

Text S1. Input features selection

The initial input features include MSI bands and various spectral indices. The spectral indices include band ratio, NDWI and chromaticity angle (alpha). Alpha is a physical quantity related to the composition and inherent optical properties of the optical deep water, with the formula can be found in [1]. The initial features were analyzed for feature importance using RFR and XGB algorithms to filter out 12 features. Furthermore, the evolution of lakes is inevitably associated with changes in water volume and area, especially in semi-arid regions, which can affect the water physicochemical properties (e.g., salinity) [2,3], so extract lake area as input feature. Eventually, the 13 input features were determined as $R_{rs}(443)$, $R_{rs}(497)$, $R_{rs}(560)$, $R_{rs}(664)$, $R_{rs}(704)$, $R_{rs}(740)$, $R_{rs}(842)$, $B4/(B4+B3)$, $B4/(B2+B3)$, $B4/B2$, NDWI, alpha, and area, the feature importance score shown in Figure S1.

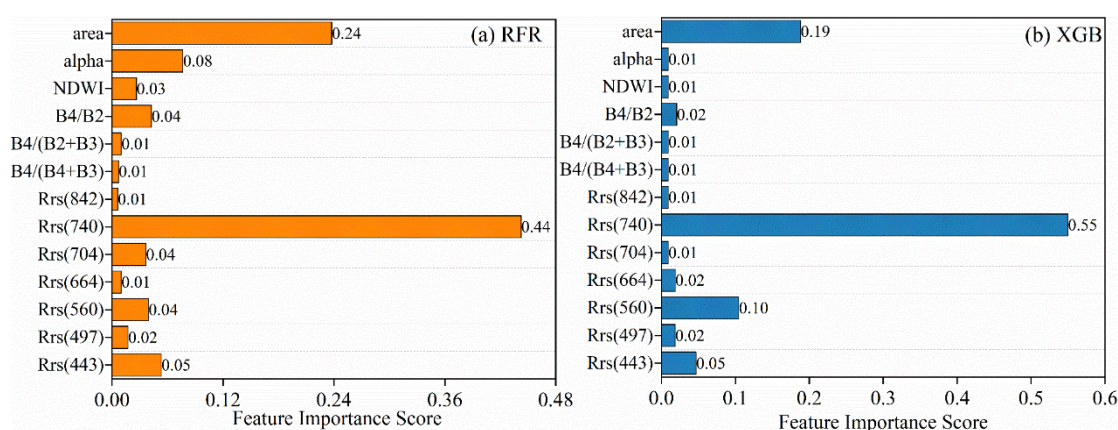


Figure S1. Feature importance score of (a) RFR, and (b) XGB.

Table S1. Software version and parameter settings for each atmospheric correction processor.

Processor	Software	Parameter setting
ACOLITE	ACOLITE v20221114.0	Atmospheric correction algorithm: Dark Spectrum Fitting
C2RCC	SNAP v8.0- C2RCC.msi v1.0	Valid-pixel expression: $B8 > 0 \ \&\& \ B8 < 0.2$ Salinity: in-situ measured mean salinity Temperature: ERA5-Land reanalysis data
POLYMER	POLYMER v4.17	NASA ancillary data other parameter default
MUMM	Seadas v7.5.3	aer_wave_short = 783 nm aer_wave_long = 865 nm mumm_alpha = 1.72 ancillary data download from NASA: GMAO MERRA2
iCOR	iCOR v3.0	Water msk detection threshold: 0.2 Cloud mask low threshold: 0.29 Cloud mask average threshold: 0.19 Cloud mask cirrus threshold: 0.13
Sen2Cor	Sen2Cor v2.11.0	Default

Table S2. Four machine learning model structures and hyperparameter settings.

Model name	Structure	Model name	Structure
XGB	Trees: 500 Learning rate: 0.05 Maximum tree depth: 7 Subsample rate: 0.9 Regularization: 0.01	RFR	Trees: 300 Maximum depth: 8 Maximum features at node splitting: 7
CNN	Convolutional layer: five one dimensional convolutional layers with kernels 16, 32, 32, 64, and 64 Five activation layers: activation function ReLU One flattening layer: 64 neurons One fully connected layer: 64 neurons Learning rate: 0.01 Optimizer: Adam Penalty factors: 0.001 Maximum training epochs: 2000 Patience (early stop mechanism): 100	DNN	Hidden layer number: 7 Each layer neurons: 64 Learning rate: adaptive Activation function: ReLU Optimizer: Adam Penalty factors: 0.001 Maximum training epochs: 2000 Patience: 100

Table S3. Accuracy statistics of in-situ measurements $R_{rs}(\lambda)$ and six AC processors derived MSI $R_{rs}(\lambda)$.

Processor	Band	R ²	RMSE	MAPE(%)	BIAS	VP(%)	Processor	Band	R ²	RMSE	MAPE(%)	BIAS	VP(%)
ACOLITE	443	0.38	0.0037	61.32	−0.0030	100	POLY	443	0.32	0.0026	44.36	−0.0010	100
	497	0.79	0.0027	26.26	−0.0015	100		497	0.63	0.0034	39.76	−0.0019	100
	560	0.78	0.0032	28.39	0.0001	100		560	0.61	0.0056	38.95	−0.0033	100
	664	0.83	0.0022	40.25	0.0010	100		664	0.59	0.0040	56.80	−0.0023	86.66
	704	0.68	0.0031	69.63	0.0020	93.33		704	0.41	0.0039	50.03	−0.0022	97.78
	740	0.59	0.0038	95.86	0.0026	93.33		740	0.04	0.0020	73.88	−0.0006	97.78
DSF	783	0.02	0.0048	145.94	0.0033	88.89	MER	783	0.03	0.0021	84.32	−0.0002	100
	842	0.01	0.0040	144.89	0.0028	95.56		842	0.05	0.0018	100.56	−0.0008	73.33
	865	0.01	0.0039	202.93	0.0027	93.33		865	0.02	0.0014	179.56	−0.0005	84.44
	443	0.28	0.0070	133.07	−0.0061	31.11	Sen2Cor	443	0.22	0.0030	79.56	0.0014	100
	497	0.46	0.0064	63.94	0.0036	100		497	0.60	0.0037	54.53	0.0021	100
	560	0.63	0.0186	135.41	0.0177	93.33		560	0.69	0.0054	46.51	0.0039	100
iCOR	664	0.67	0.0198	303.63	0.0191	97.78		664	0.72	0.0049	90.33	0.0041	100
	704	0.45	0.0255	451.35	0.0249	22.22		704	0.41	0.0068	147.73	0.0055	95.56
	740	0.01	0.0262	1479.72	0.0252	0		740	0.01	0.0070	251.69	0.0054	88.89
	783	0.01	0.0288	1707.07	0.0278	0		783	0.01	0.0075	356.68	0.0059	88.89
	842	0.01	0.0328	2080.93	0.0323	0		842	0.01	0.0071	433.15	0.0056	97.78
	865	0.02	0.0300	2570.61	0.0297	0		865	0.02	0.0063	565.87	0.0051	88.89
C2RCC	443	0.33	0.0023	56.14	−0.0003	100	MUMM	443	0.13	0.0061	129.74	0.0029	86.67
	497	0.68	0.0029	38.82	−0.0011	100		497	0.49	0.0056	91.76	0.0011	84.44
	560	0.77	0.0036	28.08	−0.0005	100		560	0.74	0.0076	62.03	0.0061	95.56
	664	0.81	0.0024	36.77	−0.0010	91.11		664	0.69	0.0037	76.75	0.0005	84.44
	704	0.80	0.0022	40.90	−0.0010	88.89		704	0.86	0.0023	54.02	0.0016	97.78
	740	0.75	0.0008	43.98	−0.0003	93.33							
	783	0.72	0.0009	46.17	−0.0003	88.89							
	865	0.64	0.0008	55.13	−0.0005	88.89							

Table S4 Results of 30% independent dataset testing and five-fold cross validation of four model.

	XGB		CNN		DNN		RFR	
	Testing	Five-fold CV	Testing	Five-fold CV	Testing	Five-fold CV	Testing	Five-fold CV
R2	0.98	0.93	0.98	0.95	0.98	0.96	0.89	0.88
RMSE	1.03	1.14	0.37	0.83	0.73	0.62	1.68	1.50
MAE	0.53	0.57	0.23	0.45	0.51	0.38	0.89	0.81
bias	1.04	1.01	0.95	1.02	0.97	1.04	1.12	1.12
MAPE	9.95	11.08	9.14	14.32	9.90	15.25	76.98	32.36

References

1. Wang, S.; Jiang, X.; Spyrakos, E.; Li, J.; Mcglinchey, C.; Constantinescu, A.M.; Tyler, A.N. Water Color from Sentinel-2 MSI Data for Monitoring Large Rivers: Yangtze and Danube. *Geo-Spat. Inf. Sci.* **2023**.
2. Wurtsbaugh, W.A.; Miller, C.; Null, S.E.; DeRose, R.J.; Wilcock, P.; Hahnenberger, M.; Howe, F.; Moore, J. Decline of the World's Saline Lakes. *Nat. Geosci.* **2017**, *10*, 816–821.
3. Liu, C.; Zhu, L.; Wang, J.; Ju, J.; Ma, Q.; Kou, Q. The Decrease of Salinity in Lakes on the Tibetan Plateau between 2000 and 2019 Based on Remote Sensing Model Inversions. *Int. J. Digit. Earth* **2023**, *16*, 2644–2659.

