A lake salinity dataset produced via microwave and optical imageries

Mingming Deng[1,2], Ronghua Ma[1,3], Lixin Wang[4], Minqi Hu[1], Kun Xue[1], and Junfeng Xiong[1]

[1]Key Laboratory of Lake and Watershed Science for Water Security, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China.
[2]University of Chinese Academy of Sciences, Beijing 100049, China.
[3]University of Chinese Academy of Sciences, Nanjing 211135, China.
[4]School of Ecology and Environment, Inner Mongolia University, Hohhot 010021, China.

Correspondence to: Ronghua Ma (rhma@niglas.ac.cn)

Abstract. Lake salinity is an important parameter to characterize physical and biogeochemical processes and a fundamental indicator to evaluate lake water quality. However, its estimation in inland waters has been challenging because passive microwave salinity satellites lack sufficient spatial resolution, and optical satellites cannot directly measure it. To address it, we constructed a framework for estimating lake salinity by combining Synthetic Aperture Radar (SAR) and Multi-Spectral Instrument (MSI) data. It can be summarized in step 1: construct a salinity mechanism model based on SAR data using the Elfouhaily spectrum, dielectric constant, and small perturbation method (SPM) models; step 2: develop four machine learning (ML) salinity algorithms using quasi-synchronous salinity and MSI with SAR imagery; and step 3: build an ensemble model to estimate salinity by coupling the mechanism and ML models via a generalized additive model. The proposed integrated algorithm (N = 84, RMSE = 0.60 ppt, and MAPE = 2.3%) outperformed single-satellite microwave mechanistic or ML models across all eleven lakes in the Inner Mongolia Xinjiang Lake zone. On this basis, we reconstructed the lake salinity dataset for 2016–2024 and conducted independent validation (N = 65, R^2 = 0.97, and RMSE = 0.89 ppt) and pixel-level histogram validation confirmed dataset quality, with no significant systematic bias across lake types. The reconstruction revealed a spatial pattern of smooth transition from the nearshore to the center and trends with significant increases in Lake Daihai and Lake Dalinor. The dataset and its development framework will facilitate exploration of salinity status and trends in inland lakes, providing scientific evidence and methodological support for salinization prevention and global lake salinity budget research. The dataset (10 m spatial resolution, TIF format) is publicly available via Zenodo (https://doi.org/10.5281/zenodo.18371515, Deng et al., 2026) and includes annual/seasonal salinity rasters and statistical files.

1 Introduction

Lakes are important reservoirs of surface water resources and can serve as indicators and regulators of the global water cycle and regional climate (Williamson et al., 2009; Gleeson et al., 2020). Salinity (total dissolved salt concentration) as a critical

parameter characterizing the physicochemical properties of lake water controls biological, physical, and chemical processes within lake ecosystems (Zhao and Temimi, 2016; Wurtsbaugh et al., 2017), including microbial community structure, species abundance, vertical mixing of water masses, water resource utilization, and nitrogen transformation (Liu et al., 2023a; Florencia Gutierrez et al., 2018; Ladwig et al., 2023; Kaushal et al., 2021; Jiang et al., 2023). Recently, climate change in the lake hydrological system caused water salinization, weakening the stability of lake ecosystems, particularly in arid and semi-arid regions (Jeppesen et al., 2015; Wurtsbaugh et al., 2017). Therefore, frequent and effective monitoring of water salinity is necessary for salinization prevention and sustainable development. However, the low density of salinity field measurements cannot reveal its spatial patterns and long-term trends, limiting the perceptions of salinity for lake communities. Existing inland lake salinity datasets (e.g., Xu et al., 2024, Tibetan Plateau: RMSE = 12.51 g/L, >1 km resolution) lack high spatial detail and arid-region specificity, whereas our dataset provides 10 m resolution and targets IMXL's multi-type lakes (freshwater to oligosaline).

Satellite remote sensing enables the frequent detection of substance concentrations in lake waters by utilizing different electromagnetic wavebands to address the issues of data sparsity and discontinuity. The response of microwave sensors to the dielectric properties of water allows it to be uniquely advantageous for salinity observation (Le Vine et al., 2022). Missions designed to measure water salinity from space have made progress in the oceanographic field, with satellite sensors including SMOS, SMAP, and Aquarius. However, their spatial resolution of 40–150 km is too coarse for small-scale inland lakes (Reul et al., 2020). The Sentinel-1 Synthetic Aperture Radar (SAR), operating at 5.405 GHz with 10 m spatial resolution, has an acquisition orbit for both ascending and descending and a revisit interval of 6 days (dual-satellite constellation), which supports the space observations of lake salinity (Torres et al., 2012). The backscatter coefficient of water surfaces measured by SAR is mainly determined by radar-related parameters (frequency and incidence angle, etc.), water surface geometry (roughness), and water physical properties (dielectric constant and temperature, etc.) (Peake, 1959; Reul et al., 2020). Changes in salinity directly affect the dielectric constant of water, which in turn influences the Fresnel reflection coefficient at the water surface and ultimately causes variations in the lake water backscatter coefficient. Clearly, when radar parameters are fixed, the contribution of the water surface geometry must be separated to solve for salinity from the SAR backscatter coefficient (Hwang et al., 2011; Meissner et al., 2014; Ma et al., 2021; Taillade et al., 2023). The microwave backscatter coefficient model combined with the wave spectrum model was commonly used to quantify the contribution of surface roughness because it effectively described the energy information of wind waves (Xie et al., 2019). The dielectric constant model was essential for inverting salinity after removing the contributions from roughness. However, the current predominant dielectric constant model was developed by seawater experiments with simple ionic compositions (Klein and Swift, 1977), whereas inland lakes receive substantial exogenous sources that result in complex ionic compositions (Zou et al., 2024). Salinity estimation at the lake inlet region or in optically complex waters (such as high variability in suspended minerals and phytoplankton) may introduce uncertainty using the present dielectric constant models

(Chen and Hu, 2017; Gonzalez-Gambau et al., 2022; Xue et al., 2025). Hence, this study attempts to supplement mechanism-based microwave salinity estimation with optical data.

65   Optical data can effectively retrieve optically active constituents (OACs) in lake water, and it has been demonstrated that the colored dissolved organic matter (CDOM) absorption coefficient [$a_g(\lambda)$, m^{-1}] and Secchi Disk Depth (SDD, m) can serve as salinity tracers for inland lakes, which confirmed the feasibility of optical data in salinity detection (Bai et al., 2013; Liu et al., 2023b). But the tracer method was constrained by the accuracy of indirect parameters and prone to errors. Furthermore, their correlations with salinity vary across different lakes and seasons to limit model transferability (Liu et al., 2014; Chen

70   and Hu, 2017). Machine learning (ML) algorithms could handle the nonlinear relationship between salinity and remote sensing reflectance [$R_{rs}(\lambda)$, sr^{-1}] to avoid errors caused by the traces. It is becoming an innovative approach for retrieving non-OACs (e.g., salinity) from optical satellite imagery (Deng et al., 2024; Guo et al., 2023; Liu et al., 2024). The Sentinel-2 Multi-Spectral Instrument (Sentinel-2 MSI) with high spatial resolution (10–60 m) and temporal resolution (5 days) (Drusch et al., 2012), enables the detection of water salinity. However, relying on optical data for salinity estimation is insufficient at

75   the mechanism level, while microwave data compensates for this deficiency as it has a clear physical mechanism. Despite its potential, the combination of microwave and optical data for regional and long-term salinity monitoring remains underexplored, restricting the advancement of space observation missions for inland lake salinity.

This study aims to: (1) develop a microwave-optical integrated framework for high-precision salinity estimation; (2) produce a 10 m resolution Inner Mongolia Xinjiang Lake zone (IMXL) lake salinity dataset (2016–2024); and (3) validate dataset

80   quality and analyze salinity spatiotemporal trends.

2 Data and Methods

2.1 Dataset coverage

The IMXL spans 70° E–120° E longitude and 30° N–50° N latitude, covering 11 arid or semi-arid inland lakes (Ma et al., 2011). The topographic patterns of alternating mountains and basins direct surface water and groundwater into the

85   depressions, forming numerous terminal lakes. The salinity of lakes in this region spans multiple magnitudes (Table 1). Referencing the classification criteria for lake salinity (Hammer, 1986; Zheng et al., 2002), the lakes were categorized as freshwater (< 1 g/L), brackish (1–3 g/L), and oligosaline (3–35 g/L) types. The salinity dataset to be produced covers the time span from 2016 to 2024 and the geographical coverage of 11 typical lakes within IMXL (Figure 1). Notably, some years are missing data for Lake Hulun and Lake Juyan due to insufficient SAR imagery or multi-source data matching pairs.

90

Table 1 Basic parameters and hydrological connectivity of the sample lakes, not all parameters were measured.

Lake name	Coordinates	Catchment area (km²)	Hydrological connectivity	Date	Samples number	Dominant ions	Salinity (ppt) Mean ± STD	Salinity (ppt) Min - Max	SDD (m) Mean ± STD	SDD (m) Min - Max
Hulun	(117°30′ E, 49°00′ N)	38,683	Open	2019/07 2020/09	35	HCO_3^-, Na^+	0.78 ± 0.08	0.54 – 0.86	0.29 ± 0.02	0.26 – 0.33
Dalinor	(116°39′ E, 43°20′ N)	4,675	Terminal	2018/09 2019/08	32	Cl^-, HCO_3^-, Na^+, K^+	6.43 ± 0.18	6.15 – 6.59	0.48 ± 0.06	0.36 – 0.54
Chagannaoer	(115°01′ E, 43°27′ N)	14,000	Terminal	2018/09 2020/09	15	Na^+, Cl^-	0.86 ± 0.03	0.83 – 0.92	/	/
Daihai	(112°41′ E, 40°35′ N)	2,341	Terminal	2017/09 2018/09 2020/07 2020/10 2024/04	60	Cl^-, HCO_3^-, Na^+, Mg^{2+}	13.52 ± 2.20	10.67 – 16.81	2.37 ± 1.1	0.63 – 4.80
Hongjiannao	(109°53′ E, 39°06′ N)	1,446	Terminal	2020/08	30	Na^+, Ca^+, HCO_3^-, SO_4^{2-}	5.94 ± 0.13	5.82 – 7.00	1.67 ± 0.29	0.70 – 2.20
Nanhaizi	(110°01′ E, 40°33′ N)	3.17	Terminal	2017/08	5	/	1.39 ± 0.01	1.39 – 1.41	0.27 ± 0.02	0.24 – 0.29
Ulansuhai	(108°45′ E, 40°50′ N)	27,515	Open	2024/10 2020/09 2017/07	29	Na^+, Mg^{2+}, Cl^-, SO_4^{2-}	1.81 ± 0.53	0.86 – 3.27	0.88 ± 0.34	0.24 – 1.30
Juyan	(101°15′ E, 42°18′ N)	142,900	Terminal	2020/09	11	Cl^-, SO_4^{2-}, Mg^{2+}, Na^{2+}	4.61 ± 0.11	4.53 – 4.93	/	/
Ulungur	(87°18′ E, 47°11′ N)	37,882	Terminal	2018/07	14	HCO_3^-, Na^+	1.15 ± 0.34	1.16 – 2.15	2.43 ± 0.11	2.21 – 2.61
Bosten	(87°05′ E, 41°55′ N)	79,204	Open	2018/04	18	Cl^-, Na^+	0.53 ± 0.02	0.50 – 0.56	2.42 ± 0.35	1.80 – 2.95
Sayram	(81°13′ E, 44°35′ N)	1,408	Terminal	2018/07	8	Mg^{2+}, SO_4^{2-}	1.15 ± 0.05	1.07 – 1.20	/	/

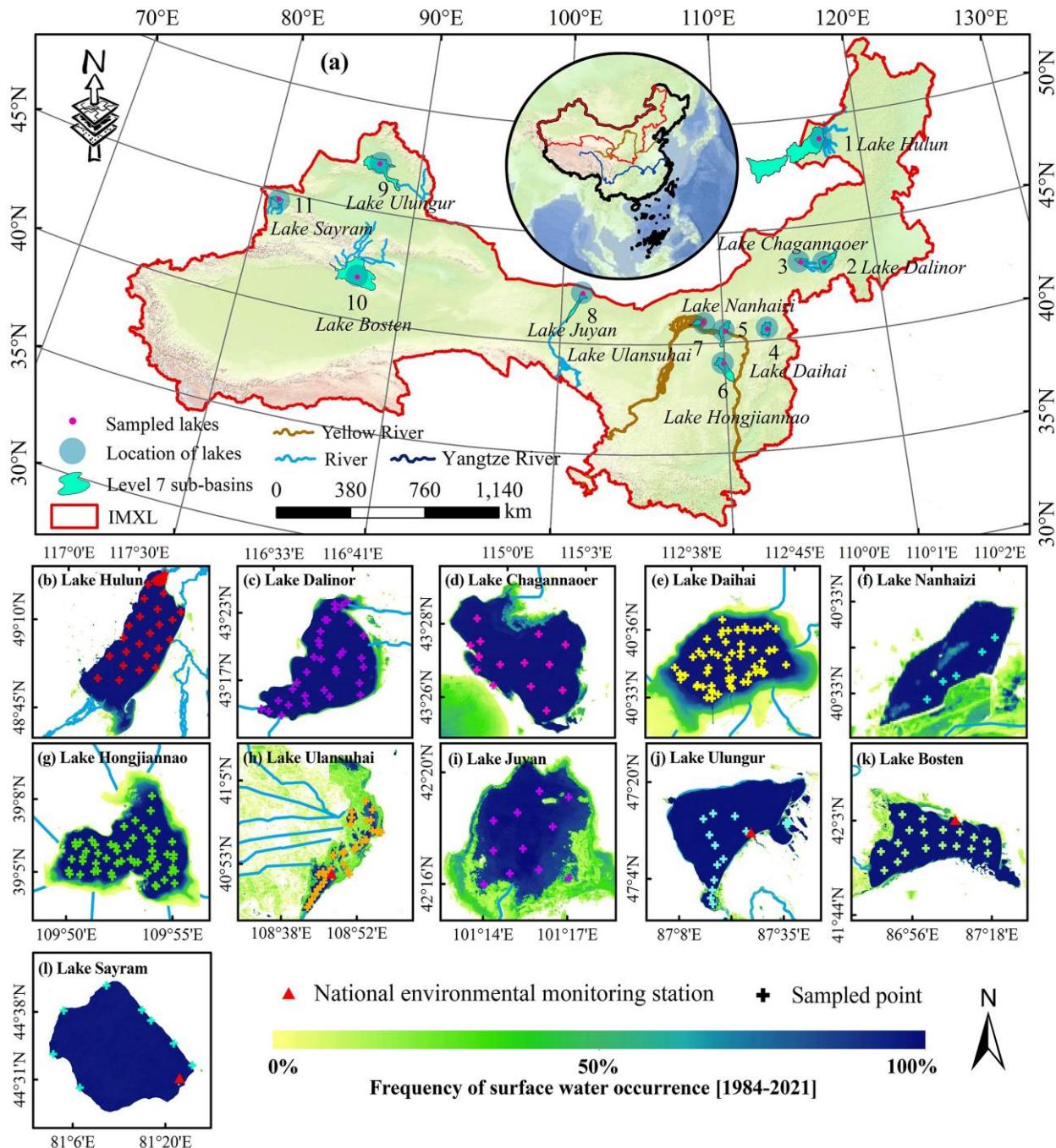



Figure 1: (a) Spatial distribution of sampled lakes covered by the dataset, inlet rivers, and sub-basins. **(b)-(l)** Eleven study lakes in the IMXL from east to west and its surface water occurrence frequency (Pekel et al., 2016), including Lake Hulun, Lake Dalinor, Lake Chagannaoer, Lake Daihai, Lake Nanhaizi, Lake Hongjiannao, Lake Ulansuhai, Lake Juyan, Lake Ulungur, Lake Bosten, and Lake Sayram. Cross symbols in each lake denote the field salinity sampling points. The frequency of surface water occurrence refers to the normalized count of pixels that were identified as water in Landsat imagery from 1984 to 2021.

95

2.2 Data products

Data products for each lake contain both raw data and derived data. The raw data includes Sentinel-1 SAR backscatter data and incidence angle, Sentinel-2 MSI remote sensing reflectance, and Landsat-8 Thermal Infrared Sensor (TIRS) temperature data. Derived data comprise daily, quarterly, yearly and all-season average salinity rasters, along with their mean and standard deviation statistical files (Table 2). All raw or derived raster data is stored in TIF format based on the WGS1984_UTM_Zone projection with a spatial resolution of 10 m. The statistical documents are compiled in Excel format.

Table 2 Attribute names and descriptions of derived data.

Attribute	Description
LakeName	Customized simplified names for each lake in this study.
Daily salinity rasters	Produced individual salinity data by the proposed framework, naming convention: [LakeName]_[YYYYMMDD]_[ProductType]_[Resolution].tif
Quarterly salinity rasters	Calculated season-average salinity data, naming convention: [LakeName]_[YYYY]_[Season]_[ProductType]_[Resolution].tif
Yearly salinity rasters	Calculated annual-average salinity data, naming convention: [LakeName]_[YYYY]_[ProductType]_[Resolution].tif
All-season salinity rasters	Calculated all-season average salinity data, naming convention: [LakeName]_[Season]_[ProductType]_[Resolution].tif
Statistical files	Compiled mean value and standard deviation of salinity images for each lake, name: [IMXLSAL]_[MeanSTD]_[2016–2024].xlsx

2.3 Metadata

2.3.1 Sensors parameters

Sentinel-1 SAR Level-1 GRD data operates at a center wavelength in the C-band, acquired in Interferometric Wide Swath (IW) mode with a pixel spacing of 10 m. It possesses both VV and VH polarization and simultaneously measures the incidence angle. Sentinel-2 MSI Level-1C images provide 13 spectral bands covering visible optical, near-infrared, and shortwave infrared, with red, green, blue, and near-infrared bands having 10 m spatial resolution, while other bands offer 20 m or 60 m resolution. Landsat-8 TIRS is a dual-band push-broom radiometer containing band 10 (10.9 μm) and band 11 (12.0 μm) with a spatial resolution of 100 m and a radiometric resolution of 12-bit.

2.3.2 Processing procedure

The Copernicus Data Space Ecosystem (CDSE) offers free Sentinel-1 SAR data that can be loaded directly at the Google Earth Engine (GEE) platform. These products were processed with thermal noise removal, radiometric calibration, terrain correction, and debelization. Additionally, we employed a 3×3 window Lee filter to restrain coherent spot noise in the radar imagery. Lastly, the backscattering coefficients of VV and VH as well as the incident angle were acquired.

Sentinel-2 MSI Level-1C images were also downloaded from CDSE. ACOLITE processors designed for water processing have demonstrated excellent performance in inland waters (Deng et al., 2024). This study derives multi-band $R_{rs}(\lambda)$ data with

a spatial resolution of 10 m from MSI images using the Dark Spectrum Fitting algorithm (DSF) already embedded within the ACOLITE processor (Knaeps et al., 2015; Vanhellemont, 2019). It produced 11 bands, namely, $R_{rs}(443)$, $R_{rs}(492)$, $R_{rs}(559)$, $R_{rs}(665)$, $R_{rs}(704)$, $R_{rs}(740)$, $R_{rs}(780)$, $R_{rs}(833)$, $R_{rs}(864)$, $R_{rs}(1610)$, and $R_{rs}(2186)$.

Landsat-8 TIRS data was provided by the U.S. Geological Survey and can be loaded at the GEE platform with the Level-2 product, namely surface temperature data. The raw values were corrected to degrees Celsius (°C) with a scaling factor. The lake surface temperature (LST) data was then resampled with 10 m and calibrated to align with SAR and MSI data.

Sentinel-1 SAR, Sentinel-2 MSI, and Landsat-8 TIRS LST images were matched with a 3-day time window to ensure salinity consistency over short periods. Eventually, a total of 385 multi-source satellite image pairs were obtained.

2.3.3 Quality controls

High-quality Sentinel-2 MSI images with few clouds covered (< 10%) were selected by visual interpretation and downloaded for all ice-free periods. Invalid pixels of clouds and cloud shadows covered in band 10 of Landsat-8 TIRS were masked with the QA_PIXEL band. Furthermore, the study extracted water and removed aquatic vegetation in the MSI image processing to avoid outliers generated by the model in sub-observed pixels. Using lake outline data from the Lake-Watershed Science Data Center (http://lake.geodata.cn) as the lake's initial boundary (Ma et al., 2011). Calculate normalized difference water index (NDWI) and combine it with the OTSU algorithm to set thresholds for extracting water from MSI imagery. The floating algae index (FAI) allows for spectral diagnosis of floating algae scum and was also useful in inland lakes (Hu, 2009). The FAI threshold was defined as 0.03 after multiple adjustments in the grassland lake Ulansuhai to exclude aquatic vegetation. A three-pixel inner buffer was created to mask out mixed pixels and minimize uncertainties in salinity estimation caused by bottom reflections (Feng et al., 2012), after which the lake area was counted.

2.4 Field dataset

From July 2017 to October 2024, 322 field survey data were collected from 11 lakes using two approaches: uniformly distributed field measurements and the China National Environment Monitoring Centre (CNEMC, https://www.cnemc.cn/). These data were divided into Dataset one and Dataset two. Dataset one contains 257 field salinity data, with 69 sets of collected surface water samples (depth < 0.5 m) and simultaneously measured SDD and in situ spectral data. Dataset one was used for training, testing, and five-fold cross-validation (CV) of the ML salinity algorithm and also as the initial driver for the salinity mechanistic model. Dataset two contains salinity data transformed from conductivity measurements by CNEMC or the latest data not involved in modeling. The conversion function was established based on synchronized measured conductivity and salinity in the IMXL (R^2 = 0.99, RMSE = 1.76 ppt, and MAPE = 28.4%) (Rusydi, 2018), as shown in Table 3. Dataset two (65 samples) was selected to ensure spatial independence (no overlap with Dataset 1's sampling points) and temporal coverage (2017–2024, including 10% of 2024 data not used in modelling) to validate generalization. The surface water salinity (units ppt) and conductivity (units µS/cm) of the lake were in situ measured using a

YSI multi-parameter water quality instrument (YSI ProDSS, USA). Spectral Evolution PSR-1100f (350–1050 nm, 1 nm spectral resolution) was used to measure water surface upward radiance (L_{sw}), sky radiance (L_{sky}), and gray plate radiance (L_p) at an observation direction of 40 degrees from the nadir and 135 degrees from the Sun (Mobley, 1999; Mueller et al., 2003). These radiance data were further used to calculate $R_{rs}(\lambda)$, with the formula given as follows (Mobley, 1999):

$$R_{rs}(\lambda) = \left[\left(L_{sw} - \rho \times L_{sky}\right) \times \rho_p\right]/\pi \times L_p \tag{1}$$

where ρ represents the water-air interface reflectance, assumed to be 0.0028 under calm conditions (wind speed < 4 m/s), and ρ_p denotes the reflectance of the reference gray plate defined as 0.30. We calculated it by setting $R_{rs}(\lambda)$ between 950 and 1050 nm to zero, since a low signal-to-noise ratio was observed within this range (Lee et al., 2016). Finally, convolve $R_{rs}(\lambda)$ through the spectral response function (SRF) to simulate bands of the MSI.

Field water samples were stored in polyethylene bottles and rapidly transported back to the laboratory for analysis at the end of the cruise. Water samples were filtered using Whatman GF/F films with a 0.7 μm pore size to extract chlorophyll a (Chl$a$, μg/L), suspended particulate matter (mg/L), and CDOM samples. Laboratory measurements of these parameter concentrations can be found in Deng et al. (2024).

Table 3 Conversion equation and factors for salinity and conductivity of lakes in the IMXL.

Salinity (ppt)	Salinity = conductivity * factors / 1000							
Conductivity (μS/cm)	<1,000	1,000–10,000	10,000–20,000	20,000–30,000	30,000–45,000	45,000–60,000	60,000–65,000	>65,000
Factors	0.50	0.55	0.58	0.60	0.65	0.70	0.725	0.75
R^2				0.99				
95% CI				0.98–1.00				

2.5 Ancillary data

ERA5 land reanalysis data was published by the European Center for Medium-Range Weather Forecasts (ECMWF, https://cds.climate.copernicus.eu/) and can be downloaded for free. This product provides hourly meteorological variables in GeoTIF format with a spatial resolution of approximately 11 km (Muñoz Sabater, 2019). It has been available since 1980 and provides a continuous record that can compensate for gaps in field meteorological observations. We downloaded hourly wind speed (WS, m/s), temperature (TEMP, °C), evaporation (EVP, mm) and precipitation (PRE, mm) data for each lake from GEE during 2016–2024 and calculated daily data on this basis. Additionally, nighttime light (NTL, nW/cm²/sr) in the Lake Daihai sub-basin was derived from the Visible Infrared Imaging Radiometer Suite (VIIRS). Population (POP) data were sourced from the statistical yearbook.

2.6 Construction of salinity model

This study proposed a brand-new framework integrating microwave and optical data to estimate lake salinity, as illustrated in Figure 2. It consists of three modules: data processing and feature building, model construction and ensemble, and lake

salinity estimation. Module 2 is the core of the framework, aiming to use wave spectrum, backscatter coefficient, and dielectric constant models as forward models to build a salinity mechanistic model, then establish ML models, and finally construct an ensemble model by coupling mechanistic and ML results via a Generalized Additive Models (GAM) model.

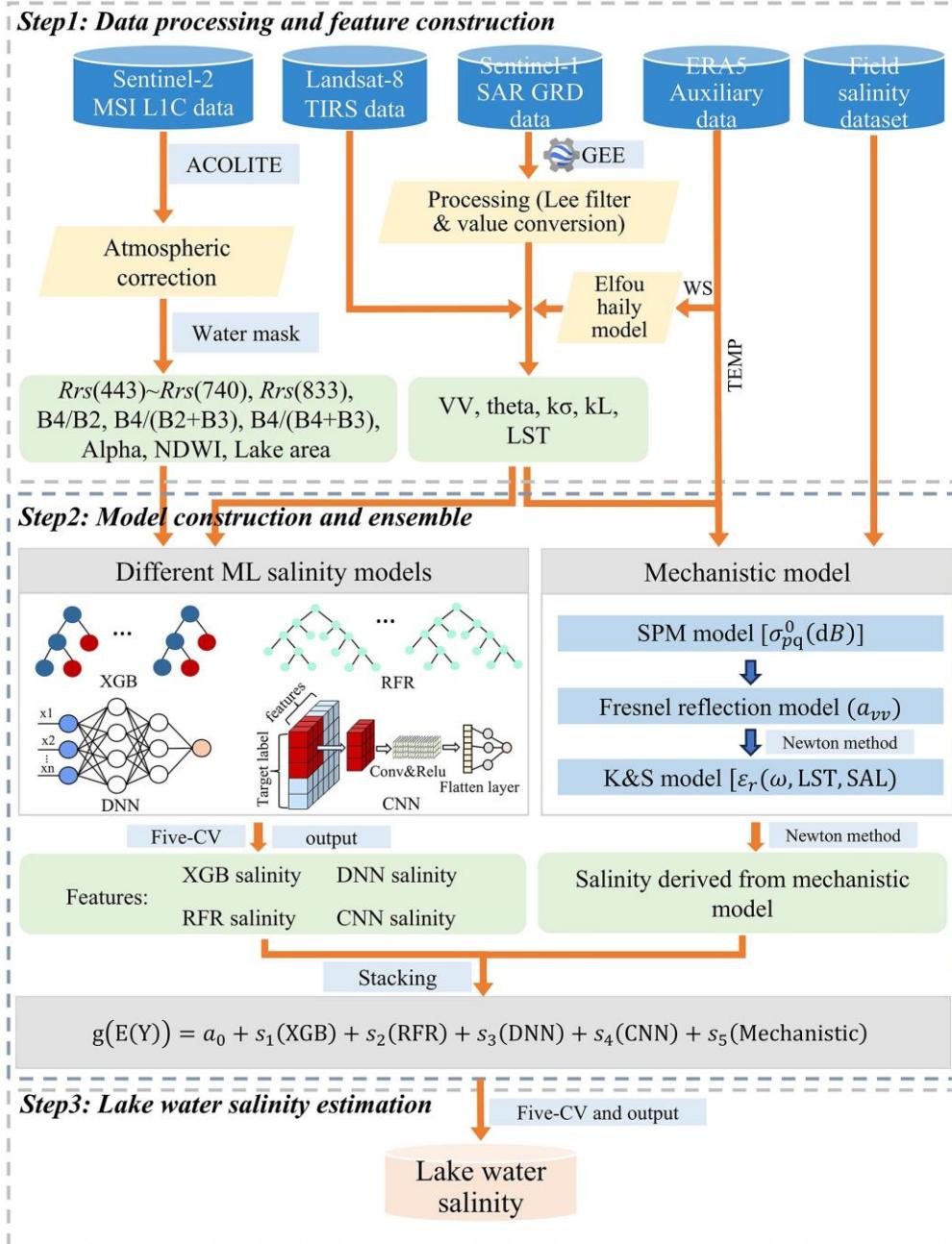



Figure 2: A brand-new lake water salinity estimation framework by a stacking salinity model, consisting of three steps: step one, data processing and feature construction; step two, model construction and ensemble; and step three, salinity estimation. Details of individual steps and the meaning of the symbols are described below.

2.6.1 Mechanistic salinity model

185 (1) Construct a lake surface roughness model. The wave spectrum was used to rapidly calculate roughness parameters. The roughness of a lake surface can be characterized by the height standard deviation (kσ) and the correlation length (kL). The Elfouhaily spectrum considers the long-wave and short-wave effects simultaneously and defines the inverse wave age as a function of wind speed and fetch. It was widely used to describe the energy information of surface wind waves, and its basic formula was given as follows (Elfouhaily et al., 1997):

190
$$\Psi(k,\varphi) = \frac{1}{k}S(k)\phi(k,\varphi) \tag{2}$$

$$S(k) = (B_l + B_h)/k^3 \tag{3}$$

where k is the wave number (rad/m), φ is the direction angle, $\Psi(k, \varphi)$ is the directional spectrum, $S(k)$ is the omnidirectional spectrum, B_l is the long-wave curvature spectrum, and B_h is the short-wave curvature spectrum. Parameter adjustments to the Elfouhaily model were required due to the significant differences in wave formation and magnitude between lakes and

195 oceans. Fetch was parameterized as a function of lake area (Fetch = $e \times \sqrt{area}$) to adjust the inverse wave age, which influences the wave mode of the Elfouhaily model. e is an empirical coefficient of 0.65 determined by several field observations and published lake experiments (Young and Verhagen, 1996). The Elfouhaily model can calculate the height variance ($\sigma\eta^2$) and mean square slope (mss) of surface waves, with the specific formulas detailed in Elfouhaily et al. (1997). The kσ and slope can be obtained by taking the square root of these two parameters, respectively. Due to the small scale of

200 lake waves, the study assumed kL as the maximum distance between wave peaks and troughs and defined it as the ratio of kσ to slope based on trigonometric relationships. Therefore, the initial lake roughness can be calculated by using the adjusted Elfouhaily model in combination with wind speed and lake area.

(2) Calculate pixel-based roughness by combining the backscatter coefficient model and cost function. The small perturbation method (SPM), a backscatter coefficient simulation model, is suitable for water surfaces with small undulations

205 dominated by capillary waves (Johnson and Zhang, 1999; Khenchaf, 2001; Shareef et al., 2016), whose fundamental formula is as follows:

$$\sigma^0_{pq}(dB) = 4kL^2k\sigma^2 cos^4\theta |\alpha_{pq}|^2 \exp\left(-(kLsin\theta)^2\right) \tag{4}$$

$$a_{VV} = \frac{(\varepsilon_r - 1)\left[\sin^2\theta - \varepsilon_r(1 + \sin^2\theta)\right]}{(\varepsilon_r \cos\theta + \sqrt{\varepsilon_r - \sin^2\theta})^2} \tag{5}$$

$$\varepsilon_r(\omega, LST, SAL) = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 + (i\omega\tau)^{1-\alpha}} - i\frac{\sigma}{\omega\varepsilon_0} \tag{6}$$

210 where $\sigma^0_{pq}(dB)$ represents the water backscattering coefficient simulated by the SPM model at different polarizations, a_{pq} is the Fresnel reflection coefficient for VV or HH polarization, θ is the incident angle, and ε_r is the water dielectric constant calculated using the Klein and Swift (K&S) model with parameter details as shown in Klein and Swift (1977). Construct a cost function and iteratively optimize it using the least squares algorithm to derive pixel-based roughness from SAR images:

$$X_i[\text{k}\sigma, \text{kL}] = \frac{[\sigma_{vv}^{measure} - \sigma_{vv}^{model}(\theta_i, \text{SAL}, \text{LST}, \text{k}\sigma, \text{kL})]^2}{\sigma_{vv_{measure}}^2}, i \in [1, N] \tag{7}$$

where $\sigma_{vv}^{measure}$ is the backscattering coefficient measured by SAR at VV polarization, σ_{vv}^{model} denotes the C-band backscattering coefficient simulated by the SPM model at VV polarization, i is the i-th pixel, N is the total number of pixels, θ_i is the i-th pixel incident angle, and $X_i[\text{k}\sigma, \text{kL}]$ is the i-th pixel roughness.

(3) Stepwise salinity retrieval based on the SPM and K&S models. Step one, the a_{vv} was calculated using the SPM model based on pixel-based roughness images. Step two, the dielectric constant was deduced by the Fresnel reflection model with the Newton method. Step three, the K&S model was employed to iteratively solve for salinity via the Newton method, using TEMP and initial salinity as inputs.

2.6.2 Machine learning salinity models

The ML salinity models will be developed utilizing four algorithms, including gradient boosting (XGB), random forest regressor (RFR), deep neural networks (DNN), and convolutional neural networks (CNN). 70% of dataset one will be used for training and 30% for testing, with the entire dataset applied to five-fold CV. The XGB and RFR are typical ensemble learning models with decision trees as the fundamental units. The XGB predictions are the weighted sum of each tree's score, while the RFR results are averages of all tree predictions (Breiman, 2001; Chen and Guestrin, 2016). DNN and CNN are typical neural network models made up of interconnected neurons. Their parameters are updated via backpropagation to optimize the loss function, and the final predictions are output by fully connected layers (LeCun et al., 2015; Alzubaidi et al., 2021).

The construction of four salinity algorithms involves feature selection, model training, model testing, and five-fold CV. A total of 18 features were selected, including $R_{rs}(443)$, $R_{rs}(497)$, $R_{rs}(560)$, $R_{rs}(664)$, $R_{rs}(704)$, $R_{rs}(740)$, $R_{rs}(842)$, B4/(B4 + B3), B4/(B2 + B3), B4/B2, NDWI, chromaticity angle (alpha), lake area, VV, theta, kσ, kL, and LST. Visible and near-infrared bands are considered sensitive to water salinity (Urquhart et al., 2012; Bayati and Danesh-Yazdi, 2021), while alpha synthesized abundant information from the visible bands (Wang et al., 2023b). Lake area as a proxy for lake water volume is closely correlated with salinity (McGrath et al., 2025). The selected microwave features are the crucial variables in the mechanistic model. Model hyperparameters were determined using grid search during the training process. The hyperparameters to be determined for each model are detailed in Table 4. For the assessment of model stability and generalization performance, five-fold CV was subsequently conducted after determining the model structures and hyperparameters (Cao et al., 2024). The entire dataset was randomly divided into five folds, each serving sequentially as the test set while the remaining folds were used for training. Model performance was assessed by averaging the statistical metrics obtained from the five rounds. To improve ML model interpretability, Shapley Additive Prediction (SHAP) values were used to quantify the contribution of each feature (Lundberg et al., 2020; Gao et al., 2025). And then four ML salinity models were applied to estimate lake salinity from MSI and SAR data.

Table 4 Listed key hyperparameters for each machine learning model, with specific parameter settings available in Supplementary Table S1 of the Zenodo repository.

Model	Key hyperparameters
XGB	Trees, Learning rate, Maximum tree depth, Colsample_bytree, Subsample rate, Min_child_weight, Regularization
RFR	Trees, Maximum depth, Maximum features at node splitting
DNN	Hidden layer number, Each layer neurons, Learning rate, Activation function, Optimizer, Alpha, Maximum training epochs, Patience
CNN	Convolutional layer kernels, Activation layers, Fully connected layer, Learning rate, Optimizer, Alpha, Maximum training epochs, Patience
GAM	Smoothing function, Link function, N_splines, Callbacks, Max_iter, Lam, Tol
BMA	GLM.family, Prior, OR, maxCol

2.6.3 Stacking salinity model

The GAM model is an interpretable statistical model that can handle nonlinear relationships between covariates and response variables by using smoothing functions (Yee and Wild, 1996). It can work with response variables of various distribution types and provides multiple link functions. Partial dependency plots (PDP) provide a highly interpretable visualization of the smoothing functions for GAM response variables, and the tipping points defined the threshold for variables contributing positively or negatively to the model. The effective degree of freedom (edf) parameter can be used to measure the nonlinear complexity of each response variable's smoothing term. An integrated salinity model was constructed based on the estimation results from mechanistic and ML models under the assumption that variables follow a Gaussian distribution, and it is structured as follows:

$$g(E(Y)) = a_0 + s_1(XGB) + s_2(RFR) + s_3(DNN) + s_4(CNN) + s_5(Mechanistic) \tag{8}$$

where $g(E(Y))$ represents the predicted salinity with a unit of ppt, a_0 is the intercept term, and $s_1(XGB)$ denotes the smoothing term constructed using thin-plate regression spline functions for the XGB-predicted salinity, with smoothing terms for other variables consistent with this approach. The key hyperparameters of the GAM model are listed in Table 4, with specific configurations shown in Supplementary Table S1. This integrated model combines the powerful diagnostic capabilities of several ML models with the prior knowledge of mechanistic models to enable collaborative estimation of lake salinity using optical and microwave data.

2.7 Accuracy evaluation

The differences between estimated and measured salinity were evaluated by using several statistical metrics named R^2, root mean square error (RMSE), mean absolute error (MAE), bias (system error), and mean absolute percentage error (MAPE). These metrics were calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{10}$$

$$\text{Bias} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{11}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{\hat{y}_i}\right| \times 100\% \tag{12}$$

where y_i is the field measured salinity, $\hat{y}_i$ is the model estimated salinity, i denotes the i-th sampling point data, and n is the number of sampling point pairs.

3 Results and analysis

270 ### 3.1 Model performance

The constructed stacking model (N = 84, RMSE = 0.60 ppt, and MAPE = 2.3%) outperformed four ML models and the mechanistic model, and the predicted salinity distributed consistently along the 1:1 line without significant underestimation or overestimation (Figure 3). The five-fold CV result for the stacking model was close to the accuracy of the 30% dataset test (N = 257, RMSE = 0.38 ppt, and MAPE = 6.9%), with the range of estimated salinity consistent with measured salinity and

275 no outliers observed, indicating that the model has good generalization and stability without significant dependence on the training set (Figure 3j). The performance of the mechanistic model was second only to the ensemble model (N = 257, RMSE = 0.80 ppt, and MAPE = 13.3%), better than the results of five-fold CV for each ML model (RMSE > 0.97 ppt and MAPE > 15.1%) (Figure 3m). The XGB model showed the best accuracy among the four ML models, followed by CNN and DNN, while the RFR model was the worst performer (Figure 3). Additionally, the bar chart shows that the ensemble model

280 outperforms the ML algorithms in terms of accuracy (RMSE < 0.55 ppt and MAPE < 8.7%) (Figure 3k,l), indicating that incorporation of the mechanistic model improves overall performance.

Figure 4 displays the SHAP values of the selected features for four ML salinity models. Among these models, lake area, theta, and VV exhibit significant contributions, as reflected in their higher SHAP values compared to other variables. The lake area contributes the most because it correlates with water volume, which directly influences the degree of salinity

285 dilution (McGrath et al., 2025). VV and theta are important parameters because of their sensitivity to water surface scattering mechanisms, while salinity affects scattering intensity by altering the dielectric constant of the water (Reul et al., 2009). B6 reflectance ranks third in contribution within the XGB, RFR, and DNN models, indicating it is an optically sensitive band for salinity. In the CNN model, the VV band contributes most significantly, while alpha was the most influential optical index.

A simpler positive correlation (edf = 2.19) with salinity was found for s_5(Mechanistic) compared to the other variables (edf >

290 2.76) in the PDP of the ensemble model (Figure 5). The PDP curve of the mechanistic model has a broader range of variation on the y-axis than the ML models, with its 95% confidence region encompassing most of the sample data, suggesting that the model contributes significantly and reliably to salinity prediction for the ensemble model. At a salinity range of 0–4.71 ppt, XGB, RFR, CNN, and Mechanistic models contribute negatively to the ensemble model, but their contributions turn positive at salinities exceeding 9.52 ppt. The contribution pattern of the DNN model differed from those of the aforementioned

295   models (Figure 5c). Overall, the integrated salinity model works admirably by effectively coupling the virtues of both mechanistic and ML models, making it successful for estimating lake salinity.

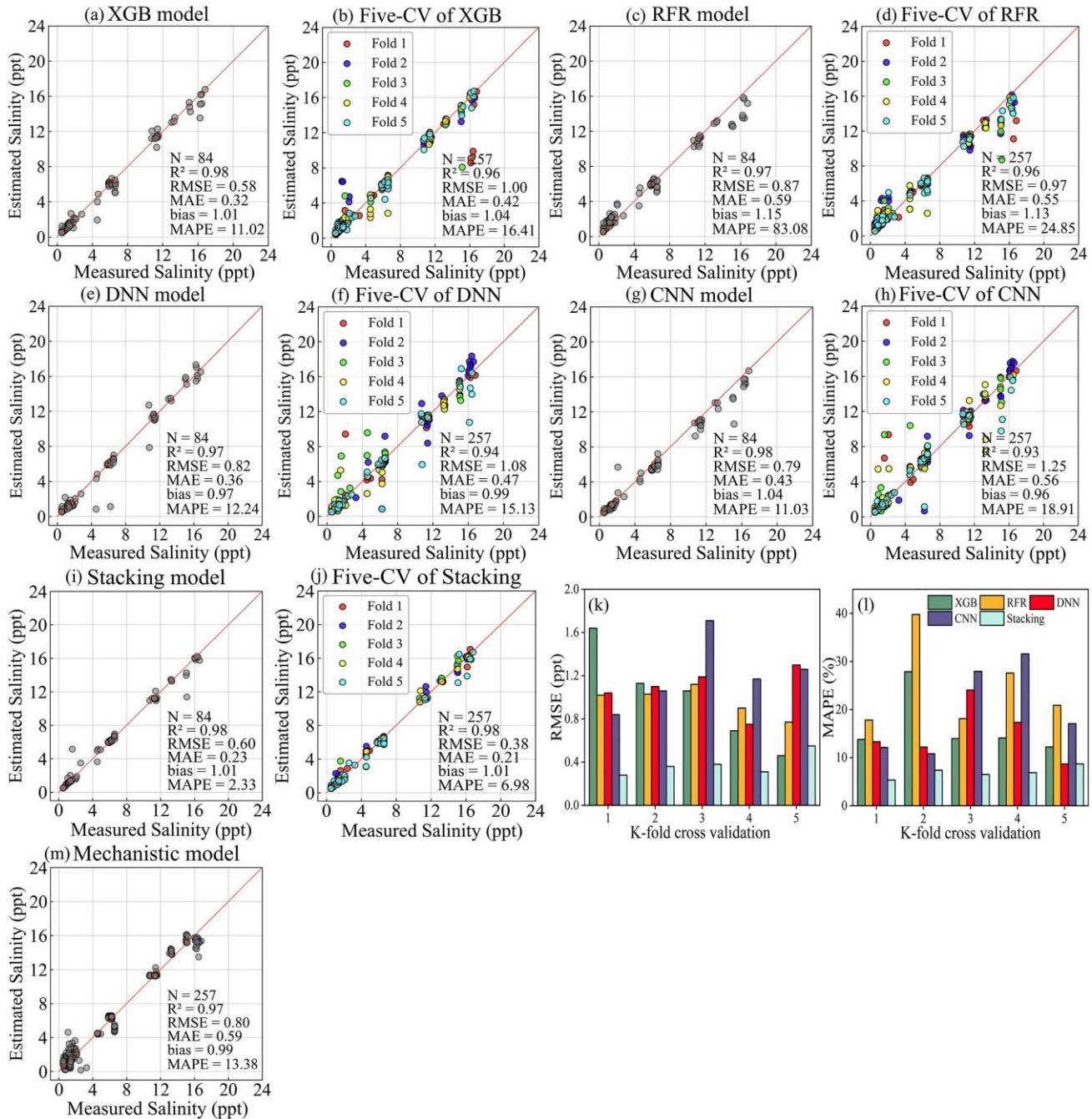



Figure 3: (a)-(i) Scatter plots of 30% test data (N = 84) for XGB (R² = 0.98 and RMSE = 0.58 ppt), RFR (R² = 0.97 and RMSE = 0.87 ppt), DNN (R² = 0.97 and RMSE = 0.82 ppt), CNN (R² = 0.98 and RMSE = 0.79 ppt), and stacking model (R² = 0.98 and RMSE = 0.60 ppt); (j) Five-fold CV for stacking model (N = 257, RMSE = 0.38 ppt, and MAPE = 6.9%); (k)-(l) RMSE/MAPE

300   **comparison; (m) Mechanistic model field validation (N = 257, RMSE = 0.80 ppt, and MAPE = 13.3%).**

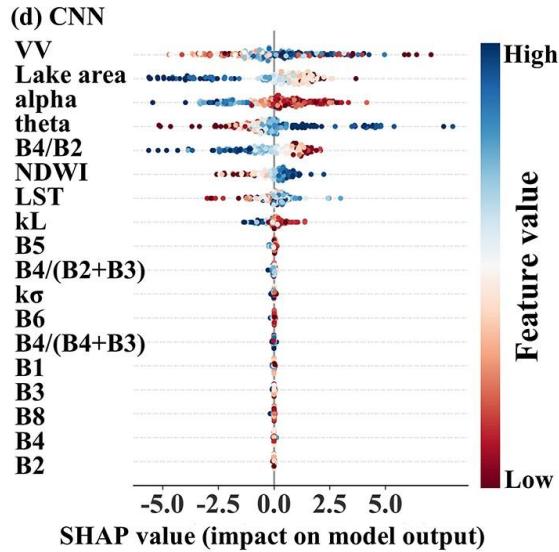
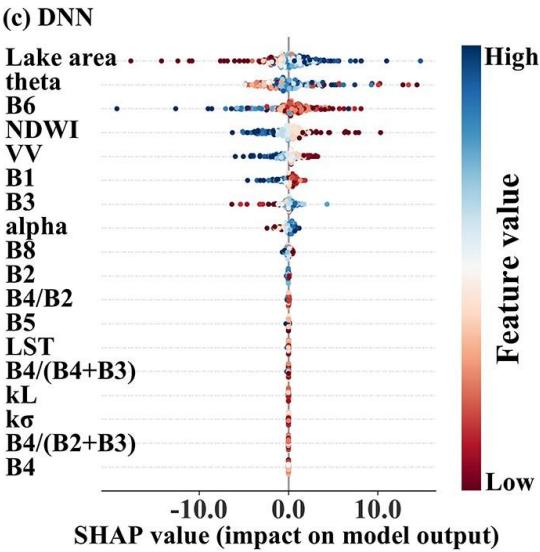
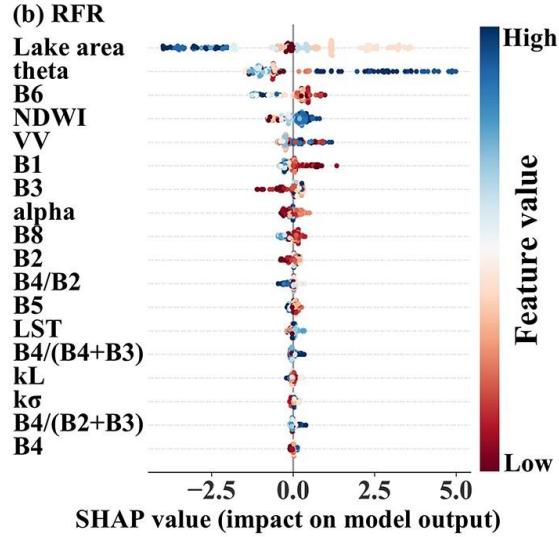
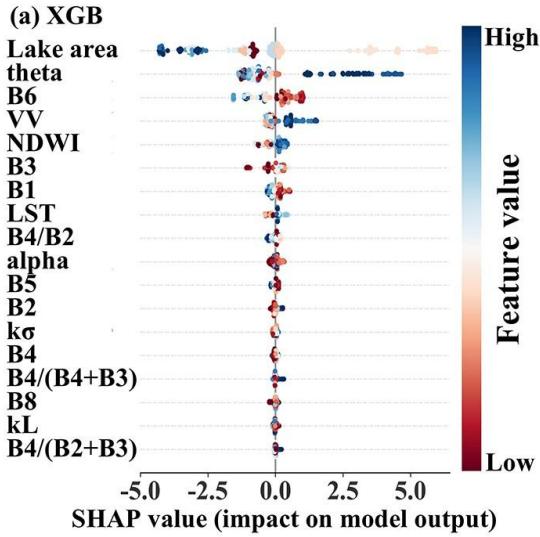



Figure 4: (a)-(d) SHAP plots for XGB, RFR, DNN, and CNN. B1–B6 and B8 correspond to $R_{rs}(443)$ $-R_{rs}(740)$ and $R_{rs}(833)$, respectively.

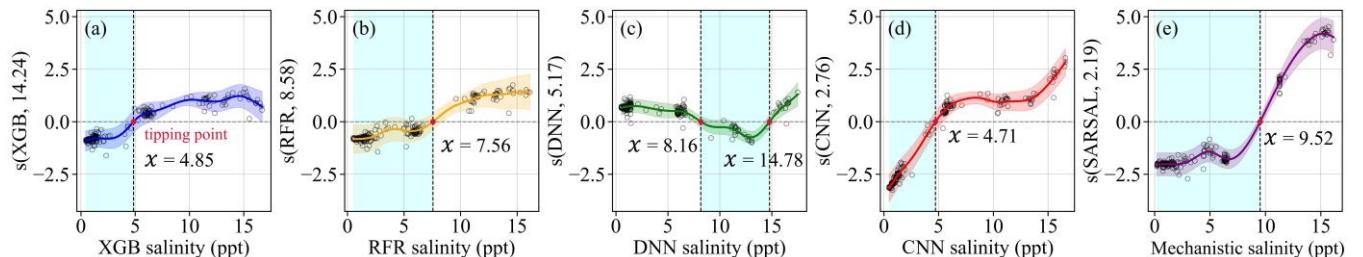



Figure 5: (a)-(e) Partial dependence plots for XGB, RFR, DNN, and CNN models, the red dots represent the tipping point for the model variables contribute positively or negatively to salinity predictions.

305

3.2 Single-scene analysis of different models

A single-scene comparison for lakes with matching field sampling points was used to examine the spatial quality of salinity data produced by the stacking model, ML model, and mechanistic model (Figure 6). Lake salinity maps generated by the stacking model show a smooth transition from the shore to the center in 10 lakes, whereas a slight discontinuity of salinity was observed in each lake using ML or mechanistic models. Especially in nearshore waters, DNN, CNN, and mechanistic models exhibit outliers due to land adjacency effects, while the stacking model corrects predictions by combining accurate salinity derived from XGB and RFR, such as in Lake Hongjiannao, as shown in Figure 6(h2,h3). However, salinity values in Lake Bosten's nearshore pixels were not fully corrected because the DNN and CNN models overestimate salinity influenced by mixed pixels (Figure 6(c4,c5,c7)). The stacking model successfully captured the spatial variations of freshwater dilute salinity by combining the capabilities of ML models in salinity estimates at river inlets into lakes, compensating for the limitations of mechanistic models, as shown in Figure 6(i7). Furthermore, the RMSE of the stacking model was observed to be lower than other algorithms across nine lakes in single-scene comparisons, with only the mechanistic model performing better in Lake Juyan. Finally, it was observed that the stacking model effectively suppressed salinity outliers to guarantee the quality of the dataset. In brief, the stacking model produced maps of lake salinity with smoother spatial variations and richer detail, yielding a higher-quality dataset.

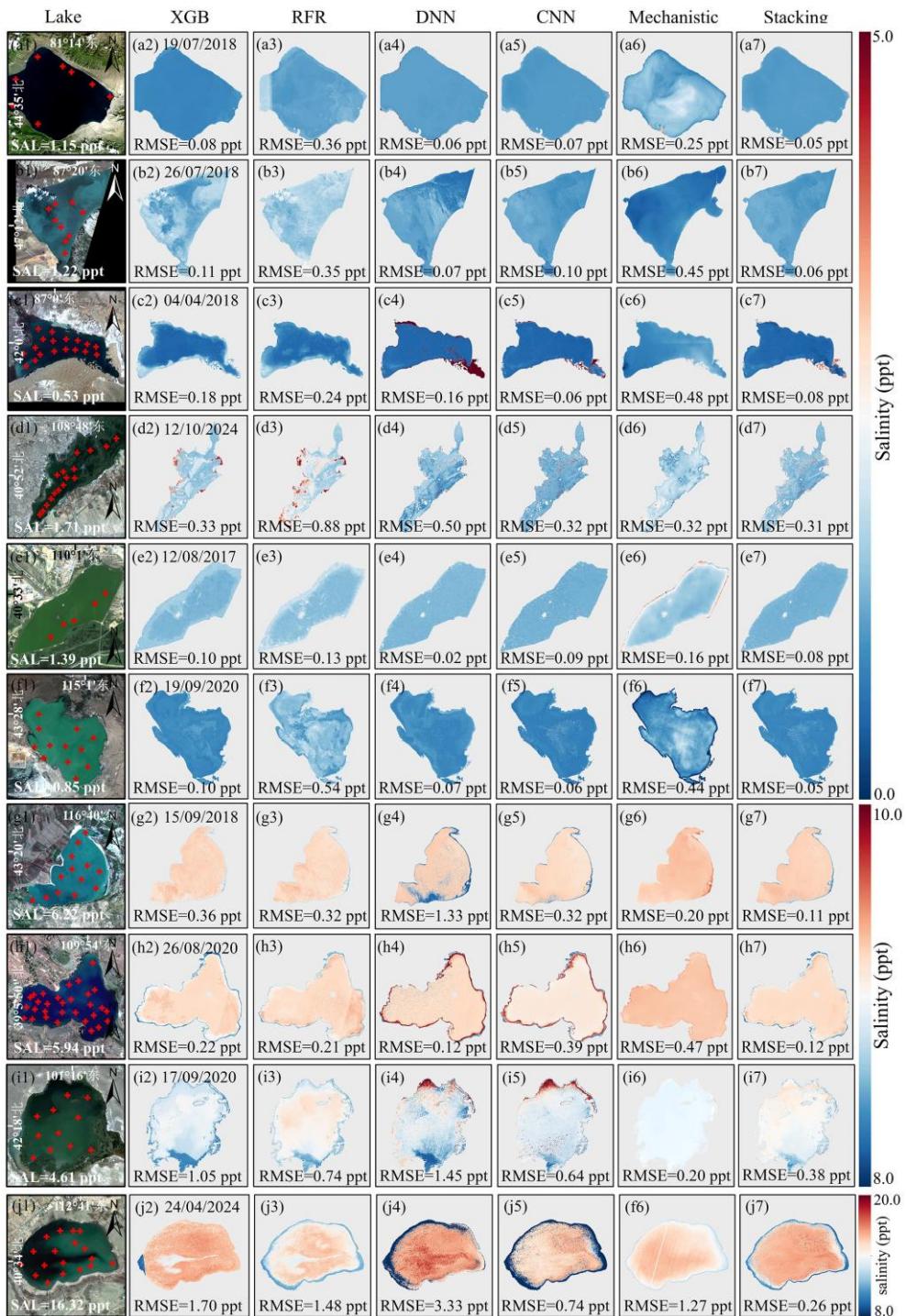



Figure 6: (a1)-(j7) Comparisons of water salinity estimated by XGB, RFR, DNN, CNN, mechanistic, and stacking models from MSI-derived $R_{rs}(\lambda)$ images and SAR data in 10 lakes, namely, Sayram, Ulungur, Bosten, Ulansuhai, Nanhaizi, Chagannaoer, Dalinor, Hongjiannao, Juyan, and Daihai, respectively. For each lake, the first column shows a true color composite generated by MSI data.

3.3 Comparison with previous single-satellite algorithms

Comparisons between the MSI-based XGB salinity algorithm (Deng et al., 2025) and the SAR-based mechanistic algorithm

330 were performed for each lake (Figure 7), and it can be revealed that the stacking algorithm has higher and more stable accuracy with an average RMSE of 0.24 ppt. Although the XGB and the mechanistic models outperformed the stacking algorithm in some lakes, including Lake Ulungur, Lake Juyan, and Lake Nanhaizi, both models (with RMSE of 0.45 ppt and 0.57 ppt, respectively) still showed slightly lower precision across the entire region. The stacking model and the mechanistic model outperformed the XGB salinity model in oligosaline-type lakes, suggesting that the mechanistic model improves the

335 accuracy of the stacking algorithm under salinity exceeding 3 ppt. In addition, the points for the three models were concentrated on the Taylor plots of several lakes. It can be deduced that the mechanistic model and XGB model also exhibit reliable performance to provide rational data support for constructing the stacking model. Overall, the proposed algorithm combines the strengths of both the physically constrained model and the data-driven model in that it avoids the single model or data source as well as improving the precision in complex inland water.

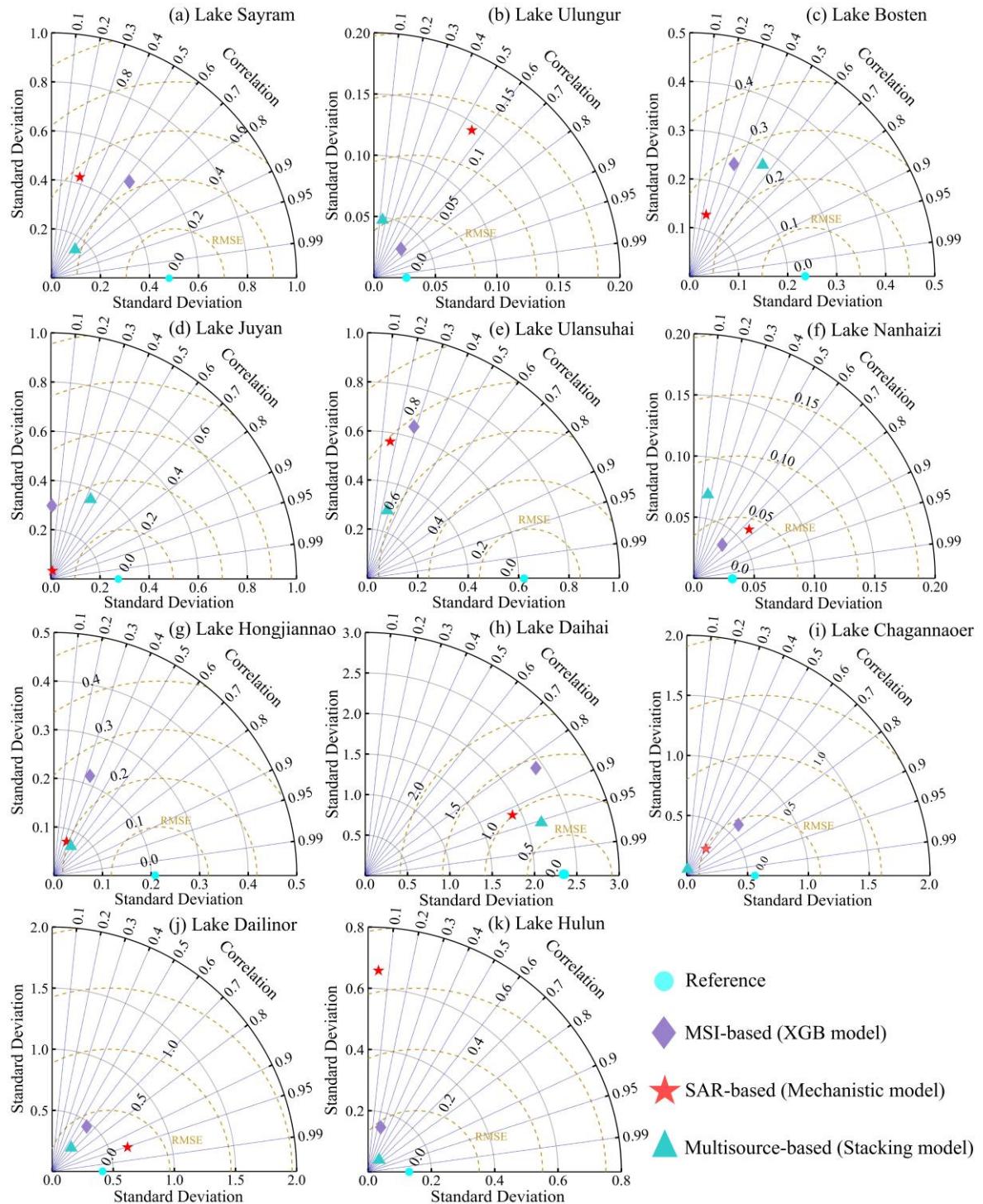
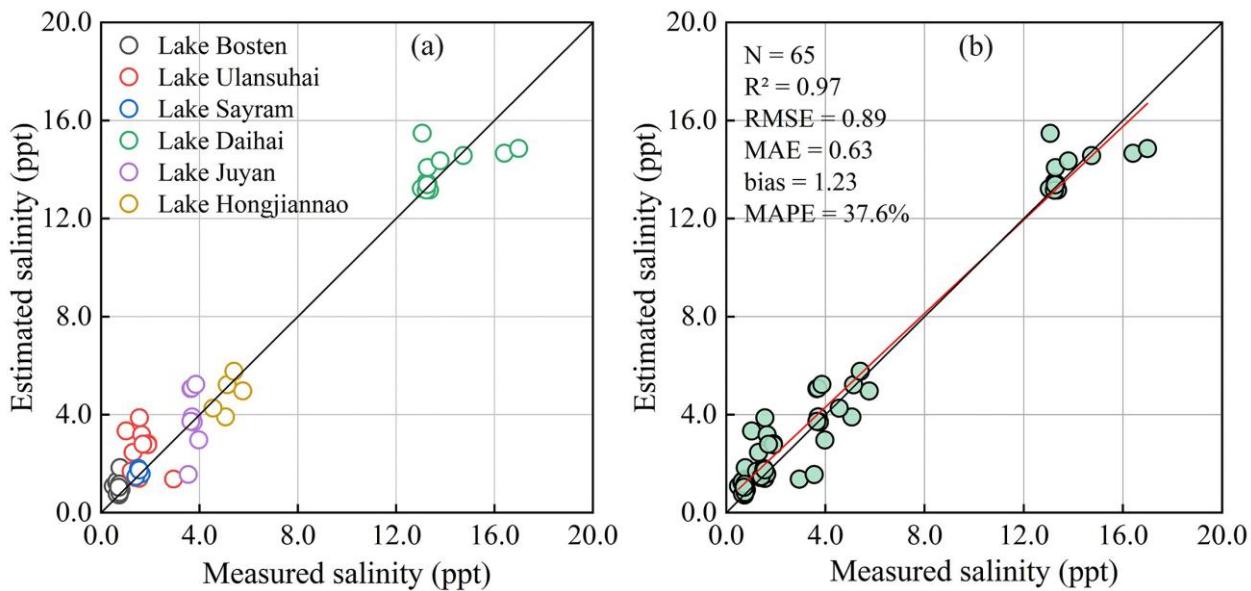



Figure 7: Comparison of multisource data-based stacking model with single-satellite model. (a)-(k) Comparison results using the salinity dataset from Lake Sayram, Lake Ulungur, Lake Bosten, Lake Juyan, Lake Ulansuhai, Lake Nanhaizi, Lake Hongjiannao, Lake Daihai, Lake Chagannaoer, Lake Dalinor, and Lake Hulun, respectively.

340

3.4 Independent validation

345 To further objectively evaluate the accuracy and scientific validity of the proposed framework, independent validation was performed using Dataset 2, which was not involved in model training, testing, or CV. The validation density was insufficient due to the absence of some lake stations and historical salinity data. The independent validation (Figure 8) demonstrates that salinity estimates from the integrated algorithm predominantly align well with measured salinity along the 1:1 line (N = 65, R² = 0.97, RMSE = 0.89 ppt, and MAPE = 37.6%). Only Lake Ulansuhai showed a slight overestimation, likely affected by

350 aquatic vegetation pixels. An underestimated validation point was observed in the southern part of the Juyan Lake, associated with the complex water characteristics in the river inlet region. No significant underestimation or overestimation was observed in other lakes. These results confirm that the integrated algorithm combining microwave and optical data has considerable accuracy in retrieving salinity in inland lakes.

355 **Figure 8: Using Dataset two for independent validation, not all lakes have independent validation data.**

3.5 All salinity images pixel-based statistical validation

Pixel-based histogram statistics were performed on the salinity raster generated by the stacking model for each lake (Figure 9). Using frequency instead of pixel counts in traditional histograms, the mean salinity, standard deviation (STD), and frequency proportions of different salinity ranges were calculated. The distribution patterns of salinity can be visualized through a frequency histogram, which helps identify outliers and objectively assess salinity map quality. Outliers are usually

360 found dispersed and dramatically changed, appearing as discontinuities in histograms. A single peak pattern of salinity histogram was observed in most freshwater or brackish lakes, including Lake Sayram, Lake Ulungur, Lake Bosten, Lake Nanhaizi, and Lake Hulun. These minimal spatial and interannual variations within the lakes align with the field measurements found. A double-peak characteristic was observed in Lake Juyan and Lake Ulansuhai, with the primary and

365 subsidiary peaks distributed consecutively, caused by the differences between the northern and southern regions for the lake salinity. However, the subpeaks in the bimodal distribution of Lake Chagannaoer showed low frequency and discontinuity, indicating that this portion of the data may be abnormal. The oligosaline-type lakes (e.g., Lake Hongjiannao, Lake Daihai, and Lake Dalinor) commonly showed multi-peak characteristics and wide fluctuation ranges, suggesting high interannual or spatial variations in salinity.

370 Further analyze image quality by examining the pixel frequency across different salinity ranges. The pixel frequency in Lake Sayram, Lake Ulungur, Lake Bosten, and Lake Nanhaizi was more than 90% within the interval of 0–2 ppt. Lake Juyan (46.6%), Lake Ulansuhai (89.1%), and Lake Daihai (88.2%) accounted for the highest pixel proportions within salinity ranges of 4–6 ppt, 0–4 ppt, and 8–16 ppt, respectively. Lake Hongjiannao and Lake Dalinor had the highest pixel proportions in the range of 4–8 ppt, with 68.5% and 70.4%. These high-proportion intervals align with the distribution of field

375 measurement salinity, and the small proportion of abnormally low-frequency pixels demonstrated that the salinity data generated by the stacking model had a good image quality with only a few outliers.

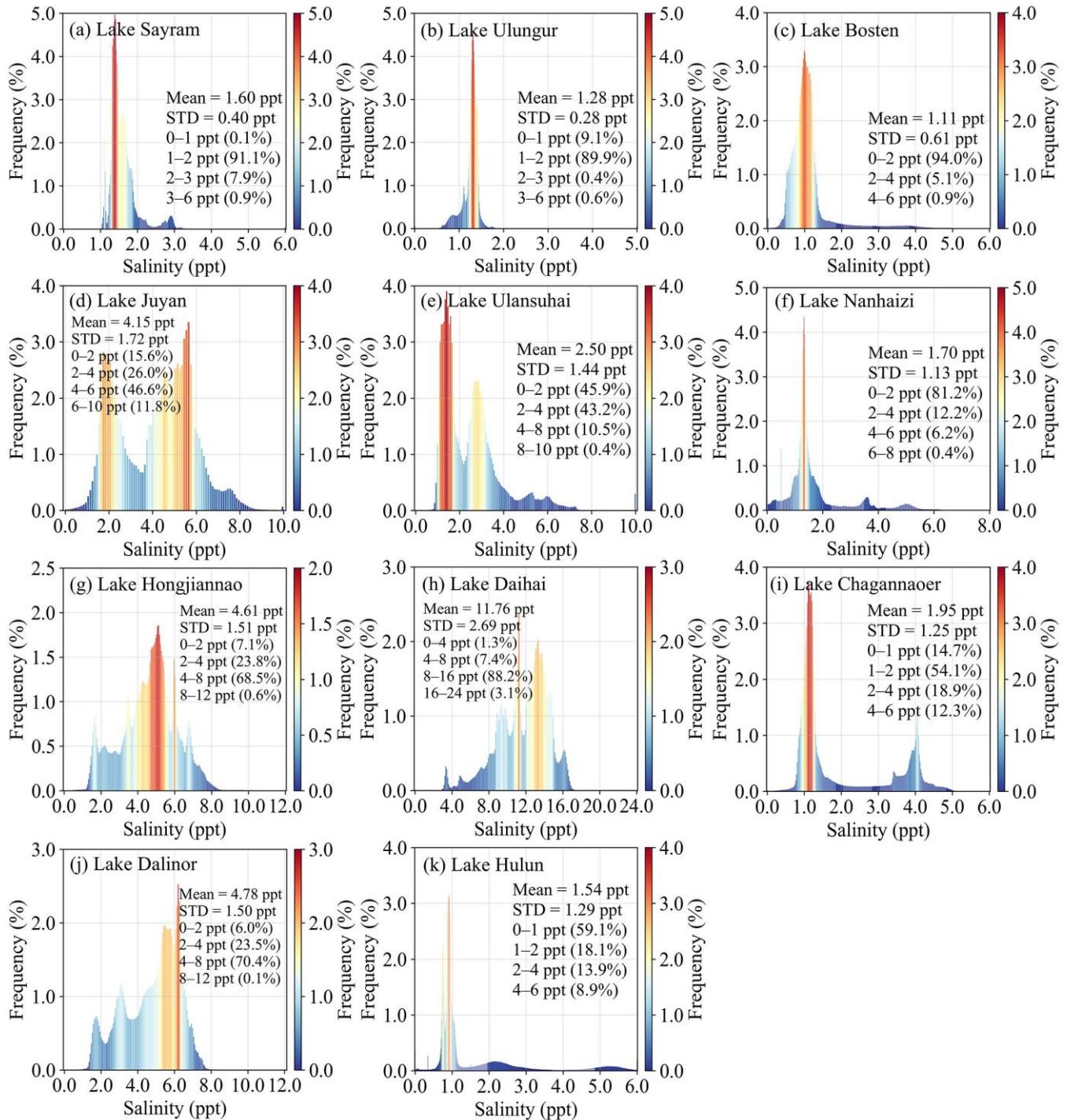



Figure 9: (a)-(k) Pixel-based frequency statistics for all salinity images in each lake, and the STD represents the degree of deviation from the mean salinity.

3.6 Spatial patterns and trends

The annual scale salinity map of 11 lakes was shown in Figure 10, with its color fluctuations effectively delineating the spatial differences in water salinity. The water salinity exhibits spatial homogeneity within several lakes, such as Lake Sayram, Lake Ulungur, and Lake Bosten. Lake salinity discontinuities were commonly observed in shore or river inflow zones, such as the western part of Lake Ulansuhai, the southern region of Lake Juyan, the northern part of Lake Chagannaoer, and the eastern part of Lake Hulun, which were primarily affected by mixing pixels or freshwater dilution (Han et al., 2021). This phenomenon with freshwater dilution of water salinity is prevalent in inland lakes receiving surface runoff, as well as lakes on the Tibetan Plateau (Wang et al., 2023a). In addition, the salinity in Lake Nanhaizi was higher in the east than in the west. Seasonal salinity patterns showed greater spatial heterogeneity than annual patterns across all lakes (Figure 11), particularly in Lake Ulansuhai, Lake Nanhaizi, and Lake Chagannaoer. It indicates that the spatial pattern of salinity is regulated by external environmental factors, such as seasonal variations in precipitation, runoff, and evaporation (Rimmer et al., 2006; Yihdego and Webb, 2012; Jiang et al., 2022).

Interannual salinity trends were calculated using the Mann-Kendall trend test (significance level $\alpha = 0.05$). Lake Daihai (slope = 0.48 ppt/year, $p < 0.01$) and Lake Dalinor (slope = 0.22 ppt/year, $p < 0.05$) showed significant increasing trends (Figure 12a), and the management agency requires attention to this occurrence. Other lakes had non-significant interannual trends, with a comparable proportion of them increasing or decreasing in salinity. Significant seasonal variations of salinity ($p < 0.05$) were observed in some lakes (Figure 12c), with higher salinity in summer and autumn compared to spring due to enhanced evaporation during summer, including Lake Ulungur, Lake Ulansuhai, Lake Hongjiannao, and Lake Chagannaoer. The other lakes did not present significant seasonal changes, and the freshwater and brackish lakes had lower seasonal salinity differences compared to oligosaline lakes, likely due to their greater water storage capacity and more reliable water supply sources (Rusuli et al., 2016).

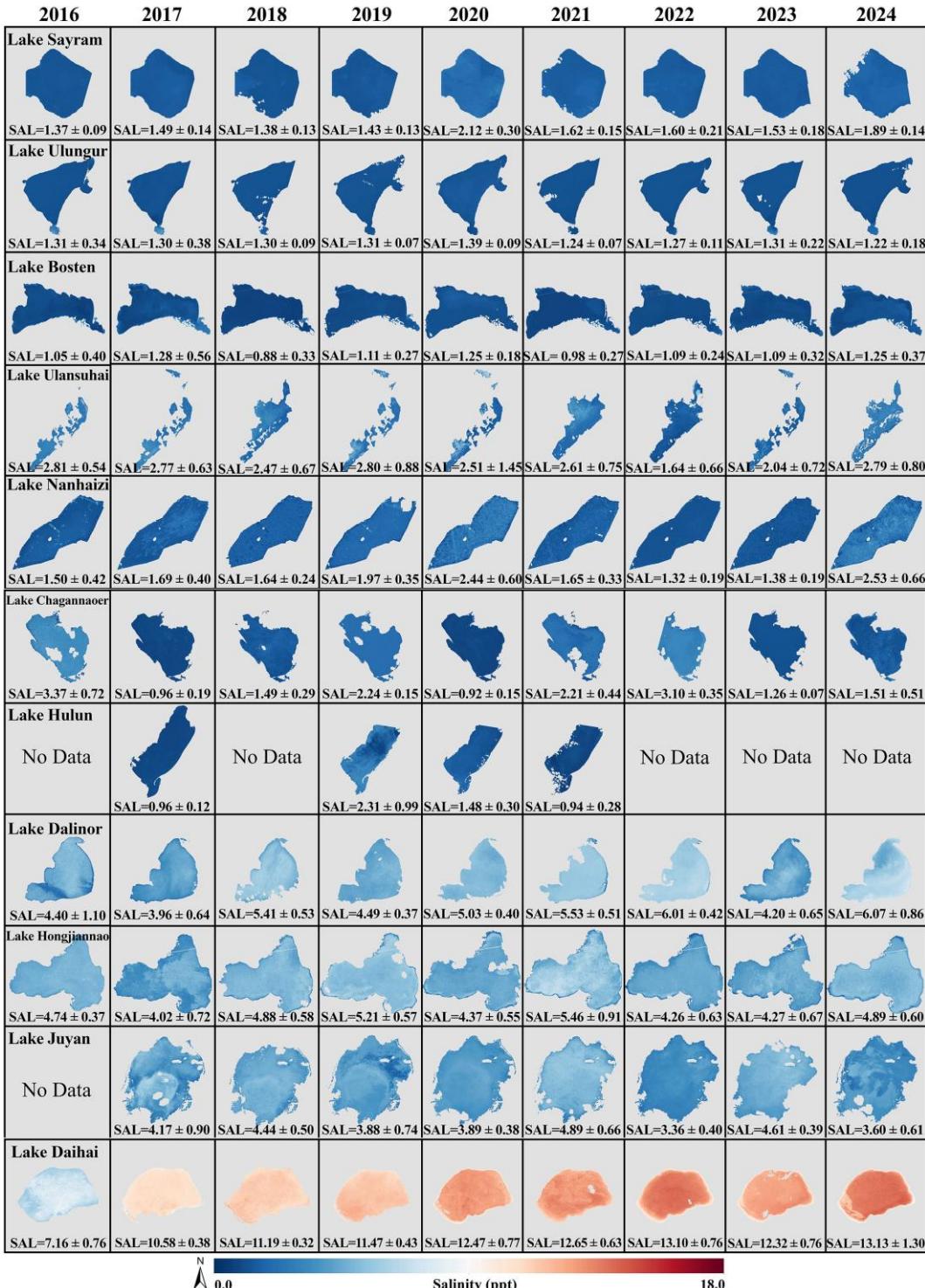
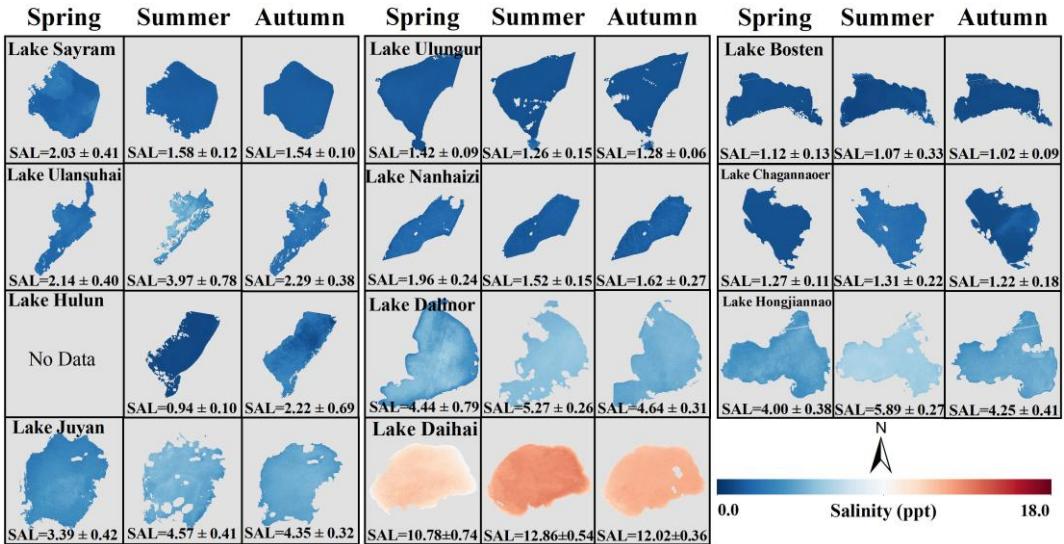



Figure 10: Annual spatial distribution and statistics values (mean ± STD) of water salinity in each lake, with some years lacking salinity maps due to insufficient matching image data.

405 **Figure 11: Seasonal spatial distribution and statistics values (mean ± STD) of water salinity in each lake. Winter data are unavailable due to ice cover.**

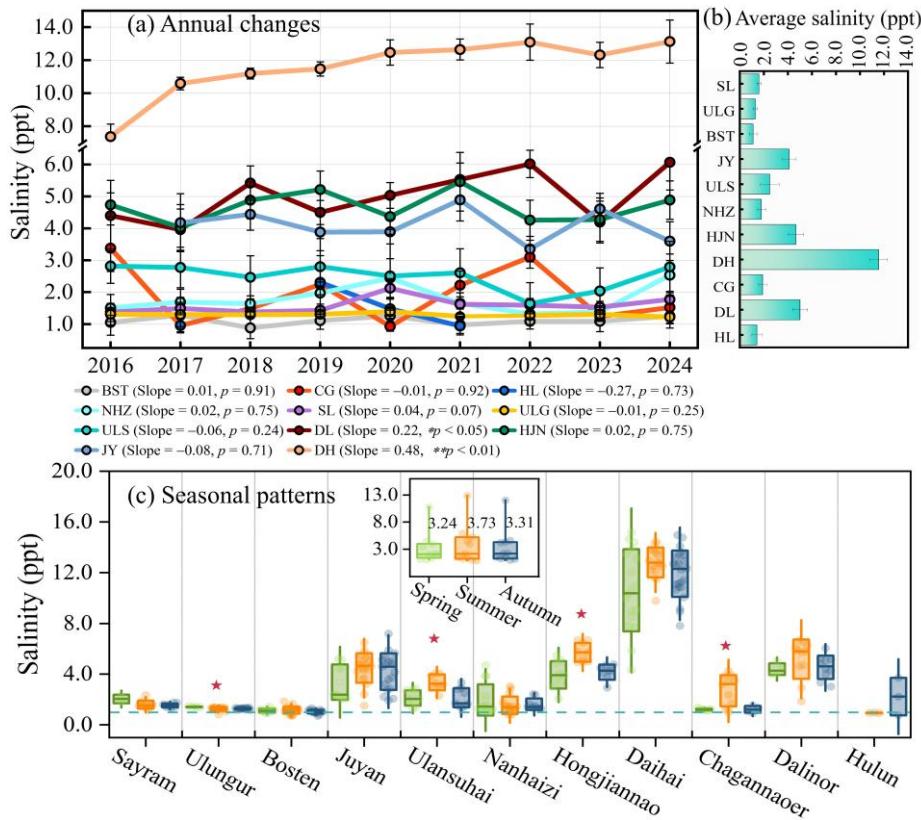



Figure 12: (a) Long-term variations of water salinity in each lake, lakes from Sayram to Hulun are simply denoted as SL to HL, * and ** indicate *p* < 0.05 and *p* < 0.01, (b) multi-year average salinity, and (c) seasonal salinity patterns in each lake, and the star
410 **denotes significant seasonal differences with *p* < 0.05.**

4 Data availability

The IMXL salinity dataset (IMXSAL) was constructed with a three-layer architecture and contains salinity images from 11 lakes during 2016–2024. Under the IMXSAL dataset, 11 data folders named for lakes and 5 tables were included. Each data folder contains 4 subfolders for storing raster data, named according to the data's temporal scale (Table 5). The dataset's total uncompressed size is about 5.35 GB. Breakdown: (1) Salinity raster for 11 lakes at different temporal scales (daily, quarterly, yearly, and all-season average), comprising 673 TIF files, each approximately 8 MB; (2) Excel files, including lake basic information table, dataset metadata table, statistical table (mean and STD), field salinity table, and Supplementary Table S1, each approximately 100 KB. The dataset was archived and publicly accessible via the Zenodo portal: https://doi.org/10.5281/zenodo.18371515 (Deng et al., 2026). Furthermore, to maintain the time-series integrity of the dataset, it will be updated yearly with new Sentinel data and expanded to cover Central Asian lakes to support transboundary water resource management. Dataset versioning follows [Major].[Minor] (e.g., v1.0: 2016–2024; v1.1: 2025 update, January 2026). Annual updates will continue for 10 years (through 2034) or until funding termination, with update notifications posted on the Zenodo repository and Lake-Watershed Science Data Center.

Table 5 The organizational architecture and file naming of the dataset.

Dataset	Folder name	Subfolder name	Filename	Format	Spatial resolution	Temporal resolution
IMXSAL	Lake name	daily	[LakeName]_[YYYYMMDD]_[ProductType]_[Resolution]	TIF	10 m	Daily
		quarterly	[LakeName]_[YYYY]_[Season]_[ProductType]_[Resolution]	TIF	10 m	Quarterly
		yearly	[LakeName]_[YYYY]_[ProductType]_[Resolution]	TIF	10 m	Yearly
		All-season	[LakeName]_[Season]_[ProductType]_[Resolution]	TIF	10 m	Nine-year
			Lake_infor	xlsx	\	\
			IMXSAL_meta	xlsx	\	\
			IMXSAL_MeanSTD_2016-2024	xlsx	\	\
			Field_salinity	xlsx	\	\
			Supplementary Table S1	xlsx	\	\

5 Discussion

5.1 Feasibilities and limitations of algorithms in observing lake salinity

The development of remote sensing algorithms for water salinity has long presented technical challenges, particularly in inland lakes. The lake salinity dataset generated using the proposed framework exhibits a spatial resolution of 10 m, significantly higher than ocean salinity products such as SMOS, SMAP, and Aquarius (> 40 km) (Hu and Zhao, 2022; Jang et al., 2022; Zhang et al., 2023), thereby enabling much greater spatial detail. It also demonstrates higher accuracy (RMSE = 0.60 ppt) compared to even a regional-scale Tibetan Plateau lake salinity product (RMSE = 12.51 g/L) (Xu et al., 2024). To further confirm the scientific validity of the GAM algorithm, it was compared with Bayesian Model Averaging (BMA)

(Hoeting et al., 1999), a method suited for addressing model uncertainty, with the BMA model's key hyperparameters listed in Table 4. The GAM model outperformed the BMA method (N = 84), with lower RMSE (0.60 ppt vs. 0.88 ppt) and MAPE (2.3% vs. 12.6%) in both the test and five-fold CV (Figure 3 and Figure 13). It suggested that the GAM algorithm more effectively handled the nonlinear relationship between the outputs of the mechanism and the ML model, with its average edf of 6.58 (n_splines = 20) implying moderate model complexity while avoiding overfitting.

In Lake Juyan's southern river estuary, the stacking model showed higher RMSE (0.41 ppt) compared to open water (0.19 ppt), due to suspended particulate matter interference. This underestimation accounts for ~5% of total pixels in estuary zones. Subsequent works will focus on integrating Sentinel-2 shortwave infrared (SWIR) band data for correcting this effect (Knaeps et al., 2015). And the spatial heterogeneity stemming from the 3-day matching interval between TIRS and Sentinel data may introduce errors (Jin et al., 2016), which future work will mitigate by combining Sentinel-3 Sea and Land Surface Temperature Radiometer (SLSTR) for temperature correction. Limited by the salinity gradient of the training set (< 35 ppt), the salinity algorithm has insufficient applicability in polysaline (35–50 g/L) and hypersaline lakes (> 50 g/L) and will expand the boundaries of the model by adding high-salinity samples in the future. The K&S model was designed for seawater, and its application to complex ionic lakes (e.g., Lake Chagannaoer) introduces about 5% uncertainty (Figure 6(f7)). This uncertainty can be reduced in the future by developing lake-specific dielectric constant models.

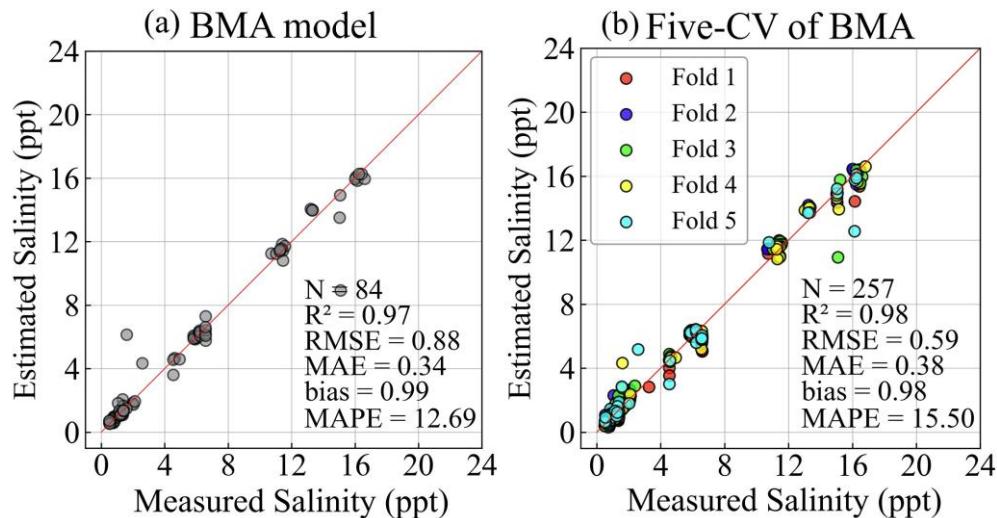



Figure 13: (a)-(b) Performance of the ensemble model constructed by using Bayesian model averaging in testing and five-fold CV.

5.2 Exploring the application potential of salinity dataset

Salinity significantly increased (slope = 0.48 ppt/year, $p < 0.01$) in Lake Daihai (Figure 12a), which was selected for driving force analysis to demonstrate the potential application of the dataset. The generalized linear model was used to quantify the relative contribution of each factor, and the correlation coefficient (r) was applied to examine the relationship between salinity and factors. The lake area accounts for 34.1% of the variation in Daihai salinity, followed by POP (32.5%) and

455 TEMP (17.0%), with other factors contributing less (Figure 14a). Lake area (r = −0.89, *p* < 0.01) and POP (r = −0.71, *p* < 0.05) showed negative correlations with salinity, while TEMP (r = 0.69, *p* < 0.05) and NTL (r = 0.95, *p* < 0.01) exhibited positive correlations, and other variables were not significant (Figure 14b). These indicate that the lake area closely related to the water volume is the dominant factor in the salinity change of Lake Daihai, while climatic warming will aggravate the salinization of inland lakes (Jeppesen et al., 2020). Overall, the dataset supports UN SDG 6.3 (improve water quality) by

460 providing high-resolution salinity data for IMXL—an arid region with 30% of China's saline lakes—enabling local policymakers to track salinization progress (e.g., Lake Daihai's ecological water replenishment with $2.57 \times 10^7 \, \text{m}^3$ in 2024 under SDG 6.6) (Liangcheng County, 2025).

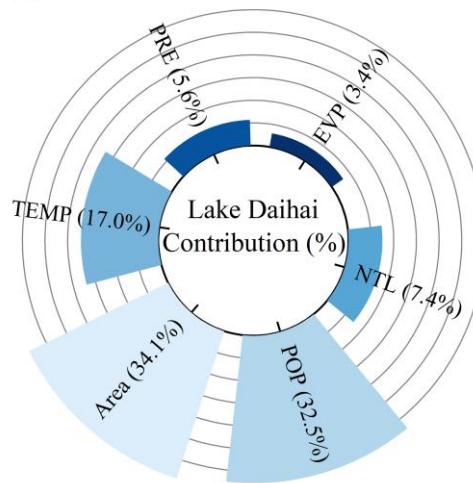
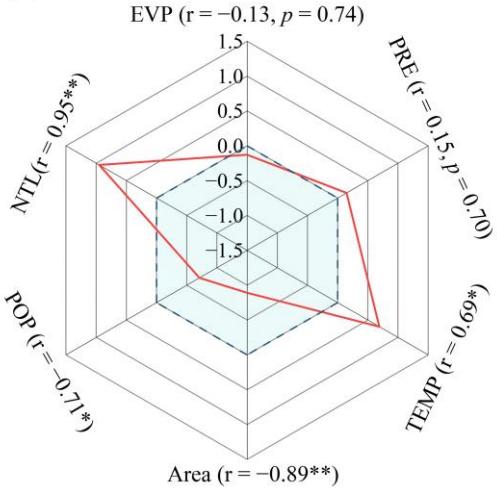



Figure 14: Driver analysis of salinity variations in Lake Daihai from 2016 to 2024, (a) relative contributions of drive factors and (b)
465 **correlations with lake salinity. Light blue colored zones indicate negative correlations. * and ** denote *p* < 0.05 and *p* < 0.01, respectively.**

6 Conclusions

This study proposes an estimation framework for lake salinity by integrating microwave and optical imagery, producing the first 10 m dataset covering 11 lakes in the IMXL (2016–2024), filling a gap in high-resolution salinity data for inland lakes
470 in arid regions. The framework's innovation lies in the development of an ensemble model, which couples microwave physical mechanisms with the nonlinear fitting capabilities of ML to overcome the deficiencies of single-satellite data monitoring in terms of spatial detail or mechanism. Moreover, several technical improvements have been made, the main ones being the adjustment of Elfouhaily spectrum parameters to suit inland lakes, the establishment of a rapidly calculated method for lake surface roughness, the definition of the SPM model for simulating the water backscattering coefficient, and
475 the integration of the mechanistic and ML models via a GAM model. Compared to the products generated from single-satellite data algorithms, this dataset shows improvements in both accuracy and mapping details (RMSE = 0.60 ppt and

MAPE = 2.3%). The histogram validation at the pixel level for all salinity images reconfirmed the satisfactory quality of the dataset. The long-term salinity dataset revealed a spatial pattern of smooth transition from the nearshore to the center and trends with significant increases in Lake Daihai and Lake Dalinor.

480 The proposed integrated algorithms provide methodological references for other lakes and help advance space-based salinity observation missions for inland waters. The created dataset supports salinization prevention (e.g., Lake Daihai water diversion planning) and global lake salinity budget research.

7 Author contributions

MD and RM designed the study; MD wrote the original paper and generated the salinity product; MD and LW collected the
485 validation data; RM, MH, KX, and JX edited this paper; MD developed the algorithm.

8 Competing interests

The contact author has declared that none of the authors has any competing interests.

9 Financial supports

This study was supported by National Natural Science Foundation of China (Grant No. 42361144002 and No. 42371371).
490

References

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaria, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L.: Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, J. Big Data, 8, 53, https://doi.org/10.1186/s40537-021-00444-8, 2021.

Bai, Y., Pan, D., Cai, W.-J., He, X., Wang, D., Tao, B., and Zhu, Q.: Remote sensing of salinity from satellite-derived CDOM in the Changjiang River dominated East China Sea, J. Geophys. Res.-Oceans, 118, 227–243, https://doi.org/10.1029/2012JC008467, 2013.

Bayati, M. and Danesh-Yazdi, M.: Mapping the spatiotemporal variability of salinity in the hypersaline Lake Urmia using Sentinel-2 and Landsat-8 imagery, J. Hydrol., 595, 126032, https://doi.org/10.1016/j.jhydrol.2021.126032, 2021.

Breiman, L.: Random forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Cao, Z., Wang, M., Ma, R., Zhang, Y., Duan, H., Jiang, L., Xue, K., Xiong, J., and Hu, M.: A decade-long chlorophyll-a data record in lakes across China from VIIRS observations, Remote Sensing of Environment, 301, 113953, https://doi.org/10.1016/j.rse.2023.113953, 2024.

Chen, S. and Hu, C.: Estimating sea surface salinity in the northern Gulf of Mexico from satellite ocean color measurements, Remote Sens. Environ., 201, 115–132, https://doi.org/10.1016/j.rse.2017.09.004, 2017.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: KDD'16: PROCEEDINGS OF THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), New York, Web of Science ID: WOS:000485529800092, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Deng, M., Ma, R., Loiselle, S. A., Hu, M., Xue, K., Cao, Z., Wang, L., Lin, C., and Gao, G.: Monitoring Salinity in Inner Mongolian Lakes Based on Sentinel-2 Images and Machine Learning, Remote Sens., 16, 3881, https://doi.org/10.3390/rs16203881, 2024.

Deng, M., Ma, R., Wang, L., Hu, M., Xue, K., and Xiong, J.: A lake salinity dataset produced via microwave and optical imageries [dataset], https://doi.org/10.5281/zenodo.18371515, 2026.

Deng, M., Ma, R., Wang, L., Hu, M., Xue, K., Cao, Z., Xiong, J., and Yu, Z.: A non-optically active lake salinity dataset by satellite remote sensing, Sci. Data, 12, 1324, https://doi.org/10.1038/s41597-025-05686-2, 2025.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P.: Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services, Remote Sensing of Environment, 120, 25–36, https://doi.org/10.1016/j.rse.2011.11.026, 2012.

Elfouhaily, T., Chapron, B., Katsaros, K., and Vandemark, D.: A unified directional spectrum for long and short wind-driven waves, J. Geophys. Res.-Oceans, 102, 15781–15796, https://doi.org/10.1029/97JC00467, 1997.

Feng, L., Hu, C., Chen, X., Tian, L., and Chen, L.: Human induced turbidity changes in Poyang Lake between 2000 and 2010: Observations from MODIS, J. Geophys. Res.-Oceans, 117, C07006, https://doi.org/10.1029/2011JC007864, 2012.

Florencia Gutierrez, M., Tavsanoglu, U. N., Vidal, N., Yu, J., Teixeira-de Mello, F., Cakiroglu, A. I., He, H., Liu, Z., and Jeppesen, E.: Salinity shapes zooplankton communities and functional diversity and has complex effects on size structure in lakes, Hydrobiologia, 813, 237–255, https://doi.org/10.1007/s10750-018-3529-8, 2018.

Gao, Z., Li, X., Zuo, L., Zou, B., Wang, B., and Wang, W. J.: Unveiling soil salinity patterns in soda saline-alkali regions using Sentinel-2 and SDGSAT-1 thermal infrared data, Remote Sensing of Environment, 322, 114708, https://doi.org/10.1016/j.rse.2025.114708, 2025.

Gleeson, T., Wang-Erlandsson, L., Porkka, M., Zipper, S. C., Jaramillo, F., Gerten, D., Fetzer, I., Cornell, S. E., Piemontese, L., Gordon, L. J., Rockström, J., Oki, T., Sivapalan, M., Wada, Y., Brauman, K. A., Flörke, M., Bierkens, M. F. P., Lehner, B., Keys, P., Kummu, M., Wagener, T., Dadson, S., Troy, T. J., Steffen, W., Falkenmark, M., and Famiglietti, J. S.:

Illuminating water cycle modifications and Earth system resilience in the Anthropocene, Water Resources Research, 56, e2019WR024957, https://doi.org/10.1029/2019WR024957, 2020.

Gonzalez-Gambau, V., Olmedo, E., Turiel, A., Gonzalez-Haro, C., Garcia-Espriu, A., Martinez, J., Alenius, P., Tuomi, L., Catany, R., Arias, M., Gabarro, C., Hoareau, N., Umbert, M., Sabia, R., and Fernandez, D.: First SMOS Sea Surface Salinity dedicated products over the Baltic Sea, Earth Syst. Sci. Data, 14, 2343–2368, https://doi.org/10.5194/essd-14-2343-2022, 2022.

Guo, H., Zhu, X., Huang, J. J., Zhang, Z., Tian, S., and Chen, Y.: An enhanced deep learning approach to assessing inland lake water quality and its response to climate and anthropogenic factors, J. Hydrol., 620, 129466, https://doi.org/10.1016/j.jhydrol.2023.129466, 2023.

Hammer, U. T.: Saline lake ecosystems of the world, Springer Science & Business Media, 1986.

Han, Y., Zhai, Y., Guo, M., Cao, X., Lu, H., Li, J., Wang, S., and Yue, W.: Hydrochemical and Isotopic Characterization of the Impact of Water Diversion on Water in Drainage Channels, Groundwater, and Lake Ulansuhai in China, WATER, 13, https://doi.org/10.3390/w13213033, 2021.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T.: Bayesian model averaging: A tutorial, Stat. Sci., 14, 382–401, https://doi.org/10.1214/ss/1009212519, 1999.

Hu, C.: A novel ocean color index to detect floating algae in the global oceans, Remote Sens. Environ., 113, 2118–2129, https://doi.org/10.1016/j.rse.2009.05.012, 2009.

Hu, R. and Zhao, J.: Sea surface salinity variability in the western subpolar North Atlantic based on satellite observations, Remote Sens. Environ., 281, 113257, https://doi.org/10.1016/j.rse.2022.113257, 2022.

Hwang, P. A., Burrage, D. M., Wang, D. W., and Wesson, J. C.: An Advanced Roughness Spectrum for Computing Microwave L-Band Emissivity in Sea Surface Salinity Retrieval, IEEE Geosci. Remote Sens. Lett., 8, 547–551, https://doi.org/10.1109/LGRS.2010.2091393, 2011.

Jang, E., Kim, Y. J., Im, J., Park, Y.-G., and Sung, T.: Global sea surface salinity via the synergistic use of SMAP satellite and HYCOM data based on machine learning, Remote Sens. Environ., 273, 112980, https://doi.org/10.1016/j.rse.2022.112980, 2022.

Jeppesen, E., Brucet, S., Naselli-Flores, L., Papastergiadou, E., Stefanidis, K., Noges, T., Noges, P., Attayde, J. L., Zohary, T., Coppens, J., Bucak, T., Menezes, R. F., Sousa Freitas, F. R., Kernan, M., Sondergaard, M., and Beklioglu, M.: Ecological impacts of global warming and water abstraction on lakes and reservoirs due to changes in water level and related changes in salinity, Hydrobiologia, 750, 201–227, https://doi.org/10.1007/s10750-014-2169-x, 2015.

Jeppesen, E., Beklioglu, M., Ozkan, K., and Akyurek, Z.: Salinization Increase due to Climate Change Will Have Substantial Negative Effects on Inland Waters: A Call for Multifaceted Research at the Local and Global Scale, Innovation-Amsterdam, 1, 100030, https://doi.org/10.1016/j.xinn.2020.100030, 2020.

Jiang, X., Fan, C., Liu, K., Chen, T., Cao, Z., and Song, C.: Centenary covariations of water salinity and storage of the largest lake of Northwest China reconstructed by machine learning, J. Hydrol., 612, 128095, https://doi.org/10.1016/j.jhydrol.2022.128095, 2022.

Jiang, X., Liu, C., Hu, Y., Shao, K., Tang, X., Zhang, L., Gao, G., and Qin, B.: Climate-induced salinization may lead to increased lake nitrogen retention, Water Res., 228, 119354, https://doi.org/10.1016/j.watres.2022.119354, 2023.

Jin, X., Zhu, Q., He, X., Chen, P., Wang, D., Hao, Z., and Huang, H.: Impact of sea surface temperature on satellite retrieval of sea surface salinity, in: REMOTE SENSING OF THE OCEAN, SEA ICE, COASTAL WATERS, AND LARGE WATER REGIONS 2016, https://doi.org/10.1117/12.2240841, 2016.

Johnson, J. T. and Zhang, M.: Theoretical study of the small slope approximation for ocean polarimetric thermal emission, IEEE Trans. Geosci. Remote Sensing, 37, 2305–2316, https://doi.org/10.1109/36.789627, 1999.

Kaushal, S. S., Likens, G. E., Pace, M. L., Reimer, J. E., Maas, C. M., Galella, J. G., Utz, R. M., Duan, S., Kryger, J. R., Yaculak, A. M., Boger, W. L., Bailey, N. W., Haq, S., Wood, K. L., Wessel, B. M., Park, C. E., Collison, D. C., Aisin, B. Y. 'aaqob, Gedeon, T. M., Chaudhary, S. K., Widmer, J., Blackwood, C. R., Bolster, C. M., Devilbiss, M. L., Garrison, D. L., Halevi, S., Kese, G. Q., Quach, E. K., Rogelio, C. M. P., Tan, M. L., Wald, H. J. S., and Woglo, S. A.: Freshwater salinization syndrome: from emerging global problem to managing risks, Biogeochemistry, 154, 255–292, https://doi.org/10.1007/s10533-021-00784-w, 2021.

Khenchaf, A.: Bistatic scattering and depolarization by randomly rough surfaces: application to the natural rough surfaces in X-band, Waves Random Media, 11, 61–89, https://doi.org/10.1088/0959-7174/11/2/301, 2001.

Klein, L. and Swift, C.: Improved Model for Dielectric-Constant of Sea-Water at Microwave-Frequencies, IEEE Trans. Antennas Propag., 25, 104–111, https://doi.org/10.1109/TAP.1977.1141539, 1977.

Knaeps, E., Ruddick, K. G., Doxaran, D., Dogliotti, A. I., Nechad, B., Raymaekers, D., and Sterckx, S.: A SWIR based algorithm to retrieve total suspended matter in extremely turbid waters, Remote Sensing of Environment, 168, 66–79, https://doi.org/10.1016/j.rse.2015.06.022, 2015.

Ladwig, R., Rock, L. A., and Dugan, H. A.: Impact of salinization on lake stratification and spring mixing, Limnol. Oceanogr. Lett., 8, 93–102, https://doi.org/10.1002/lol2.10215, 2023.

Le Vine, D. M., Lang, R. H., Zhou, Y., Dinnat, E. P., and Meissner, T.: Status of the Dielectric Constant of Sea Water at L-Band for Remote Sensing of Salinity, IEEE Trans. Geosci. Remote Sensing, 60, 4210114, https://doi.org/10.1109/TGRS.2022.3207944, 2022.

LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, nature, 521, 436–444, 2015.

Lee, Z., Shang, S., Lin, G., Chen, J., and Doxaran, D.: On the modeling of hyperspectral remote-sensing reflectance of high-sediment-load waters in the visible to shortwave-infrared domain, Appl. Optics, 55, 1738–1750, https://doi.org/10.1364/AO.55.001738, 2016.

Liangcheng County: The surface area of Lake Daihai in Liangcheng county has reached a 10-year high, 2025.

Liu, C., Wu, F., Jiang, X., Hu, Y., Shao, K., Tang, X., Qin, B., and Gao, G.: Climate Change Causes Salinity To Become Determinant in Shaping the Microeukaryotic Spatial Distribution among the Lakes of the Inner Mongolia-Xinjiang Plateau, Microbiol. Spectr., 11, e03178-22, https://doi.org/10.1128/spectrum.03178-22, 2023a.

Liu, C., Zhu, L., Wang, J., Ju, J., Ma, Q., and 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, 16, 2644–2659, https://doi.org/10.1080/17538947.2023.2233469, 2023b.

Liu, D., Shi, K., Chen, P., Yan, N., Ran, L., Kutser, T., Tyler, A. N., Spyrakos, E., Woolway, R. I., Zhang, Y., and Duan, H.: Substantial increase of organic carbon storage in Chinese lakes, Nat. Commun., 15, 8049, https://doi.org/10.1038/s41467-024-52387-2, 2024.

Liu Y., Bao A., and Chen X.: Measuring salinity of low salinity lake by optical remote sensing: A case study of Bosten Lake, Journal of Remote Sensing, 18, 902–911, 2014.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I.: From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell., 2, 56–67, https://doi.org/10.1038/s42256-019-0138-9, 2020.

Ma, R., Yang, G., Duan, H., Jiang, J., Wang, S., Feng, X., Li, A., Kong, F., Xue, B., Wu, J., and Li, S.: China's lakes at present: Number, area and spatial distribution, Sci. China-Earth Sci., 54, 283–289, https://doi.org/10.1007/s11430-010-4052-6, 2011.

Ma, W., Liu, G., Yu, Y., and Du, Y.: Roughness correction method for salinity remote sensing using combined active/passive observations, Acta Oceanol. Sin., 40, 189–195, https://doi.org/10.1007/s13131-021-1744-z, 2021.

McGrath, G. S., Huntley, B., and Venarsky, M. P.: Variation of salinity with water level in shallow lakes is complex and rich in information, J. Hydrol., 660, 133347, https://doi.org/10.1016/j.jhydrol.2025.133347, 2025.

620 Meissner, T., Wentz, F. J., and Ricciardulli, L.: The emission and scattering of L-band microwave radiation from rough ocean surfaces and wind speed measurements from the Aquarius sensor, J. Geophys. Res.-Oceans, 119, 6499–6522, https://doi.org/10.1002/2014JC009837, 2014.

Mobley, C. D.: Estimation of the remote-sensing reflectance from above-surface measurements, Appl. Optics, 38, 7442–7455, https://doi.org/10.1364/AO.38.007442, 1999.

625 Mueller, J. L., Morel, A., Frouin, R., Davis, C., Arnone, R., Carder, K., Lee, Z. P., Steward, R. G., Hooker, S., and Mobley, C. D.: Ocean Optics Protocols For Satellite Ocean Color Sensor Validation, Revision 4. Volume III: Radiometric Measurements and Data Analysis Protocols., 2003.

Muñoz Sabater, J.: ERA5-Land monthly averaged data from 1950 to present., https://doi.org/10.24381/cds.68d2bb30, 2019.

Peake, W.: The Interaction of Electromagnetic Waves with Some Natural Surfaces, 111 pp., 1959.

630 Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S.: High-resolution mapping of global surface water and its long-term changes, Nature, 540, 418–422, https://doi.org/10.1038/nature20584, 2016.

Reul, N., Saux-Picart, S., Chapron, B., Vandemark, D., Tournadre, J., and Salisbury, J.: Demonstration of ocean surface salinity microwave measurements from space using AMSR-E data over the Amazon plume, Geophys. Res. Lett., 36, L13607, https://doi.org/10.1029/2009GL038860, 2009.

635 Reul, N., Grodsky, S. A., Arias, M., Boutin, J., Catany, R., Chapron, B., D'Amico, F., Dinnat, E., Donlon, C., Fore, A., Fournier, S., Guimbard, S., Hasson, A., Kolodziejczyk, N., Lagerloef, G., Lee, T., Le Vine, D. M., Lindstrom, E., Maes, C., Mecklenburg, S., Meissner, T., Olmedo, E., Sabia, R., Tenerelli, J., Thouvenin-Masson, C., Turiel, A., Vergely, J. L., Vinogradova, N., Wentz, F., and Yueh, S.: Sea surface salinity estimates from spaceborne L-band radiometers: An overview of the first decade of observation (2010-2019), Remote Sens. Environ., 242, 111769, 640 https://doi.org/10.1016/j.rse.2020.111769, 2020.

Rimmer, A., Boger, M., Aota, Y., and Kumagai, M.: A lake as a natural integrator of linear processes: Application to Lake Kinneret (Israel) and Lake Biwa (Japan), J. Hydrol., 319, 163–175, https://doi.org/10.1016/j.jhydrol.2005.07.018, 2006.

Rusuli, Y., Li, L., Li, F., and Eziz, M.: Water-level regulation for freshwater management of Bosten Lake in Xinjiang, China, Water Sci. Technol.-Water Supply, 16, 828–836, https://doi.org/10.2166/ws.2016.002, 2016.

645 Rusydi, A. F.: Correlation between conductivity and total dissolved solid in various type of water: A review, in: GLOBAL COLLOQUIUM ON GEOSCIENCES AND ENGINEERING 2017, 1st Global Colloquium on GeoSciences and Engineering (GCGE), Bristol, journalAbbreviation: IOP Conf. Ser. Earth Envir. Sci.number-of-pages: 6collection-title: IoP Conference Series-Earth and Environmental ScienceWeb of Science ID: WOS:000445923300019, 012019, https://doi.org/10.1088/1755-1315/118/1/012019, 2018.

650 Shareef, M. A., Toumi, A., and Khenchaf, A.: Estimating Of Water Quality Parameters Using SAR And Thermal Microwave Remote Sensing Data, in: 2016 2ND INTERNATIONAL CONFERENCE ON ADVANCED TECHNOLOGIES FOR SIGNAL AND IMAGE PROCESSING (ATSIP), 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), New York, Web of Science ID: WOS:000390841700108, 586–590, 2016.

Taillade, T., Engdahl, M., and Fernandez, D.: Can We Retrieve Sea Surface Salinity With Polarimetric Radar 655 Measurements?, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, 20, https://doi.org/10.1109/LGRS.2023.3286436, 2023.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I. N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., and Rostan, F.: GMES Sentinel-1 mission, Remote Sens. Environ., 120, 9–24, 660 https://doi.org/10.1016/j.rse.2011.05.028, 2012.

Urquhart, E. A., Zaitchik, B. F., Hoffman, M. J., Guikema, S. D., and Geiger, E. F.: Remotely sensed estimates of surface salinity in the Chesapeake Bay: A statistical approach, Remote Sens. Environ., 123, 522–531, https://doi.org/10.1016/j.rse.2012.04.008, 2012.

665 Vanhellemont, Q.: Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives, Remote Sens. Environ., 225, 175–192, https://doi.org/10.1016/j.rse.2019.03.010, 2019.

Wang, D., Huang, Y., and Yang, H.: Seasonal Differences of Lake Bacterial Community Structures and Their Driving Mechanisms in the Northeastern of the Qinghai-Tibet Plateau, Journal of Lake Sciences, 35, 267–282, 2023a.

Wang, S., Jiang, X., Spyrakos, E., Li, J., Mcglinchey, C., Constantinescu, A. M., and Tyler, A. N.: Water color from Sentinel-2 MSI data for monitoring large rivers: Yangtze and Danube, Geo-Spat. Inf. Sci., 670 https://doi.org/10.1080/10095020.2023.2258950, 2023b.

Williamson, C. E., Saros, J. E., and Schindler, D. W.: Sentinels of Change, Science, 323, 887–888, https://doi.org/10.1126/science.1169443, 2009.

Wurtsbaugh, W. A., Miller, C., Null, S. E., DeRose, R. J., Wilcock, P., Hahnenberger, M., Howe, F., and Moore, J.: Decline of the world's saline lakes, Nat. Geosci., 10, 816–821, https://doi.org/10.1038/ngeo3052, 2017.

675 Xie, D., Chen, K.-S., and Yang, X.: Effects of Wind Wave Spectra on Radar Backscatter From Sea Surface at Different Microwave Bands: A Numerical Study, IEEE Trans. Geosci. Remote Sensing, 57, 6325–6334, https://doi.org/10.1109/TGRS.2019.2905558, 2019.

Xu, P., Liu, K., Shi, L., and Song, C.: Machine learning modeling reveals the spatial variations of lake water salinity on the endorheic Tibetan Plateau, J. Hydrol.-Reg. Stud., 56, 102042, https://doi.org/10.1016/j.ejrh.2024.102042, 2024.

680 Xue, K., Ma, R., Wang, M., Wei, X., Liu, H., Hu, M., Jiang, L., Shen, M., and Cao, Z.: A state-of-art algorithm to retrieve particulate organic carbon concentration in optically complex waters via multiple satellite missions, Remote Sens. Environ., 329, 114914, https://doi.org/10.1016/j.rse.2025.114914, 2025.

Yee, T. W. and Wild, C. J.: Vector generalized additive models, J. R. Stat. Soc. Ser. B-Stat. Methodol., 58, 481–493, 1996.

Yihdego, Y. and Webb, J.: Modelling of seasonal and long-term trends in lake salinity in southwestern Victoria, Australia, J. 685 Environ. Manage., 112, 149–159, https://doi.org/10.1016/j.jenvman.2012.07.002, 2012.

Young, I. R. and Verhagen, L. A.: The growth of fetch limited waves in water of finite depth. Part 1. Total energy and peak frequency, Coastal Engineering, 29, 47–78, https://doi.org/10.1016/S0378-3839(96)00006-3, 1996.

Zhang, L., Zhang, Y., and Yin, X.: Aquarius sea surface salinity retrieval in coastal regions based on deep neural networks, Remote Sens. Environ., 284, 113357, https://doi.org/10.1016/j.rse.2022.113357, 2023.

690 Zhao, J. and Temimi, M.: An Empirical Algorithm for Retreiving Salinity in the Arabian Gulf: Application to Landsat-8 Data, in: 2016 IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 36th IEEE International Geoscience and Remote Sensing Symposium (IGARSS), New York, Web of Science ID: WOS:000388114604153, 4645–4648, https://doi.org/10.1109/IGARSS.2016.7730212, 2016.

Zheng, X., Zhang, M., Xu, C., and Li, B.: China Salt Lakes Record, Beijing: SciencePress, 2002.

695 Zou, J., Sun, B., Zhao, S., Pan, X., and Ye, F.: Analysis of changes in water chemistry characteristics and influencing factors of three major lakes in Inner Mongolia in the last decade, Journal of Environmental Engineering Technology, 14, 1247–1259, 2024.