Supplemental Materials

Global aerosol retrieval over land from Landsat imagery integrating Transformer and Google Earth Engine

Contents

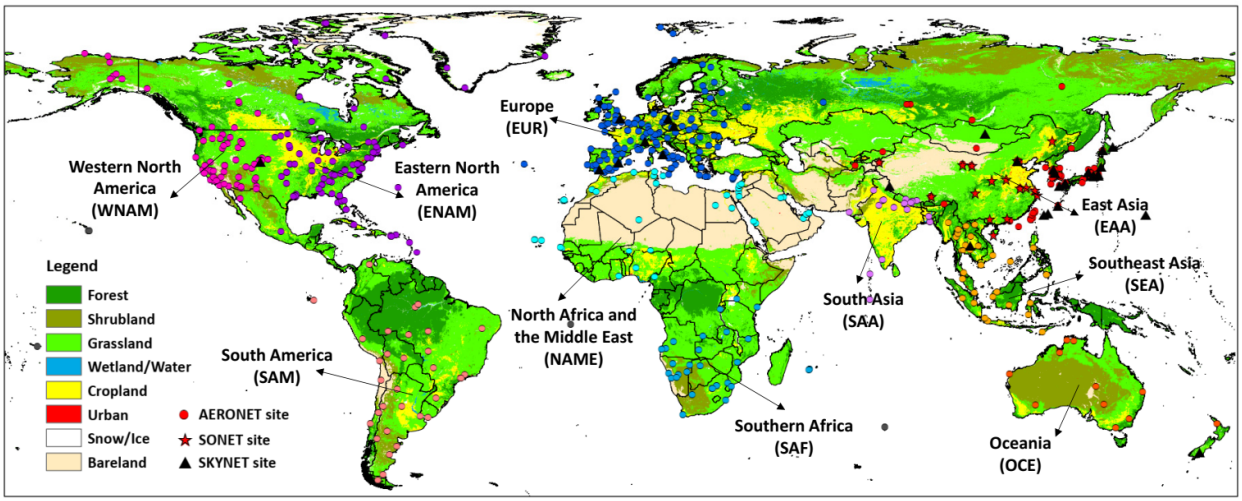

Figure S1. Spatial distributions of ground-based AERONET (dots), SONET (stars), and SKYNET (triangles) sites. The color of the symbol indicates which of the ten defined continental regions of interest the site belongs to. The background colors in the image represent surface types sourced from the MODIS Land Use Cover Types (MCD12C1) product (500 m) for 2022 (available at https://lpdaac.usgs.gov/products/mcd12c1v006/), utilizing the International Geosphere-Biosphere Programme classification scheme.

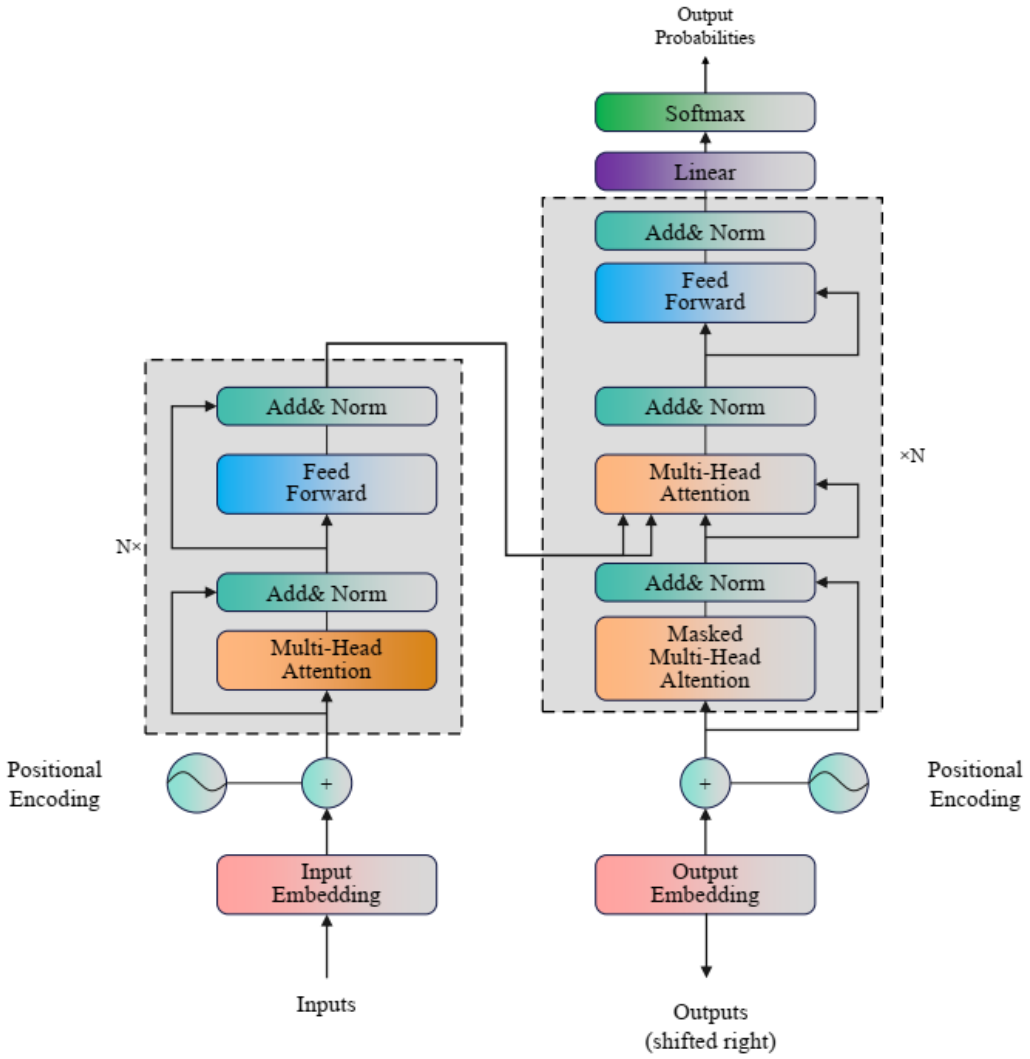

Figure S2. Schematic diagram of the original Transformer model structure.

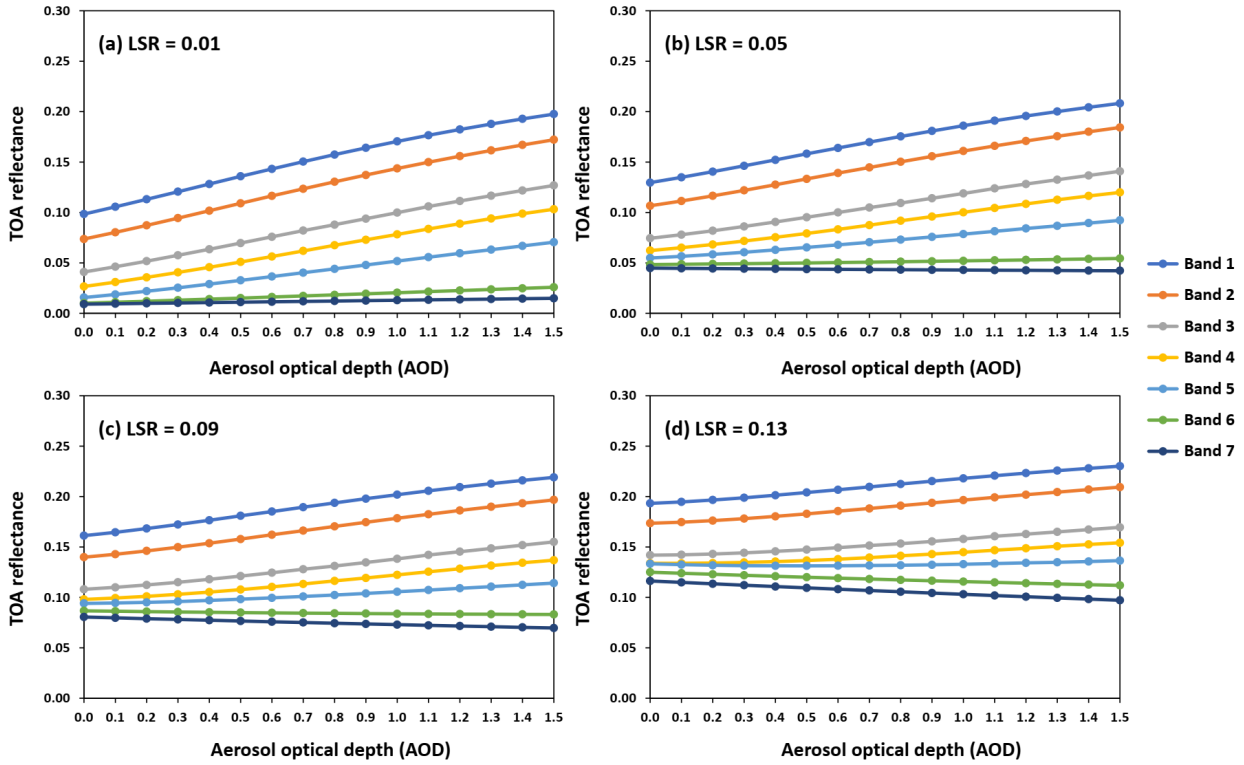

Figure S3. Simulated variations in top-of-the-atmosphere (TOA) reflectances of different Landsat channels with increasing aerosol optical depth (AOD) at a specific viewing geometry (θ_s = 30°, θ_v = 30°, and ϕ = 100°) under different land surface reflectance (LSR) conditions.

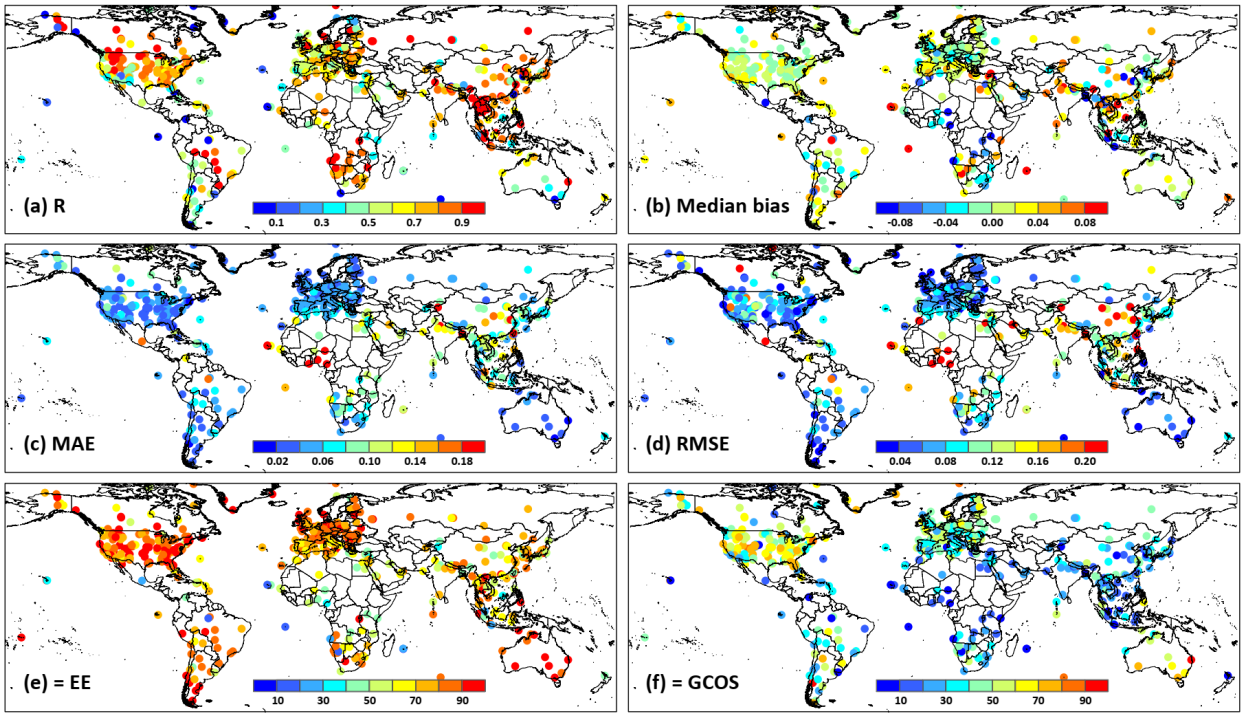

Figure S4. Spatial (station-based) cross-validated AOD retrievals from Landsat imagery compared to ground-based measurements at the individual-site scale from 2013 to 2022 using the AeroTrans-Landsat model: (a) correlation coefficient (R), (b) median bias, (c) mean absolute error (MAE), (d) root-mean-square error (RMSE), and the percentages of retrievals falling within the (e) expected error envelope (= EE) of the MODIS Deep Blue product and (f) the goal uncertainty defined by the Global Climate Observing System (GCOS).

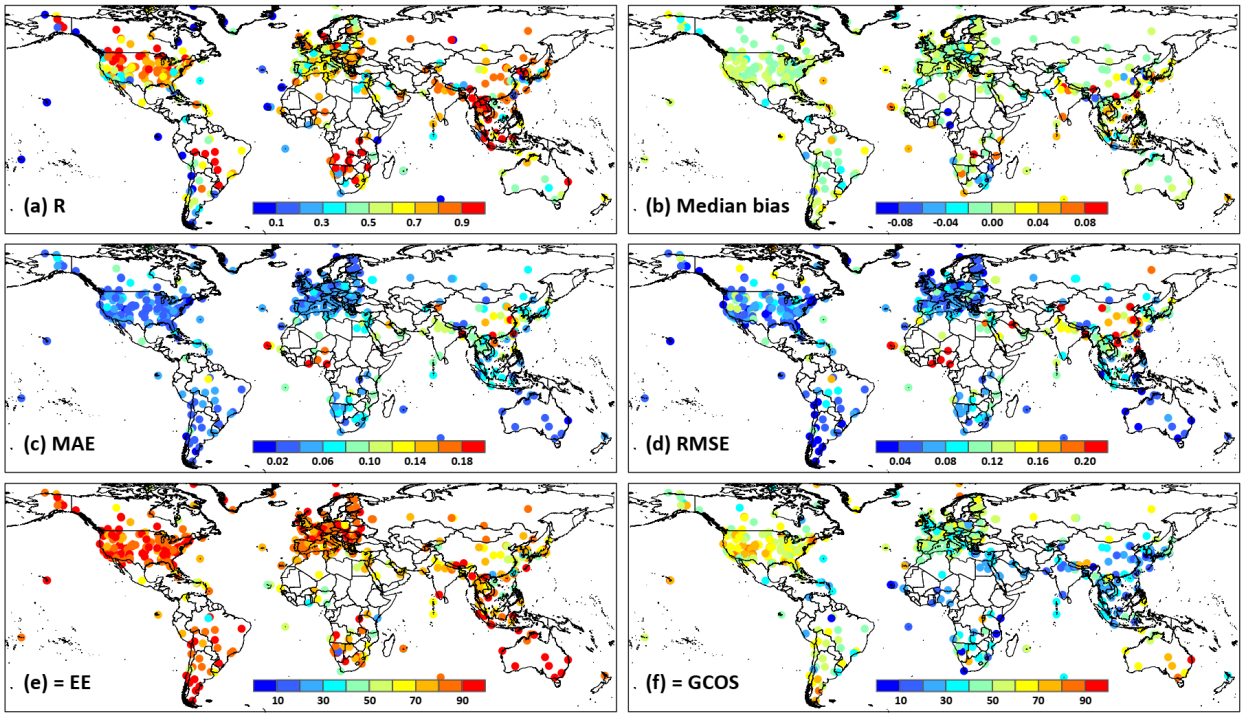

Figure S5. Temporal (month-based) cross-validated AOD retrievals from Landsat imagery compared to ground-based measurements at the individual-site scale from 2013 to 2022 using the AeroTrans-Landsat model: (a) correlation coefficient (R), (b) median bias, (c) mean absolute error (MAE), (d) root-mean-square error (RMSE), and the percentages of retrievals falling within the (e) expected error envelope (= EE) of the MODIS Deep Blue product and (f) the goal uncertainty defined by the Global Climate Observing System (GCOS).

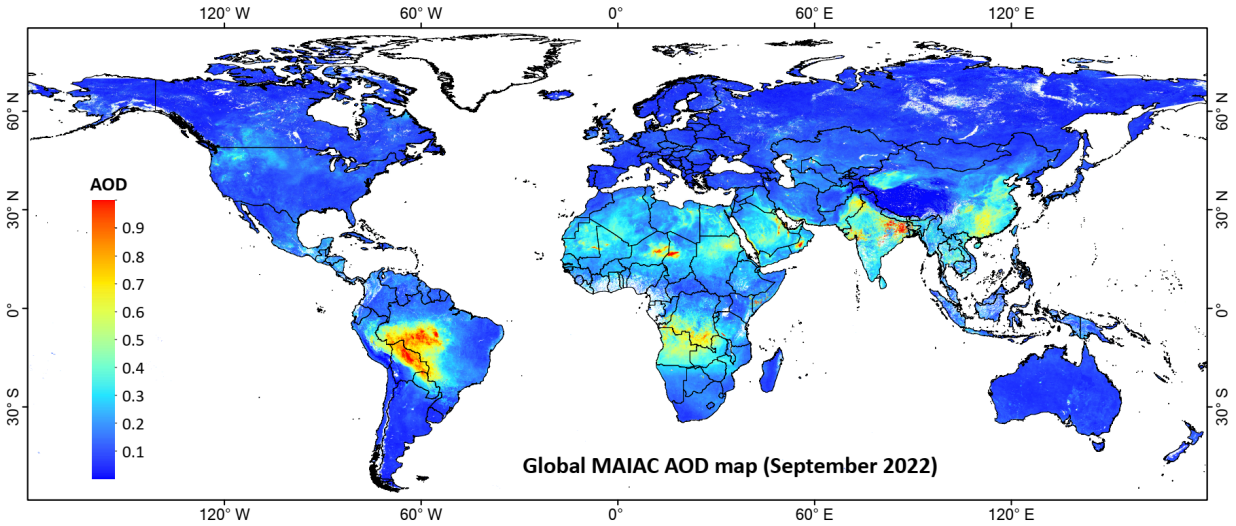

Figure S6. Same as Figure 7 but generated from the MODIS MAIAC AOD (550 nm) product at a 1-km resolution.

Table S1. Detailed information about the Landsat 8 and 9 Collection 2 Tier 1 TOA reflectance products in the Google Earth Engine data catalog.

Band	Wavelength	Center	Band description	Spatial resolution
B1*	0.433–0.453	0.443	Coastal aerosol	30 m
B2*	0.450–0.515	0.482	Blue	30 m
B3*	0.525–0.600	0.562	Green	30 m
B4*	0.630–0.680	0.655	Red	30 m
B5*	0.845–0.885	0.865	Near infrared	30 m
B6*	1.560–1.660	1.610	Shortwave infrared	30 m
B7*	2.100–2.300	2.200	Middle infrared	30 m
B8	0.500–0.680	0.590	Panchromatic	15 m
B9	1.360–1.390	1.375	Cirrus	15 m
B10	10.30–11.30	10.800	Thermal infrared 1	Resampled from 100 m to 30 m
B11	11.50–12.50	12.000	Thermal infrared 2	Resampled from 100 m to 30 m
QA_PIXEL	-	-	Collection 2 QA bitmask	30 m
QA_RADSAT	-	-	Radiometric saturation QA	30 m
SAA	-	-	Solar azimuth angle	30 m
SZA	-	-	Solar zenith angle	30 m
VAA	-	-	View azimuth angle	30 m
VZA	-	-	View zenith angle	30 m

* Indicates spectral channels adopted in the AOD retrieval using deep learning
Source form: https://landsat.gsfc.nasa.gov/satellites/landsat-9/landsat-9-instruments/landsat-9-spectral-specifications/

Table S2. Summary of global ground-based sites, matched images, and collocated data samples aligned with Landsat imagery for each year from 2013 to 2022 in this study.

Year	Number of ground-based sites	Number of matched images	Number of matched data samples
2013	250	1260	1391
2014	300	1846	2040
2015	330	2133	2453
2016	363	2353	2758
2017	364	2482	2866
2018	388	2475	2927
2019	389	2502	2931
2020	379	2467	2860
2021	364	2507	2994
2022	235	2255	2549
All	552	22,280	25,769

Table S3. Station-based cross-validated AOD retrievals from Landsat imagery against ground-based measurements in specific continental regions of interest and on a global scale from 2013 to 2022 using our developed model.

Region	R	Median Bias	MAE	RMSE	= EE (%)	= GCOS (%)
EAA	0.830	0.010	0.088	0.144	69.04	31.10
SAA	0.858	0.002	0.124	0.185	70.28	30.77
SEA	0.925	-0.001	0.095	0.127	68.75	26.25
EUR	0.621	0.013	0.051	0.073	78.79	41.74
ENAM	0.705	0.006	0.041	0.072	86.01	57.37
WNAM	0.748	0.006	0.050	0.097	81.90	52.36
SAM	0.785	0.015	0.052	0.079	81.49	41.16
NAME	0.720	-0.004	0.101	0.157	64.55	28.64
SAF	0.796	0.015	0.069	0.098	69.21	34.85
OCE	0.517	0.015	0.035	0.049	88.89	55.84
Global	0.845	0.009	0.061	0.104	77.63	43.42

EAA: Eastern Asia; SAA: South Asia; SEA: Southeast Asia; EUR: Europe; ENAM: Eastern North America; WNAM: Western North America; SAM: South America; NAME: North Africa and the Middle East; SAF: Southern Africa; OCE: Oceania.

Table S4. Month-based cross-validated AOD retrievals from Landsat imagery against ground-based measurements in specific continental regions of interest and on a global scale from 2013 to 2022 using our developed model.

Region	R	Median Bias	MAE	RMSE	= EE (%)	= GCOS (%)
EAA	0.843	0.004	0.084	0.138	71.62	35.67
SAA	0.895	0.004	0.108	0.160	76.57	35.31
SEA	0.941	0.011	0.081	0.115	77.92	32.36
EUR	0.643	0.005	0.047	0.071	83.02	48.00
ENAM	0.743	0.004	0.039	0.068	87.49	60.23
WNAM	0.829	0.001	0.041	0.081	87.04	61.46
SAM	0.875	0.000	0.040	0.061	87.69	57.08
NAME	0.773	0.002	0.089	0.143	70.40	32.68
SAF	0.894	0.005	0.049	0.073	82.25	47.68
OCE	0.627	0.003	0.029	0.045	93.53	64.98
Global	0.876	0.003	0.054	0.094	82.35	50.15

EAA: Eastern Asia; SAA: South Asia; SEA: Southeast Asia; EUR: Europe; ENAM: Eastern North America; WNAM: Western North America; SAM: South America; NAME: North Africa and the Middle East; SAF: Southern Africa; OCE: Oceania.

Table S5. Spatial performance of global AOD retrievals from Landsat imagery using our developed model, with data samples from each defined continental region used as testing data and data samples from other continental regions used as training data.

Region	R	Median Bias	MAE	RMSE	= EE (%)	= GCOS (%)
EAA	0.753	-0.005	0.109	0.188	64.64	30.40
SAA	0.862	0.008	0.128	0.183	68.53	32.05
SEA	0.907	-0.001	0.098	0.141	69.44	30.42
EUR	0.562	0.006	0.051	0.077	79.13	44.58
ENAM	0.634	-0.001	0.047	0.080	82.38	55.81
WNAM	0.726	0.011	0.054	0.099	75.58	47.68
SAM	0.779	-0.003	0.051	0.083	81.80	54.91
NAME	0.511	-0.048	0.136	0.215	52.20	22.97
SAF	0.638	-0.002	0.083	0.129	62.37	36.99
OCE	0.613	-0.001	0.030	0.046	92.12	65.12

EAA: Eastern Asia; SAA: South Asia; SEA: Southeast Asia; EUR: Europe; ENAM: Eastern North America; WNAM: Western North America; SAM: South America; NAME: North Africa and the Middle East; SAF: Southern Africa; OCE: Oceania.

Table S6. Temporal performance of global AOD retrievals from Landsat imagery using our developed model, with data samples from each year used as testing data and data samples from remaining years used as training data.

Year	R	Median Bias	MAE	RMSE	= EE (%)	= GCOS (%)
2013	0.903	0.001	0.055	0.091	83.54	50.32
2014	0.899	-0.001	0.051	0.092	84.66	54.80
2015	0.891	0.003	0.053	0.093	83.86	52.06
2016	0.913	0.007	0.051	0.086	84.12	53.84
2017	0.885	0.007	0.051	0.089	84.30	53.11
2018	0.898	0.002	0.052	0.090	83.67	51.45
2019	0.882	0.005	0.052	0.086	83.45	49.37
2020	0.884	0.000	0.055	0.099	82.20	50.84
2021	0.876	0.004	0.056	0.098	82.40	51.70
2022	0.804	0.009	0.045	0.076	84.50	53.51

Table S7. Comparison of 10-fold cross-validation (10-CV) results between various machine- and deep-learning models for the AOD retrieval from Landsat imagery.

Category	Model	10-CV by sample			10-CV by station			10-CV by month		
		R	RMSE	MAE	R	RMSE	MAE	R	RMSE	MAE
Tree-based machine learning	Extra Trees	0.771	0.133	0.074	0.711	0.145	0.084	0.726	0.141	0.079
	Random Forest	0.826	0.114	0.063	0.750	0.133	0.077	0.788	0.124	0.069
	CatBoost	0.884	0.091	0.052	0.791	0.120	0.072	0.844	0.105	0.058
	XGBoost	0.884	0.091	0.051	0.782	0.122	0.073	0.845	0.105	0.058
	LightGBM	0.894	0.087	0.049	0.789	0.120	0.072	0.851	0.102	0.058
Deep learning	DBN	0.873	0.095	0.058	0.806	0.116	0.070	0.831	0.108	0.065
	LSTM	0.887	0.090	0.054	0.832	0.108	0.065	0.857	0.101	0.059
	MLP	0.893	0.088	0.050	0.832	0.108	0.064	0.859	0.100	0.057
	ResNet	0.893	0.088	0.051	0.835	0.107	0.064	0.866	0.098	0.056
	Transformer	**0.905**	**0.083**	**0.048**	**0.845**	**0.104**	**0.061**	**0.876**	**0.094**	**0.054**