Water Research 282 (2025) 123665

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

An approach for mapping phytoplankton communities in freshwater lakes based on phytoplankton absorption features

Yuxin Zhu [a], Qingxia Miao [a], Heng Lyu [a,b,*], Yiling Zheng [a], Wenyu Liu [a], Yunmei Li [a,b], Junda Li [a], Fangfang Chen [a], Song Miao [c]

[a] *School of Geography, Nanjing Normal University, Key Laboratory of Virtual Geographic Environment of Education Ministry, PR China*
[b] *Jiangsu Center for Collaboration Invocation in Geographical Information Resource Development and Application, Nanjing, 210023, PR China*
[c] *School of Geomatics and Municipal Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou, 310018, PR China*

ARTICLE INFO

Keywords:
Phytoplankton communities
Freshwater lakes
Absorption data
Remote sensing

ABSTRACT

Phytoplankton communities play a crucial role in the lake ecosystem due to their varying characteristics, functions, and impacts of different phytoplankton groups. Understanding the composition of phytoplankton groups in freshwater lakes is essential for comprehending geochemical processes and managing water quality. In this study, an improved Diagnostic Pigment Analysis method for freshwater lakes was developed and the proportion of five major phytoplankton groups—Dinophyta, Cryptophyta, Chlorophyta, Cyanophyta, and Bacillariophyta—was derived through the absorption-decomposition method. The validation results demonstrated that the developed algorithm had satisfactory estimation accuracy for all five groups. Among all the phytoplankton groups, Cyanophyta achieved the best performance, with Median Absolute Percentage Error (MAPE) of 14.22 %, and Bias of 8.37 %. In contrast, Cryptophyta exhibited the poorest accuracy, with MAPE as high as 40.24 %. The MAPE values ranged from 10.91 % to 33.65 %, and the Bias values ranged from 1.06 % to 9.38 %. Meanwhile, the developed algorithm was successfully applied to the Ocean and Land Color Instrument (OLCI) images for mapping the spatial distribution of phytoplankton communities in Lake Taihu, demonstrating its ability to be applied to satellite imagery. This proposed algorithm provided a new approach to quantitatively determine the composition of phytoplankton communities in freshwater lakes, which can obtain valuable insights from observing the composition and succession patterns of these communities from satellite platforms.

1. Introduction

Phytoplankton play a crucial role in oceanic, and freshwater ecosystems and their associated biogeochemical cycles. It has been demonstrated that phytoplankton communities can significantly influence their ecological roles and biogeochemical functions with temporal and spatial taxonomic group diversity (Uitz et al., 2015). Aquatic environments, including inland, coastal, and open-ocean waters, are rarely composed of a single phytoplankton taxonomic group (Sathyendranath and IOCCG, 2014). Especially in freshwater lakes, phytoplankton composition significantly varies with time and space, closely related to local environmental changes and climatic shifts (Brewin et al., 2015, 2014; Mouw et al., 2017; Xi et al., 2017). At the same time, different phytoplankton groups often possess distinct characteristics and functions, leading to varying ecological impacts. For instance, some Cyanophyta produce phytoplankton toxins, while certain Chlorophyta provide food for fish. Understanding the composition of phytoplankton is essential for comprehending biogeochemical processes in water and managing water quality (Zhu et al., 2023).

Recently, High-performance liquid chromatography (HPLC) techniques have been widely adopted to extract specific pigments, and then get phytoplankton community information (Brewin et al., 2010; Mouw et al., 2017). This method is based on the theory that those specific pigments can serve as markers for specific phytoplankton groups. For example, Chlorophyll *b* (Chl-b) is a characteristic pigment of Chlorophyta (Xu et al., 2001). Therefore, A method of diagnostic pigmentation analysis (DPA) was developed for estimating phytoplankton taxonomic composition based on pigment concentration information obtained from HPLC (Vidussi et al., 2001). Uitz et al. (2006) applied the DPA method to establish a multiple linear regression algorithm between characteristic pigments and Chl-a and then calculated the weighting coefficients for each pigment, ultimately the phytoplankton size

* Correspondence author.
E-mail address: heng.lyu@gmail.com (H. Lyu).

https://doi.org/10.1016/j.watres.2025.123665
Received 11 November 2024; Received in revised form 14 April 2025; Accepted 16 April 2025
Available online 17 April 2025

composition was quantified by using Chl-a. Similarly, Li et al. (2023) refined the DPA calculation by introducing adjustments that allowed for a more accurate estimation of the composition of major phytoplankton groups in the global ocean.

While these DPA-based approaches demonstrate the effectiveness of pigment analysis in determining phytoplankton composition, traditional methods such as HPLC are time-consuming and labor-intensive, making them impractical for large-scale or long-term monitoring. Consequently, there have been increasing efforts to derive phytoplankton group composition using satellite remote sensing technology. These efforts include abundance-based approaches, such as quantifying phytoplankton biomass through cell counts or chlorophyll a concentrations (Sun et al., 2023); spectral-based approaches, which obtain phytoplankton communities using optical spectral characteristics (Mouw et al., 2017); and ecological-based approaches which derive phytoplankton communities through several ecological parameters combined with the other environmental parameters that can be obtained through remote sensing, such as sea surface temperature (SST) (Brewin et al., 2014; Sun et al., 2023). In exploring these approaches, it was found that there was a close relationship between the absorption coefficient of phytoplankton and phytoplankton community structure and biomass. As a result, several studies have focused on the phytoplankton community estimation based on absorption features. Chase et al. (2017) developed a method to retrieve phytoplankton pigment content in coastal regions from absorption data using Gaussian decomposition. Hirata et al. (2008) used the optical absorption by phytoplankton at 443 nm, $a_{ph}(443)$ to classify phytoplankton size classes in tropical and subtropical oceanic waters. Similarly, Devred et al. (2011) employed such methods to classify phytoplankton size classes, with particular attention to the North Atlantic and Southern Ocean. In the South China Sea, Sun et al. (2019) used hyperspectral absorption data and integrated spectral decomposition techniques to enhance the accuracy of phytoplankton functional type retrievals. Uitz et al. (2015) developed a derivative analysis method to differentiate various phytoplankton functional types on a global scale. Hierarchical cluster analysis was also adopted by Zhang et al. (2018) to improve phytoplankton group retrievals based on absorption characteristics in the East China Sea. These studies successfully assessed phytoplankton communities and their ecological roles using absorption-based approaches.

However, most of these studies have primarily focused on marine or near-coastal waters. There is still a lack of approaches to map the phytoplankton community composition of freshwater lakes with varying optical features. This is because the methods developed for marine environments cannot be directly applied to freshwater lakes due to their greater diversity and variability in phytoplankton communities compared to the relatively uniform distribution patterns found in the ocean (Bi and Hieronymi, 2024; Zhu et al., 2023). Additionally, varying nutrient inputs, water flow regimes, and human activities further contribute to this diversity (Zhu et al., 2023). Furthermore, existing methods for identifying phytoplankton groups in freshwater lakes remain limited. While some approaches based on laboratory-cultured algae data have been attempted for phytoplankton classification(Zhu et al., 2023), these methods exhibit significant errors when applied to field data from natural lakes. Therefore, there is still a pressing need to develop new methods specifically tailored for freshwater lakes. The DPA algorithm, which has been widely used in marine and coastal waters, has not yet been thoroughly evaluated for its applicability to freshwater lakes. Therefore, a new method combining DPA and phytoplankton absorption features was finally attempted for estimating phytoplankton communities information in freshwater lakes, and the objectives of this study are: (1) to establish an improved DPA method (IDPA) for freshwater lakes; (2) to evaluate the applicability of the developed method on mapping phytoplankton communities in freshwater lakes using satellite imagery; (3) to map phytoplankton composition characteristics in typical freshwater lakes.

2. Materials and methods

2.1. Study area and sampling sites

In this study, to ensure the diversity of data and the applicability of the methodology, Lake Dianchi, Lake Erhai, Lake Qiandaohu, and Lake Taihu in China with different eutrophication levels and rich phytoplankton compositions are selected as research lakes, with Lake Taihu being hyper-eutrophic (Trophic Level Index (TLI) $>$ 70) and Lake Dianchi exhibiting severe eutrophication (TLI $>$70). Meanwhile, Lake Erhai is classified as mesotrophic (TLI between 40 and 50) and Lake Qiandaohu is characterized by its oligotrophic conditions and cleaner water (TLI $<$ 40). The field campaigns were conducted in May, June, August, and November of 2023, and a total of 105 sampling sites were finally obtained (Fig. 1). At each site, 5 L of water was collected to measure the absorption coefficients of pigment particles, and the Chl-a concentration and analyze pigment composition through Agilent High-Performance Liquid Chromatography (HPLC). Additionally, another 1 L of water was sampled to determine the biomass and species composition, with cellular morphology fixed using Lugol's reagent.

Pigments analysis was conducted using an Agilent HPLC 1260 Series with a quaternary pump, an auto-sampler, a Poroshell 120 EC—C18 column (150 × 3.0 mm; 2.7 μm particle size), and a photodiode array detector. The extraction and separation pigment procedure modified by Chen et al. (2001) was adopted. Pigments were calibrated using an authentic standard (Denmark DHI). The pigments, Chlorophyllide b, Chlorophyllide a, Chlorophyll-c3, Chlorophyll-c2, Peridinin, Pheophorbide a, Fucoxanthin, Aphanizophyll, Neoxanthin, Violaxanthin, Diadinoxanthin, Myxoxanthophyll, Astaxanthin, Alloxanthin, Diatoxanthin, Lutein, Canthaxanthin, Chlorophyll-b (Chl-b), Echinenone, Chlorophyll-a (Chl-a), Echinenone-Chl-a, Divinyl Chl-a, Pheophytin b, Pheophytin b', Pheophytin a, Pheophytin a', β-carotene, Total Chl-a (T Chl-a), 28 types in total (in μg/L) were finally detected (Li et al., 2022).

The absorption coefficients of all particles ($a_p(\lambda)$), algal particles ($a_{ph}(\lambda)$), and non-algal particles ($a_{nap}(\lambda)$) (in m^{-1}) were determined using the Transmittance-Reflectance method by UV-3600Plus series (UV 3600Plus/UV 3600iPlus) (Cai et al., 2023). Also, the Chl-a concentration was analyzed spectrophotometrically at 750 and 665 nm with a correction for phaeopigments by UV-3600Plus series (UV 3600Plus/UV 3600iPlus). The Chl-a specific absorption coefficient ($a^*_{pH}(\lambda)$) was then calculated using $a_{ph}(\lambda)$ divided by the concentration of T-Chla measured by HPLC.

The phytoplankton biomass was determined using the cell volume conversion method. The phytoplankton community composition was analyzed through microscopy and identified to genus. Each genus of phytoplankton was counted using the random field method. This enabled the calculation of the abundance of each species, which was then summed to obtain the total phytoplankton abundance. As the phytoplankton cell density is close to that of water, 1 mm^3 is approximately equal to 1 mg of fresh-weight biomass. Therefore, the biomass (in mg/L) can be derived through the product of volume (in μm^3) and abundance (in cells/L). The proportion of the specific phytoplankton group is then determined by the biomass (in mg/L) of the specific group divided by the total biomass.

2.2. Satellite images data

The Ocean and Land Color Instrument (OLCI) sensor, which has 21 bands in the range of 400–1020 nm and a spatial resolution of 300 m, is a successor to the ENVISAT Medium Resolution Imaging Spectrometer Instrument (MERIS). It is one of the payloads on the Sentinel-3 satellite and is suitable for monitoring freshwater lakes. It has been proven to have higher performance than its predecessors (Bi et al., 2019, 2018; Shen et al., 2017).

OLCI level 1B images covering Lake Taihu were obtained from the European Space Agency data hub (https://scihub.copernicus.eu/).

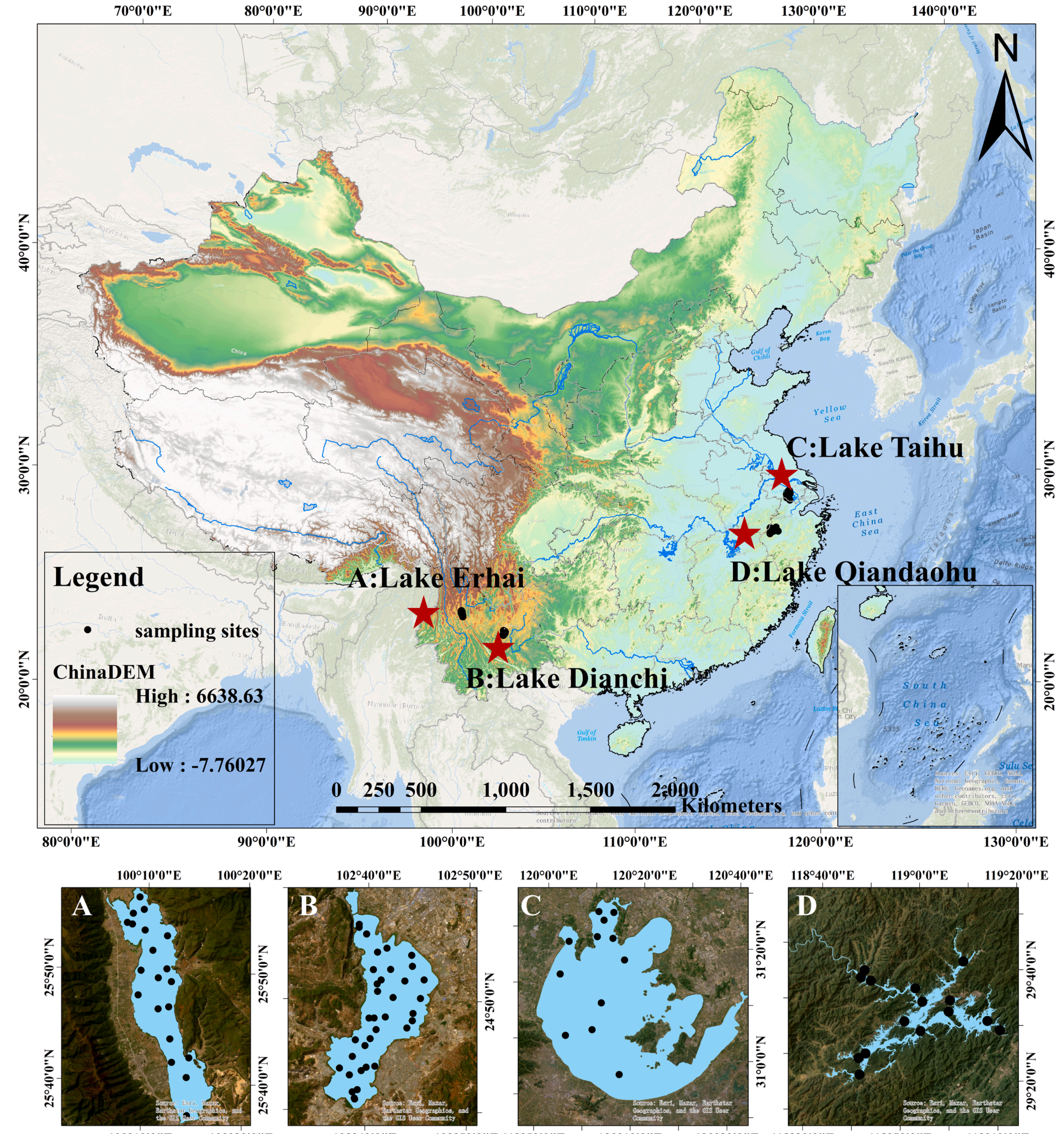

Fig. 1. Diagram of the study area.

Image processing, including subsetting, Rayleigh scattering correction, aerosol scattering correction, and reprojection, was performed using the SeaWiFS data analysis system (SeaDAS, version 7.4). The Management Unit Mathematical Models (MUMM) was used to correct aerosol scattering for the Rayleigh-corrected OLCI images. This algorithm has been widely used for turbid freshwater lakes (Bi et al., 2018; Doron et al., 2011; Ruddick et al., 2006, 2000). The output data mainly included remote sensing reflectance Rrs(λ) and Rayleigh-corrected reflectance Rrc(λ). In this study, Rrc(λ) was used for masking the cloud and blooms. The cloud index, CI method (Zhai et al., 2018) was employed for cloud masking, utilizing Rrc(443), Rrc(560), Rrc(660), and Rrc(754) to compute the CI. An adjustment coefficient of 1/15 was applied to determine the threshold, enabling effective identification of cloud-contaminated pixels across diverse imagery. Additionally, the CSI method proposed by Zhu et al. (2018) was adopted to eliminate algal blooms, with a threshold set at 0.11. This method maintains high accuracy even in highly turbid waters.

2.3. Diagnostic pigment analysis method: DPA

Diagnostic Pigment Analysis (DPA), originally proposed by Vidussi et al. (2001) and further improved by Uitz et al. (2006), is a well-established approach for classifying phytoplankton groups based on their specific pigment signatures. Phytoplankton communities comprise diverse taxonomic groups, each distinguished by unique pigment compositions. The DPA approach establishes a multiple linear regression relationship between characteristic pigments and chlorophyll-a (Chl-a), enabling the determination of weighting coefficients for each characteristic pigment. In this study, the DPA method was applied to calculate the composition of phytoplankton communities using Eq. (1):

$$C_{T\ Chl-a} = \sum_{i}^{n} w_i\, p_i \tag{1}$$

where $C_{T\ Chl-a}$ is the concentration of total Chl-a (T Chl-a), n is the number of groups of phytoplankton involved in the calculation, w_i is the weight of each characteristic pigment, and p_i is the concentration of different characteristic pigments.

Then the w_i determined from the rate of Eq. (1) could then be used to calculate the percentage of Chl-a for different phytoplankton groups (F_i) using Eq. (2):

$$F_i = \frac{w_i p_i}{C_E} \tag{2}$$

where w_i is the weight of each characteristic pigment, p_i is the concentration of different characteristic pigments, and C_E is the Chl-a concentration obtained from the weighted estimation of different weighted character pigments.

2.4. The method for obtaining phytoplankton community information

In this study, the absorption-decomposition method was adopted to obtain the composition of phytoplankton communities. The absorption coefficient of algal particles ($aph(\lambda)$) is the sum of the absorption coefficients contributed by different phytoplankton groups. The absorption coefficient contributed by each group can be obtained by multiplying the Chl-a-specific absorption coefficient of different phytoplankton groups ($a_i^*(\lambda)$) by its corresponding concentration (C_i). Therefore, in this study, $aph(\lambda)$ was calculated by linearly summing the products of $a_i^*(\lambda)$ and the corresponding concentration (C_i) for different phytoplankton communities, as shown in the following equation:

$$aph(\lambda) = \sum_{i=1}^{n} C_i a_i^*(\lambda) \tag{3}$$

By dividing the T Chl-a on both sides, Eq. (3) can be further expressed as:

$$a^*ph(\lambda) = \sum_{i=1}^{n} F_i a_i^*(\lambda) \tag{4}$$

where F_i is the fraction of total Chl-a of the ith phytoplankton group obtained from the DPA method.

The F_i obtained through the DPA method and the $a^*ph(\lambda)$ measured by UV-3600Plus were substituted into Eq. (4). By applying the bootstrap method and a constrained least squares approach with a non-negative lower bound, the Chl-a-specific absorption coefficient ($a_i^*(\lambda)$) can be derived. Then, the calculated $a_i^*(\lambda)$ and the $aph(\lambda)$ measured by UV-3600Plus were incorporated into Eq. (3) Employing the same computational framework, the chl-a concentration of different phytoplankton groups (C_i) can be obtained. Finally, the proportion of different phytoplankton groups (F_{i-simu}) can be calculated using Eq. (5).

$$F_{i-simu} = \frac{C_i}{\sum_{i=1}^{5} ci} \tag{5}$$

2.5. The estimation method of a_{ph}: QAA-750E

The IOP inversion algorithm (QAA-750E) developed by Xue et al. (2019) for lakes was applied to calculate the phytoplankton absorption. This method has satisfactory performance for estimating the phytoplankton absorption coefficient using 750 nm as the reference wavelength. In this study, $a_{ph}(674)$ was derived based on Sentinel-3 OLCI imagery by using the QAA-750E algorithm.

2.6. The evaluation of the algorithm performance

The performance of the model is evaluated by using accuracy evaluation metrics, including MAE (Mean Absolute Error), RMSE (Root Mean Square Error), MAPE (Median Absolute Percentage Error), and Bias. The coefficient of determination, R^2, was also adopted to evaluate the fitness of the model.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\overline{y_i} - y_i| \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\overline{y_i} - y_i)^2} \tag{7}$$

$$MAPE = Median \sum_{i=1}^{n} \left|\frac{\overline{y_i} - y_i}{y_i}\right| * 100\% \tag{8}$$

$$Bias = bias = 10^Y - 1, Y = \frac{\sum_{i=1}^{n} (\overline{y_i} - y_i)}{n} \tag{9}$$

where n is the number of sample points, $\overline{y}_i$ is the simulated data, and y_i is the measured data.

3. Results

3.1. The improved DPA method for freshwater lakes (IDPA)

Since the previous DPA coefficients were determined based on phytoplankton communities' characteristics of the near coast or the ocean, thus, these coefficients cannot be applied to the freshwater lakes. The extreme variability of phytoplankton structure in freshwater lakes indicates that it is challenging to represent the phytoplankton composition of different lakes using a single formula. Therefore, in this study, the water bodies were categorized into Cyanophyta-dominated and non-Cyanophyta-dominated water types for better performance. The Cyanophyta biomass percentage greater than 60 % was determined as the Cyanophyta-dominated type (BinGuo et al., 2011; Zhu et al., 2023), and the other water bodies were labeled as non-Cyanophyta-dominated type. Consequently, the weights of each pigment for two water types were calculated using multivariate linear fitting and the bootstrap method (100 iterations). The final weight coefficients of IDPA were presented in Table 1 (R^2>0.95).

After the different w_i for the two types of water bodies were obtained, Eq. (2) was used to calculate the F_i for the different phytoplankton groups. Once F_i was determined the compositional information of different phytoplankton groups can then be derived using the method mentioned in Section 2.4. Additionally, the wavelengths of 400 nm, 460 nm, 500 nm, 520 nm, 560 nm, 590 nm, 600 nm, 620 nm, 650 nm, 665 nm, 680 nm, and 700 nm were used in performing the calculations.

Table 1
w_i for the IDPA method.

	Phytoplankton	Type1 Cyanophyta dominated	Type2 non-Cyanophyta dominated
Peridinin+ Diadinoxanthin	Dinophyta	0.53	0.01
Alloxanthin	Cryptophyta	0.01	0.01
Diatoxanthin + Fucoxanthin	Bacillariophyta	0.27	0.09
Echinenone + Myxoxanthophyll	Cyanophyta	2.41	0.78
Chl b	Chlorophyta	0.16	3.06

3.2. The accuracy evaluation of the approach based on phytoplankton absorption features

3.2.1. Evaluation of the accuracy of $\mathbf{a}_i^*(\boldsymbol{\lambda})$

From Section 2.4, it was known that $a_i^*(\lambda)$ played an important role in deriving proportion of different phytoplankton groups, therefore, it is essential to evaluate the accuracy of the calculation of $a_i^*(\lambda)$. However, it is rare to find samples that are entirely dominated by a single phytoplankton group in natural aquatic ecosystems. Thus, the sample point with the highest biomass proportion of the specific phytoplankton group was determined as a water body predominantly dominated by that phytoplankton group (Sun et al., 2022). Specifically, the biomass proportion of Cyanophyta, Chlorophyta, Bacillariophyta, Dinophyta and Cryptophyta at these points was greater than 60 %, Also, $a_i^*(\lambda)$ were compared with the absorption curves of cultured pure algae, as shown in Fig. 2. Since only Cyanophyta, Chlorophyta, and Bacillariophyta were cultured in the laboratory, no comparisons were made between pure algae and the in situ samples for Dinophyta and Cryptophyta. The magnitude of $a_i^*(\lambda)$ may vary depending on cell size and concentration during culture, so only the shapes of the of $a_i^*(\lambda)$ curves were compared. After normalizing the curves to a range between 0 and 1. Afterward the Pearson correlation coefficients (r) between the derived $a_i^*(\lambda)$ and cultured pure algae and those water bodies predominantly dominated by the specific group were calculated. In addition, MAPE, MAE, RMSE and Bias were also calculated using the values of the dominant water as provisional true values, as shown in Fig. 2.

From the results shown in Fig. 2A, From the results we can see that, it can be seen that the $a_i^*(\lambda)$ of Bacillariophyta exhibited high concordance with the dominant water (r_1=0.97, r_2=0.97, RMSE=0.34m^2/mg, MAE=0.06m^2/mg, MAPE=32 %, Bias=−0.07m^2/mg). Similarly, the derived $a_i^*(\lambda)$ of Dinophyta and Cryptophyta illustrated in Fig. 2D and E, separately, showed excellent agreement with the measured $a_i^*(\lambda)$ of specific phytoplankton group dominated water with r_2>0.96, RMSE<0.30m^2/mg, MAE<0.06m^2/mg, MAPE<25 %, and |Bias|< 0.07m^2/mg. However, a noticeable difference was observed in the specific absorption at 500 nm when compared to the pure algae, which displayed a more pronounced shoulder at this wavelength, as seen in Cyanophyta and Chlorophyta (shown in Fig. 2B, C). Nonetheless, the correlation between these results remained pretty high, with r_1>0.91, r_2>0.95, RMSE<0.27m^2/mg, MAE<0.05m^2/mg, MAPE<36 %, and |Bias|<0.09m^2/mg. The pronounced shoulder at 500 nm observed in pure algal samples, particularly in Cyanophyta and Chlorophyta, was typically attributed to specific pigments (like phycoerythrin) that were more prevalent in cultured environments, making these features more frequently detected in laboratory-grown algae. In natural water bodies, however, various factors—including mixed populations, varying light conditions, and nutrient availability—can lead to shifts in these spectral features. These observed discrepancies at 500 nm between the inverted $a_i^*(\lambda)$ and the measured pure algae absorption curves suggested that environmental influences may play a significant role in altering the optical properties of algae.

Although some of the spectral features present in the pure algae were not captured by the $a_i^*(\lambda)$ obtained in this study, the key characteristic features of the different groups were successfully inverted. The peak at 620 nm for Cyanophyta and the peak at 650 nm for Chlorophyta were accurately constructed. These characteristic bands directly affect the reliability of the estimation results when using $a_i^*(\lambda)$ to obtain phytoplankton communities, and it demonstrated capturing key taxonomic features was crucial to the robustness of the algorithm in capturing key taxonomic features.

3.2.2. Evaluation of the accuracy of $\mathbf{F}_{\mathbf{i-simu}}$

The derived specific phytoplankton group proportion using Eq. (3) and Eq. (5) was evaluated by comparing with the measured proportion based on an independent validation dataset (comprising one-third of the sampling data from each lake). The scatterplot of measured proportion and derived proportion is shown in Fig. 3. It can be seen from Fig. 3 that the accuracy was pretty satisfactory across all five groups.

For Dinophyta (shown in Fig. 3A), the sample points were mostly distributed near the 1:1 line, except for a few points showing significant underestimation. The precision metrics indicated good performance with an RMSE of 1.49 %, MAE of 3.48 %, MAPE of 33.60 %, and a Bias of 8.22 %. It was found that the performance for Cryptophyta estimation shown in Fig. 3B was the poorest among the five groups, likely due to their pretty lower proportion, with the highest proportion being less than 0.05 %. Despite this, the performance was acceptable with the precision metrics of an RMSE of 0.04 %, MAE of 0.06 %, MAPE of 40.24 %, and a Bias of 0.005 %. For Bacillariophyta shown in Fig. 3C, the results showed better accuracy than Cryptophyta but were slightly worse than Cyanophyta and Chlorophyta, with an RMSE of 3.83 %, MAE of 6.70 %, MAPE of 25.60 %, and a Bias of 9.38 %. Fig. 3D exhibited excellent performance for estimating Cyanophyta with an RMSE of 7.61 %, MAE of 10.28 %, MAPE of 14.22 %, and a Bias of 8.37. Fig. 3E also demonstrated relatively satisfactory accuracy with an RMSE of 7.96 %, MAE of 13.52 %, MAPE of 10.91 %, and a Bias of 1.06 % for Chlorophyta.

In summary, the algorithm performed quite satisfactorily across all groups, with minor discrepancies likely due to specific characteristics such as concentration levels and the inherent variability in the dataset. It was found that those outliers were generally located in the region with exceptionally high Chl-a concentrations, particularly in water bodies where surface scum or blooms were present.

3.3. The application of the developed algorithm in lake taihu based on satellite data

To assess the potential applicability of our algorithm on satellite images, the developed algorithm was finally applied to Lake Taihu, a typical eutrophic lake. Here, the QAA-750E method was adopted to derive *aph* at 674 nm. The *aph* at other bands were derived based on the empirical relationship with *aph* at 674 nm. And the calculation formulas were listed in Table 2, and these relationships were developed based on the dataset with a total of 1836 samples collected from freshwater lakes in recent years.

The OLCI image acquired on August 12, 2023, was used to map phytoplankton communities in Lake Taihu. The in situ data from Lake Taihu indicated that Cryptophyta and Dinophyta accounted for less than 10 % of the total biomass and exhibited slight spatial variation, while the three dominant groups: Cyanophyta, Chlorophyta, and Bacillariophyta account for a significant portion of the total biomass, therefore, the proportions of these three dominant phytoplankton groups were finally analyzed for Lake Taihu. For further evaluating the algorithm performance when applied to real satellite images, 12 satellite-ground synchronous validation points were used to calculate the algorithm estimation accuracy. The atmospheric correction precision was ensured with a MAPE of 27.18 %. The comparison between satellite-derived values and measured values is illustrated in Fig. 4. Based on the

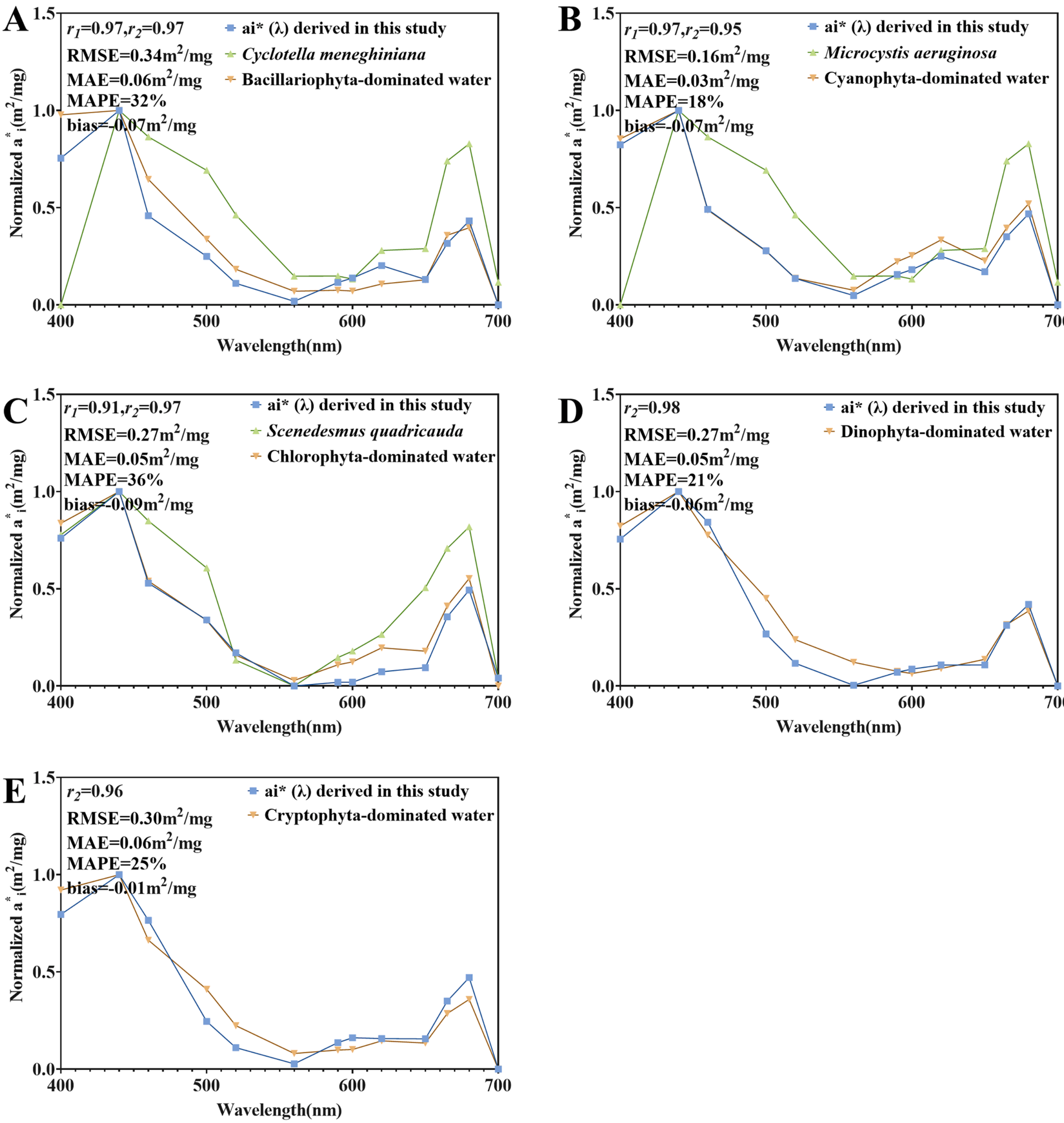

Fig. 2. Comparison of the $a_i^*(\lambda)$ derived in this study with those of pure algae cultivated in the laboratory and with those of water bodies absolutely dominated by this group (r_1 represents the r between the derived $a_i^*(\lambda)$ and cultured pure algae, and r_2 represents the r between the derived $a_i^*(\lambda)$ and water predominantly dominated by the specific group).

calculated accuracy metrics, Bacillariophyta showed acceptable accuracy, with an RMSE of 2.42 %, MAE of 17.15 %, MAPE of 43.65 %, and Bias of 0.021 %. The low Bias indicated that there was no substantial overestimation or underestimation, while the relatively high MAPE was likely due to two outlier points that contributed to the larger overall error. In contrast, Cyanophyta demonstrated pretty high precision, with the lowest MAPE of the three groups. The accuracy metrics demonstrated that the performance in deriving Chlorophyta was also acceptable with a MAPE of 31.53 %. Overall, despite some variation in the derived accuracy for different phytoplankton groups, the validation results showed that the general change trends of phytoplankton communities can be well captured by satellite observation.

Consequently, the developed algorithm was applied to satellite imagery of Lake Taihu, and the spatial distribution of phytoplankton communities is shown in Fig. 5. It can be seen that Cyanophyta was the dominant phytoplankton group in Lake Taihu, followed by Bacillariophyta and chlorophytes, with other phytoplankton groups occupying a minimal portion of the total biomass, which was consistent with previous research (Chen, 2003; Li et al., 2019). Spatially, the percentage of Cyanophyta was higher in Zhushan Bay, Meiliang Bay, and the western

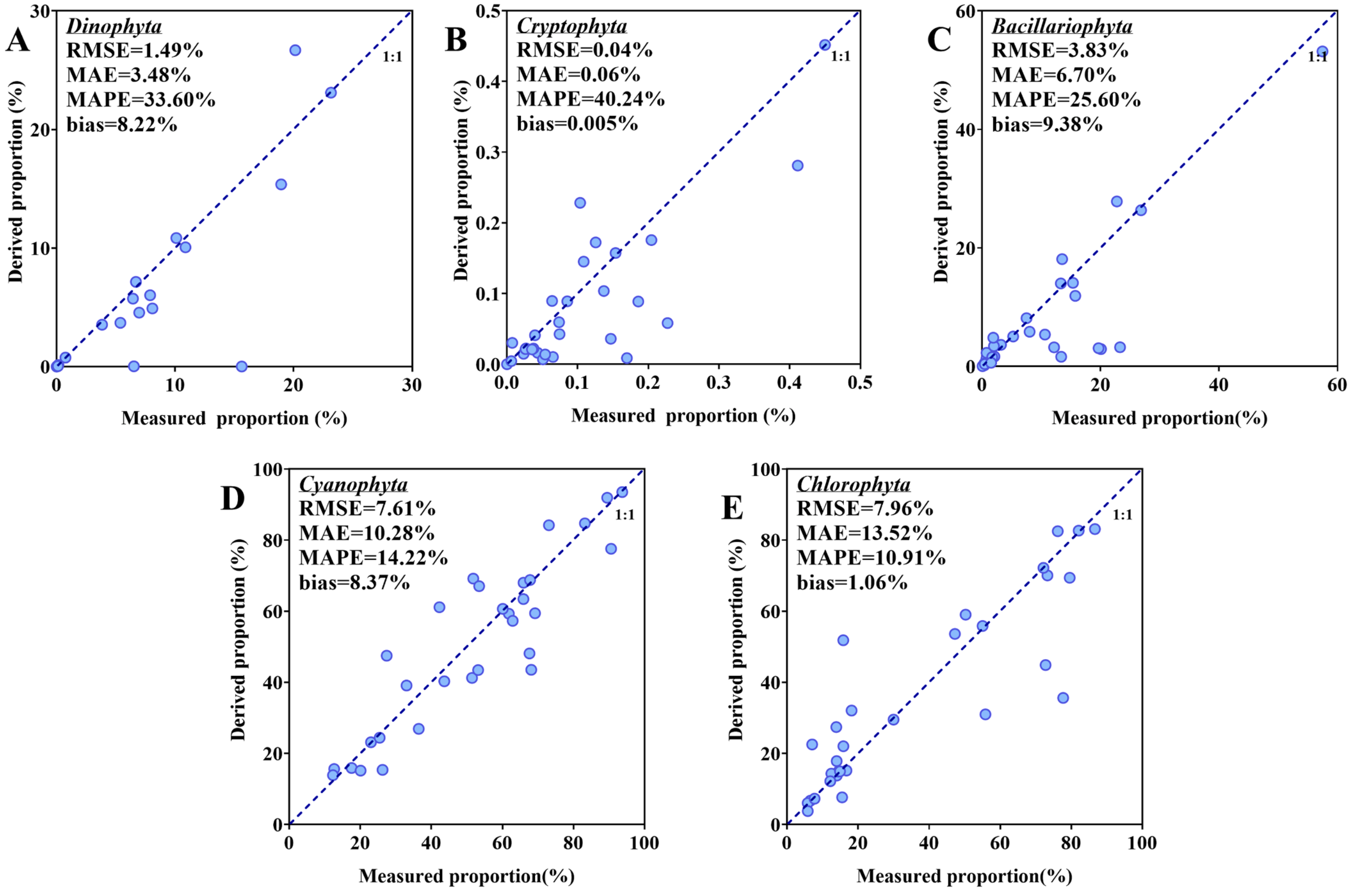

Fig. 3. Comparison of derived and measured proportion (%) for the validation dataset.

Table 2
Formulas for calculating $aph(\lambda)$ in different bands from $aph(674)$.

Wavelength	Formula
520nm	$y = 0.8682x + 0.0407$ $R^2 = 0.94$
560nm	$y = 0.4368x + 0.0613$ $R^2 = 0.86$
620nm	$y = 0.7061x - 6E\text{-}05$ $R^2 = 0.96$
665nm	$y = 0.8284x + 0.0167$ $R^2 = 0.99$
681nm	$y = 0.9991x - 0.0083$ $R^2 = 0.99$

where $x = aph(674)$ was inverted from QAA-750E (Xue et al., 2019).

part of Lake Taihu than in other regions, and this spatial distribution was also found in a previous study (Zhu et al., 2018), the high values in the eastern region may be due to high winds and waves causing floating algal particles to hang onto the littoral zone, and, this is a phenomenon that we have observed in many previous field campaigns. The spatial distribution of Chlorophyta was more homogeneous, higher in the southwestern littoral and eastern part, and slightly lower in other parts of the lake. The percentage of Bacillariophyta was mostly concentrated in Zhushan Bay, and in the eastern part of Lake Taihu.

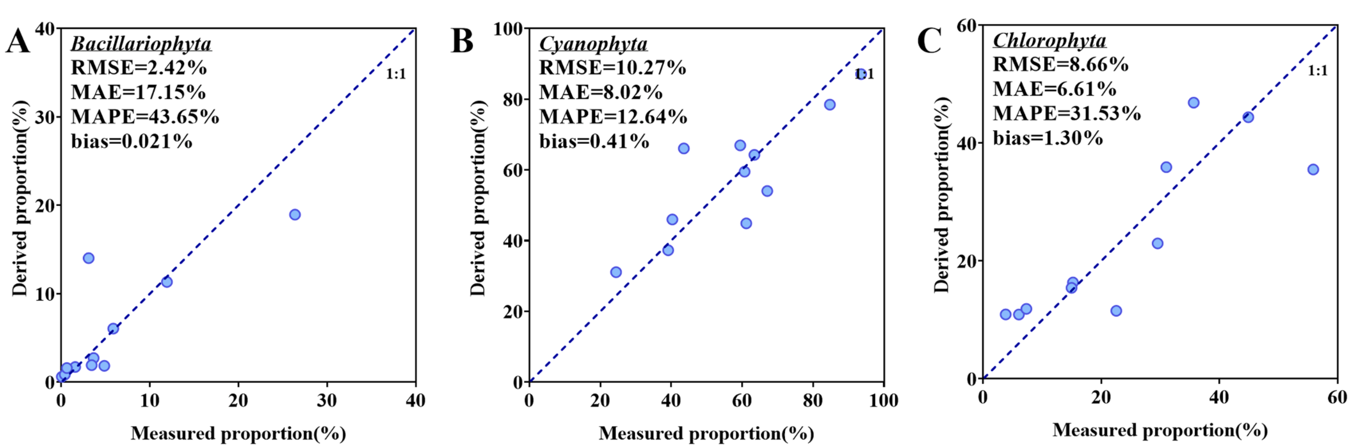

Fig. 4. Satellite-ground synchronization validation results.

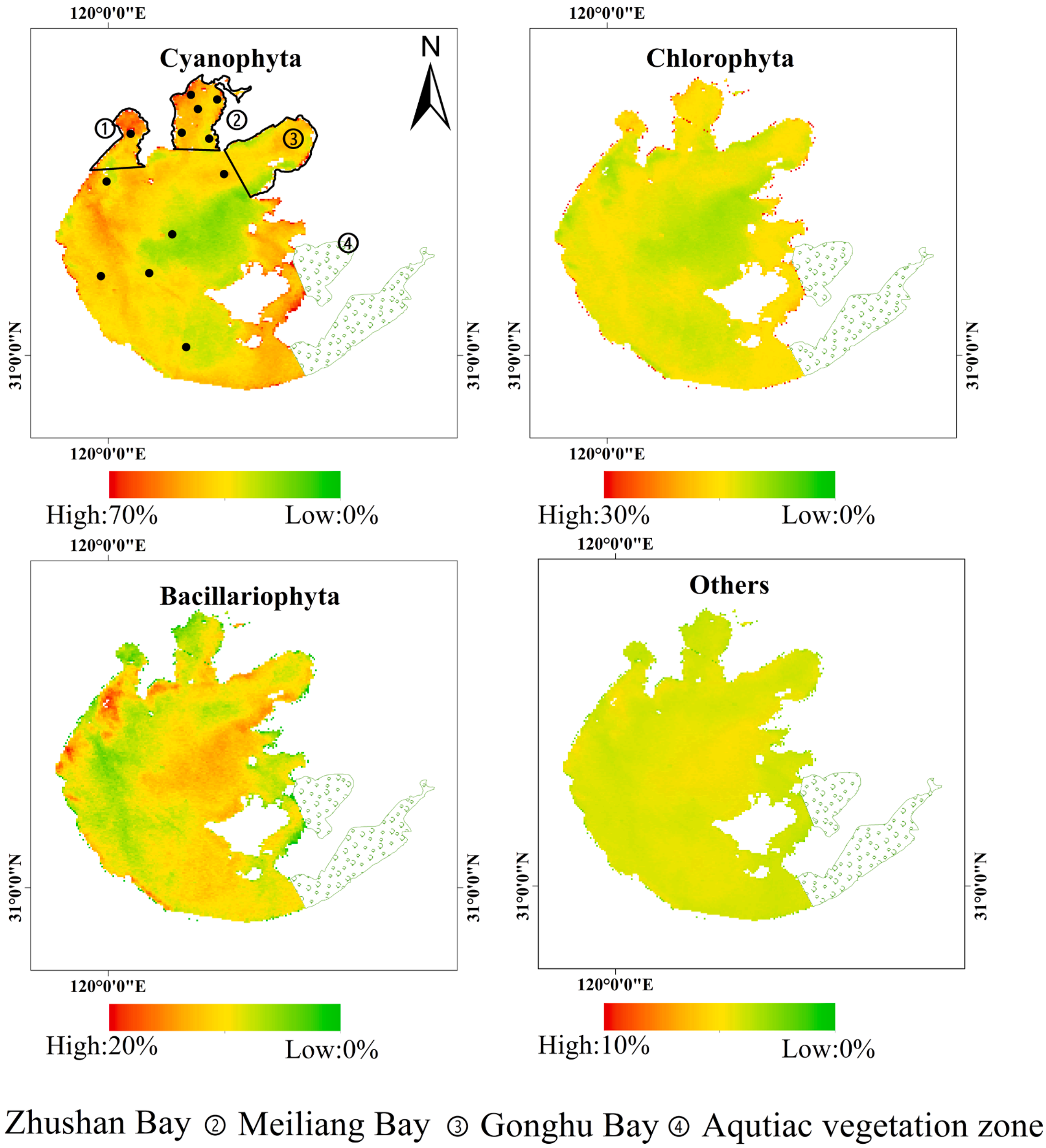

Fig. 5. Spatial distribution of the proportion of different phytoplankton groups in Lake Taihu derived from OLCI images on August 12th, 2023.

4. Discussion

4.1. Why water classification is necessary for obtaining phytoplankton communities information?

As mentioned before, the phytoplankton communities in freshwater lakes are highly variable, differing not only from lake to lake but also within different zones of the same lake. The dominant algae in various water bodies can vary significantly, for example, *Microcystis* dominates Lake Biwa in Japan (Ozawa et al., 2005), while Bacillariophyta dominate Lake Finjasjön in Sweden (Parsons, 1985; Pridmore and Etheredge, 1987), *and Microcystis* is absolutely dominant in Lake Taihu (Chen, 2003; Deng et al., 2014). This variability indicates that it is challenging to represent the phytoplankton composition of different lakes with a single algorithm. In this study, the water bodies were classified into two categories: Cyanophyta-dominated and non-Cyanophyta-dominated water types. Based on the measured data, it was found that the relationship between diagnostic pigments (e.g., echinenone for Cyanophyta) and total chlorophyll a (Chl a) exhibited distinct trends when the phytoplankton community was dominated by Cyanophyta compared to non-Cyanophyta groups. The possible reasons for this phenomenon may be that the unique pigments of Cyanophyta are not only taxonomically diagnostic but also play a role in their adaptation to specific environmental conditions, such as high light intensity or nutrient limitation (Jeffrey et al., 1997; MacIntyre et al., 2000). Driven by these factors, the relationship between marker pigments and T chl-a may undergo changes. To evaluate the estimation performance after classification, the results obtained after categorization were compared with those calculated directly without categorization. Eventually, the relationship between the Chl-a concentration calculated using different coefficients and the measured Chl-a concentration was shown in Fig. 6. It was evident that the results obtained using two sets of coefficients were more accurate as shown in Fig. 6B and C. It can be seen from Fig. 6A, that several validation points with large errors (circled in red in Fig. 6A) can be obviously observed, which may lead to significant Bias in subsequent calculations. The water classification made an attempt to avoid large deviations during the process of estimating phytoplankton communities.

In addition, when Eq. (5) was used to derive $a_i^*(\lambda)$, it was found that the issue of convergence is still a challenge. Therefore, optimal inputting of spectral bands is critical to derive the results. Previous studies have

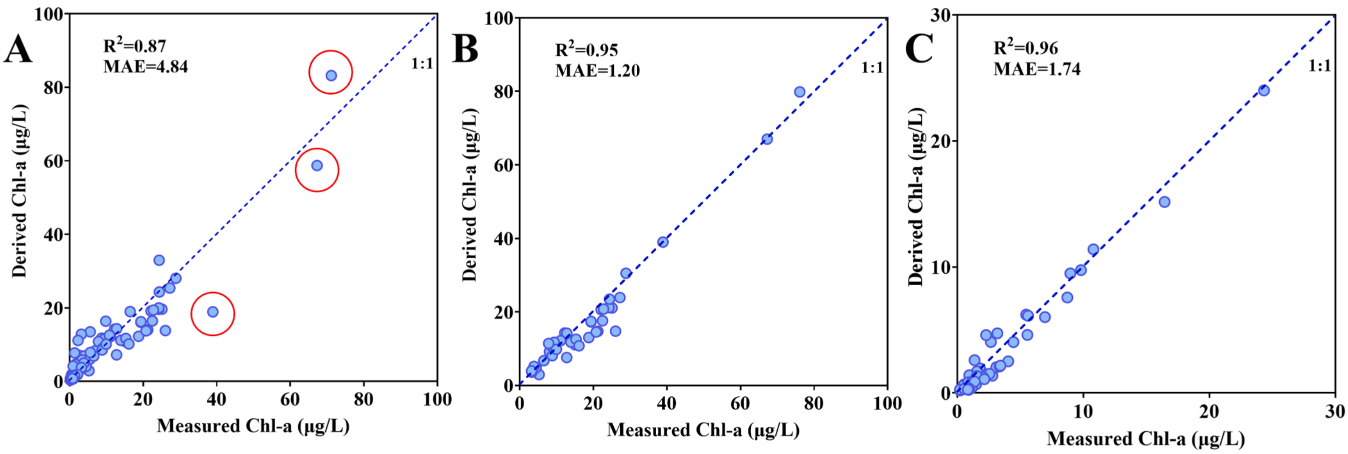

Fig. 6. Derived Chl-a concentration vs weighted derived results of different weighted characteristic pigments (A is the unclassified result, B is the result of Type1, and C is the result of Type2).

demonstrated that second-order derivative spectra are highly effective for the qualitative identification of pigments, providing better performance than standard spectra (Bidigare et al., 2002). Additionally, second-order derivative spectra have been shown to yield superior results in distinguishing different phytoplankton groups (Uitz et al., 2015). Based on these findings, the second-order derivatives of $aph(\lambda)$ were calculated, and their r with Chl-a concentration was assessed. The Pearson r between second-order derivatives of $aph(\lambda)$ and Chl-a concentrations of different phytoplankton groups were drawn in Fig. 7, and it can be read that significant correlations were obviously observed primarily 460 nm, 500 nm, 520 nm, 560 nm, 590 nm, 600 nm, 620 nm, 650 nm, and 680 nm. Considering the bands configuration of ocean or land satellites, the commonly used band at 665 nm was also included. In addition, in order not to miss features at 460 nm and 680 nm, the bands at 400 nm and 700 nm, the characteristic band of peridinin and Chl-a, were finally added. All input spectral bands were finally determined based on their capacity to capture distinct signatures of phytoplankton groups. These bands were then used as input for Eqs. (3) and (5) to estimate the relative percentage distribution of different phytoplankton groups. This approach ensures that the most characteristic wavelengths corresponding to the specific absorption features of each phytoplankton group can be fully employed, thus the accuracy of estimating group composition can be significantly improved.

4.2. Stability of IDPA coefficients for freshwater lakes

The impact of variations in the IDPA coefficients on the derived results was extensively assessed by incorporating systematic noise into

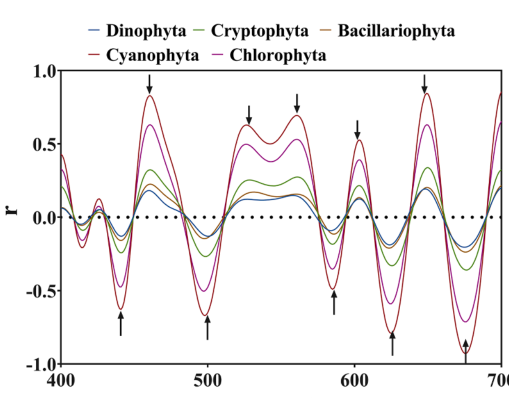

Fig. 7. Pearson r between second-order derivatives of $aph(\lambda)$ and Chl-a concentrations of different phytoplankton groups.

these coefficients. Specifically, the noise levels of ±10 % and ±20 % were introduced to evaluate the sensitivity of our results to changes in the coefficients. The analysis was conducted using multiple linear regression to train the coefficients, and the subsequently derived Chl-a concentration comparison between adding noise and not adding noise was depicted in Fig. 8.

The results revealed that the derived results were relatively stable, with RMSE<0.34μg/L, MAE = 0.10μg/L, MAPE = 0.52 %, and |Bias |< 0.16 when the systematic noise was ±10 %. The validation points are still distributed close to the 1:1 line, indicating that minor variations in the coefficients do not significantly affect the accuracy. However, when the noise level increased to ±20 %, there was a noticeable impact on the results with RMSE<1.72μg/L, MAE< 1.37, MAPE < 9.59 %, and |Bias |< 6.70.

Notably, the impact of noise was most pronounced when it was applied to the coefficients related to Cyanophyta, suggesting that the coefficient of Cyanophyta was more sensitive to variations and may require more precise calibration. To address potential issues arising from variations in coefficient values, a bootstrap method during the training phase of the coefficients was employed. This self-sampling technique mitigated the risk that minor changes in individual factors will disproportionately influence the final results. By adopting this method, the derived results remained reliable and resilient against minor fluctuations in the weighting coefficients.

To further investigate the stability of the derived weight coefficients, the lakes were also classified into two categories: eutrophic lakes, i.e., Lake Taihu and Lake Dianchi (TLI>70), and relatively clearer lakes, i.e., Lake Erhai and Lake Qiandaohu (TLI<50), for which the weight coefficients were computed separately. The recalculated weight coefficients were then compared with the w_i calculated in Section 3.1 to assess the stability of the weight coefficients across different trophic levels. Due to the significant variation among the coefficients of different groups, all the data were log-transformed for easier display, and the results are shown in Fig. 9. Overall, the differences between the calculated w_i for different trophic levels lakes and the calculated w_i for the whole lakes were relatively weak, and the two sets of coefficients obtained from the two datasets were relatively consistent. This demonstrated that the coefficients calculated for IDPA were relatively stable and that the strategy of calculating the coefficients for the Cyanophyta-dominated and non-Cyanophyta-dominated water types, separately, was completely feasible and could improve the accuracy and stability of the estimation algorithm.

The broad applicability of our methodology across various lake types has been demonstrated. Lakes with diverse environmental conditions, ranging from eutrophic to oligotrophic, were selected, and a flexible framework adaptable to diverse freshwater lake ecosystems was

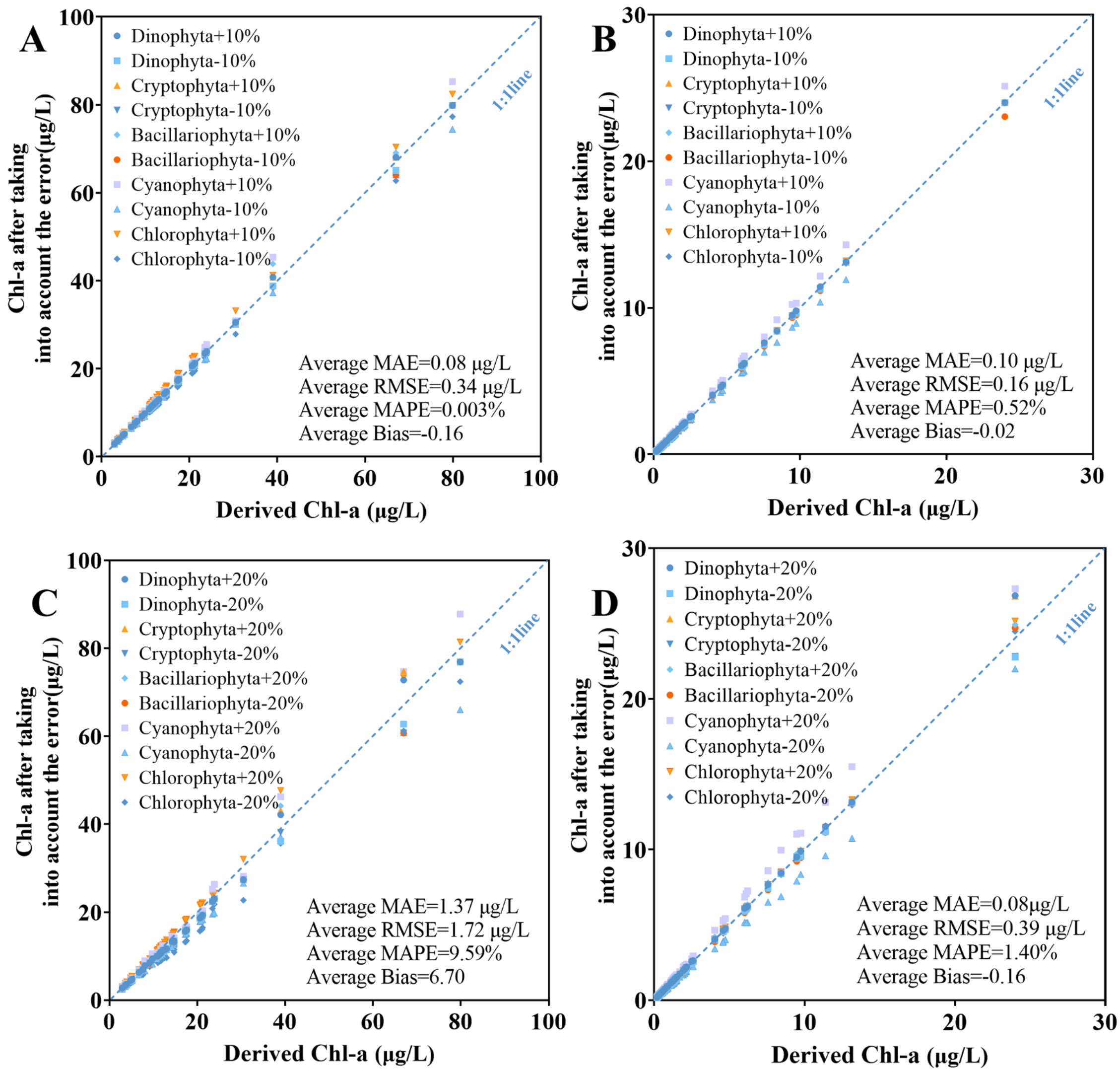

Fig. 8. Comparison of noise-added derived Chl-a and original derived Chl-a after introducing ±10 % (A and B) and ±20 % (C and D) systematic error (A and C are results for Type1, B and D are results for Type2).

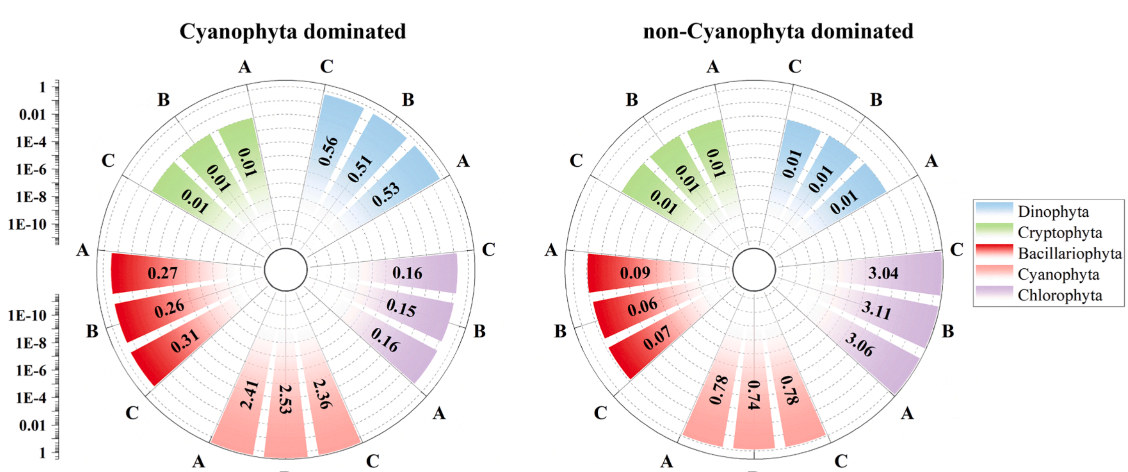

Fig. 9. Log-transformed weight coefficients for different trophic levels (A represents w_i calculated in Section 3.1, B represents w_i calculated for eutrophic lakes, C represents w_i calculated for relatively clearer lakes).

established. Noise-injection methods and the categorization of lakes into distinct groups for recalibration purposes were utilized to confirm the relative stability of coefficients of IDPA. What's more, the suitability of the absorption decomposition method for broader applications is further enhanced by its inherent comprehensiveness. The effectiveness of our methodology is validated, and its potential for future studies across various aquatic environments is highlighted by these findings, which offer valuable insights for environmental monitoring and management strategies.

4.3. Differences in phytoplankton composition between freshwater lakes and oceans

Phytoplankton composition varies significantly between freshwater lakes and marine environments due to differences in nutrient availability, light conditions, and hydrodynamics. Freshwater systems are often dominated by cyanobacteria, green algae, and diatoms, while marine ecosystems tend to support a higher abundance of diatoms, dinoflagellates, and coccolithophores (Reynolds, 2006; Smayda, 1997). These distinct phytoplankton communities exhibit different bio-optical properties, particularly in their absorption spectra, which are influenced by the unique pigment compositions of each group.

Building on these differences, Cyanobacteria, common in freshwater environments, typically have lower absorption in the blue and red spectral regions compared to marine diatoms due to the presence of phycobiliproteins rather than Chl-c (Paerl and Otten, 2013). Moreover, the variability of phytoplankton groups in freshwater lakes is much significant spatially and temporally (Bi and Hieronymi, 2024). Therefore, applying these algorithms directly to freshwater lakes without adequate recalibration can lead to inaccurate estimation results, especially in regions where cyanobacteria blooms are prevalent.

Given these variations in optical properties, it is critical to recalibrate algorithms and absorption coefficients for freshwater lakes systems to account for the specific phytoplankton compositions. Recent studies have emphasized the importance of region-specific parameter adjustments, particularly in dynamic and eutrophic freshwater environments where cyanobacteria often dominate (Hunter et al., 2010; Kutser, 2009). To further validate the optical differences between freshwater lakes and oceans, the $a^{*}_{pH}(\lambda)$ of dominant phytoplankton groups in Lake Taihu was compared to that from marine environments, and the data for marine waters was downloaded from (Lomas et al., 2023). As shown in Fig. 10, in Lake Taihu, where cyanobacteria frequently dominate, the spectra showed significantly lower values in the blue (440–495 nm) and red (660–700 nm) regions compared to marine diatom-dominated waters .Specifically, water bodies dominated by diatoms exhibit a more prominent peak in the blue and red regions, likely due to the presence of Chl-c, which contrasts with the phycobiliproteins found in cyanobacteria. Additionally, some differences in the specific absorption curves of different phytoplankton taxa in freshwater lakes and oceans, both in magnitude and shape, were also clearly observed (Sun et al., 2022). These findings underscore the necessity of tailoring bio-optical models to the unique characteristics of freshwater systems.

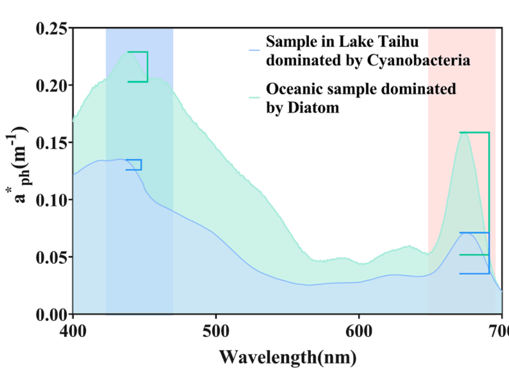

Fig. 10. Comparison of a^{*}_{pH} between the sample point in Lake Taihu and the oceanic sample point.

Further supporting this need, the DPA (Diagnostic Pigment Analysis) coefficients used for the global ocean (Li et al., 2023) were found to be quite different from the results of this study, as shown in Table 3. From the differences in the coefficients, it can be clearly seen that, from the perspective of the global ocean, the contribution of different phytoplankton groups to the Chl-a concentration did not vary much. However, the several lakes involved in this study were mostly dominated by Cyanophyta or Chlorophyta, and such a difference suggested that the coefficients needed to be recalibrated, and improvements in the methodology should be made. In summary, the large differences in phytoplankton community composition between freshwater and marine systems can lead to significant changes in bio-optical characteristics, making it highly desirable to improve the DPA method and recalculate the specific absorption coefficients.

*4.4. Influence of **aph**(λ) estimation error on deriving proportions of different phytoplankton groups*

As mentioned before, it was known that $aph(\lambda)$ is a critical parameter for determining phytoplankton composition in this study. Therefore, the estimation accuracy of $aph(\lambda)$ determines whether this algorithm can be applied to satellite images. To assess the $aph(\lambda)$ estimation accuracy on the final estimation result, system errors of ±10 % and ±20 % were introduced into the $aph(\lambda)$ values at wavelengths of 510 nm, 560 nm, 620 nm, 665 nm, and 681 nm, respectively. The results obtained after introducing these errors were then compared to the results calculated directly using the measured $aph(\lambda)$. The RMSE, MAPE, MAE and Bias are shown in Fig. 11. It can be clearly observed that introducing a −10 % error resulted in the lowest estimation error with RMSE = 2.63 %, MAE = 5.27 %, MAPE = 24.24 %, and Bias = 0.51. While introducing a + 10 % error, the algorithm still had a satisfactory estimation performance, with RMSE = 8.66 %, MAE = 6.61 %, MAPE = 24.47 %, and Bias = 1.3. However, when a ± 20 % error was introduced, the estimation errors became significantly larger, with RMSE exceeding 10 %, MAE surpassing 11 %, MAPE greater than 39 %, and Bias above 21.5. These results indicated that when the $aph(\lambda)$ estimation error reaches 20 %, the algorithm's performance deteriorates substantially.

The QAA-750E method was adopted in this study for the $aph(\lambda)$ estimation, and the estimation error was calculated by using the measured data from Lake Taihu. The evaluation result demonstrated the MAPE was under 15 %, which proved that QAA-750E can be applied to estimate the phytoplankton absorption of Lake Taihu. In conclusion, the impact of $aph(\lambda)$ estimation error on final phytoplankton composition estimation results was significant and cannot be ignored. To obtain reliable outcomes, the error in estimating $aph(\lambda)$ should be controlled within ±20 %. It was well known that $aph(\lambda)$ estimation errors may arise from several factors, including limitations in satellite image resolution, and inaccuracies in atmospheric correction. A more accurate and robust $aph(\lambda)$ estimation method was expected to be developed in future research.

Table 3
w_i for the freshwater lakes in this study and the global ocean in Li et al. (2023)'s study.

Phytoplankton	Type1 Cyanophyta dominated	Type2 non-Cyanophyta dominated	Li et al. (2023)
Dinophyta	0.53	0.01	1.11
Cryptophyta	0.01	0.01	1.71
Bacillariophyta	0.27	0.09	1.43
Cyanophyta	2.41	0.78	0.96
Chlorophyta	0.16	3.06	1.96

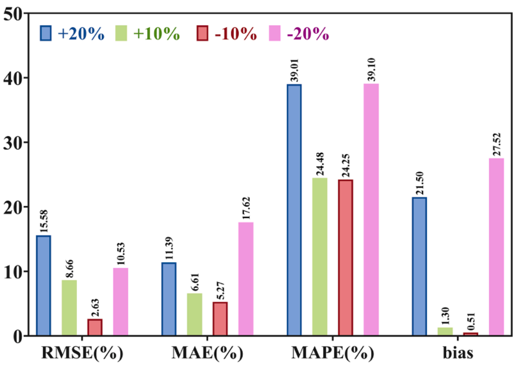

Fig. 11. Results of accuracy evaluation index after the introduction of ±10 %–20 % error.

4.5. Implications for long-term observations of phytoplankton communities based on satellite images

The successful demonstration of this algorithm's application to Lake Taihu highlights its potential for long-term monitoring of phytoplankton communities using satellite data. This algorithm provides a direct method to estimate the composition characteristics of phytoplankton groups, beyond functional or particle-sized groups (PSCs). However, its performance can be influenced by several factors, such as atmospheric correction accuracy, the retrieval of $aph(\lambda)$, and the presence of optically active substances like CDOM (colored dissolved organic matter) and suspended sediments. These substances can significantly mask phytoplankton absorption features, making it difficult to accurately retrieve $aph(\lambda)$.

Given these challenges, future studies should focus on refining atmospheric correction methods and developing estimation method of $aph(\lambda)$, particularly in complex waters where multiple optically active components are present. With the development of big data models, machine learning techniques, such as random forests and artificial neural networks, could be considered to obtain more accurate $aph(\lambda)$ results. Furthermore, to meet the demands of long-term and large-scale applications, more data from lakes with varying trophic levels are needed to validate the stability of the coefficients, ensuring that the method can be reliably applied across different regions.

In conclusion, these efforts will contribute to successful long-term observations of phytoplankton communities based on satellite imagery, providing insights into phytoplankton succession patterns, the response of phytoplankton groups to eutrophication, climate change, and human activities.

5. Conclusion

In this study, an improved DPA method for freshwater lakes was developed, enabling efficient derivation of phytoplankton community composition, which can provide detailed phytoplankton phyla information by calculating the proportions of different phytoplankton groups, not merely focusing on PSCs. The coefficients obtained from this method differ significantly from those for the ocean. This method demonstrated robust performance, achieving an RMSE of less than 8 %, an MAE of less than 13.52 %, a MAPE of less than 14.22 %, and a |Bias| of less than 9 %. Subsequently, this algorithm was successfully applied to Lake Taihu on August 12, 2023, based on Sentinel-3 OLCI images, to map the distribution of Chlorophyta, Cyanophyta, and Bacillariophyta, demonstrating the feasibility of using satellite imagery to monitor phytoplankton community composition in freshwater lakes. This study offers a novel and effective approach to obtaining the proportional distribution of different phytoplankton groups in optically complex freshwater lakes. The results proved the potential of satellite data in monitoring phytoplankton community structures, offering valuable insights into phytoplankton succession patterns and providing a solid foundation for understanding their responses to eutrophication, climate change human activities and other factors.

CRediT authorship contribution statement

Yuxin Zhu: Writing – original draft, Validation, Methodology, Formal analysis, Data curation. **Qingxia Miao:** Validation, Investigation, Data curation. **Heng Lyu:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization. **Yiling Zheng:** Validation, Data curation. **Wenyu Liu:** Methodology, Investigation. **Yunmei Li:** Resources, Funding acquisition, Conceptualization. **Junda Li:** Methodology, Investigation. **Fangfang Chen:** Validation, Data curation. **Song Miao:** Validation, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by National Natural Science Foundation of China (grant numbers 42271341, 42201423, and U2102207) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (grant numbers KYCX23_1702 and KYCX24_1833); We also thank the ESA for providing the Sentinel-3 OLCI data.

Data availability

The data that has been used is confidential.

References

Bi, S., Hieronymi, M., 2024. Holistic optical water type classification for ocean, coastal, and inland waters. Limnol. Oceanogr. 69, 1547–1561. https://doi.org/10.1002/lno.12606.

Bi, S., Li, Y., Lyu, H., Mu, M., Xu, J., Lei, S., Miao, S., Hong, T., Zhou, L., 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 Trans. Geosci. Remote Sensing 57, 7447–7459. https://doi.org/10.1109/TGRS.2019.2913401.

Bi, S., Li, Y., Wang, Q., Lyu, H., Liu, G., Zheng, Z., Du, C., Mu, M., Xu, J., Lei, S., Miao, S., 2018. Inland Water Atmospheric Correction Based on Turbidity Classification Using OLCI and SLSTR Synergistic Observations. Remote Sens. (Basel) 10, 1002. https://doi.org/10.3390/rs10071002.

Bidigare, R.R., Van Heukelem, L., Trees, C.C., 2002. HPLC phytoplankton pigments: sampling, laboratory methods, and quality assurance procedures. Ocean optics protocols for satellite ocean color sensor validation. revision 3, 258–268.

BinGuo, Z., JiBiao, Z., XingZhang, L., Zheng, Z., Qun, L., FeiHong, G., 2011. Inhibition of overwinter cyanobacteria recovery in Taihu water by gamma-ray irradiation. China Environ. Sci. 31, 316–320.

Brewin, R., Sathyendranath, S., Hirata, T., Lavender, S., Barciela, R., Hardman-Mountford, N., 2010. A three-component model of phytoplankton size class for the Atlantic Ocean. Ecol. Modell. 221, 1472–1483. https://doi.org/10.1016/J.ECOLMODEL.2010.02.014.

Brewin, R., Sathyendranath, S., Jackson, T., Barlow, R., Brotas, V., Airs, R., Lamont, T., 2015. Influence of light in the mixed-layer on the parameters of a three-component model of phytoplankton size class. Remote Sens. Environ. 168, 437–450. https://doi.org/10.1016/J.RSE.2015.07.004.

Brewin, R., Sathyendranath, S., Tilstone, G., Lange, P.K., Platt, T., 2014. A multicomponent model of phytoplankton size structure. J. Geophys. Res. 119, 3478–3496. https://doi.org/10.1002/2014JC009859.

Cai, X., Li, Y., Lei, S., Zeng, S., Zhao, Z., Lyu, H., Dong, X., Li, J., Wang, H., Xu, J., Zhu, Y., Wu, L., Cheng, X., 2023. A hybrid remote sensing approach for estimating chemical oxygen demand concentration in optically complex waters: a case study in inland lake waters in eastern China. Science of The Total Environment 856, 158869. https://doi.org/10.1016/j.scitotenv.2022.158869.

Chase, A.P., Boss, E., Cetinić, I., Slade, W., 2017. Estimation of Phytoplankton Accessory Pigments From Hyperspectral Reflectance Spectra: toward a Global Algorithm. JGR Oceans 122, 9725–9743. https://doi.org/10.1002/2017JC012859.

Chen, N., Bianchi, T.S., McKee, B.A., Bland, J.M., 2001. Historical trends of hypoxia on the Louisiana shelf: application of pigments as biomarkers. Org. Geochem. 32, 543–561. https://doi.org/10.1016/S0146-6380(00)00194-7.

Chen, Y., 2003. Long-term dynamics of phytoplankton assemblages: microcystis-domination in Lake Taihu, a large shallow lake in China. J. Plankton Res. 25, 445–453. https://doi.org/10.1093/plankt/25.4.445.

Deng, J., Qin, B., Paerl, H.W., Zhang, Y., Ma, J., Chen, Y., 2014. Earlier and warmer springs increase cyanobacterial (Microcystis spp.) blooms in subtropical Lake Taihu, China. Freshw. Biol. 59, 1076–1085. https://doi.org/10.1111/fwb.12330.

Devred, E., Sathyendranath, S., Stuart, V., Platt, T., 2011. A three component classification of phytoplankton absorption spectra: application to ocean-color data. Remote Sens. Environ. 115, 2255–2266. https://doi.org/10.1016/J.RSE.2011.04.025.

Doron, M., Bélanger, S., Doxaran, D., Babin, M., 2011. Spectral variations in the near-infrared ocean reflectance. Remote Sens. Environ. 115, 1617–1631. https://doi.org/10.1016/j.rse.2011.01.015.

Hirata, T., Aiken, J., Hardman-Mountford, N., Smyth, T., Barlow, R., 2008. An absorption model to determine phytoplankton size classes from satellite ocean colour. Remote Sens. Environ. 112, 3153–3159. https://doi.org/10.1016/J.RSE.2008.03.011.

Hunter, P.D., Tyler, A.N., Carvalho, L., Codd, G.A., Maberly, S.C., 2010. Hyperspectral remote sensing of cyanobacterial pigments as indicators for cell populations and toxins in eutrophic lakes. Remote Sens. Environ. 114, 2705–2718. https://doi.org/10.1016/j.rse.2010.06.006.

Jeffrey, S.W., Mantoura, R.F.C., Wright, S.W., International Council of Scientific Unions, Unesco, 1997. Phytoplankton Pigments in oceanography: Guidelines to Modern methods, Monographs on Oceanographic Methodology. UNESCO Publishing, Paris.

Kutser, T., 2009. Passive optical remote sensing of cyanobacteria and other intense phytoplankton blooms in coastal and inland waters. Int. J. Remote Sens. 30, 4401–4425. https://doi.org/10.1080/01431160802562305.

Li, C., Feng, W., Chen, H., Li, X., Song, F., Guo, W., Giesy, J.P., Sun, F., 2019. Temporal variation in zooplankton and phytoplankton community species composition and the affecting factors in Lake Taihu-a large freshwater lake in China. Environ. Pollut. 245, 1050–1057. https://doi.org/10.1016/j.envpol.2018.11.007.

Li, Z., Gao, Y., Wang, S., Zhang, K., Lin, Q., Jia, J., Lu, Y., 2022. Lake pigment characteristics and their applicability in reconstructing phytoplankton communities under irregular hydrological regulation in a floodplain lake system. J Hydrol 614, 128575. https://doi.org/10.1016/j.jhydrol.2022.128575.

Li, Z., Sun, D., Wang, S., Huan, Y., Zhang, H., Liu, J., He, Y., 2023. A global satellite observation of phytoplankton taxonomic groups over the past two decades. Glob. Chang. Biol. 29, 4511–4529. https://doi.org/10.1111/gcb.16766.

Lomas, M., Neeley, A., Vandermeulen, R., Mannino, A., Thomas, C., Novak, M., Freeman, S., 2023. Phytoplankton Optical Fingerprint Libraries For Development of Phytoplankton Ocean Color Satellite Products. https://doi.org/10.5061/DRYAD.RBNZS7HFG.

MacIntyre, H.L., Kana, T.M., Geider, R.J., 2000. The effect of water motion on short-term rates of photosynthesis by marine phytoplankton. Trends Plant Sci. 5, 12–17. https://doi.org/10.1016/S1360-1385(99)01504-6.

Mouw, C., Hardman-Mountford, N., Alvain, S., Bracher, A., Brewin, R., Bricaud, A., Ciotti, Á., Devred, E., Fujiwara, A., Hirata, T., Hirawake, T., Kostadinov, T., Roy, S., Uitz, J., 2017. A Consumer's Guide to Satellite Remote Sensing of Multiple Phytoplankton Groups in the Global Ocean. Front. Mar. Sci. 4. https://doi.org/10.3389/fmars.2017.00041 null.

Ozawa, K., Fujioka, H., Muranaka, M., Yokoyama, A., Katagami, Y., Homma, T., Ishikawa, K., Tsujimura, S., Kumagai, M., Watanabe, M.F., Park, H.-D., 2005. Spatial distribution and temporal variation of Microcystis species composition and microcystin concentration in Lake Biwa. Environ. Toxicol. 20, 270–276. https://doi.org/10.1002/tox.20117.

Paerl, H.W., Otten, T.G., 2013. Harmful Cyanobacterial Blooms: causes, Consequences, and Controls. Microb. Ecol. 65, 995–1010. https://doi.org/10.1007/s00248-012-0159-y.

Parsons, M.J., 1985. Biosystematics of the cryptogamic flora of New Zealand: algae. N.Z. J. Bot. 23, 663–675. https://doi.org/10.1080/0028825X.1985.10434235.

Pridmore, R.D., Etheredge, M.K., 1987. Planktonic cyanobacteria in New Zealand inland waters: distribution and population dynamics. N. Z. J. Mar. Freshwater Res. 21, 491–502. https://doi.org/10.1080/00288330.1987.9516245.

Reynolds, C.S., 2006. The Ecology of Phytoplankton, Ecology, Biodiversity and Conservation. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511542145.

Ruddick, K.G., De Cauwer, V., Park, Y.-J., Moore, G., 2006. Seaborne measurements of near infrared water-leaving reflectance: the similarity spectrum for turbid waters. Limnol. Oceanogr. 51, 1167–1179. https://doi.org/10.4319/lo.2006.51.2.1167.

Ruddick, K.G., Ovidio, F., Rijkeboer, M., 2000. Atmospheric correction of SeaWiFS imagery for turbid coastal and inland waters. Appl. Opt. 39, 897–912. https://doi.org/10.1364/AO.39.000897. AO.

Sathyendranath, S., IOCCG, 2014. Phytoplankton Functional Types from Space. (Report). International Ocean Colour Coordinating Group (IOCCG). https://doi.org/10.25607/OBP-106.

Shen, M., Duan, H., Cao, Z., Xue, K., Loiselle, S., Yesou, H., 2017. Determination of the Downwelling Diffuse Attenuation Coefficient of Lake Water with the Sentinel-3A OLCI. Remote Sens. (Basel) 9, 1246. https://doi.org/10.3390/rs9121246.

Smayda, T.J., 1997. Harmful algal blooms: their ecophysiology and general relevance to phytoplankton blooms in the sea. Limnol. Oceanogr. 42, 1137–1153. https://doi.org/10.4319/lo.1997.42.5_part_2.1137.

Sun, X., Brewin, R.J.W., Sathyendranath, S., Dall'Olmo, G., Airs, R., Barlow, R., Bracher, A., Brotas, V., Kheireddine, M., Lamont, T., Marañón, E., Morán, X.A.G., Raitsos, D.E., Shen, F., Tilstone, G.H., 2023. Coupling ecological concepts with an ocean-colour model: phytoplankton size structure. Remote Sens. Environ. 285, 113415. https://doi.org/10.1016/j.rse.2022.113415.

Sun, X., Shen, F., Brewin, R., Liu, D., Tang, R., 2019. Twenty-Year Variations in Satellite-Derived Chlorophyll-a and Phytoplankton Size in the Bohai Sea and Yellow Sea. J. Geophys. Res. 124, 8887–8912. https://doi.org/10.1029/2019jc015552.

Sun, X., Shen, F., Brewin, R.J.W., Li, M., Zhu, Q., 2022. Light absorption spectra of naturally mixed phytoplankton assemblages for retrieval of phytoplankton group composition in coastal oceans. Limnology & Oceanography 67, 946–961. https://doi.org/10.1002/lno.12047.

Uitz, J., Claustre, H., Morel, A., Hooker, S.B., 2006. Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll. J. Geophys. Res. 111, C08005. https://doi.org/10.1029/2005JC003207.

Uitz, J., Stramski, D., Reynolds, R.A., Dubranna, J., 2015. Assessing phytoplankton community composition from hyperspectral measurements of phytoplankton absorption coefficient and remote-sensing reflectance in open-ocean environments. Remote Sens. Environ. 171, 58–74. https://doi.org/10.1016/j.rse.2015.09.027.

Vidussi, F., Claustre, H., Manca, B.B., Luchetta, A., Marty, J.-C., 2001. Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. J. Geophys. Res. 106, 19939–19956. https://doi.org/10.1029/1999JC000308.

Xi, H., Hieronymi, M., Krasemann, H., Röttgers, R., 2017. Phytoplankton Group Identification Using Simulated and In situ. Hyperspectral Remote Sensing Reflectance. Front. Mar. Sci. 4, 272. https://doi.org/10.3389/fmars.2017.00272.

Xu, H., Vavilin, D., Vermaas, W., 2001. Chlorophyll *b* can serve as the major pigment in functional photosystem II complexes of cyanobacteria. Proceedings of the National Academy of Sciences 98, 14168–14173. https://doi.org/10.1073/pnas.251530298.

Xue, K., Ma, R., Duan, H., Shen, M., Boss, E., Cao, Z., 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 Sens. Environ. 225, 328–346. https://doi.org/10.1016/j.rse.2019.03.006.

Zhai, H., Zhang, H., Zhang, L., Li, P., 2018. Cloud/shadow detection based on spectral indices for multi/hyperspectral optical remote sensing imagery. ISPRS Journal of Photogrammetry and Remote Sensing 144, 235–253. https://doi.org/10.1016/j.isprsjprs.2018.07.006.

Zhu, Q., Li, J., Zhang, F., Shen, Q., 2018. Distinguishing Cyanobacterial Bloom From Floating Leaf Vegetation in Lake Taihu Based on Medium-Resolution Imaging Spectrometer (MERIS) Data. IEEe J. Sel. Top. Appl. Earth. Obs. Remote Sens. 11, 34–44. https://doi.org/10.1109/JSTARS.2017.2757006.

Zhu, Y., Li, Y., Bi, S., Lyu, H., Cai, X., Wang, H., Li, Junda, Li, Jianzhong, Xu, J., 2023. Spatial and temporal distribution analysis of dominant algae in Lake Taihu based on ocean and land color instrument data. Ecol. Indic. 155, 110959. https://doi.org/10.1016/j.ecolind.2023.110959.