

Large-scale algal biomass mapping from OLCI observations for lakes in both basins of the Changjiang (Yangtze) River and the Huaihe River, China*

Zhengyang YU^{1,2}, Ronghua MA^{2,3,**}, Chuanmin HU⁴, Minqi HU², Kun XUE², Junfeng XIONG²

¹ School of Geography and Planning, Huaiyin Normal University, Huai'an 223300, China

² Key Laboratory of Lake and Watershed Science for Water Security, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China

³ University of Chinese Academy of Sciences, Nanjing 211135, China

⁴ College of Marine Science, University of South Florida, St. Petersburg 33701, USA

Received Nov. 27, 2025; accepted in principle Mar. 22, 2026; accepted for publication May 9, 2026

© Chinese Society for Oceanology and Limnology, Science Press and Springer-Verlag GmbH Germany, part of Springer Nature 2026

Abstract While existing studies on the remote estimation of algal biomass have focused predominantly on case-specific analyses, systematic assessments of large-scale spatial patterns remain underexplored. To address this knowledge gap, we developed a novel classification-based framework utilizing Ocean and Land Color Instrument (OLCI) satellite imagery, enabling the first multiyear (2016–2023) mapping of algal biomass within the euphotic depth (B_{eu}) across 24 lakes in the middle-lower Changjiang (Yangtze) and Huaihe rivers basins, China. The methodology employed the adjusted Davies-Bouldin index (ADBI) as a metric for assessing classification efficacy and was validated through the match-up dataset (root mean square error=13.24 mg/m², mean absolute percentage error=36.3%, $n=40$). Spatiotemporal analysis revealed that the total algal biomass within the euphotic zone (total B_{eu}) of the hyper-sized lakes was lower in 2018 and 2023 than in the other years and remained relatively stable for four years (2019–2022). For large lakes, the interannual total B_{eu} showed a distinct single-peak trend over the study period, with the highest value occurring in 2019. In contrast, medium-sized lakes exhibited relatively subtle interannual fluctuations without a clear temporal pattern. Our work reveals previously undocumented spatial heterogeneity in the distribution of algal biomass and provides a foundational dataset for monitoring eutrophication dynamics under climate change.

Keyword: algal biomass; Ocean and Land Color Instrument (OLCI) satellite imagery; eutrophication

1 INTRODUCTION

Many lakes have been or are currently grappling with phytoplankton blooms, a pressing environmental problem that has catastrophic impacts on food and drinking water supplies, biodiversity, and economic sustainability (Hu et al., 2022, 2024). Excessive nitrogen, phosphorus, and other nutrients in these water bodies fuel the proliferation of phytoplankton, leading to more frequent blooms that severely imperil the normal functioning of aquatic ecosystems (Ho et al., 2019). Even less encouraging is the increased frequency, intensity, and duration of cyanobacterial bloom events (Ma et al., 2023). To

evaluate the water quality of lakes effectively, a series of indicators has been proposed, such as the concentration of chlorophyll *a* (Chl *a*) and the bloom area.

However, the indicators are limited to the water surface and thus fail to offer a comprehensive interpretation from the perspective of the entire water column (Xue et al., 2017). Recent studies

* Supported by the National Natural Science Foundation of China (Nos. 42361144002, 42471353, 42301406, 42371371) and the Project of Huai'an Municipal Bureau of Science and Technology (No. HAB2025015)

** Corresponding author: rhma@niglas.ac.cn

challenge the reliability of using surface measurements alone because of the vertical migration of phytoplankton (Xue et al., 2017; Qi et al., 2018). Cyanobacteria possess a remarkable ability to regulate their position in the water column through gas vesicles, enabling them to optimize their access to light and CO_2 for physiological activities (Stockwell et al., 2020). Their vertical migration patterns are governed by both this unique physiological structure and external environmental factors. Wind conditions play a particularly crucial role: under gentle breezes (<3 m/s), cyanobacteria tend to move upward and accumulate at the surface, whereas stronger winds (>3 m/s) cause them to sink (Xu et al., 2021). Moreover, wind-induced waves can alter water column hydrodynamics, triggering vertical migration and restructuring the vertical distribution patterns of phytoplankton communities (Wu et al., 2013, 2015). Therefore, the surface Chl-*a* concentrations and bloom areas tend to fluctuate within a short period of time. The vertical migration of algae fundamentally challenges the reliability of surface information, particularly surface Chl-*a* concentrations or bloom areas, as indicators for eutrophication assessment (Silulwane et al., 2001). To address the uncertainties resulting from vertical migration, column-integrated algal biomass (CAB), which accounts for algal biomass throughout a specific water depth, has emerged as a more comprehensive alternative.

CAB quantifies the total Chl-*a* weight per unit area, expressed in mg/m^2 (Lai et al., 2023). Spatially, a distinct discrepancy is evident between the distribution patterns of CAB and surface Chl *a* (Li et al., 2020; Cao et al., 2022a; Lai et al., 2024). The temporal variations in CAB and surface Chl *a* also exhibit inconsistency (Cao et al., 2022a; Yu et al., 2025). Yu et al. (2025) highlighted that the annual trends of CAB and Chl *a* followed distinct patterns. However, understanding the algal biomass dynamics of Chinese lakes remains challenging because of the lack of long-term and large-scale field data. Typically, Chl-*a* vertical profiles are collected through field measurements, but this method provides limited information and is labor intensive. Remote sensing technology is highly advantageous for algal biomass estimation, as its ability to conduct large-scale, simultaneous observations offers essential support. Methods for algal biomass estimation via remote sensing can be categorized into two main approaches: indirect and direct schemes. The indirect schemes further comprise two subtypes: result-oriented methods (Li et al., 2017;

Yu et al., 2024) and process-oriented methods (Liu et al., 2021; Li et al., 2022). The result-oriented method relies on establishing empirical functions to determine depth-dependent parameters without considering the actual vertical distribution of Chl-*a* profiles. First, statistical modeling is performed based on measured surface Chl-*a* concentrations and algal biomass at various depths, and the coefficients of these statistical models are recorded; then, using empirical models, regression analysis on water depth and the coefficients of these models is performed; finally, the algal biomass is estimated based on the water depth data. The process-oriented method prioritizes the accurate fitting of Chl-*a* vertical profiles. First, the method establishes a classifier based on surface Chl-*a* concentrations to predict vertical distribution patterns of Chl-*a* profiles; next, for each vertical distribution pattern, it determines the coefficients of the corresponding fitting function; then, it estimates these coefficients using surface Chl-*a* concentrations; finally, it calculates the algal biomass using an integral based on water depth data. Once the vertical distribution is determined, the parameters that govern the profile shape can be derived through empirical functions. The direct scheme is not as multistep a scheme as the indirect scheme is. By establishing a direct relationship between remote sensing reflectance (or remote sensing indices) and measured algal biomass, the direct scheme eliminates the need for complex intermediate steps (Hu et al., 2021). However, owing to insufficient Chl-*a* vertical profiles, earlier studies have failed to accurately represent the complex optical properties and simplified mechanistic descriptions of lake water across diverse geographical settings, thereby restricting the scientific community's ability to understand algal biomass on a broader scale.

Currently, studies on algal biomass in Chinese lakes are confined to local areas, and large-scale research is lacking. In this paper, we propose a direct scheme to map B_{eu} (the algal biomass within the euphotic zone) for lakes located in the middle and lower reaches of the Changjiang (Yangtze) and Huaihe rivers basins and to analyze the annual trends of B_{eu} . Given the substantial variability in optical characteristics among the studied lakes, it is challenging to estimate algal biomass using a single model trained on a combined dataset. To address this issue, this study applies spectral clustering to optically classify water bodies, enabling the development of group-specific models and thereby improving the accuracy of algal biomass estimation.

The proposed direct scheme comprises spectral clustering (Von Luxburg, 2007) and a generalized additive model (GAM) (Hastie and Tibshirani, 2014). GAM is employed for algal biomass retrieval because it leverages penalized splines to regulate model complexity, striking a balance between avoiding overfitting and achieving robust fitting performance. This study presents a large-scale mapping of algal biomass within the euphotic zone across 24 lakes in the middle and lower reaches of the Changjiang and Huaihe rivers basins, covering the period 2016–2023.

2 MATERIAL AND METHOD

2.1 Study area

The Changjiang and Huaihe rivers basins in China have long been a focus of environmental management efforts. These regions are heavily populated, have complex river networks, and are China's most economically developed areas, contributing significantly to the country's GDP.

Twenty-four lakes with areas larger than 100 km² located in the middle and lower reaches of the Changjiang River and Huaihe rivers were included and analyzed in this study (Fig.1). The lakes were grouped by size: medium-sized (100–500 km²), large (500–1 000 km²), and hyper-sized lakes

(>1 000 km²). More detailed information about these lakes is provided in Supplementary Table S1.

2.2 Field survey and laboratory measurement

Data-gathering cruises were launched from 2013 to 2024 on clear, cloudless, and windless days in case of undesirable measurements of water-leaving reflectance (R_{rs}), with 146 locations involved (Supplementary Table S2). R_{rs} was measured via an ASD field spectrometer (FieldSpec Pro Dual VNIR, Analytical Spectra Devices, Inc.) (Xue et al., 2019). We quantified the water surface radiance ($L_{sf}(\lambda)$) and the sky radiance ($L_{sky}(\lambda)$) at a 135° azimuth relative to the sun and a nadir viewing angle of 40°, alongside the radiance of a reference panel ($L_p(\lambda)$) to determine $R_{rs}(\lambda)$ (Mueller et al., 2003). The water samples at each location were collected at discrete depths (0, 0.1, 0.3, 0.5, 0.7, 1.0, 1.5, 2.0, and 3.0 m) with 1-L Niskin bottles. The wind speed, coordinates, and Secchi disk depth (SDD) were recorded concurrently using a portable anemometer, a global positioning system (GPS), and a Secchi disk, respectively. The water samples were transported to the laboratory and stored under light-proof conditions for subsequent analysis. The water samples were filtered through Whatman GF/F filters with a pore size of 0.70 µm. Chl *a* was extracted

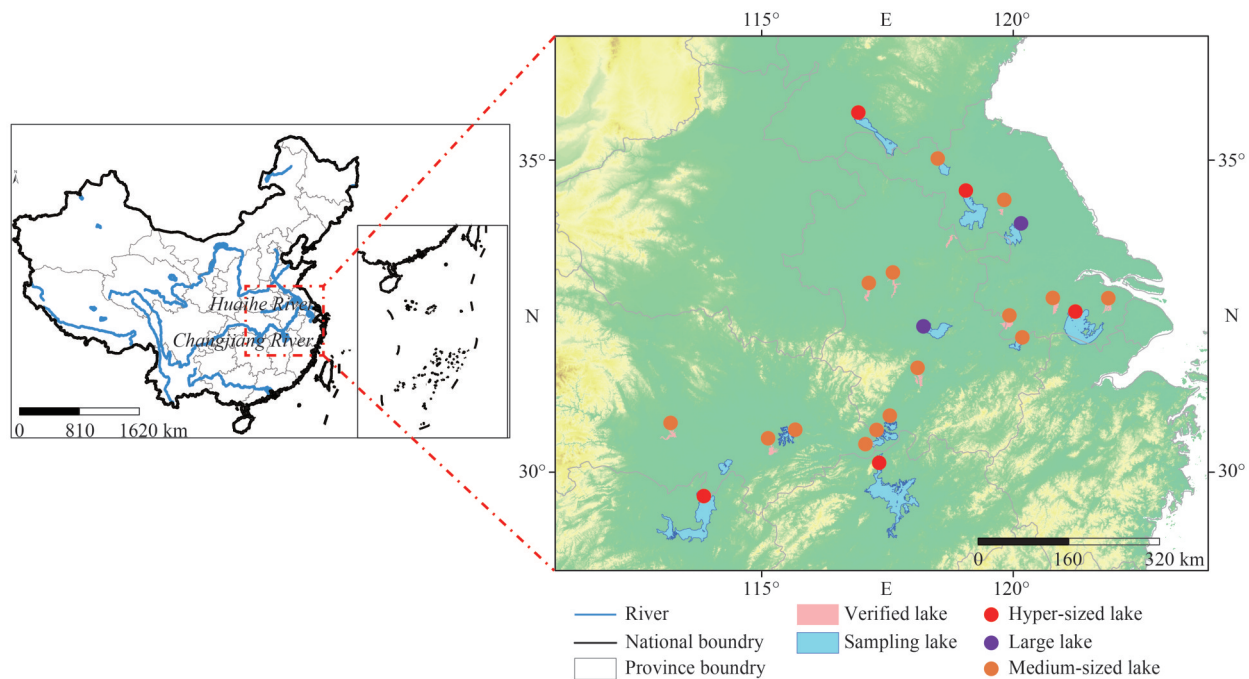


Fig.1 The spatial distribution of lakes, each with an area greater than 100 km²

The base map of China used in this figure is sourced from the Standard Map Service System of the Ministry of Natural Resources of China, with the map review No. GS(2022)4310.

with 90% acetone and subsequently quantified with a Shimadzu UV-2700 spectrophotometer. The absorption coefficient of colored dissolved organic matter ($a_{\text{CDOM}}(\lambda)$) was measured via a Shimadzu UV-3600 spectrophotometer. Suspended particulate inorganic matter (SPIM) was measured using the same filters used for Chl-*a* extraction, following combustion at 550 °C for 4 h to remove organic matter. SPIM concentration was calculated as the weight difference before and after combustion normalized by the filtered water volume. The specifics regarding the measurement methodology are detailed in community-accepted protocols (Ma et al., 2006). The statistics of the water quality parameters are given in Table 1.

The collected data were randomly partitioned into a calibration dataset ($n=110$) and a validation dataset ($n=36$). In accordance with NASA's criterion (Bailey and Werdell, 2006), an independent match-up dataset ($n=40$) was constructed. Currently, if the time interval between satellite overpasses and field surveys is less than 3 h, the two are considered a match-up pair. A 3×3 pixel window centered on each in-situ sampling location was extracted from the OLCI imagery, and the median R_{rs} value of the valid pixels within the window was used to represent the satellite-derived R_{rs} for that station.

2.3 Satellite data processing

The Ocean and Land Color Instrument (OLCI)

Table 1 Statistics of the water quality parameters including SDD, Chl *a*, $a_g(440)$, and SPIM for the studied lakes

Lake	SDD (cm)	Chl <i>a</i> (μg/L)	$a_g(440)$ (1/m)	SPIM (mg/L)
Gaoyou Lake	16.12±7.77	35.93±16.13	1.68±0.60	29.14±9.00
Hongze Lake	15.95±9.84	31.90±14.03	2.06±2.59	78.67±56.23
Luoma Lake	62.0±12.0	11.30±3.98	0.82±.80	21.20±2.80
Nansi Lake	30.10±4.90	79.23±9.80	0.66±0.42	28.50±30.10
Taihu Lake	56.17±12.77	25.73±18.11	0.89±0.33	25.84±9.43
Chaohu Lake	23.26±7.07	33.96±9.43	0.87±0.71	63.50±16.56
Dongting Lake	68.20±8.78	8.90±8.78	0.43±0.50	18.25±13.35
Poyang Lake	35.00±13.69	7.00±13.35	0.49±0.20	29.12±20.63
Nanyi Lake	62.15±9.58	8.96±3.10	0.53±0.24	10.21±7.79
Liangzi Lake	41.66±8.50	24.21±6.41	0.37±0.18	10.37±3.85
Honghu Lake	37.50±15.62	33.32±23.36	0.56±0.30	23.18±7.66
Longgan Lake	7.50±3.60	22.41±1.27	0.39±0.11	15.80±6.55
Pohu Lake	20.00±0	12.70±0	0.22±0	12.30±0
Huangda Lake	10.00±0	34.21±0	0.28±0	15.60±0

onboard Sentinel-3A/B, launched by the European Space Agency (ESA), has been in operation since 2016. With 21 bands from 400 to 1 020 nm, a 300-m spatial resolution, and an approximately two-day revisit cycle, OLCI offers improved spectral resolution and signal-to-noise ratio compared to previous sensors such as MERIS and MODIS. Its bands centered at 620, 665, and 681 nm are particularly effective for retrieving optical properties in complex aquatic environments, supporting accurate algal biomass estimation in this study. L1B products were downloaded from <https://oceancolor.gsfc.nasa.gov>, and calibrated using the dark spectrum function (DSF) to yield atmospherically corrected reflectance. Compared to C2RCC and POLYMER, DSF has been demonstrated to perform more robustly in optically complex waters, particularly in highly turbid lakes (Vanhellemont, 2019, 2020). Afterward, we visually checked each image to exclude those with visible cloud contamination. The total number of cloud-free OLCI images used in this study was 3 616 (annual distribution is shown in Fig.2), with seasonal distribution as follows: spring: 958, summer: 942, autumn: 950, and winter: 766. Although the number of observations is somewhat lower in winter, each season still contains a sufficient number of valid retrievals (minimum 766) to ensure reliable seasonal averages.

Leveraging the existing lake-boundary dataset furnished by Ma et al. (2011), we decreased the boundary of each lake by 3 pixels. This was done to mitigate the influence of near-shore mixed pixels. Additionally, for lakes experiencing notable boundary variations, such as Poyang Lake and Dongting Lake, the normalized difference water index (NDWI) (Xu, 2006) was employed to extract their boundaries, which is expressed as follows:

$$\text{NDWI} = \frac{R_{\text{GREEN}} - R_{\text{NIR}}}{R_{\text{GREEN}} + R_{\text{NIR}}}, \quad (1)$$

where R_{GREEN} and R_{NIR} are the reflectance of the green band (560 nm) and near-infrared band (865 nm) for OLCI, respectively.

The Rayleigh-corrected reflectance (R_{rc}) at the 665, 865, and 1 020 nm was used to construct the floating algae index (FAI) (Hu, 2009). With a time-independent FAI threshold (1.194×10^{-4}), aquatic vegetation areas and pixels affected by algal scums were identified and masked following the community-accepted criterion (Hu et al., 2010). The threshold of the NDWI was determined using the modified histogram bimodal method (Zhang et al., 2018).

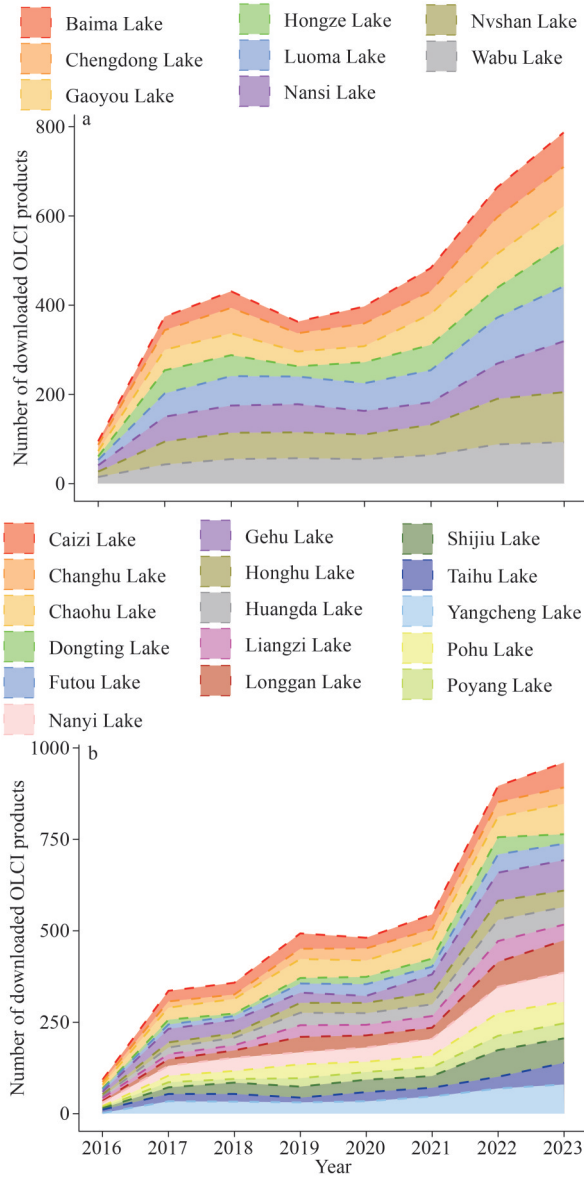


Fig.2 Statistics of the downloaded OLCI images

a. number of downloaded OLCI images of lakes in the Huaihe River Basin; b. number of downloaded OLCI images of lakes in the Changjiang River Basin.

2.4 Auxiliary data and statistical metrics

In the course of our analysis, we employed the root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), and median absolute error (MDAE) to quantify the biases, as defined by Eqs.2–5, respectively:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2}, \quad (2)$$

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{x_i - y_i}{y_i} \right|, \quad (3)$$

$$\text{MAE} = 10^Y, \quad Y = \sum_{i=0}^N |\log_{10}(x_i) - \log_{10}(y_i)| / N, \quad (4)$$

$$\text{MDAE} = \text{median}(|x_i - y_i|), \quad (5)$$

where N is the total number of samples and where x_i and y_i are the i^{th} estimate and measurement, respectively.

2.5 A classification-based scheme for algal biomass mapping

The B_{cu} inversion process from OLCI images is shown in Fig.3. In the absence of in-situ diffuse attenuation coefficient of the photosynthetically available radiation ($K_d(\text{PAR})$) for validation, we used HydroLight simulations based on measured inherent optical properties, with the reliability confirmed by good agreement between simulated and in-situ R_{rs} . First, the $K_d(\text{PAR})$ was derived through radiative transfer simulation performed using the Hydrolight-Ecolight 5.0 software (Supplementary Text S1). The four-component model was selected to simulate $K_d(\text{PAR})$, as it provides a comprehensive representation of the underwater light field. Vertical

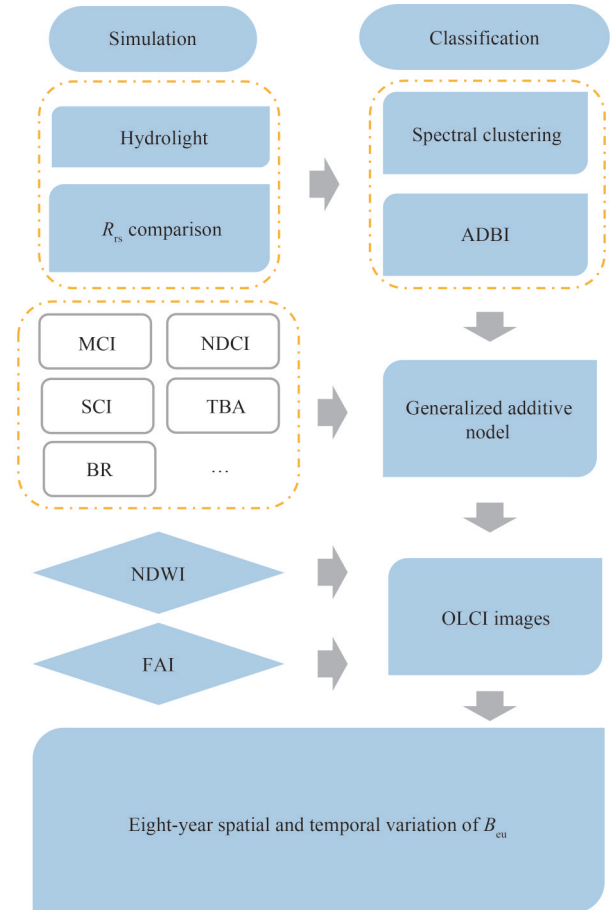


Fig.3 Flowchart of the B_{cu} inversion method

profiles of Chl *a*, SPIM, and $a_{\text{CDOM}}(440)$ were imported into the model as input data. The default values for the pure water absorption and backscattering coefficients were used in the simulations. Given the low transparency of the water body, an infinite water depth was assumed throughout the simulation. Additionally, the coordinates, wind speeds, and Greenwich Mean Time were incorporated as inputs to enable more precise simulations. The statistics of the input parameters are provided in Supplementary Table S3. The spectral angle mapper (SAM) (Kruse et al., 1993) was used to quantitatively assess the simulation results. It measures the spectral similarity between the measured and simulated spectra by computing the spectral angle distance. It is generally accepted that the smaller the angle between two spectra is, the greater the similarity. Each measured spectrum was paired with its corresponding simulated spectrum to form a spectral pair. We sorted these spectral pairs according to their SAM-derived angles. The band-average biases (band-average RMSE and band-average MAPE) were used for quality control of the simulations. Supplementary Text S2 provides a more in-depth description of the band-average biases, and the band-average biases increase as the similarity between the measured and simulated spectra decreases. We iteratively calculated the band-average biases for the first n spectral pairs ($n=1, 2, 3, \dots, 99$) and realized that the band-average biases increased abruptly after the first 44 pairs (Fig.4). This suggests that, starting from the 45th spectral pair, the similarity between the simulated and measured spectra significantly decreased, which

further indicated that only the first 44 simulated spectra were of acceptable quality. Therefore, the simulations corresponding to the first 44 spectral pairs were used to estimate the euphotic depth (Z_{eu}), and the remaining simulations were termed simulation failures. The SAM-derived angles of the 44th and 45th spectral pairs were 0.06π and 0.28π , respectively.

Owing to the lack of mature theoretical and technological conditions, direct measurement of Z_{eu} remains an unsolved problem. Therefore, an indirect strategy, as described in Eq.6, is generally used to calculate Z_{eu} (Kirk, 2010):

$$Z_{\text{eu}} = \frac{4.605}{K_d(\text{PAR})}. \quad (6)$$

The 44 calculated Z_{eu} values were related to the corresponding SDD values to establish an empirical model for estimating Z_{eu} in the case of simulation failure (Supplementary Fig.S1). B_{eu} was calculated by applying adjusted trapezoidal integrals to the measured Chl-*a* vertical profiles (Chl *a*(z)). The adjusted trapezoidal integral calculation was performed in two steps: the first step followed the standard trapezoidal integration (Bi et al., 2019; Hu et al., 2021), and the second step calculated the remaining biomass on the basis of the principle of triangle similarity. To facilitate understanding, Supplementary Fig.S2 demonstrates the calculation process. The maximum sampling depth within the euphotic depth was defined as D_1 , and its corresponding Chl-*a* concentration was defined as C_1 . Similarly, the deeper sampling depth near D_1 and its corresponding Chl-*a* concentration were defined

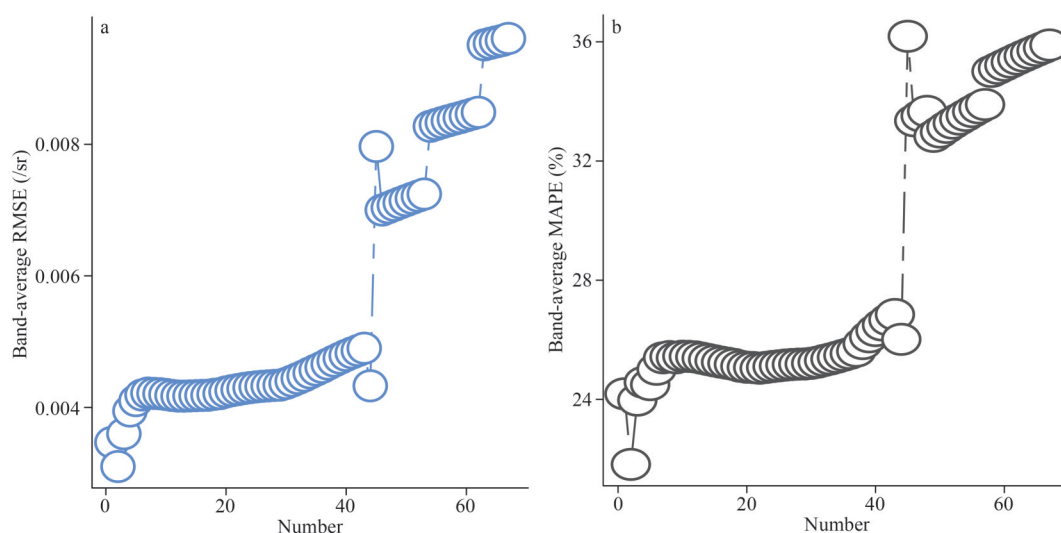


Fig.4 Band-average RMSE (a) and band-average MAPE (b) of the first n sorted spectral pairs

as D_2 and C_2 , respectively. Thus, B_{eu} was computed as the sum of B_{eu}' and B_{eu}'' (Eqs. 7–9):

$$B_{\text{eu}}' = \sum_{i=1}^N S_i, \quad (7)$$

$$B_{\text{eu}}'' = \frac{1}{2} \left(\frac{(Z_{\text{eu}} - D_1)(C_2 - C_1)}{D_2 - D_1} + 2C_1 \right) (Z_{\text{eu}} - D_1), \quad (8)$$

$$B_{\text{eu}} = B_{\text{eu}}' - B_{\text{eu}}'', \quad (9)$$

where S_i is the area of the i^{th} trapezoid, which is equivalent to the algal biomass within two adjacent sampling depths. Here, B_{eu}' , namely, the plain trapezoidal integral, is the sum of the areas of several trapezoids. In Supplementary Fig. S2, B_{eu}' and B_{eu}'' denote the algal biomass within D_1 and the remaining algal biomass, respectively. This scheme is effective in cases where Chl- a profiles cannot be fitted perfectly (Yu et al., 2025). For a more in-depth analysis of B_{eu}'' , please refer to Supplementary Text S3.

Given the variability in the optically active constituents (OACs) of lakes, a proper data classification scheme is necessary. The spectral clustering algorithm was selected in this context to divide the calibration dataset. The number of clusters was determined ($k=2$) by the eigenvalues of the Laplacian matrix. Finally, the calibration dataset was divided into two subsets, hereafter referred to as Type I ($n=43$) and Type II ($n=67$) (Fig. 5). The Type I spectra were almost bimodal, with a prominent peak at approximately 700 nm, which is characteristic of chlorophyll fluorescence. In contrast, the spectral profiles of Type II predominantly exhibited single-peaked or flat configurations ($n=67$). A community-accepted view is that for low-productivity and low-turbidity waters, a peak at approximately 560 nm can be observed, whereas for turbid waters, R_{rs} in the 400–650-nm range is elevated because of low suspended particulate matter (SPM) absorption (Li et al., 2012). Consistent with these spectral differences, Type I had a higher Chl a -to-SPM ratio and elevated B_{eu} levels than did Type II (Supplementary Table S4).

Spectral clustering, an unsupervised classification method, can operate without the need for true labels. Consequently, a metric independent of true labels is required to assess the efficacy of spectral clustering. In this context, the adjusted Davies-Bouldin index (ADBI) was proposed to evaluate the performance of spectral clustering. The ADBI is an adjusted form

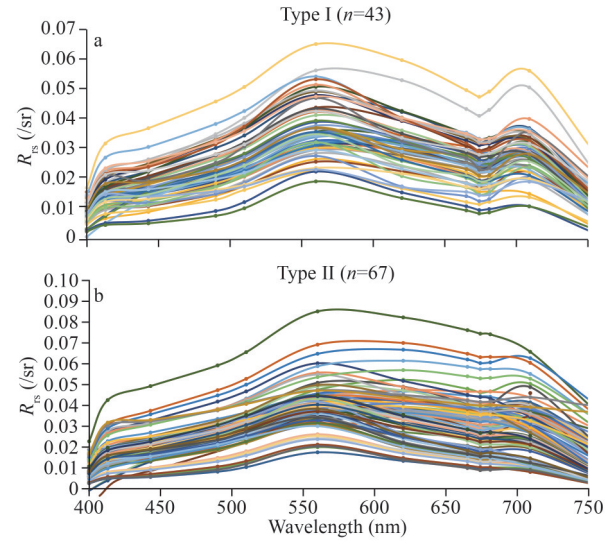


Fig. 5 OLCI-simulated spectra of Type I (a) and Type II (b) samples

Type I had a higher Chl a -to-SPM ratio and elevated B_{eu} levels than did Type II.

of the Davies-Bouldin index (Davies and Bouldin, 1979). The original Davies-Bouldin index is a validation metric that calculates the intra-cluster diameter by using the distance between cluster centers. In our analysis, we expanded the evaluation of point-based clustering to vector-based clustering (Supplementary Text S4). The results indicated that spectral clustering performed well on our dataset (ADBI ($\in[0, +\infty)$)=0.53).

Directly relating the band ratios or remote sensing indices to B_{eu} has the advantage of reducing uncertainties introduced by intermediate processes, such as the use of Chl a or other OACs (Lai et al., 2023). We evaluated the performance of four well-established remote sensing indices for estimating B_{eu} (Table 2): the normalized difference chlorophyll

Table 2 Mathematical formulations of the four indices and band ratios

Algorithm	Formulation	Wavelength
NDCI	$[R_{\text{rs}}(\lambda_1) - R_{\text{rs}}(\lambda_2)] / [R_{\text{rs}}(\lambda_1) + R_{\text{rs}}(\lambda_2)]$	$\lambda_1=708 \text{ nm}, \lambda_2=665 \text{ nm}$
SCI	$H = [R_{\text{rs}}(\lambda_4) + (\lambda_4 - \lambda_3) \times (R_{\text{rs}}(\lambda_2) - R_{\text{rs}}(\lambda_4))] / (\lambda_4 - \lambda_2) - R_{\text{rs}}(\lambda_3)$	$\lambda_1=560 \text{ nm},$
	$H\Delta = R_{\text{rs}}(\lambda_2) - [R_{\text{rs}}(\lambda_4) + (\lambda_4 - \lambda_2) \times (R_{\text{rs}}(\lambda_1) - R_{\text{rs}}(\lambda_4))] / (\lambda_4 - \lambda_1)$	$\lambda_2=620 \text{ nm},$
	$\text{SCI} = H - H\Delta$	$\lambda_3=665 \text{ nm},$ $\lambda_4=681 \text{ nm}$
TBA	$[R_{\text{rs}}(\lambda_1)^{-1} - R_{\text{rs}}(\lambda_2)^{-1}] \times R_{\text{rs}}(\lambda_3)$	$\lambda_1=665 \text{ nm}, \lambda_2=709 \text{ nm},$ $\lambda_3=754 \text{ nm}$
MCI	$R_{\text{rs}}(\lambda_2) - R_{\text{rs}}(\lambda_1) - (\lambda_2 - \lambda_1) \times [R_{\text{rs}}(\lambda_3) - R_{\text{rs}}(\lambda_1)] / (\lambda_3 - \lambda_1)$	$\lambda_1=681 \text{ nm}, \lambda_2=708 \text{ nm},$ $\lambda_3=753 \text{ nm}$
Band ratio	$R_{\text{rs}}(\lambda_1) / R_{\text{rs}}(\lambda_2)$	

index (NDCI) (Mishra and Mishra, 2012), synthetic chlorophyll index (SCI) (Shen et al., 2010), three-band algorithm (TBA) (Dall’Olmo et al., 2003), and maximum chlorophyll-*a* index (MCI) (Gower et al., 2005). Additionally, we computed pairwise band ratios using the 17 OLCI spectral bands (400–865 nm) to determine the optimal band combinations for algorithm calibration.

The performance of the remote sensing indices and band ratios was evaluated using Pearson’s correlation coefficient (Kirch, 2008). The indices and band ratios that strongly correlated (Pearson’s $r \geq 0.70$) with B_{eu} were selected as explanatory variables for the GAM. For Type I, we selected TBA (Pearson’s $r = 0.70$, $P < 0.01$) and the band ratio ($\lambda_1 = 709$ nm, $\lambda_2 = 681$ nm, Pearson’s $r = 0.70$, $P < 0.01$) as explanatory variables. For Type II, two band ratios ($\lambda_1 = 681$ nm, $\lambda_2 = 674$ nm, Pearson’s $r = 0.71$, $P < 0.01$; $\lambda_1 = 709$ nm, $\lambda_2 = 674$ nm, Pearson’s $r = 0.70$, $P < 0.01$) were selected. The GAM was used as the regression model to fit the selected explanatory variables (remote sensing indices and band ratios) to the response variable (B_{eu}). To calibrate the GAM model, we specified the “OptimizeHyperparameters” parameter as “all-univariate” to ensure optimal values of interactions, initial learning rates, maximum splits, and the count of trees. We used “BayesianOptimizer” as the optimizer for the hyperparameter tuning process. Additionally, we compacted the size of the final GAM model to reduce the runtime.

3 RESULT

3.1 Accuracy and reasonability of the scheme

The performance of the GAM was evaluated using various metrics. For Type I, the GAM exhibited an RMSE of 9.43 mg/m² and a MAPE of 17.2%. For Type II, the RMSE was 7.40 mg/m², and the MAPE was 30.3%. For the validation dataset, the RMSE was 13.24 mg/m², and the MAPE was 36.3% (Fig. 6). Furthermore, the results of the match-up validation demonstrated acceptable limits of accuracy, with an RMSE of 12.62 mg/m² and an MAPE of 36.5% (Supplementary Fig. S3a). To understand the relationship between optical data and biological data, we selected and presented the spectra and vertical structures of two Chl-*a* profiles from the validation dataset, which were labeled differently by the spectral clustering (Supplementary Fig. S4).

To validate the rationality of the classification-

based scheme, we compared it with the plain GAM. The proposed classification-based scheme outperformed the plain GAM, as illustrated in Fig. 7a. The proposed classification-based scheme was compared with the process-oriented algorithm (Liu et al., 2021) and the result-oriented algorithm (Li et al., 2017) (Fig. 7b). The former relies on establishing empirical functions to determine depth-dependent parameters, while the latter prioritizes the accurate fitting of Chl-*a* vertical profiles. The parameters of the conventional algorithms were recalibrated (Table 3). We found that the classification-based scheme achieved higher accuracy than the non-classification approach, with an R^2 of 0.85 and an RMSE of 12.3 mg/m², compared to 0.72 and 18.7 mg/m² for the non-classification approach (Fig. 6).

3.2 Spatial pattern of the pixel-based B_{eu}

For each pixel, the seasonal B_{eu} was calculated as the mean of all valid B_{eu} retrievals from cloud-free OLCI images acquired during the corresponding season. Annual B_{eu} was calculated as the mean of all valid retrievals within each calendar year. No additional weighting was applied, and pixels with no valid observations were treated as no data.

Taihu Lake demonstrated more pronounced seasonal variability. In summer, high B_{eu} values were clearly observed in the western nearshore waters, whereas in winter, high values were barely discernible across the entire lake. The spatial distribution patterns of B_{eu} in two large lakes, Chaohu Lake and Gaoyou Lake, exhibited distinct characteristics Fig. 8. For Chaohu Lake, the spatial distribution pattern of B_{eu} can be summarized as “high in the west, low in the east”. The spatial distribution pattern of B_{eu} for Gaoyou Lake can be summarized as “high at the shore, low in the center” (Supplementary Fig. S5).

The spatial patterns of the annual average B_{eu} indicated that, from 2016–2023, the near-shore waters of certain lakes constantly sustained high B_{eu} levels (Fig. 9). For example, the waters along the northern shores of Luoma Lake and Longgan Lake consistently had high biomass. Similarly, the waters along the western shores of Chaohu Lake and Gaoyou Lake also continuously presented high B_{eu} levels. Nevertheless, significant annual variability in spatial distribution patterns was detected in several other lakes, namely, Yangcheng Lake, Honghu Lake,

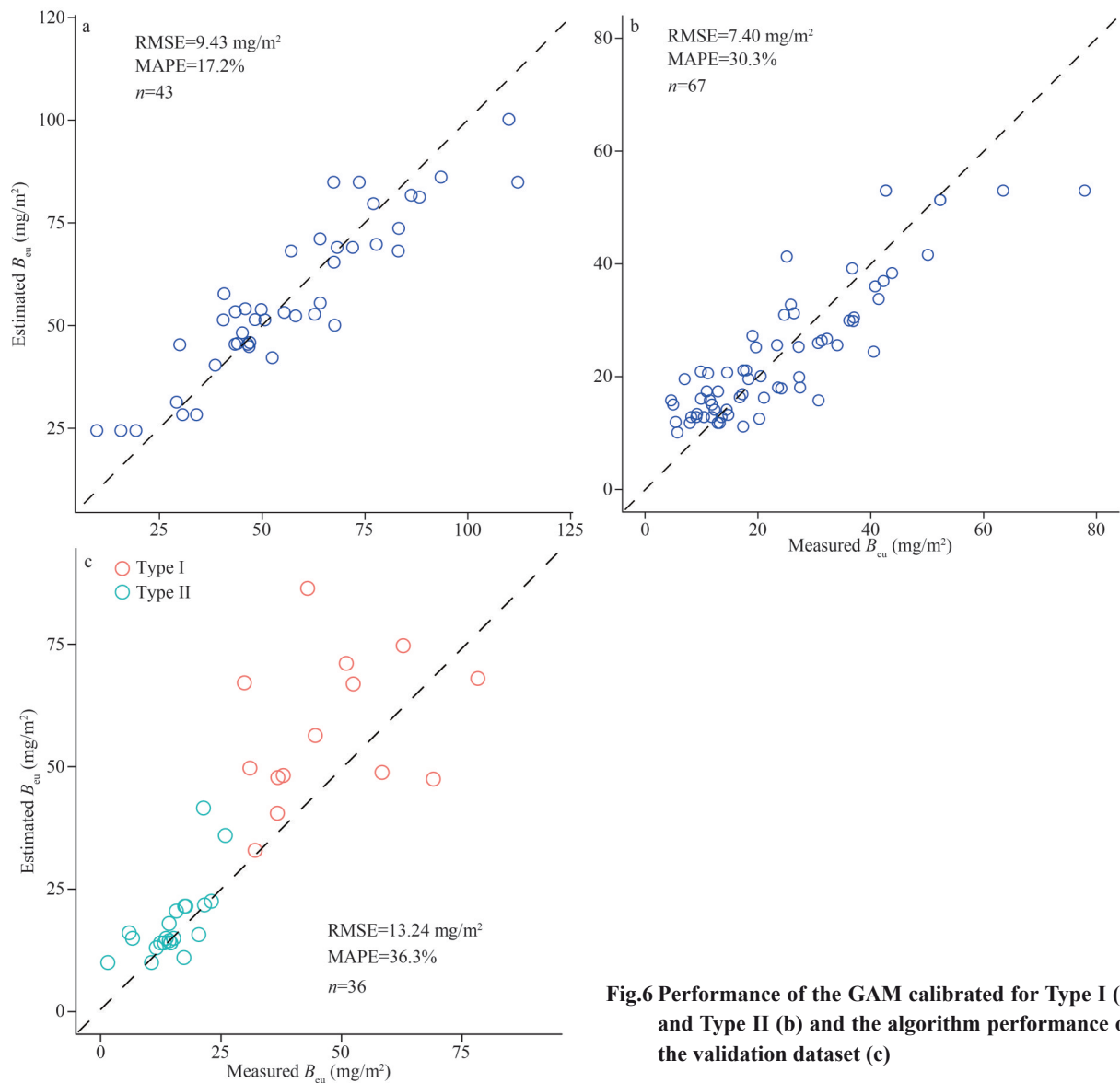


Fig.6 Performance of the GAM calibrated for Type I (a) and Type II (b) and the algorithm performance on the validation dataset (c)

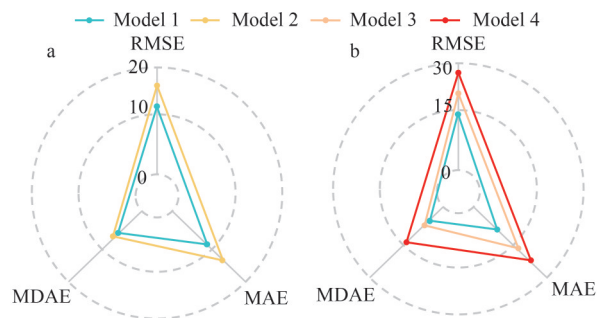


Fig.7 Performance of the classification-based scheme and the plain GAM (a); comparison with the conventional algorithm (the process-oriented algorithm and the result-oriented algorithm) (b)

Models 1–4 represent the classification-based scheme, the plain GAM, the process-oriented algorithm and the result-oriented algorithm, respectively.

Table 3 Key parameter settings in result-oriented algorithm

Algorithm	Empirical model	Parameters' value
Result-oriented	$a = a_1 \times \text{depth} + \beta_1$	$a_1 = 0.78, \beta_1 = 0.026$
	$b = a_2 \times \text{depth} + \beta_2$	$a_2 = 4.20, \beta_2 = -2.37$
	$\delta = a_1 \times R_{rs}(709) + \beta_1$	$a_1 = -54.61, \beta_1 = 2.77$
Process-oriented	$h = a_2 \times R_{rs}(709) + \beta_2$	$a_2 = -4\,312.05, \beta_2 = 180.03$
	$C_0 = a_3 \times \text{Chl } a + \beta_3$	$a_3 = 0.47, \beta_3 = 3.94$

a and b in result-oriented algorithm are two key parameters for algal biomass estimation; for the process-oriented algorithm, we listed three parameters (δ , h , and C_0) for integral. In algorithm comparison, we classify Chl- a profiles into two categories: mixed type and Gaussian type. The three parameters define the shape of a Gaussian Chl- a vertical profile.

Gehu Lake, Gaoyou Lake, and Baima Lake. Intriguingly, this variability did not seem to be

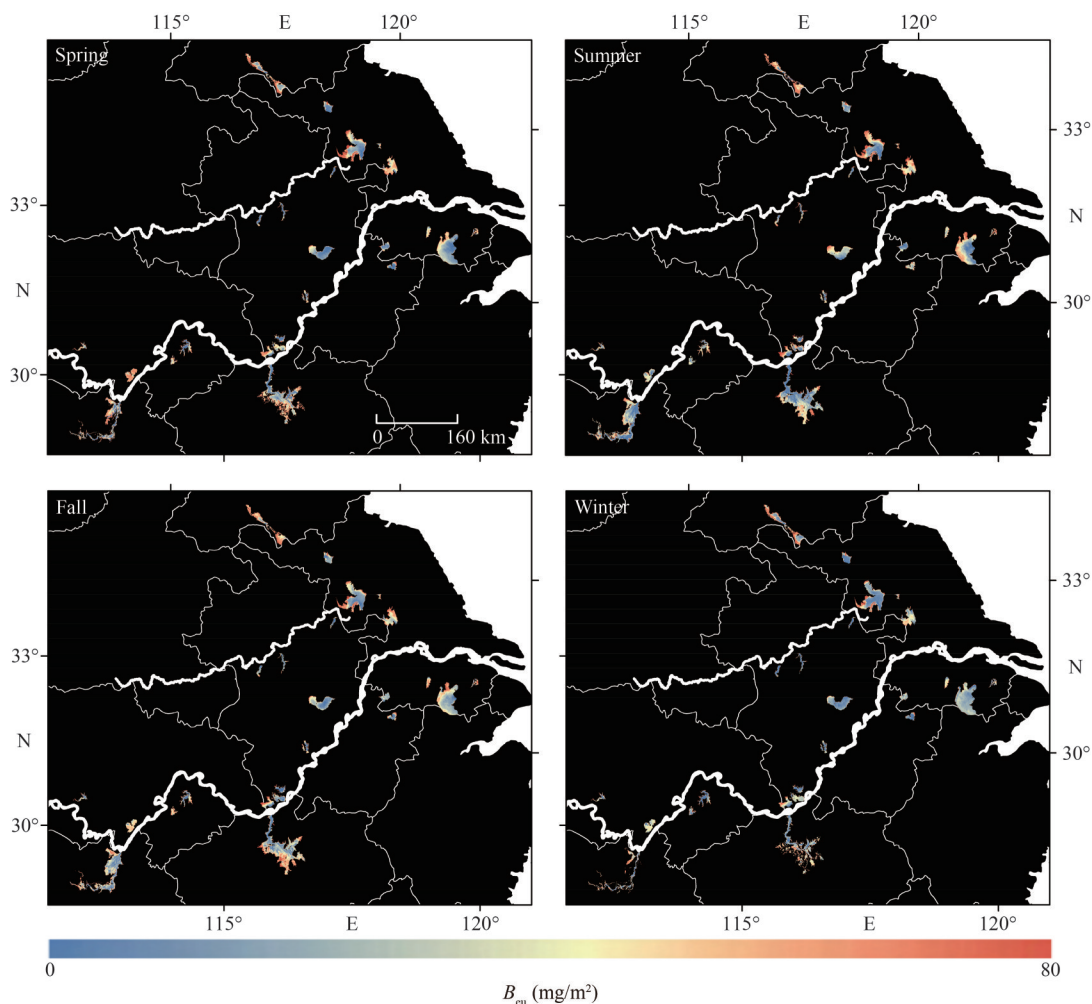


Fig.8 Seasonal spatial patterns of pixel-based B_{eu} for the studied lakes

Spring: from March to May; summer: from June to August; autumn: from September to November; winter: from December to February.

associated with lake size or basin. In addition, a few lakes, such as Chengdong Lake and Wabu Lake, presented low B_{eu} levels across the entire lake, with only restricted areas of high biomass.

3.3 Characterization of algal biomass dynamics

The concept of the total algal biomass within the euphotic depth (referred to as total B_{eu} , quantified in tons) is straightforwardly determined by summing the pixel-based B_{eu} weights, which is defined as follows:

$$\text{total } B_{eu} = 10^{-9} \times \text{resolution}^2 \times \sum_{i=1}^N B_{eu}(i), \quad (10)$$

where resolution, N , $B_{eu}(i)$, and 10^{-9} are the spatial resolution of the pixel (here, 300 m×300 m), the total number of pixels of a lake, the i^{th} pixel-based B_{eu} , and the factor that converts mg to tons, respectively.

In our analysis, we used the values per lake instead of the total values to mitigate the influence of differences in lake area and counts across the various bins. For example, the total area of lakes in the Changjiang River Basin is much larger than that in the Huaihe River Basin. Similarly, the total number and area of large lakes are generally smaller than those of medium-sized lakes. By using the values per lake, we were able to account for these differences and focus the analysis on the underlying trends.

The total B_{eu} of hyper-sized lakes in 2018 and 2023 was significantly lower than that in other years and remained relatively unchanged over the four-year period from 2019–2022 (Fig.10c). For large lakes, the trend in the total B_{eu} displayed a “single peak”, with total B_{eu} increasing from 2016–2019 before showing an overall decline from 2019–2023 (Fig.10b). Another important finding was that the

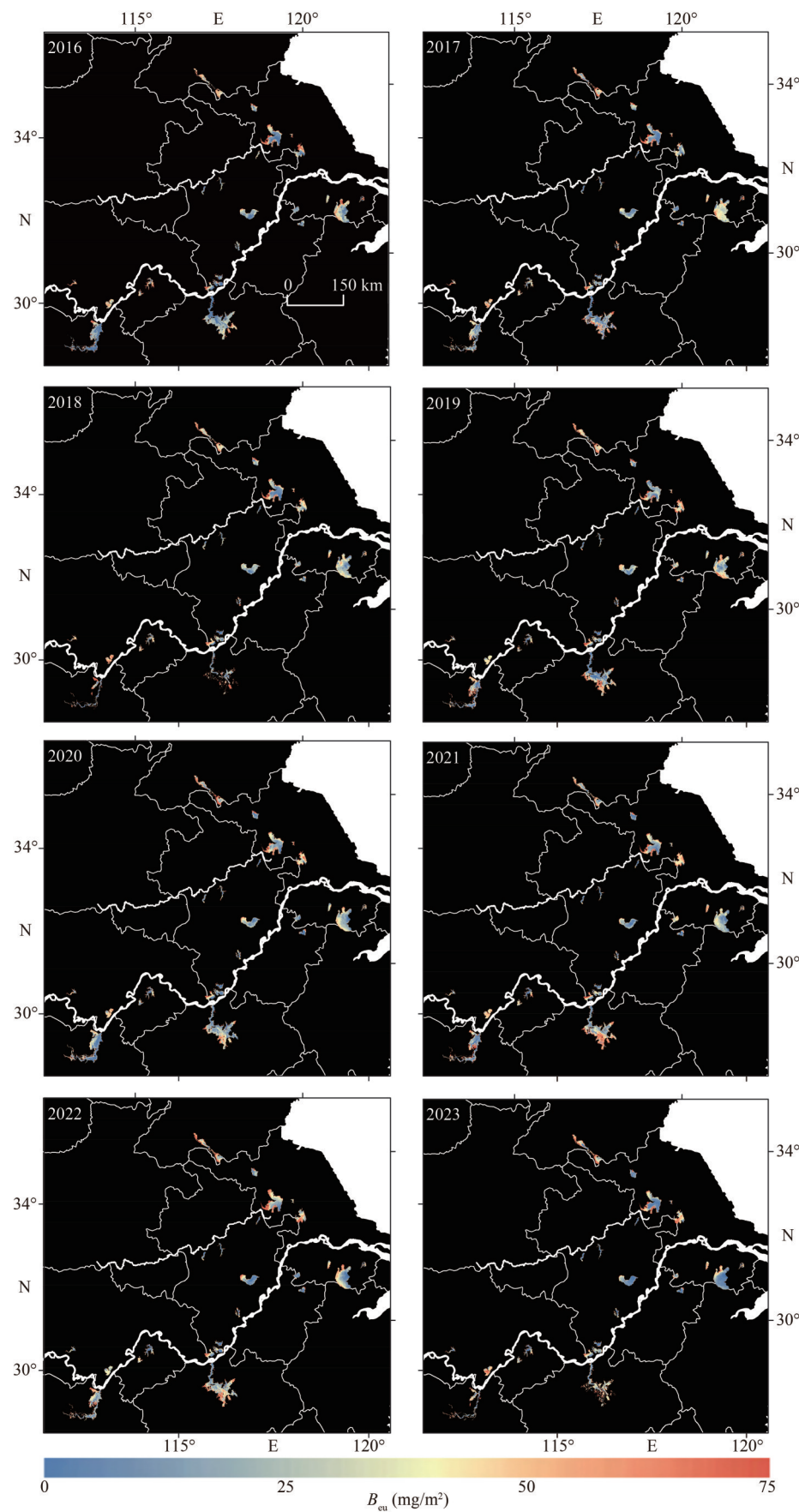


Fig.9 Annual spatial patterns of pixel-based B_{eu} for the studied lakes from 2016–2023

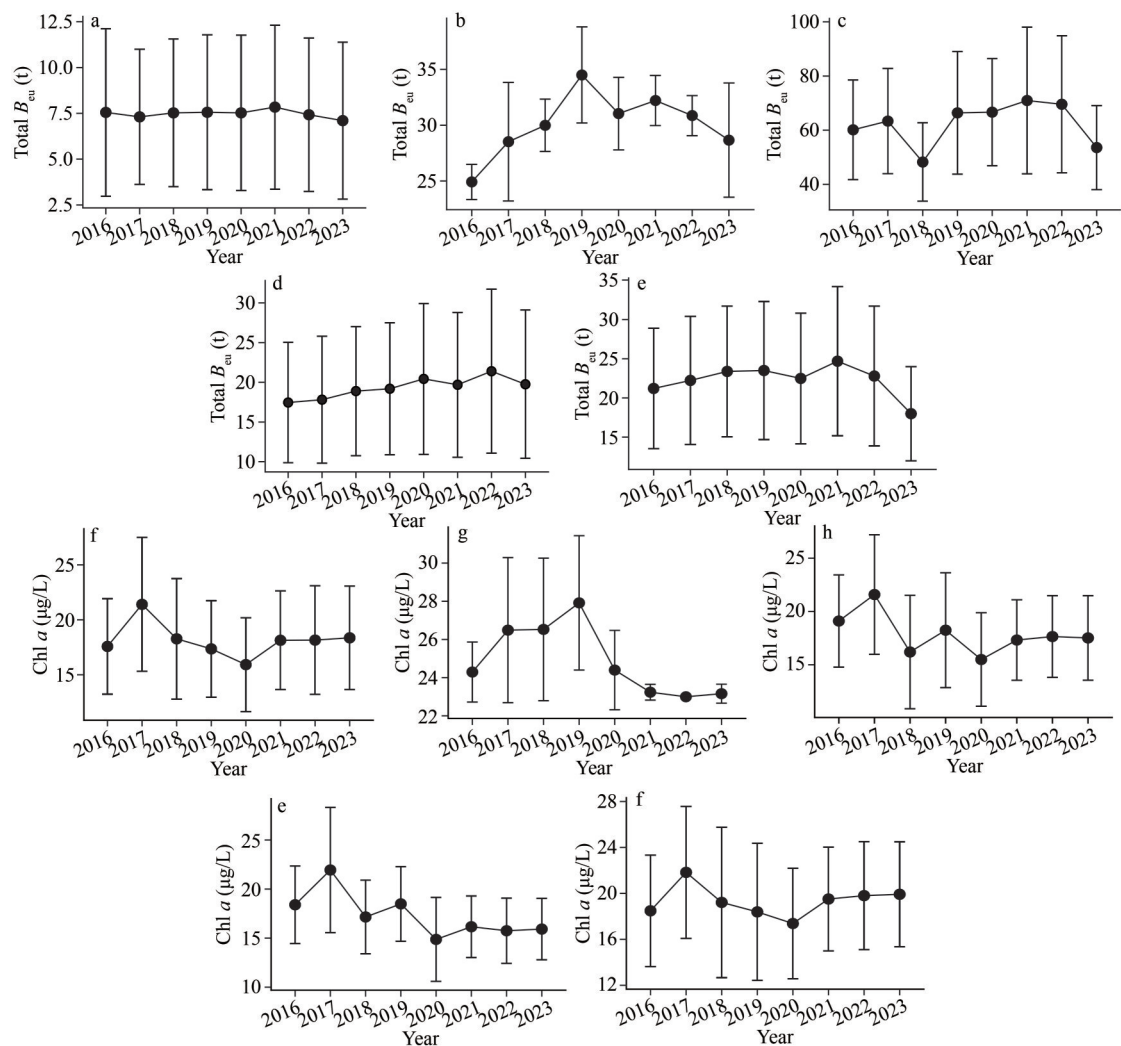


Fig.10 Interannual dynamics of the total B_{eu} of medium-sized lakes (a), large lakes (b), hyper-sized lakes (c), lakes in the Huaihe River Basin (d), and lakes in the Changjiang River Basin (e); interannual dynamics of surface Chl- a concentration of medium-sized lakes (f), large lakes (g), hyper-sized lakes (h), lakes in the Huaihe River Basin (i), and lakes in the Changjiang River Basin (j)

interannual total B_{eu} of medium-sized lakes remained stable for eight years, with no significant variations observed (Fig.10a).

In the Changjiang River Basin, the first notable increase in the total B_{eu} occurred between 2020 and 2021, after which the total B_{eu} decreased (Fig.10e). In contrast, a slow growth trend in the total B_{eu} was observed in the Huaihe River Basin, culminating in a relatively clear peak in 2022 (Fig.10d).

When the analysis was carried out on the basis of lake size, hyper-sized lakes demonstrated the most substantial change in the total B_{eu} , with a coefficient of variation (CV) of 12.7%. Moreover, the magnitude of the changes in the total B_{eu} increased with the size of the lakes. This was indicated by the increase in the CV from 2.9% to

12.7% (Table 4). The changes in total B_{eu} of lakes in the Changjiang River Basin, with a CV of 9.0%, were slightly greater than those in the Huaihe River Basin, where the CV was 6.7%. It was suggested that the variability in the total B_{eu}

Table 4 CV (%) results

Division	CV of total B_{eu} (%)
Middle-sized lake	2.9
Large lake	9.5
Hyper-sized lake	12.7
Huaihe River Basin	6.7
Changjiang River Basin	9.0

The statistics are given in different categories (middle-sized lakes, large lakes, hyper-sized lakes, Huaihe River Basin, and Changjiang River Basin).

differed not only according to the sizes of the lakes but also between different basins. Larger lakes and those in the Changjiang River Basin seemed to experience more significant fluctuations in the total B_{eu} over time.

4 DISCUSSION

4.1 Mechanism underlying the size-dependent variability of total B_{eu}

The observed pattern that the magnitude of interannual total B_{eu} variability increases with lake size can be explained by three interacting mechanisms: hydrological buffering, nutrient loading and internal release, and anthropogenic disturbance.

Large lakes typically have longer water residence times and greater water volumes compared to small lakes. This hydrological buffering capacity allows large lakes to integrate and accumulate the effects of interannual variations in external inputs, leading to amplified year-to-year fluctuations in algal biomass (Kelly et al., 2001). In contrast, small lakes with shorter residence times respond more rapidly to short-term disturbances but may exhibit lower interannual variability due to faster flushing and reduced cumulative effects (Cao et al., 2022b).

Large lakes often drain larger catchments and consequently receive higher external nutrient loads (Pozdnyakov et al., 2019). Moreover, they tend to have greater sediment nutrient stocks and higher potential for internal phosphorus release under favorable conditions (Wang et al., 2021). These characteristics can drive more pronounced interannual variability in algal biomass, particularly when external loads fluctuate or when redox conditions at the sediment-water interface change interannually (Smith et al., 2011).

Large lakes in the middle and lower reaches of the Changjiang River and Huaihe River Basins are typically subject to more intensive human activities, including aquaculture, navigation, water transfer projects, and shoreline modifications (He et al., 2019). These activities often exhibit interannual variability and can directly or indirectly affect algal biomass dynamics (Roozen et al., 2007). The cumulative effects of such disturbances may contribute to the greater interannual variability observed in larger lakes.

4.2 Disagreement of interannual surface Chl-*a* trends and total B_{eu} trends

A well-validated model designed for Chinese

lakes was employed to derive surface Chl-*a* concentrations (Cao et al., 2024, 2025). The temporal differences between B_{eu} and surface Chl-*a* highlight the necessity of using B_{eu} as an indicator for water quality assessment. Figure 10 shows that the interannual trends in surface Chl-*a* concentrations in medium-sized and hyper-sized lakes significantly differed from those of the total B_{eu} . For example, the surface Chl-*a* trend of medium-sized lakes exhibited a single peak, whereas the total B_{eu} was almost unchanged from 2016–2023 (Fig.10a & f). The surface Chl-*a* trend of the hyper-sized lakes was distinctly bimodal (Fig.10h), whereas the corresponding total B_{eu} showed no pronounced peak (Fig.10c). At the basin scale, the trend of surface Chl-*a* in the Huaihe River Basin was markedly different from that of B_{eu} . Specifically, the surface Chl-*a* concentration exhibited a bimodal pattern, whereas the total B_{eu} showed an almost consistently increasing trend (Fig.10d & i). In the Changjiang River Basin, the trends of surface Chl-*a* and total B_{eu} also exhibited significant differences. The surface Chl-*a* concentration peaked in 2017 and remained nearly unchanged after 2021, whereas the total B_{eu} peaked in 2021 and then declined rapidly (Fig.10e & j). The significant divergence between the interannual surface Chl-*a* trends and total B_{eu} trends suggested that B_{eu} captured distinct aspects of water quality dynamics that were not fully reflected by surface Chl-*a* alone. This highlights the necessity of incorporating B_{eu} as a complementary indicator for a more comprehensive assessment of aquatic ecosystems, particularly medium-sized and hyper-sized lakes, as well as lakes in the Huaihe River Basin and the Changjiang River Basin, where these differences are pronounced.

Table 5 indicates that the interannual surface Chl-*a* concentration and the interannual total B_{eu} were either negatively correlated or uncorrelated. Among the findings, the negative correlation between the

Table 5 Correlations between the interannual surface Chl-*a* concentrations and the interannual total B_{eu} , both of which are normalized

Bins	Pearson's <i>r</i>
Huaihe River Basin	-0.78, $P<0.05$
Changjiang River Basin	-0.15, $P<0.05$
Hyper-sized lake	0.12, $P<0.01$
Large lake	0.28, $P<0.05$
Medium-sized lake	-0.26, $P<0.01$

surface Chl-*a* concentration and the total B_{eu} was most pronounced in the Huaihe River Basin (Pearson's $r=-0.78$, $P<0.05$). As the surface Chl-*a* concentration increased, the total B_{eu} tended to decrease. A Pearson's r with an absolute value below 0.3 generally indicates a very weak or negligible linear relationship between two variables. In these cases, changes in the surface Chl-*a* concentration did not strongly predict changes in the total B_{eu} . Our findings underscore the necessity and importance of studying the total B_{eu} , as it provides a more comprehensive understanding of phytoplankton biomass dynamics within the euphotic zone, which cannot be fully captured by surface Chl *a* alone. By focusing on the total B_{eu} , we can gain deeper insights into the factors regulating phytoplankton growth and their ecological implications.

The negative correlation between surface Chl *a* and total B_{eu} likely arises from the complex vertical dynamics of phytoplankton. Vertical migration of buoyant cyanobacteria can elevate surface Chl *a* without increasing depth-integrated biomass, while photoacclimation and seasonal changes in water column stratification further decouple surface signals from whole-water-column biomass. These mechanisms collectively explain why surface Chl *a* alone is insufficient to capture the full variability of phytoplankton biomass in the euphotic zone.

4.3 Limitation

One limitation of this study concerns the temporal consistency of satellite data availability. The number of available OLCI images increased notably after 2021, primarily due to the full operational status of the Sentinel-3A and Sentinel-3B constellation. Consequently, the early period of the time series (2016–2018) is characterized by relatively sparse observations compared to later years. This temporal heterogeneity in data volume may introduce uncertainty in the estimated interannual trends, as the earlier years have lower sampling density. While the observed patterns—such as the single-peak trend in large lakes—are broadly consistent with known hydrological and meteorological events during the study period, future studies with longer time series or multi-sensor data fusion would help further constrain the trends and reduce potential biases associated with uneven data distribution.

Although the overall validation results (Fig.6) demonstrate satisfactory performance of the classification-based B_{eu} retrieval model, certain

outliers with large discrepancies were observed. For example, one Type I sample showed an in situ B_{eu} of approximately 42 mg/m² but an estimated value exceeding 80 mg/m². Such discrepancies can arise from several sources. Uncertainties in the in situ B_{eu} calculation, including depth-resolved Chl-*a* profile integration and K_d (PAR) estimation, may propagate into the reference values. Additionally, the spectral clustering process may misclassify a small number of samples that lie near the boundaries between optical types, leading to suboptimal model performance for those cases. These factors collectively contribute to the observed discrepancies and represent inherent challenges in remote sensing of complex inland waters. Future efforts to improve retrieval accuracy could focus on water type classification and refinement of in situ measurements.

5 CONCLUSION

For the first time, eight-year B_{eu} maps were generated for lakes in the middle and lower reaches of the Changjiang River Basin and Huaihe River Basin using OLCI images. This study first identified the spatial patterns of pixel-based B_{eu} and then explored the trends of total B_{eu} for lakes in the Changjiang and Huaihe rivers basins. The most dramatic variations in the total B_{eu} were found in hyper-sized lakes (CV=12.7%). The variations in B_{eu} in the Changjiang River Basin (CV=9.0%) were somewhat more significant than those in the Huaihe River Basin (CV=6.7%). In addition, the variations in the surface Chl-*a* concentration did not align with changes in the total B_{eu} , implying the necessity and importance of studying the total B_{eu} . The generated B_{eu} maps provide indispensable products and new insights for lake ecosystem management. They can potentially assist in understanding the temporal and spatial changes in algal biomass over a long-term period, enabling more effective management strategies for maintaining the health and balance of these ecosystems. However, this model was developed based on in situ data from the Changjiang River and Huaihe River basins. Due to the lack of in situ data from other lake regions, the model's performance in these areas will require further evaluation in the future.

6 DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

7 ACKNOWLEDGMENT

The authors would like to thank the editor and anonymous reviewers for their valuable comments that greatly improved our manuscript. We thank Chuanmin HU, College of Marine Science, University of South Florida, and Zhigang CAO, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, for their insightful comments.

References

- Bailey S W, Werdell P J. 2006. A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sensing of Environment*, **102**(1-2): 12-23, <https://doi.org/10.1016/j.rse.2006.01.015>.
- Bi S, Li Y M, Lyu H et al. 2019. Quantifying spatiotemporal dynamics of the column-integrated algal biomass in nonbloom conditions based on OLCI data: a case study of Lake Dianchi, China. *IEEE Transactions on Geoscience and Remote Sensing*, **57**(10): 7447-7459, <https://doi.org/10.1109/TGRS.2019.2913401>.
- Cao Z G, Ma R H, Pahlevan N et al. 2022a. Evaluating and optimizing VIIRS retrievals of chlorophyll-*a* and suspended particulate matter in turbid lakes using a machine learning approach. *IEEE Transactions on Geoscience and Remote Sensing*, **60**: 4211417, <https://doi.org/10.1109/TGRS.2022.3220529>.
- Cao Z G, Melack J M, Liu M et al. 2022b. Spatial and seasonal variability of chlorophyll *a* in different-sized lakes across eastern China. *Inland Waters*, **12**(2): 205-214, <https://doi.org/10.1080/20442041.2021.1970462>.
- Cao Z G, Wang M H, Ma R H et al. 2025. Seamless observations of chlorophyll-*a* from OLCI and VIIRS measurements in inland lakes. *Water Research*, **270**: 122825, <https://doi.org/10.1016/j.watres.2024.122825>.
- Cao Z G, Wang M H, Ma R H et al. 2024. 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>.
- Dall'Olmio G, Gitelson A A, Rundquist D C. 2003. Towards a unified approach for remote estimation of chlorophyll-*a* in both terrestrial vegetation and turbid productive waters. *Geophysical Research Letters*, **30**(18): 2003GL018065, <https://doi.org/10.1029/2003GL018065>.
- Davies D L, Bouldin D W. 1979. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-1**(2): 224-227, <https://doi.org/10.1109/TPAMI.1979.4766909>.
- Gower J, King S, Borstad G et al. 2005. Detection of intense plankton blooms using the 709 nm band of the MERIS imaging spectrometer. *International Journal of Remote Sensing*, **26**(9): 2005-2012, <https://doi.org/10.1080/01431160500075857>.
- Hastie T, Tibshirani R. 2014. Generalized Additive Models. In: Kenett R S, Longford N T, Piegorsch W W et al. eds. Wiley StatsRef: Statistics Reference Online. John Wiley & Sons Inc., Hoboken, <https://doi.org/10.1002/9781118445112.stat03141>.
- He Z S, Huo S L, Ma C Z et al. 2019. The contributions of climate changes and human activities to long-term variations in lake sediments based on results from generalized additive models. *Water Resources Management*, **33**(3): 1069-1085, <https://doi.org/10.1007/s11269-018-2167-2>.
- Ho J C, Michalak A M, Pahlevan N. 2019. Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature*, **574**(7780): 667-670, <https://doi.org/10.1038/s41586-019-1648-7>.
- Hu C M. 2009. A novel ocean color index to detect floating algae in the global oceans. *Remote Sensing of Environment*, **113**(10): 2118-2129, <https://doi.org/10.1016/j.rse.2009.05.012>.
- Hu C M, Lee Z, Ma R H et al. 2010. Moderate Resolution Imaging Spectroradiometer (MODIS) observations of cyanobacteria blooms in Taihu Lake, China. *Journal of Geophysical Research: Oceans*, **115**(C4): 2009JC005511, <https://doi.org/10.1029/2009JC005511>.
- Hu M Q, Ma R H, Xiong J F et al. 2022. Eutrophication state in the Eastern China based on Landsat 35-year observations. *Remote Sensing of Environment*, **277**: 113057, <https://doi.org/10.1016/j.rse.2022.113057>.
- Hu M Q, Ma R H, Xue K et al. 2024. Eutrophication evolution of lakes in China: four decades of observations from space. *Journal of Hazardous Materials*, **470**: 134225, <https://doi.org/10.1016/j.jhazmat.2024.134225>.
- Hu M Q, Zhang Y C, Ma R H et al. 2021. Optimized remote sensing estimation of the lake algal biomass by considering the vertically heterogeneous chlorophyll distribution: study case in Lake Chaohu of China. *Science of the Total Environment*, **771**: 144811, <https://doi.org/10.1016/j.scitotenv.2020.144811>.
- Kelly C A, Fee E, Ramlal P S et al. 2001. Natural variability of carbon dioxide and net epilimnetic production in the surface waters of boreal lakes of different sizes. *Limnology and Oceanography*, **46**(5): 1054-1064, <https://doi.org/10.4319/lo.2001.46.5.1054>.
- Kirch W. 2008. Pearson's Correlation Coefficient. In: Kirch W ed. Encyclopedia of Public Health. Springer, Dordrecht. p.1090-1091, https://doi.org/10.1007/978-1-4020-5614-7_2569.
- Kirk J T O. 2010. Light and Photosynthesis in Aquatic Ecosystems. 3rd edn. Cambridge University Press, Cambridge. 638p, <https://doi.org/10.1017/CBO9781139168212>.
- Kruse F A, Lefkoff A B, Boardman J W et al. 1993. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, **44**(2-3): 145-163, [https://doi.org/10.1016/0034-4257\(93\)90013-N](https://doi.org/10.1016/0034-4257(93)90013-N).
- Lai L, Zhang Y C, Cao Z et al. 2023. Algal biomass mapping of eutrophic lakes using a machine learning approach with MODIS images. *Science of The Total Environment*, **880**: 163357, <https://doi.org/10.1016/j.scitotenv.2023.163357>.

- Lai L, Zhang Y C, Han T et al. 2024. Satellite mapping reveals phytoplankton biomass's spatio-temporal dynamics and responses to environmental factors in a eutrophic inland lake. *Journal of Environmental Management*, **360**: 121134, <https://doi.org/10.1016/j.jenvman.2024.121134>.
- Li J, Ma R H, Xue K et al. 2020. Drivers to spatial and temporal dynamics of column integrated phytoplankton biomass in the Shallow Lake of Chaohu, China. *Ecological Indicators*, **109**: 105812, <https://doi.org/10.1016/j.ecolind.2019.105812>.
- Li J, Zhang Y C, Ma R H et al. 2017. Satellite-based estimation of column-integrated algal biomass in nonalgae bloom conditions: a case study of lake Chaohu, China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **10**(2): 450-462, <https://doi.org/10.1109/JSTARS.2016.2601083>.
- Li Y, Shi K, Zhang Y L et al. 2022. Remote sensing of column-integrated chlorophyll *a* in a large deep-water reservoir. *Journal of Hydrology*, **610**: 127918, <https://doi.org/10.1016/j.jhydrol.2022.127918>.
- Li Y M, Wang Q, Wu C Q et al. 2012. Estimation of chlorophyll *a* concentration using NIR/Red bands of MERIS and classification procedure in inland turbid water. *IEEE Transactions on Geoscience and Remote Sensing*, **50**(3): 988-997, <https://doi.org/10.1109/TGRS.2011.2163199>.
- Liu D, Yu S J, Cao Z G et al. 2021. Process-oriented estimation of column-integrated algal biomass in eutrophic lakes by MODIS/Aqua. *International Journal of Applied Earth Observation and Geoinformation*, **99**: 102321, <https://doi.org/10.1016/j.jag.2021.102321>.
- Ma J G, Loiselle S, Cao Z G et al. 2023. Unbalanced impacts of nature and nurture factors on the phenology, area and intensity of algal blooms in global large lakes: MODIS observations. *Science of The Total Environment*, **880**: 163376, <https://doi.org/10.1016/j.scitotenv.2023.163376>.
- Ma R, Tang J, Dai J et al. 2006. Absorption and scattering properties of water body in Taihu Lake, China: absorption. *International Journal of Remote Sensing*, **27**(19): 4277-4304, <https://doi.org/10.1080/01431160600851835>.
- Ma R H, Yang G S, Duan H T et al. 2011. China's lakes at present: number, area and spatial distribution. *Science China Earth Sciences*, **54**(2): 283-289, <https://doi.org/10.1007/s11430-010-4052-6>.
- Mishra S, Mishra D R. 2012. Normalized difference chlorophyll index: a novel model for remote estimation of chlorophyll-*a* concentration in turbid productive waters. *Remote Sensing of Environment*, **117**: 394-406, <https://doi.org/10.1016/j.rse.2011.10.016>.
- Mueller J L, Morel A, Frouin R et al. 2003. Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 4. Volume III: Radiometric Measurements and Data Analysis Protocols. Goddard Space Flight Space Center, Greenbelt, <https://doi.org/10.25607/OBP-62>.
- Pozdnyakov S R, Kondratyev S A, Minakova E A et al. 2019. Estimation of nutrient load on the Kuibyshev reservoir from the catchment area. *Geography and Natural Resources*, **40**(3): 237-246, <https://doi.org/10.1134/S1875372819030065>.
- Qi L, Hu C M, Visser P M et al. 2018. Diurnal changes of cyanobacteria blooms in Taihu Lake as derived from GOCI observations. *Limnology and Oceanography*, **63**(4): 1711-1726, <https://doi.org/10.1002/lno.10802>.
- Roozen F C J M, Lüring M, Vlek H et al. 2007. Resuspension of algal cells by benthivorous fish boosts phytoplankton biomass and alters community structure in shallow lakes. *Freshwater Biology*, **52**(6): 977-987, <https://doi.org/10.1111/j.1365-2427.2007.01729.x>.
- Shen F, Zhou Y X, Li D J et al. 2010. Medium resolution imaging spectrometer (MERIS) estimation of chlorophyll-*a* concentration in the turbid sediment-laden waters of the Changjiang (Yangtze) Estuary. *International Journal of Remote Sensing*, **31**(17-18): 4635-4650, <https://doi.org/10.1080/01431161.2010.485216>.
- Silulwane N F, Richardson A J, Shillington F A et al. 2001. Identification and classification of vertical chlorophyll patterns in the Benguela upwelling system and Angola-Benguela front using an artificial neural network. *South African Journal of Marine Science*, **23**(1): 37-51, <https://doi.org/10.2989/025776101784528872>.
- Smith L, Watzin M C, Druschel G. 2011. Relating sediment phosphorus mobility to seasonal and diel redox fluctuations at the sediment-water interface in a eutrophic freshwater lake. *Limnology and Oceanography*, **56**(6): 2251-2264, <https://doi.org/10.4319/lo.2011.56.6.2251>.
- Stockwell J D, Doubek J P, Adrian R et al. 2020. Storm impacts on phytoplankton community dynamics in lakes. *Global Change Biology*, **26**(5): 2756-2784, <https://doi.org/10.1111/gcb.15033>.
- Vanhellemont Q. 2019. Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives. *Remote Sensing of Environment*, **225**: 175-192, <https://doi.org/10.1016/j.rse.2019.03.010>.
- Vanhellemont Q. 2020. Sensitivity analysis of the dark spectrum fitting atmospheric correction for metre- and decametre-scale satellite imagery using autonomous hyperspectral radiometry. *Optics Express*, **28**(20): 29948-29965, <https://doi.org/10.1364/OE.397456>.
- Von Luxburg U. 2007. A tutorial on spectral clustering. *Statistics and Computing*, **17**(4): 395-416, <https://doi.org/10.1007/s11222-007-9033-z>.
- Wang Y T, Zhang T Q, Zhao Y C et al. 2021. Characterization of sedimentary phosphorus in Lake Erie and on-site quantification of internal phosphorus loading. *Water Research*, **188**: 116525, <https://doi.org/10.1016/j.watres.2020.116525>.
- Wu T F, Qin B Q, Brookes J D et al. 2015. The influence of changes in wind patterns on the areal extension of surface cyanobacterial blooms in a large shallow lake in China. *Science of The Total Environment*, **518-519**: 24-30, <https://doi.org/10.1016/j.scitotenv.2015.02.090>.
- Wu T F, Qin B Q, Zhu G W et al. 2013. Dynamics of cyanobacterial bloom formation during short-term hydrodynamic fluctuation in a large shallow, eutrophic,

- and wind-exposed Lake Taihu, China. *Environmental Science and Pollution Research*, **20**(12): 8546-8556, <https://doi.org/10.1007/s11356-013-1812-9>.
- Xu H, Qin B Q, Paerl H W et al. 2021. Environmental controls of harmful cyanobacterial blooms in Chinese inland waters. *Harmful Algae*, **110**: 102127, <https://doi.org/10.1016/j.hal.2021.102127>.
- Xu H Q. 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, **27**(14): 3025-3033, <https://doi.org/10.1080/01431160600589179>.
- Xue K, Ma R H, Duan H T et al. 2019. Inversion of inherent optical properties in optically complex waters using sentinel-3A/OLCI images: a case study using China's three largest freshwater lakes. *Remote Sensing of Environment*, **225**: 328-346, <https://doi.org/10.1016/j.rse.2019.03.006>.
- Xue K, Zhang Y C, Ma R H et al. 2017. An approach to correct the effects of phytoplankton vertical nonuniform distribution on remote sensing reflectance of cyanobacterial bloom waters. *Limnology and Oceanography: Methods*, **15**(3): 302-319, <https://doi.org/10.1002/lom3.10158>.
- Yu Z Y, Ma R H, Hu M Q et al. 2024. An improved algorithm for the column-integrated algal biomass retrieval in Lake Chaohu, a large eutrophic lake. *International Journal of Applied Earth Observation and Geoinformation*, **127**: 103670, <https://doi.org/10.1016/j.jag.2024.103670>.
- Yu Z Y, Ma R H, Hu M Q et al. 2025. A brand-new algorithm for mapping algal biomass in lakes. *Journal of Remote Sensing*, **5**: 0436, <https://doi.org/10.34133/remotesensing.0436>.
- Zhang F F, Li J S, Zhang B et al. 2018. A simple automated dynamic threshold extraction method for the classification of large water bodies from landsat-8 OLI water index images. *International Journal of Remote Sensing*, **39**(11): 3429-3451, <https://doi.org/10.1080/01431161.2018.1444292>.

Electronic supplementary material

Supplementary material (Supplementary Tables S1–S4, Figs.S1–S5, and Text S1–S4) is available in the online version of this article at <https://doi.org/10.1007/s00343-026-5485-x>.