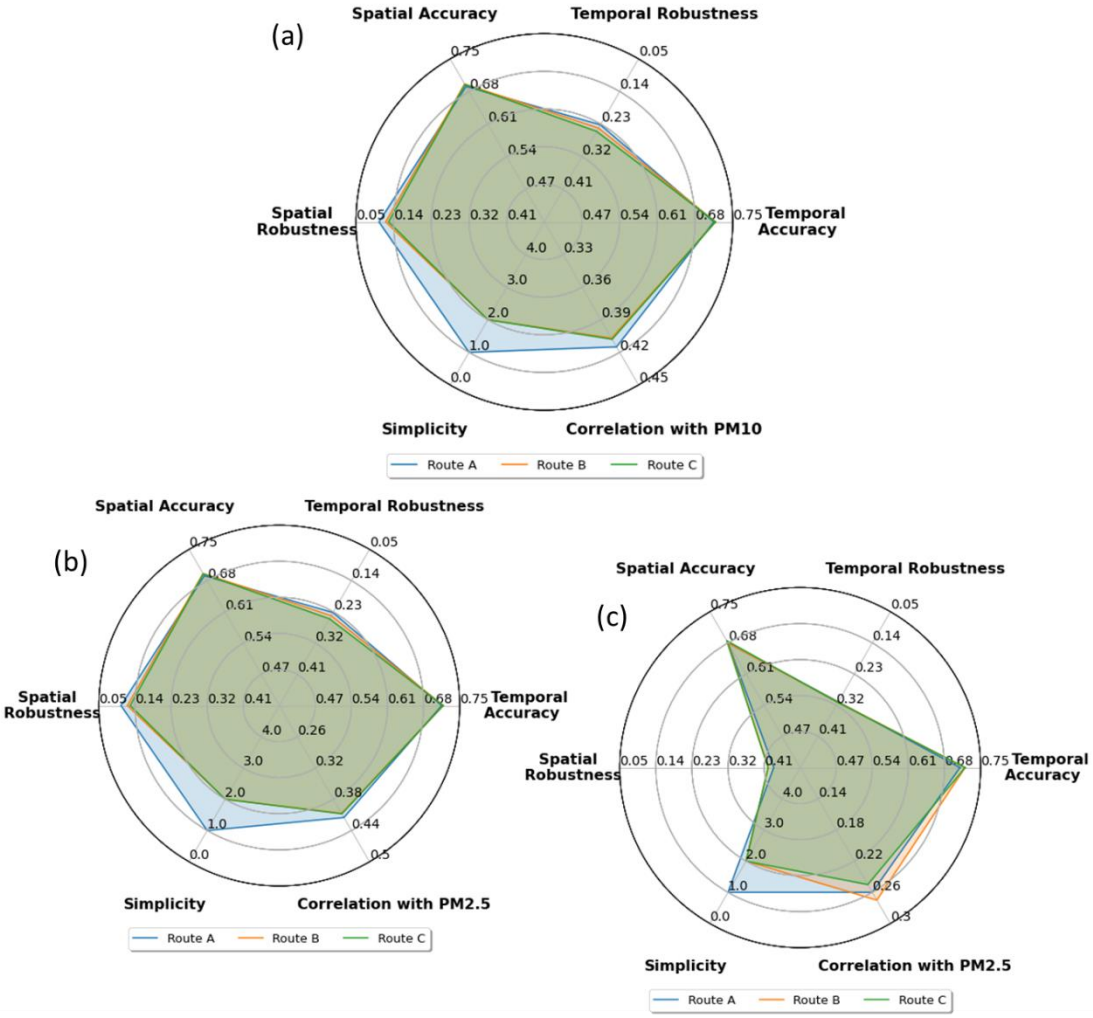

Figure s16. the radar plot for estimated AOD by three different routes (Route A (blue): without Satellite AOD; Route B (orange): Satellite AOD Included; Route C (green): missing-filled Satellite AOD included) under three different AOD product (tAOD (a), fAOD (b) and cAOD (c)). The Six dimensions in radar plot: Spatial and Temporal Accuracy (Spatial and Temporal CV R^2); Spatial and Temporal Robustness (the variation coefficients of Spatial CV R^2 among different locations and Temporal CV R^2 among different years); simplicity (the steps required); Correlation with $PM_{2.5}$.

To further figure out the reason of the small contribution from satellite AOD (Figure s17), we divided the validation results into two scenarios: "Sat scenario" and "Non-Sat scenario" (satellite MAIAC AOD are available or not). In the Sat scenarios, using the satellite AOD (Route B and Route C) can bring 0.061-0.069 improvement (Spatial CV R^2 $_{\text{Route A VS B VS C}}$=0.727 VS 0.796 VS 0.788), but three model routes have almost same performance (Spatial CV R^2 $_{\text{Route A VS B VS C}}$=0.623 VS 0.623 VS 0.622) during the Non-Sat scenario. However, the Sat scenario only accounts for 36.3% of 224,369 matching datasets, so model improvements in the Sat scenario cannot make a big change in overall performance. Furthermore, the fraction of Sat scenarios in the whole domain (24.21%) is smaller than AERONET matching datasets (36.30%) the weather condition, as the requirement between AERONET site and Satellite is a bit similar: sunlight cannot be blocked by clouds during measurements (for AERONET) or clear sky (for satellite) (Flag et al., 2015; Holben et al., 2006; GLOBE 2010). In another word, the differences among different routes (whether using satellite AOD as input) will be even smaller when they are applied in the whole domain.

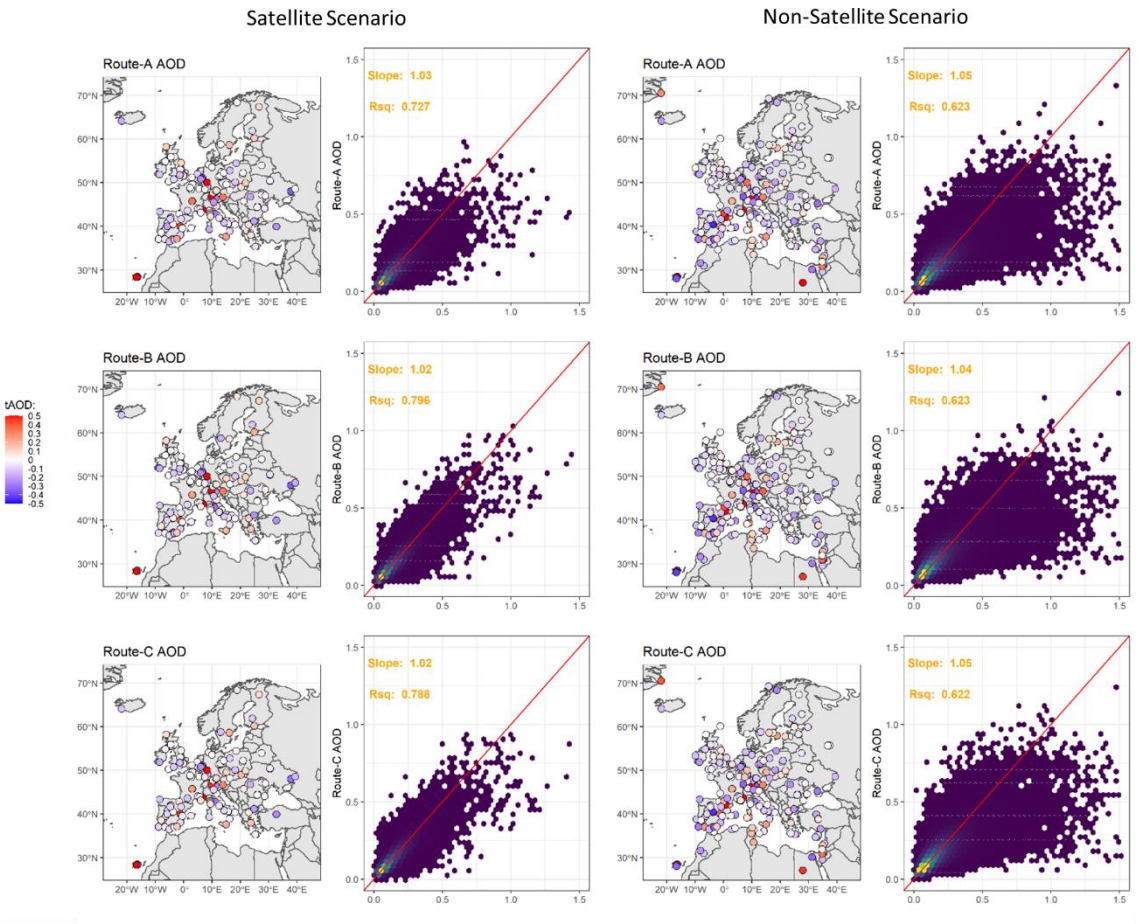

Figure s17. the spatial distribution maps of normalize mean bias (NMB) and scatter plots of tAOD estimation bias (Estimation minus AERONet data) for different data sources in 2003-2020.

And Route A have a bit more robust (lower variation coefficients) validation results among different locations and years (Temporal Robustness $_{\text{Route A VS B VS C}}$=0.232 VS 0.241 Vs 0.250; Spatial Robustness $_{\text{Route A VS B VS C}}$=0.105 VS 0.121 VS 0.128). The potential reason why adding satellite AOD as inputs will weaken the result robustness is that the Route B and C only improve the performance in Sat scenarios, and different locations or years generally have different missing rates, so it will bring more variance in the results. As for the correlation between ground-level PM2.5 and tAOD estimation from three different model routes, the Route A has slightly higher correlation than other two routes (R=0.418 VS 0.415 Vs 0.413).

For fAOD and cAOD (Fig s16 (b) and (c)), they have similar results as the tAOD product: three different models routes have similar performance but Route A is simpler (takes one less model than other routes). Thus, we will mainly discuss and compare the Route A with other aerosol products in the following section. In the subgroup analysis, Route B and C still mainly improve the performance in Sat scenarios for fAOD and cAOD (Figure 18 and 19), but the smaller fraction of Sat scenarios is still the main reason for small differences between three model routes.

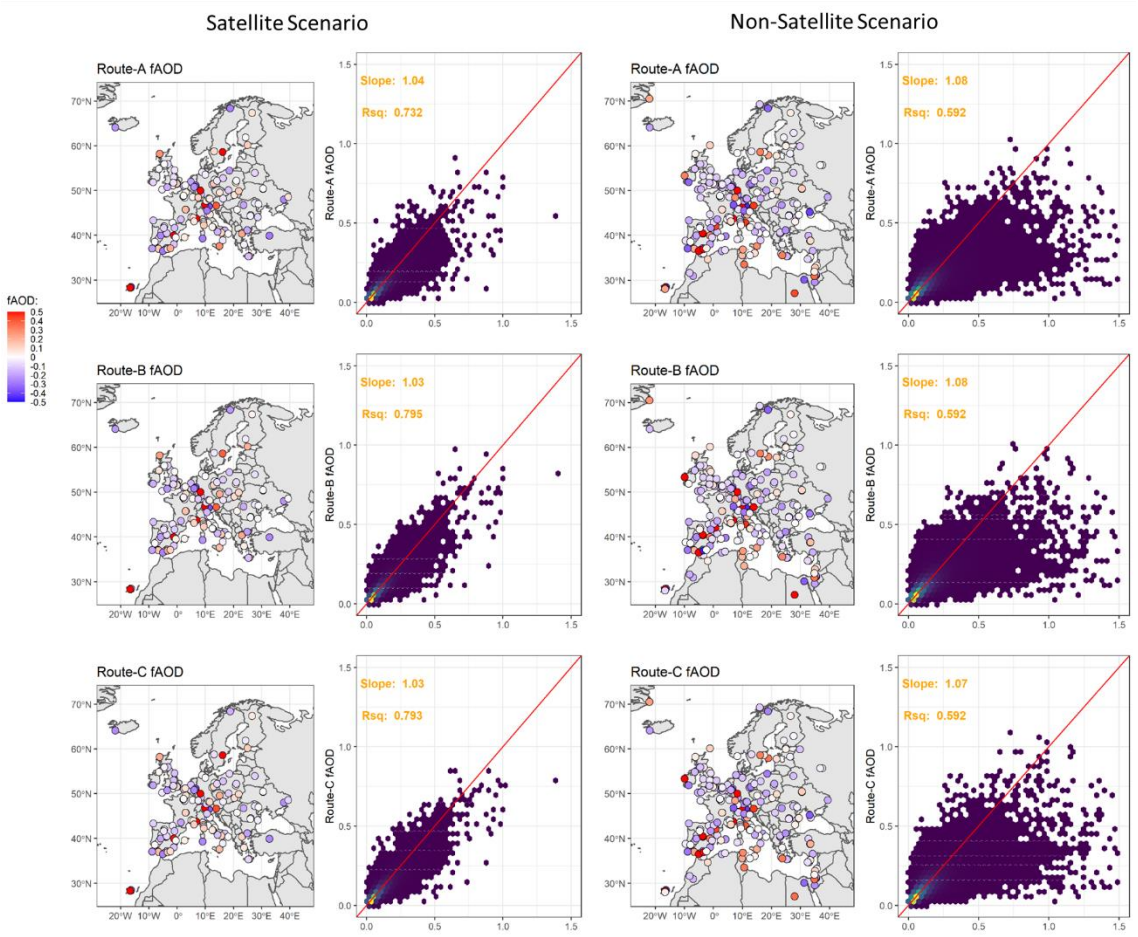

Figure s18. the spatial distribution maps of normalize mean bias (NMB) and scatter plots of fAOD estimation bias (Estimation minus AERONet data) for different data sources in 2003-2020.

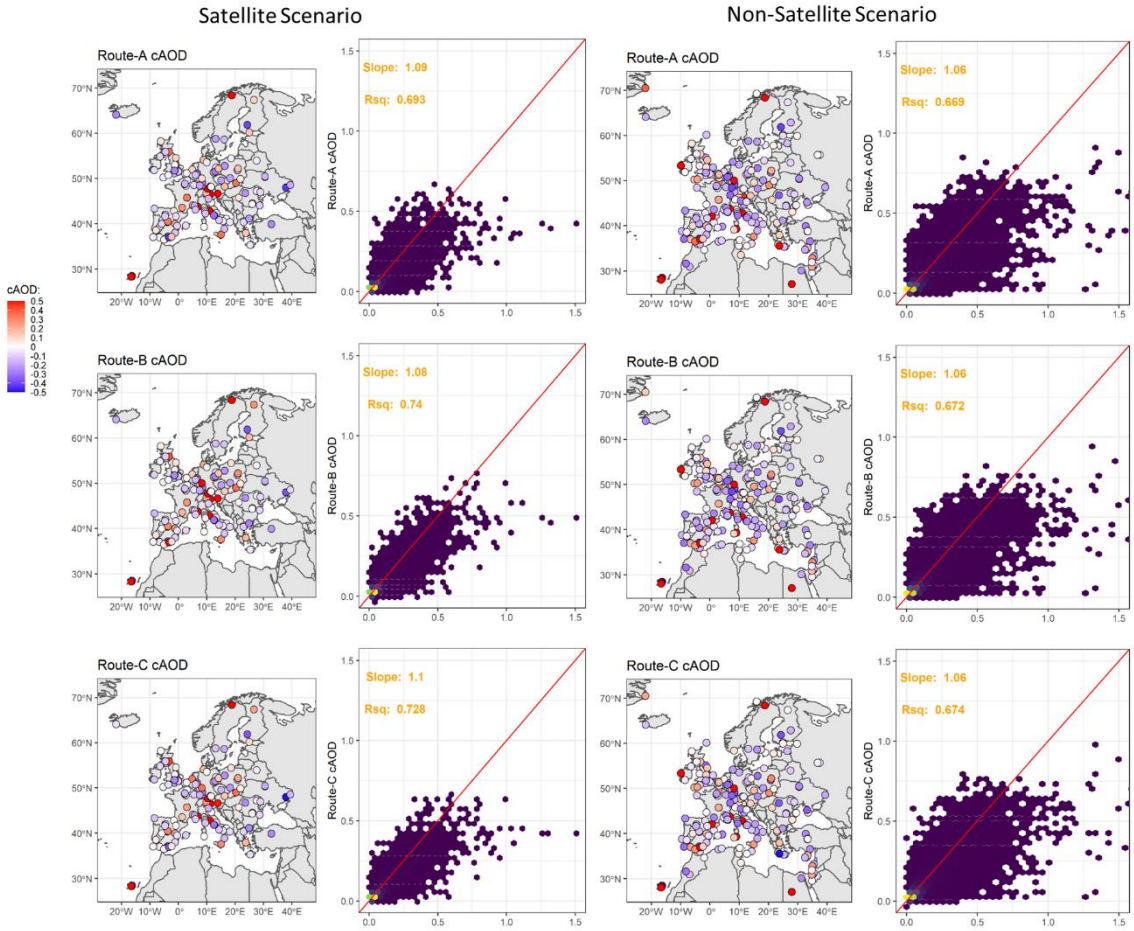

Figure s19. the spatial distribution maps of normalize mean bias (NMB) and scatter plots of cAOD estimation bias (Estimation minus AERONet data) for different data sources in 2003-2020.