



A ground-based remote sensing system for high-frequency and real-time monitoring of phytoplankton blooms

Weijia Wang^{a,b}, Kun Shi^{a,c,*}, Yibo Zhang^{a,c}, Na Li^{a,b}, Xiao Sun^{a,b}, Dong Zhang^{a,b}, Yunlin Zhang^a, Boqiang Qin^{a,c}, Guangwei Zhu^{a,c}

^a Taihu Laboratory for Lake Ecosystem Research, State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China

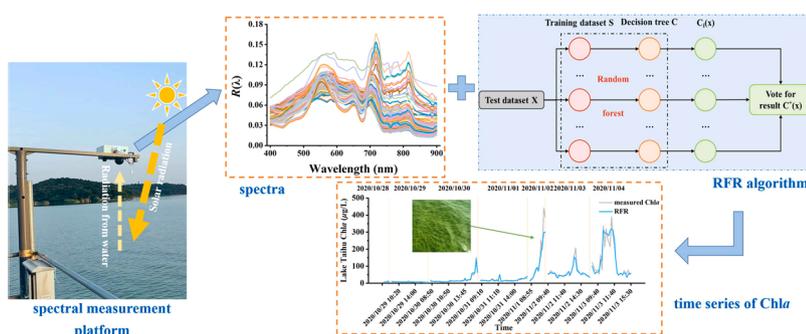
^b University of Chinese Academy of Sciences, Beijing 100049, China

^c Nanjing Zhongke Deep Insight Technology Research Institute Co., Ltd, Nanjing 211899, China

HIGHLIGHTS

- Phytoplankton blooms has severely threatened ecosystem and human health.
- The GRSS compensates for the existing methods for monitoring phytoplankton blooms.
- GRSS can accurately monitor real-time Chla dynamics with high frequency.
- The AI chip in GRSS combined with the RFR algorithm can automatically monitor Chla.
- The AI chip supports remote access, along with algorithm upgrades and replacements.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Lingxin CHEN

Keywords:

Chlorophyll a
Inland waters
Eutrophication
Hyperspectrometer
Machine learning

ABSTRACT

The worldwide expansion of phytoplankton blooms has severely threatened water quality, food webs, habitat stability and human health. Due to the rapidity of phytoplankton migration and reproduction, high-frequency information on phytoplankton bloom dynamics is crucial for their forecasting, treatment, and management. While several approaches involving satellites, *in situ* observations and automated underwater monitoring stations have been widely used in the past several decades, they cannot fully provide high-frequency and continuous observations of phytoplankton blooms at low cost and with high accuracy. Thus, we propose a novel ground-based remote sensing system (GRSS) that can monitor real-time chlorophyll *a* concentrations (Chla) in inland waters with a high frequency. The GRSS mainly consists of three platforms: the spectral measurement platform, the data-processing platform, and the remote access control, display and storage platform. The GRSS is capable of obtaining a remote sensing irradiance ratio ($R(\lambda)$) of 400–1000 nm at a high frequency of 20 s. Eight different Chla retrieval algorithms were calibrated and validated using a dataset of 481 pairs of GRSS $R(\lambda)$ and *in situ* Chla measurements collected from four inland waters. The results showed that random forest regression achieved the best performance in deriving Chla ($R^2 = 0.95$, root mean square error = 13.40 $\mu\text{g/L}$, and mean relative error = 25.7%). The GRSS successfully captured two typical phytoplankton bloom events in August 2021 with rapid changes in Chla from 20 $\mu\text{g/L}$ to 325 $\mu\text{g/L}$ at the minute level, highlighting the critical role that this GRSS can

* Correspondence to: Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, 73 East Beijing Road, Nanjing 210008, China.

E-mail address: kshi@niglas.ac.cn (K. Shi).

<https://doi.org/10.1016/j.jhazmat.2022.129623>

Received 1 May 2022; Received in revised form 12 July 2022; Accepted 15 July 2022

Available online 16 July 2022

0304-3894/© 2022 Elsevier B.V. All rights reserved.

play in the high-frequency monitoring of phytoplankton blooms. Although the algorithm embedded into the GRSS may be limited by the size of the training dataset, the high-frequency, continuous and real-time data acquisition capabilities of the GRSS can effectively compensate for the limitations of traditional observations. The initial application demonstrated that the GRSS can capture rapid changes of phytoplankton blooms in a short time and thus will play a critical role in phytoplankton bloom management. From a broader perspective, this approach can be extended to other carriers, such as aircraft, ships and unmanned aerial vehicles, to achieve the networked monitoring of phytoplankton blooms.

1. Introduction

Phytoplankton blooms in global inland waters have demonstrated a noticeably increasing trend, which is likely attributable to eutrophication, global warming and increasing hydrologic variability (Ho et al., 2019; Huisman et al., 2018; Shi et al., 2019). The worldwide expansion of phytoplankton blooms has severely threatened water quality, food webs, habitat stability and human health (Paerl and Barnard, 2020; Paerl and Huisman, 2008; Taranu et al., 2012). In particular, phytoplankton blooms can increase turbidity, thereby decreasing light availability and suffocating submerged aquatic vegetation (Bricker et al., 2008; Shi et al., 2014). The microbial degradation of phytoplankton blooms may deplete oxygen, inducing hypoxia and anoxia and causing the deaths of fish and benthic invertebrates while regenerating sediment-associated nutrients, thus perpetuating the blooms (Xu et al., 2021; Zhang et al., 2016a). Moreover, a variety of toxic materials can be produced by phytoplankton blooms, leading to serious and occasionally fatal liver, digestive, and neurological diseases in humans, birds and mammals (Carmichael, 2001; Huisman et al., 2018). Thus, phytoplankton blooms are becoming one of the world's most pressing environmental and ecological concerns. Information on phytoplankton dynamics is vital for forecasting, early warning, and controlling phytoplankton blooms (Hu, 2009; Paerl and Otten, 2013). Thus, reliable monitoring that can accurately and rapidly quantify phytoplankton blooms are critical to mitigating the detrimental impacts of phytoplankton blooms on water quality, aquatic ecosystems, and human health.

At present, phytoplankton blooms can be monitored in three ways, including *in situ* observations, underwater sensor automatic observation stations, and satellite remote sensing. Compared with other monitoring approaches, *in situ* observations can provide more accurate information on phytoplankton blooms (O'Reilly et al., 1998). However, this approach is time-consuming, laborious, and lagging because water sampling and laboratory analysis require considerable labor and time. In addition, the approach relies on manual water sampling at preset sampling sites and at regular (biweekly or monthly) intervals or via shorter-term ship cruises and thus can provide only limited data on phytoplankton dynamics (Kutser, 2004). Moreover, using a limited number of measurements through *in situ* observations to represent the overall spatiotemporal distribution of phytoplankton blooms has proven to be problematic in lake waters (Tebbs et al., 2013). The underwater sensor automatic observation station, usually equipped with various instruments, e.g., YSI (Xylem Inc., USA), Hydrolab DS5 multiparameter sondes (HACH, USA), PHYTO-PAM chlorophyll fluorometer (Walz, Germany), and FluoroProbe (bbe-Moldaenke, Kiel, Germany), can measure the changes in phytoplankton blooms online and provide high-frequency real-time data (Bowling et al., 2016; Van Beusekom et al., 2009). Nevertheless, its applicability to the monitoring of phytoplankton blooms in lakes is limited by the following two aspects. First, the monitoring accuracy is extremely susceptible to environmental factors, such as temperature, pH, and concentrations of water constituents, because the instruments must be installed below the water surface. For example, in turbid waters, a large number of suspended sediments and plankton easily adhere to the instrument probes, which seriously affects the monitoring accuracy of the instruments (Qin et al., 2015). Second, the high operation and maintenance costs of automatic

observation stations inhibit their ability to support long-term observations (Li et al., 2022; Sun et al., 2022). Normally, the underwater sensors need to be manually corrected every one or two weeks. *In situ* observations and underwater sensor automatic observation stations both collect information on phytoplankton blooms at specific sites and are unable to offer the geographical distribution of phytoplankton blooms.

With its ability to perform large-scale, long-duration, and periodic monitoring at a lower cost, satellite remote sensing has become a powerful approach for inland water observations, as it allows detailed observations of important biogeochemical parameters related to water quality dynamic processes to be obtained (Bonansea et al., 2015; Mouw et al., 2015; Sagan et al., 2020; Shi et al., 2014; Wang et al., 2011). The unique spectral absorption characteristics of phytoplankton pigments enable the use of satellite remote sensing to detect phytoplankton dynamics (Devred et al., 2011). Over the past three decades, satellite remote sensing has been widely used to reveal the spatial and temporal distributions of several metrics of phytoplankton dynamics, e.g., water surface chlorophyll *a* concentrations (Chl_a), bloom areas, and phenology in inland and oceanic waters (Gons et al., 2002; Kahru et al., 2011; Kutser, 2004; Matthews et al., 2012; O'Reilly et al., 1998; Palmer et al., 2015; Sagan et al., 2020; Shi et al., 2019; Tebbs et al., 2013). In particular, Chl_a is most widely used to quantify phytoplankton bloom, as this photosynthetic pigment exists in all algal species (Le et al., 2009; Neil et al., 2019; Song et al., 2013). At present, numerous algorithms, including (I) empirical algorithms, (II) analytical algorithms, (III) semi-analytical algorithms, and (IV) machine learning algorithms, have been developed to derive Chl_a from remote sensing data (Cao et al., 2020; Dall'Olmo and Gitelson, 2005; Duan et al., 2012; Gitelson et al., 2008; Gitelson et al., 2007; Gons et al., 2008; Le et al., 2009; Neil et al., 2019). With these developed algorithms, several satellites have been confirmed for use in monitoring the dynamics of phytoplankton blooms in inland waters, such as the Landsat series (TM (Thematic Mapper)/ETM+ (Enhanced Thematic Mapper Plus)/OLI (Operational Land Imager)), Moderate Resolution Imaging Spectroradiometer (MODIS, Aqua and Terra), Medium Resolution Imaging Spectrometer (MERIS), and Sentinel 2-A/B (Cao et al., 2022; Gitelson et al., 2008, 2007; Gons et al., 2008; Pahlevan et al., 2022; Zhang et al., 2021).

However, satellite remote sensing has been inhibited as an effective tool to routinely monitor phytoplankton blooms for the following reasons. First, the highly accurate monitoring of phytoplankton blooms remains a great challenge due to the lack of reliable reflectance ($R_{rs}(\lambda)$) products for inland waters (Mouw et al., 2015). Accurate atmospheric corrections for inland waters are known to be more difficult than those for lands and oceans due to a smaller fraction of the signals from waters received by satellite sensors and the greater optical complexity of inland waters (Pahlevan et al., 2021; Warren et al., 2019). Although several methods of atmospheric corrections for inland waters, such as the near-infrared (NIR)-based algorithm, the improved shortwave infrared (SWIR)-based algorithm, and the dark pixel-based algorithm, have been developed, the derived $R_{rs}(\lambda)$ products still contain errors in excess of 10%, which increase the uncertainty in quantifying Chl_a based on satellite images (Matthews et al., 2010; Wang et al., 2011). Second, the limited availability of high-quality satellite images due to cloud and satellite temporal resolutions is insufficient for capturing the rapidly changing patterns of phytoplankton blooms in inland waters. For example, images from the Landsat series have a fine spatial resolution of

30 m, but the revisit time of 16 days limits their application in monitoring short-term phytoplankton dynamics (Cao et al., 2022). MODIS is the most popular multispectral sensor for monitoring inland water quality due to its high temporal resolution. Nevertheless, a previous study reported that the percentage of usable observations for MODIS-Aqua is less than 25% for the $R_{rs}(645)$ product in Lake Taihu on the Yangtze Plain of China, meaning that fewer than 100 useable images can be obtained per year in this region (Shi et al., 2015). Furthermore, Chla usually changes dynamically within a day or even hours, as phytoplankton can vertically and horizontally migrate under certain meteorological and hydrological conditions (Wang et al., 2022). Finally, low spatial and spectral resolutions limit the use of satellite remote sensing data for monitoring phytoplankton blooms in small inland waters, e.g., rivers, ponds, and small lakes.

There are already several systems for real-time monitoring of water quality in inland waters through wireless sensor network transmissions (Nasirudin et al., 2011) or by adding anomaly detection algorithms (Geetha and Gouthami, 2017). However, the underwater probes of these systems are also susceptible to water erosion and contamination, which can be effectively addressed by ground-based remote sensing systems (GRSS). Several GRSSs were developed previously to collect continuous, autonomous and long-term measurements of the reflectance properties of a semiarid savanna (Cogliati et al., 2015) and plant canopies (sun-induced fluorescence) (Huber et al., 2014) and to establish a radiation budget (Zapadka et al., 2020) to validate and correct Earth satellite observations. However, regular water quality measurements from ground-based systems are rare for inland waters. Currently, only a very small number of sites exist for which automated inland water quality data have been collected from ground-based systems (Li et al., 2022; Sun et al., 2022).

Thus, there is currently a great need to develop an effective alternative to compensate for the weaknesses in the existing methods for monitoring phytoplankton blooms in inland waters. To address this unmet need, we attempted to develop a GRSS that could continuously and accurately monitor the real-time long-term trend and short-term dynamics in phytoplankton blooms with a high frequency in inland waters. The objectives of this study were to (1) outline the measurement scheme and setup of the GRSS, (2) establish the best algorithm for the GRSS to derive Chla and (3) present the initial application of the GRSS as well as discuss future improvements in the GRSS.

2. Materials and methods

2.1. Framework setup of the GRSS

Here, we developed a GRSS for inland waters by assembling a commercial-grade hyperspectrometer and an AI chip processor (jointly developed by Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing Zhongke Deep insight Technology Research Institute Co., Ltd, and Hangzhou Hikvision Digital Technology Co., Ltd). The design concept is based on the use of real-time high-frequency hyperspectral data to derive Chla using remote sensing algorithms. The GRSS mainly consists of three platforms: the spectral measurement platform, the data-processing platform, and the remote access control, display and storage platform (Fig. S1). The water spectral reflectance signal determined by the spectral measurement platform is transformed to Chla information through its estimation algorithm embedded in the AI chip processor (data-processing platform). The spectral reflectance and Chla data are remotely transmitted to the remote access control, display and storage platform via a wireless transmission device.

The core of our system includes the platforms of spectral measurement and data processing equipped with a hyperspectrometer and an AI chip processor. The hyperspectrometer and AI chip processor are hosted in a compact instrument box. A cooling system is set for controlling the temperature within the instrument box to reduce temperature-related

effects, such as spectral drift and detector noise. The technical design facilitates the easy field installation of the GRSS. Here, we positioned the hyperspectrometer at a height of 5–10 m above the water surface and a distance of 2–5 m offshore to eliminate the influences of the lake, reservoir and river shore (Fig. S1). The height and distance are the optimal choices after many experiments and field investigations, which can be adjusted according to the field installation conditions. The hyperspectrometer can also be installed on a fixed or mobile platform in an open water area. The hyperspectrometer covers the spectral range of 400–1000 nm with a full width at half maximum of 1 nm, a spectral resolution of 1 nm and a signal-to-noise ratio of > 1800:1. The minimum measurement interval of the hyperspectrometer is 20 s, providing an opportunity to continuously determine the water reflectance with a high frequency for inland waters. We set the observation time range between 9 am and 4 pm to reduce the effects of the zenith angle. The hyperspectrometer has two optical detectors (upper and lower) to simultaneously determine the upwelling irradiance from water ($E_u(\lambda)$, $W m^{-2} nm^{-1}$) and the downwelling irradiance from sky ($E_d(\lambda)$, $W m^{-2} nm^{-1}$), respectively. The determined $E_u(\lambda)$ and $E_d(\lambda)$ were transmitted to the AI chip processor through an optical fiber to calculate the water irradiance ratio ($R(\lambda)$) according to the following equation:

$$R(\lambda) = \frac{E_u(\lambda)}{E_d(\lambda)} \quad (1)$$

The Chla information was immediately derived from the determined $R(\lambda)$ with the algorithm embedded in the AI chip processor. The Chla estimation algorithm used was developed previously based on the determined $R(\lambda)$ and the *in situ* Chla measurements in these waters with different optical environments. The details about the development and validation of the Chla estimation algorithm can be found in the section “Algorithm development and validation embedded in the GRSS for Chla retrieval”. The monitoring ball equipped in the instrument box records videos and photos of phytoplankton blooms, and sensors for rainfall and temperature provide complementary environmental data for phytoplankton bloom monitoring. The remote access control, display and storage platform provides users with access to remote control and data extraction, storage and display. The remote control function includes the setup of basic configurations, such as observation start and end times, observation intervals (temporal resolutions), and early warnings of phytoplankton blooms according to specific requirements.

2.2. Sampling waters and data collection

Four inland waters, namely, Lake Taihu, Liangxi River, Lake Qiandaohu and Fuchunjiang Reservoir, including lake, river and artificial reservoirs with different nutrient states and physical properties were considered (Fig. 1). These waters provide drinking water sources, flood control and energy storage, aquaculture, shipping, tourism and other ecological services for large and medium neighboring cities, and they play pivotal roles in the socioeconomic development of eastern China. Lake Taihu, a typically shallow, eutrophic lake with an area of 2338 km² and an average depth of 1.9 m, is located in the southern area of the Yangtze River Delta (Zhang et al., 2016c). The Liangxi River, which connects the Beijing-Hangzhou Grand Canal and Meiliang Bay, is an important river connecting the most eutrophic bays in Lake Taihu. Lake Qiandaohu, also known as the Xin’anjiang Reservoir, is a vast manmade lake with a surface area of 580 km² and a mean depth of approximately 34 m located in Zhejiang Province (Zhou et al., 2016). This lake is a key drinking water source that serves at least ten million people (Zhang et al., 2016b). The Fuchunjiang Reservoir is a vital centralized drinking water source downstream of Lake Qiandaohu and upstream of Hangzhou. However, with rapid economic development and population increases over the past few decades, these inland waters have experienced severe threats from phytoplankton blooms (Zhang et al., 2016b, 2016c; Wang et al., 2011), resulting in water supply shortages and a significant decline in ecological services.

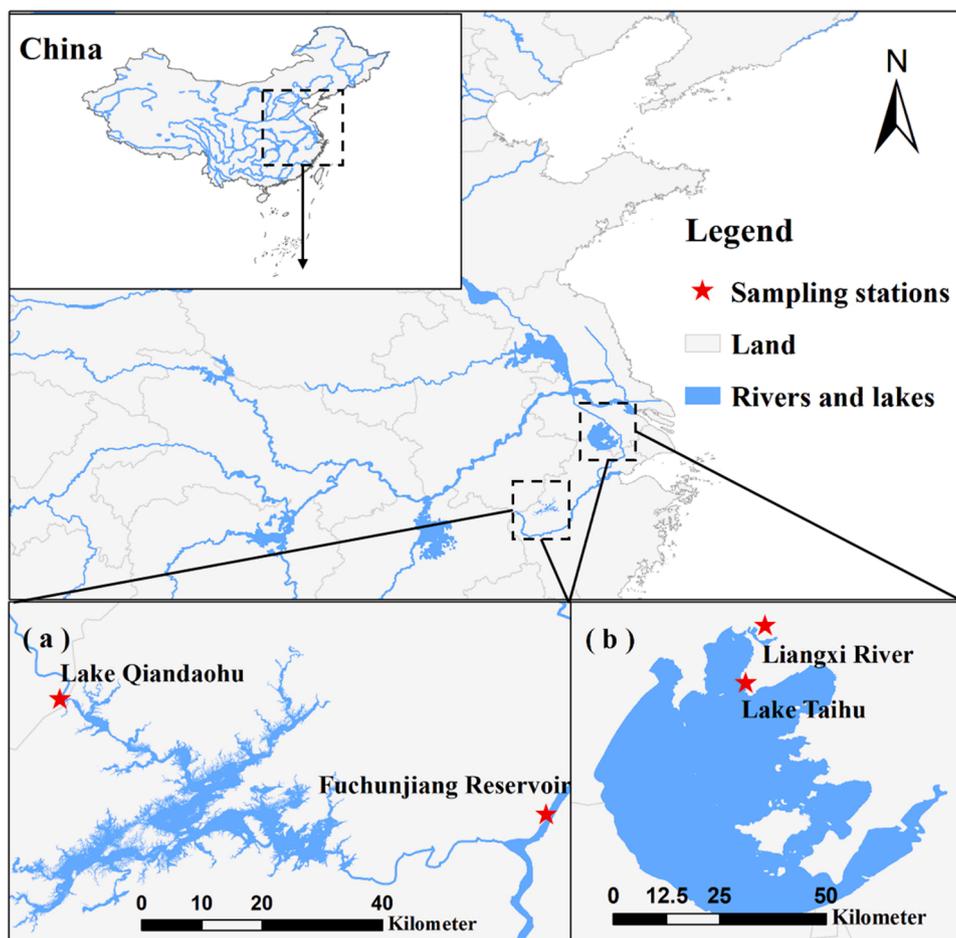


Fig. 1. Sampling sites in this study. (a) Location of Lake Qiandaohu and the Fuchunjiang Reservoir. (b) Location of Lake Taihu and the Liangxi River.

Four field data collections were conducted in the four inland waters from October to November 2020 (Table 1). Simultaneously, the hyperspectrometer of the GRSS continuously obtained $R(\lambda)$ at 1-minute intervals, while pretreatment was implemented for the collected samples. Water clarity was measured as Secchi disk depth (SDD) while sampling. Other parameters were measured in the laboratory. At 80 °C, the pigment was extracted with 90% thermal ethanol owing to its fat-soluble property, and then Chla was determined by spectrophotometry with absorbance at 750 nm and 665 nm (Haeder et al., 2020). Total suspended matter (TSM) was determined gravimetrically from samples

collected on preweighed GF/F filters (Zhang et al., 2014); the total nitrogen (TN) and total phosphorus (TP) concentrations were determined with combined persulfate digestion (Ebina et al., 1983); the ammonium (NH_4^+ -N) concentration was measured by the indophenol blue method (APHA, 1995); and the chemical oxygen demand (COD) was determined by permanganate titration (Liu et al., 2012).

2.3. Candidate algorithms for deriving Chla

Three different types of algorithms that have been most widely used

Table 1

The performances of the empirical, semiempirical and machine learning algorithms were compared for the training dataset, test dataset and total dataset. The evaluation indicators include the slope, MRE, RMSE and R^2 .

Model	Dataset	Bands (x)	Formula	Linear slope	R^2	MRE (%)	RMSE ($\mu\text{g/L}$)	References
Band-ratio	Total data	709 nm/665 nm	$\text{Chla} = 18.87x^2 + 79.30x - 65.66$	0.79	0.79	227.22	24.73	Gons et al. (2002)
Baseline (MCI)	Total data	709–665–754 nm	$\text{Chla} = 113.90x^2 - 41.78x - 1.80$	0.79	0.79	90.36	25.10	Mittenzwey et al., (1992)
NDCI	Total data	709–665 nm	$\text{Chla} = 706.55x^2 - 264.95x + 24.83$	0.83	0.83	50.04	22.58	Mishra and Mishra (2012)
Semi-analytical (Gons05)	Total data	709–665–775 nm	$\text{Chla} = -6E-04x^2 - 1.03x - 34.40$	0.68	0.68	235.54	30.85	Gons et al. (2008)
Three-band	Total data	671–710–740 nm	$\text{Chla} = 172.57x + 26.85$	0.81	0.81	140.43	23.72	Dall’Omo and Gitelson (2005)
	Total data	675–695–730 nm	$\text{Chla} = 165.78x + 13.92$	0.74	0.74	166.53	27.83	Gitelson et al. (2007)
BP	Train data	–	–	0.94	0.92	86.73	16.46	Rumelhart et al. (1986)
	Test data	–	–	1.17	0.95	86.24	15.73	
	Total data	–	–	0.98	0.92	86.60	16.27	
SVR	Train data	–	–	0.86	0.91	53.18	17.13	Cortes and Vapnik (1995)
	Test data	–	–	1.00	0.97	55.70	7.86	
	Total data	–	–	0.89	0.92	53.87	15.18	
RFR	Train data	–	–	0.87	0.92	14.59	16.00	Breiman (2001)
	Test data	–	–	1.02	0.97	21.55	8.25	
	Total data	–	–	0.90	0.93	16.49	14.31	

in recent decades were evaluated for retrieving Chla from the GRSS. These algorithms are (i) empirical algorithms, including the band-ratio algorithm, normalized difference chlorophyll index (NDCI), and MERIS maximum chlorophyll indices (MCI) algorithm; (ii) semi-analytical algorithms, including the semi-analytical algorithm proposed by Gons et al. in 2008 and three-band algorithms; and (iii) machine learning algorithms, including backpropagation neural network (BP), support vector regression (SVR) and random forest regression (RFR) algorithms.

(1) Empirical algorithms.

The band ratio, NDCI, and MCI are the three empirical algorithms most used to derive Chla in inland waters. The critical step of band-ratio algorithm development is to determine the reflectance at two bands that are sensitive to Chla (Duan et al., 2012). The basic principle for the band choice is that the absorption peak at λ_1 near 665 nm is generally assigned to phytoplankton, and the absorption at λ_2 near 700 nm presents a valley due to the combined effect of phytoplankton and pure water. Therefore, the reflectance ratio of λ_1 and λ_2 is suitable for deriving Chla (Gons et al., 2002; O'Reilly et al., 1998). The formula is defined as Eq. (2).

$$Chla \propto \frac{R_{rs}(\lambda_2)}{R_{rs}(\lambda_1)} \quad (2)$$

The NDCI algorithm was proposed by Mishra and Mishra (2012), similarly using the optical properties of Chla at approximately 700 nm and 665 nm to derive Chla in estuarine and turbid productive waters, thus avoiding the heterogeneous influences of TSM and chromophoric dissolved organic matter (CDOM) at shorter wavelengths on the reflectance spectra (Eq. (3)). Their results showed that this algorithm could improve the accuracy of Chla estimation in turbid inland waters of widely varying geographic regions.

$$Chla \propto \frac{R_{rs}(\lambda_2) - R_{rs}(\lambda_1)}{R_{rs}(\lambda_2) + R_{rs}(\lambda_1)} \quad (3)$$

Gower et al. (2005) proposed an MCI algorithm for retrieving Chla in coastal turbid water (Eq. (4)). The core idea of MCI is to construct a baseline through R_{rs} of two bands (λ_1 and λ_3) and to select λ_2 , which is sensitive to Chla within the baseline wavelength range. Then, Chla can be estimated through the difference in R_{rs} between λ_2 and the baseline. The subtraction calculation of MCI largely avoids the overestimation of Chla in the division principle of the band-ratio algorithm.

$$MCI = R_{rs}(\lambda_2) - R_{rs}(\lambda_1) - (R_{rs}(\lambda_3) - R_{rs}(\lambda_1)) * \left(\frac{\lambda_2 - \lambda_1}{\lambda_3 - \lambda_1}\right) \quad (4)$$

(2) Semi-analytical algorithms.

In turbid coastal and inland waters, the absorption of non-algal particles and CDOM in the red and NIR bands cannot be ignored, and thus, the radiative transfer theory regarding the backscattering and absorption properties of water materials is incorporated into the Chla retrieval algorithm. According to radiative transfer theory, $R_{rs}(\lambda)$ is proportional to $b_b/(a + b_b)$, where b_b is the total backscattering coefficient and a is the total absorption (Lee et al., 1998) (Eq. (5)). Based on the assumption that b_b is wavelength-independent in the NIR, and the absorption of other components in the water, namely, phytoplankton (a_{Chla}), non-algal particles (a_d) and CDOM (a_{CDOM}), is much smaller than that of pure water (a_w), b_b can be calculated as follows (Eq. (6)):

$$R_{rs} \propto \frac{b_b}{a + b_b} \quad (5)$$

$$b_b(779) = 1.61 \times \frac{R_{rs}(779)}{0.082 - 0.6 \times R_{rs}(779)} \quad (6)$$

Assuming that the absorption of inland waters at 665 nm is dominated by Chla and pure water and absorption at 709 nm is dominated by pure water, Gons et al. (2005) developed the following novel semi-analytical algorithm for Chla retrieval in inland and coastal waters

(hereafter referred to as the Gons05 algorithm):

$$a_{Chla}(665) = \frac{R_{rs}(709)}{R_{rs}(665)} \times (a_w(709) + b_b) - b_b^p - a_w(665) \quad (7)$$

$$Chla = \frac{a_{Chla}(665)}{a_{Chla}^*(665)} \quad (8)$$

where p is 1.062, $a_w(709)$ and $a_w(665)$ are 0.70 m^{-1} and 0.40 m^{-1} , respectively, and the Chla mass-specific absorption coefficient $a_{Chla}^*(665)$ is $0.0161 \text{ m}^2 \cdot \text{mg}^{-1}$.

A three-band algorithm was developed by Dall'Olmo and Gitelson (2005) for estimating Chla in turbid, productive waters and has been validated to be effective for MERIS data (Eq. (9)). The algorithm is a further exploration of radiative transfer theory based on the following assumptions: (a) λ_1 is located near the absorption peak of phytoplankton, and $a_d(\lambda_1) \approx a_d(\lambda_2)$, $a_{CDOM}(\lambda_1) \approx a_{CDOM}(\lambda_2)$; (b) the absorption at λ_3 is dominated by pure water, and other components can be ignored; and (c) the b_b values for three bands are approximately equal. Here, $a_d(\lambda)$ represents the absorption of nonalgal particles, and $a_{CDOM}(\lambda)$ represents the absorption of CDOM.

$$Chla \propto [R_{rs}^{-1}(\lambda_1) - R_{rs}^{-1}(\lambda_2)] \times R_{rs}(\lambda_3) \quad (9)$$

λ_1 , λ_2 and λ_3 were chosen to be 671 nm, 710 nm and 740 nm, respectively (Dall'Olmo and Gitelson, 2005). In subsequent studies, after comparing the results of several models with different bands, 675 nm, 695 nm, and 730 nm were chosen as specific values for λ_1 to λ_3 , respectively (Gitelson et al., 2007) (Eq. 5). Compared with empirical algorithms, the three-band algorithm can minimize the effects of non-algal particles and CDOM in retrieving Chla when their absorption is negligible compared with that of pure water.

(3) Machine learning algorithms.

In recent years, machine learning algorithms have been increasingly used for remote sensing investigations in marine, coastal and inland waters for they can utilize complex structures to capture sufficient features of input variables and to obtain more accurate output results (Cao et al., 2020b; Pyo et al., 2019). Several of the most mature and widely used machine learning algorithms, e.g., neural networks, support vector machines (SVMs) and random forest (RF), have been proven to be powerful tools in Chla retrieval (Bazi and Melgani, 2007; Du et al., 2015; Keiner and Yan, 1998; Sun et al., 2009; Topouzelis and Psyllos, 2012; Zeng et al., 2020).

The neural network is a mathematical model that imitates the structure and function of a biological neural network and comprises a mass of connected nodes (or 'neurons') (Krizhevsky et al., 2017). BP is a multilayer feedforward network trained by the error backpropagation algorithm proposed by Rumelhart et al. (1986) that can express extremely complex nonlinear relationships (Boccaletti et al., 2006) and is currently the most commonly encountered type of neural network (Wang et al., 2015). The forward calculation, local gradient feedback calculation, and transfer weight correction between neurons and error calculation constitute the core of BP. The basic structure of BP is comprised of an input layer, output layer and hidden layer.

SVMs, originally proposed to address dichotomous problems (Cortes and Vapnik, 1995), have been widely used for multiple classification and regression. The principle of SVR is that the sample in the low-dimensional space is mapped to a high-dimensional feature space through a nonlinear transformation function, and the regression linear hyperplane is sought in the feature space such that the "distance" to the farthest sample point in the hyperplane is minimized, thus solving the highly nonlinear problem in the low-dimensional space (Cortes and Vapnik, 1995). SVMs avoid the problems associated with high dimensionality and overfitting and can achieve high prediction accuracy and fast convergence (Vapnik, 1998). SVR is particularly advanced in addressing the problem in which the feature dimension is greater than the sample size (Pal and Foody, 2010). By minimizing the structural and

empirical risks and the confidence range, the generalization ability of the algorithm is improved; thereby, good prediction results can be obtained under a small statistical sample size (Suykens et al., 2001).

RF is a representative ensemble classifier comprised of several tree structure classifiers and is an extended variant of bagging. Based on the construction of the bagging ensemble of decision tree learners, random attribute selection is further introduced in the training process of decision trees (Breiman, 2001). The RFR algorithm automatically creates a random decision tree group, selects the random variable set from the training dataset, and constructs each tree using random reposition sampling (Mutanga et al., 2012). Each decision tree is generated independently, and each node is segmented by a predefined number of features (Mtry) by random selection. RFR creates trees with high variance and low bias by growing forests into a predefined number of classification trees (Ntrees) (Odindi et al., 2014) and finally estimates the results by equalizing all of the trees. It establishes parallel processing for mutually independent trees, has prominent robustness against missing or unbalanced data and can commendably predict the results of up to thousands of variables (Iverson et al., 2008).

2.4. Development of Chla algorithms

To minimize the impact of the solar elevation on $R(\lambda)$, the average value of $R(\lambda)$ from 575 nm to 585 nm was taken as the normalization parameter, and $R_{norm}(\lambda)$ is the spectrum after normalization by the following equation:

$$R_{norm}(\lambda) = \frac{R(\lambda)}{\sum_{575}^{585} R(\lambda) / 11} \quad (10)$$

For the empirical and semi-analytical algorithms, the independent variables were calculated using the optimum wavelengths according to spectral characteristics, and the dependent variable was Chla. The wavelengths of 665 nm and 709 nm were selected as λ_1 and λ_2 of the band-ratio algorithm and NDCI algorithm, respectively, and 665 nm, 709 nm and 754 nm were chosen to be λ_1 to λ_3 , respectively, for the MCI algorithm. The λ_1 to λ_3 combinations of the three-band algorithms were 671–710–740 nm and 675–695–730 nm, respectively.

For machine learning algorithms, a total of 501 wavelengths from 400 nm to 900 nm were selected as input variables, and the output was Chla. From a total dataset of 481 data pairs consisting of $R(\lambda)$ and Chla, 350 pairs (approximately 70% of the total dataset) were randomly selected for model training, and the remaining 131 pairs were used for testing. The machine learning algorithms were implemented by the mathematical software MATLAB 2021a. In the BP algorithm, the tansig function, purelin function, mse function, trainlm function, and learnngdm function were set as the hidden layer transfer function, output layer function, performance function, training function, and learning function, respectively. The maximum number of iterations, learning rate, target error and momentum were set to 1000, 0.001, 0.0001 and 0.95, respectively, after tuning through a step-by-step grid strategy. The optimal number of hidden layers was determined to be 2 after several experiments. To improve the convergence speed of the network and optimize the generalization ability, the dataset needs to be normalized before training. The SVR algorithm was supported by the LIBSVM-3.25 program package, a software package for pattern recognition and regression of SVM developed by Chang and Lin (2011) (<https://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html>). The Gaussian radial basis function (RBF), which can perform excellent nonlinear fitting, was selected as the kernel function of SVR. The 'grid search method' was used to determine the best parameter values of penalty factor c and kernel parameter g because parameter optimization can improve the prediction accuracy of SVR and reduce the influence of training sample size on the generalization ability (Bazi and Melgani, 2007). In the RFR algorithm, the number of regression trees (Ntree) was set to 50.

2.5. Statistical analysis and accuracy assessment

SPSS 24.0 software was used for statistical analysis, including calculation of average, median, maximum and minimum values. Pearson's correlation analysis was utilized to study the relationship between variables. The significance level was divided into highly significant ($p \leq 0.01$), significant ($0.01 < p \leq 0.05$) and nonsignificant ($p > 0.05$). The coefficient of determination (R^2), root mean square error (RMSE) and mean relative error (MRE) were used to perform the statistical analyses. The performance metrics are defined as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_{m,i} - Y_{e,i})^2} \quad (11)$$

$$MRE = \frac{\sum_{i=1}^N \left| \frac{Y_{m,i} - Y_{e,i}}{Y_{m,i}} \right|}{N} \times 100 \quad (12)$$

where N is the number of data pairs and Y_m and Y_e are the measured values and estimated values of the model, respectively. The subscript i denotes individual data points.

3. Algorithm development and validation embedded in the GRSS for Chla retrieval

3.1. Biogeochemical and optical characteristics of the sampled waters

The water quality parameters of the sampling experiments summarized in Table S1 covered a wide range. Chla ranged from 0.70 to 442.94 $\mu\text{g/L}$, and the average (mean \pm standard deviation, the same below) was $30.03 \pm 54.53 \mu\text{g/L}$, which mostly covered the distribution of Chla in inland waters. The highest average Chla was found in Lake Taihu, with a value of $53.61 \pm 78.36 \mu\text{g/L}$, while a similar average of $45.09 \pm 28.07 \mu\text{g/L}$ was observed in the Liangxi River. Chla was significantly low in Lake Qiandaohu and the Fuchunjiang Reservoir, with averages of $2.36 \pm 0.63 \mu\text{g/L}$ and $1.11 \pm 0.19 \mu\text{g/L}$, respectively.

TN, TP, COD and TSM all displayed high values in Lake Taihu and Liangxi River but low values in Lake Qiandaohu and the Fuchunjiang Reservoir, similar to the distribution of Chla. In contrast, SDD presents the opposite distribution, indicating that Lake Taihu and the Liangxi River could be considered turbid waters, while Lake Qiandaohu and the Fuchunjiang Reservoir are characterized as clear or slight turbid waters. To acquire a deeper understanding of the properties, the comprehensive trophic level index (TLI) was assessed based on Chla, SDD, TN, TP and COD (Wang, 2002). Lake Qiandaohu and the Fuchunjiang Reservoir were oligotrophic and mesotrophic waters, respectively, while Lake Taihu and the Liangxi River were hypereutrophic waters.

$R(\lambda)$ spectra were significantly different in terms of the shape and values for the four sampling waters (Fig. 2) and can be roughly divided into two types: Type I contained Lake Taihu and the Liangxi River, which were dominated by Chla. The absorption characteristics of Chla caused it to form a reflection valley at 665 nm and a strong reflection peak with pure water at approximately 700 nm. Type II included Lake Qiandaohu and the Fuchunjiang Reservoir, with low Chla. The spectral characteristics of $R(\lambda)$ were not obvious between 650 nm and 750 nm, and there was only a reflection peak near 570 nm, which was formed by the weak absorption of Chla and the scattering of cells. With the increase in Chla, the position of the reflection peak of Type II water near 570 nm moves toward the shortwave direction; thus, the peak positions of Type I water were close to 560 nm. The $R(\lambda)$ value of the Fuchunjiang Reservoir was higher than that of the Liangxi River near 570 nm because the former was dominated by TSM and the concentration of Chla was low; therefore, the optical characteristics mainly involved scattering, which increased the reflectance value, while the latter was dominated by Chla, with obvious optical absorption and a reduced reflectance value. The $R(\lambda)$ spectra of the Fuchunjiang Reservoir and Lake Qiandaohu in water type II were quite different in value because Chla and suspended matter

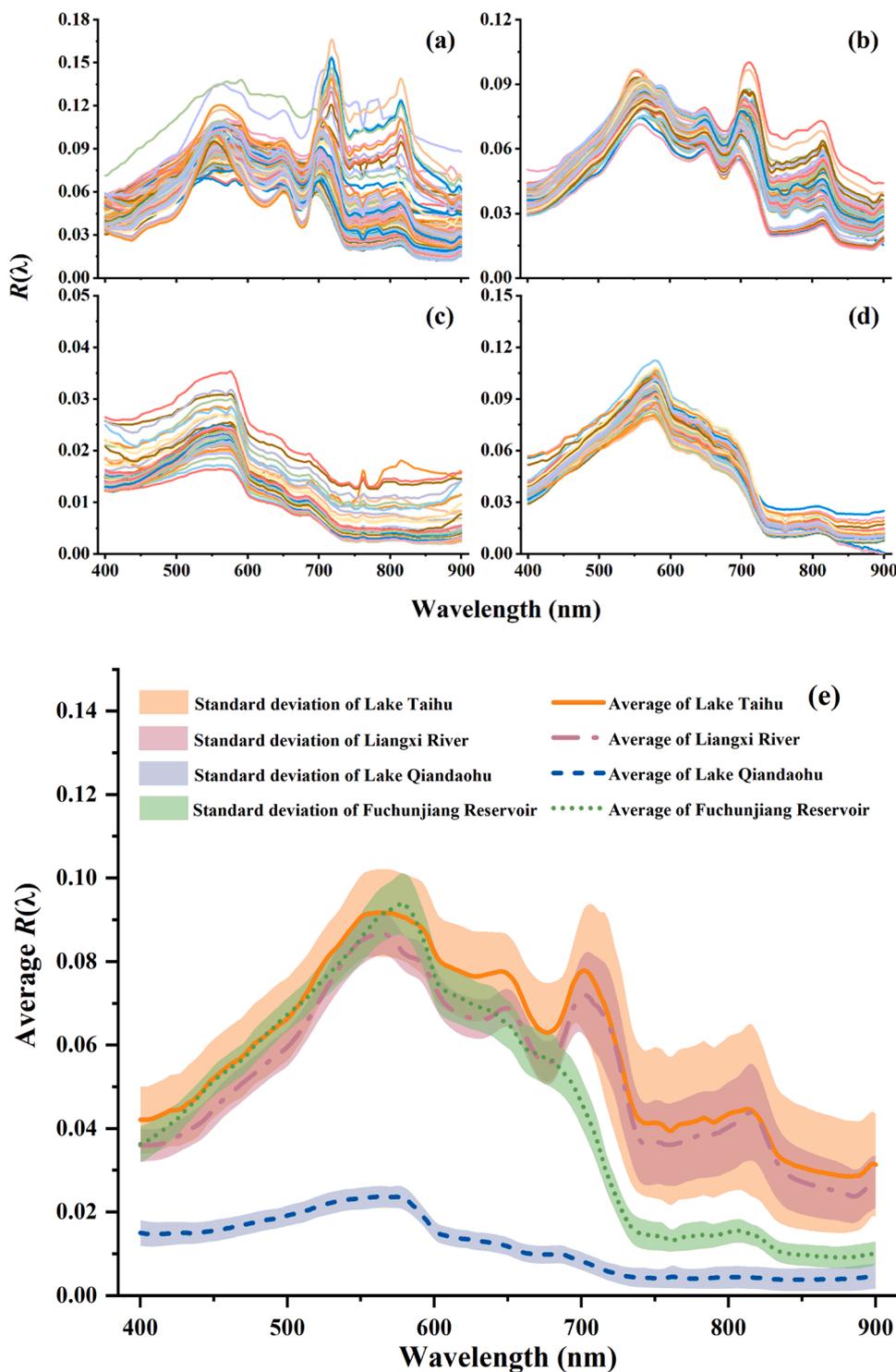


Fig. 2. The $R(\lambda)$ spectra of (a) Lake Taihu, (b) the Liangxi River, (c) Lake Qiandaohu and (d) the Fuchunjiang Reservoir and the average spectra of (e) the four sampling waters.

were extremely low in Lake Qiandaohu; therefore, the value of $R(\lambda)$ was close to that of pure water.

3.2. Validation of the Chla retrieval algorithms

The results for Chla validation with each algorithm are illustrated in Fig. S2 and Table 1. Comparing the *in situ* measured and estimated data of Chla empirical algorithms, the NDCI algorithm yielded low error (Fig. S2c: MRE = 50.04%, RMSE = 22.58 $\mu\text{g/L}$) in the case of a wide

range of datasets. The correlation (R^2) between measured and estimated data of 0.83, presented reasonable validation results, which demonstrated the accuracy and applicability of $R(\lambda)$ spectra measured by GRSS. The performance of the MCI algorithm was inferior to that of the NDCI in terms of low value estimation (Fig. S2b: MRE = 90.36%, RMSE = 25.10 $\mu\text{g/L}$, $R^2 = 0.79$), and the band-ratio algorithm performed the worst, with an MRE of 227.22%, remaining the highest among the empirical algorithms (Fig. S2a).

The performance of semi-analytical algorithms was marginally

worse than that of the empirical algorithms, with the MRE ranging from 140.43% to 235.54% and the RMSE ranging from 23.72 $\mu\text{g/L}$ to 30.85 $\mu\text{g/L}$. The three-band algorithm proposed by Dall'Olmo and Gitelson in 2005 performed best among the semi-analytical algorithms, with $R^2 = 0.81$, MRE = 140.43%, and RMSE = 23.72 $\mu\text{g/L}$ (Fig. S2e). The widest scatter of estimated values was found for the Gons05 algorithm (Fig. S2d: MRE = 235.54%, RMSE = 30.85 $\mu\text{g/L}$, $R^2 = 0.68$), followed by the three-band algorithm proposed by Gitelson et al. (2007) (Fig. S2f: MRE = 166.53%, RMSE = 27.83 $\mu\text{g/L}$, $R^2 = 0.74$). Notably, a small number of negative Chla retrievals were obtained when using empirical and semi-analytical algorithms, which typically occur at Chla below 10 $\mu\text{g/L}$. This error regarding low Chla has also been discussed in previous studies (Liu et al., 2020). As these empirical and semi-analytical algorithms were developed based on turbid and productive waters, while the dataset used in this study contains non-negligible samples with low Chla, large errors may occur in estimating low Chla.

The scatter plots in Fig. S2 highlight the finding that all of the machine learning algorithms performed reasonably in terms of estimating Chla (Figs. S2e-4o), even at low Chla, when compared with empirical and semi-analytical algorithms. Most notably, these algorithms yielded R^2 values in excess of 0.90 and regression slopes close to 1 compared with the *in situ* measurements. This finding is not surprising considering that machine learning is equipped with more complex algorithm structures and mathematical calculations that can fully utilize spectral information (Cao et al., 2020b; Pyo et al., 2019). The RFR algorithm performs reasonably well (Fig. S2n, $R^2 = 0.97$, slope = 1.02, RMSE = 8.25 $\mu\text{g/L}$, and MRE = 21.55%) but slightly overestimates high Chla values, which may be related to the small amount of data with high Chla values in the training dataset. The SVR algorithm shows a similar excellent performance, with $R^2 = 0.97$, slightly lower than that of RFR, and slope = 1.00, RMSE = 7.86 $\mu\text{g/L}$, and MRE = 55.70% (Fig. S2k). The Chla estimation result of the BP algorithm was marginally inferior to the two mentioned above, showing the highest uncertainties in the three machine learning statistical measures (Fig. S2h: $R^2 = 0.95$, slope = 1.17, RMSE = 15.73 $\mu\text{g/L}$, and MRE = 86.24%).

In light of the wide range of Chla values in the dataset, the performance of most of the algorithms validated here should be considered satisfactory within their scope of application. The empirical algorithms are simple to calculate and have high applicability in certain regions. The semi-analytical algorithms integrate optical principles, which are more convincing in mechanism (Duan et al., 2012b; Duan et al., 2010; Gons et al., 2005). However, the spectral information of Chla may be ignored due to the limited parameters of the independent variables of empirical and semi-analytical algorithms. The machine learning algorithms in this study used the $R(\lambda)$ information for all wavelengths in the calculations, thus obtaining more accurate results than the empirical algorithms through complex mathematical calculations. In summary, for the GRSS with a spectral interval of 1 nm, machine learning algorithms offer obvious advantages over other algorithms, and RFR was the best algorithm for the accurate quantitative monitoring of Chla using GRSS.

4. Initial application and discussion

4.1. Robustness and stability of machine learning algorithms

4.1.1. Robustness evaluation of machine learning algorithms

To further determine the applicability of these three machine learning algorithms in the GRSS, we completed a robustness evaluation. Gaussian noise of 2%, 5% and 10% was added to the total dataset to evaluate the performance of each algorithm under different levels of interference. Similar to the method in Section 2.4, 350 pairs of noisy data were randomly selected as the training dataset, and the remaining 131 pairs of data were used as the test dataset. The experimental results of the test dataset are listed in Table 2.

With increasing noise, the comparisons between the Chla estimates

Table 2

Stability and robustness after adding noise, including the evaluation indicators of the three machine learning methods in retrieving Chla and the descriptive statistics of the data.

	Noise	Linear slope	MRE (%)	RMSE ($\mu\text{g/L}$)	R^2
BP	2%	1.17	144.31	15.74	0.95
	5%	1.01	376.8	12.38	0.94
	10%	0.85	389.6	19.29	0.88
SVR	2%	0.90	59.82	8.96	0.97
	5%	0.92	269.09	9.06	0.97
	10%	1.02	143.31	8.49	0.97
RFR	2%	1.04	57.47	8.60	0.97
	5%	1.02	259.14	8.91	0.97
	10%	1.03	178.28	9.57	0.97

from the three machine learning algorithms and the measured Chla presented obvious differences. The R^2 between the measured and estimated values in the test dataset of the BP algorithm gradually decreased, while the MRE and RMSE increased with increasing noise. Here, the accuracy of the BP algorithm depended largely on the input data. SVR performed relatively stable and effectively in the experiment with added noise. The fitting R^2 remained at 0.97, and the RMSE of the estimated data varied little. The result of the RFR algorithm showed similar performance with SVR, with the MRE remaining relatively low. The robustness outcomes of the SVR and RFR algorithms were better than that of the BP algorithm. However, an obvious defect of the SVR and BP algorithms is that the low Chla measurements were estimated as negative values, which conflicts with the fact that Chla is always positive. Therefore, in terms of robustness and accuracy, the RFR algorithm achieved the optimal performance among the machine learning algorithms.

4.1.2. Stability over time series

Based on the evaluation of overall accuracy and robustness, we expanded the estimates of three algorithms (BP, RFR and SVR) into the original time series to intuitively analyze the fitting performance. Chla was high and widely distributed in Lake Taihu but low with little variation in Lake Qiandaohu. Therefore, selecting these two sites for the time series evaluation of the algorithm is typical and reasonable.

In the total dataset with large fluctuations, the performance of the three machine learning algorithms was satisfactory in Lake Taihu because they were able to fit the temporal variation trend of Chla extremely well (Fig. 3). However, SVR and RFR slightly underestimated the high value of Chla mutation caused by phytoplankton blooms at Lake Taihu from November 1 to November 3, 2020. Comparatively, the BP algorithm obviously overestimated low values, even the peak that occurred on November 4, 2020.

Since we performed algorithmic modeling on the total dataset, the time series fitting analysis of Lake Qiandaohu can be used as a test of the estimation performance of low-value samples in the total dataset with a large fluctuation range (Fig. 4). The BP algorithm tended to overestimate Chla less than 10 $\mu\text{g/L}$. The estimated value of the SVR algorithm was close to the measured value but was not consistent with the original time series. As shown in Fig. 4, the error between the estimated and measured values of RFR was much smaller than those of the SVR and BP algorithms for the low Chla in the entire dataset.

In the research on Chla retrieval for various inland waters based on the $R(\lambda)$ from the GRSS, the performances, robustness and characteristics of the three machine learning methods above were comprehensively considered. The results showed that RFR is the optimal algorithm for estimating Chla combined with the GRSS.

4.2. Advantages and limitations of the proposed GRSS

Following the rigorous comparison and analysis presented above, the

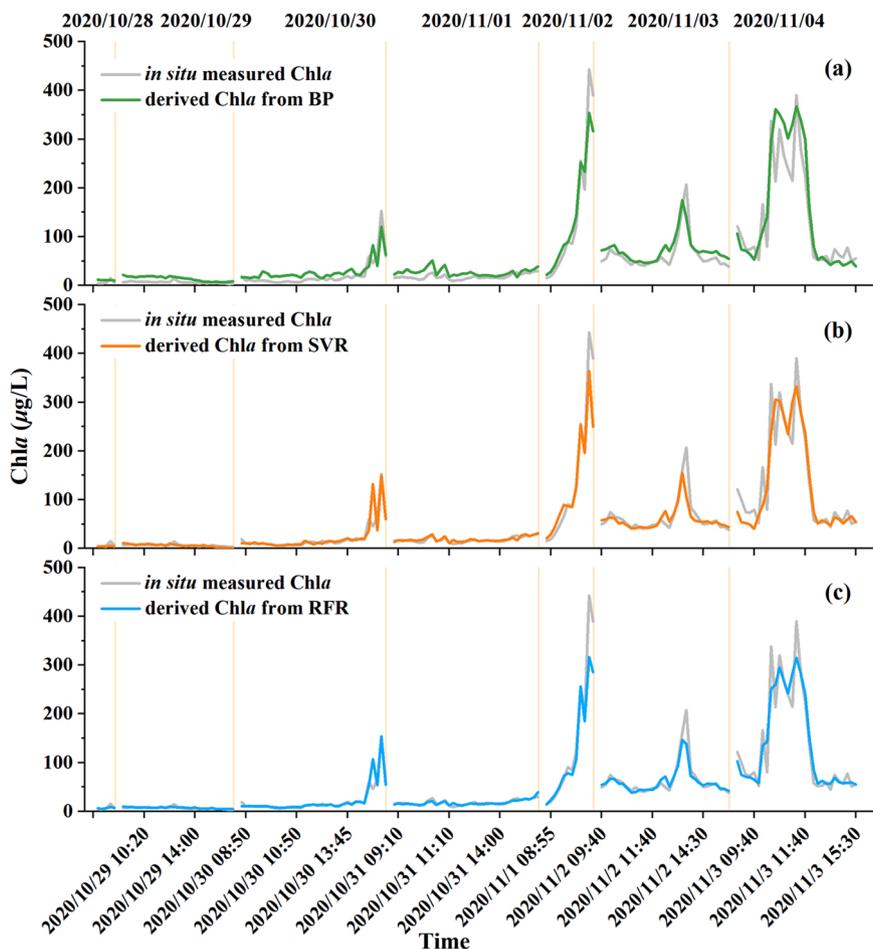


Fig. 3. Temporal results of the Chla estimated by BP, RFR and SVR in Lake Taihu. The gray curve represents the measured Chla, and the green, orange and blue curves represent the Chla estimated by the (a) BP, (b) SVR, and (c) RFR algorithms, respectively.

GRSS combined with the RFR algorithm has been shown to be capable of the high-frequency and real-time estimation of Chla in waters with different optical properties, with excellent estimation performance. Compared with previous research on Chla measurement (Bazi and Melgani, 2007; Du et al., 2015), the GRSS undoubtedly expands our understanding of the remote sensing monitoring of Chla in inland waters and contributes to a better method of discerning the temporal variation of Chla with a high frequency. In general, the GRSS owns several major advantages as follows.

First, the GRSS can instantaneously and continuously provide high-frequency data on phytoplankton blooms in an environmentally friendly manner and at low cost. Compared with conventional *in situ* sampling, the GRSS saves considerable labor and material resources, and its monitoring process does not produce wastewater, thus avoiding causing secondary environmental pollution. Moreover, continuous and high-frequency observations compensate for the insufficient amount of *in situ* sampling data and the inability to obtain data continuously in real time, especially during extreme weather events such as typhoons and summer heatwaves, which strongly restructure phytoplankton blooms (Bartosiewicz et al., 2019; Johnk et al., 2008; Zhu et al., 2014). More importantly, the operation and maintenance of the GRSS are less expensive than those of traditional automatic instruments. This feature makes the GRSS more economical and automatic, and it has a longer service life than automatic monitoring stations equipped with YSI or FluoroProbe. This also provides a good premise for the long-term monitoring of phytoplankton dynamics at a broad spatial scale after the arrangement of multiple GRSS points.

Second, the water irradiance signal obtained by the GRSS is free of

atmospheric influences. The water signal obtained by satellite remote sensing is greatly affected by the atmospheric correction method (Chavez, 1996). The near-surface observation position of the GRSS compensates for the disadvantages arising from the high reflectance uncertainty in satellite remote sensing due to atmospheric interference and thus avoids the need for a complex atmospheric correction. To verify the accuracy of the irradiance signal of the GRSS, we compared the $R(\lambda)$ obtained by the surface object spectrometer ASD (FieldSpec 4 Hi-Res spectroradiometer, Analytical Spectra Devices, Inc.) and the GRSS at the same time and from the same sampling site. The results showed satisfactory consistency (Fig. 5a). This finding further indicates that the GRSS can accurately and effectively obtain water information with different optical properties and distinguish water signals from phytoplankton blooms (Fig. 5b).

Third, the GRSS has strong practicality for monitoring phytoplankton bloom dynamics at high frequency in small rivers and lakes. Unlike traditional satellite remote sensing, the monitoring capability of the GRSS is not limited by its spatial resolution, which enables the continuous observation of small waters such as streams and urban waterways. Affected by the cloud and return visit cycle, the number of available images provided by satellite remote sensing within one year is limited. The GRSS can conduct second-level high-frequency observations under different weather conditions (sunny, cloudy and even light rain), which greatly increases the amount of available data. Phytoplankton blooms have strong spatiotemporal dynamics, and the GRSS can monitor aquatic environments at a high frequency in real time, thus capturing the occurrences of and changes in phytoplankton blooms in a timely manner (Fig. 5b).

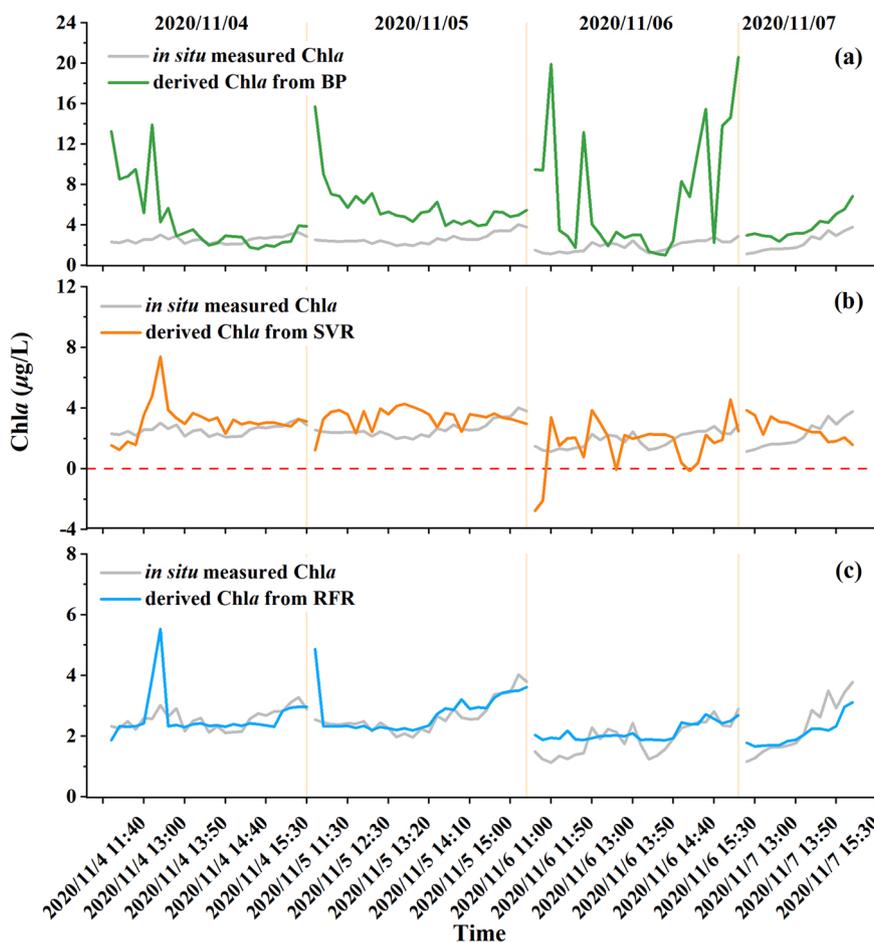


Fig. 4. Temporal results of the Chla estimated by BP, RFR and SVR in Lake Qiandaohu. The gray curve represents the measured Chla, and the green, orange and blue curves represent the Chla estimated by the (a) BP, (b) SVR, and (c) RFR algorithms, respectively.

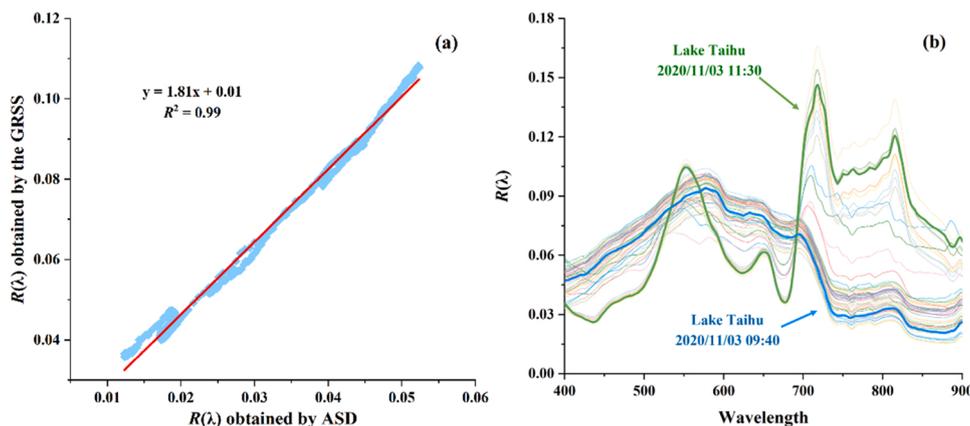


Fig. 5. (a) Comparison of the $R(\lambda)$ obtained by ASD and the GRSS. The blue points refer to the reflectivity of the ASD and GRSS at the same time and the same site, and the red line is the fitting line. (b) $R(\lambda)$ curve of different water states recognized by the GRSS within a short time. The blue curve is the water signal in Lake Taihu obtained at 09:04 a.m. on November 3, 2020, and the green curve is the phytoplankton bloom signal in Lake Taihu obtained at 11:00 a.m. on the same day.

Finally, the AI chip embedded in the GRSS supports algorithm replacement and upgrade, as well as remote control and access to data. An important breakthrough of the GRSS compared to existing monitoring methods is the integration of AI chip. The powerful capabilities of the AI chip include real-time calculation of the collected spectra in combination with algorithms, real-time feedback to the user via wireless transmission, and data storage. Users can remotely access all data from the calculation process, including raw irradiance, $R(\lambda)$, and Chla

estimation results as well as photos and videos of the monitoring site. The algorithms embedded in the AI chip can be upgraded or replaced, providing strong technical support for the subsequent development and application of GRSS.

There are still some limitations in this study. Water constituents generally have complex compositions in inland waters, resulting in substantial differences in the optical properties. Although our training and validating dataset has a wide range of concentrations and

compositions of water constituents, there are still many water types that have not been considered. The RFR algorithm is a data-driven machine learning algorithm and generally lacks physical mechanisms; thus, its transferability may be restricted by our training dataset.

In the future, we will collect more samples from various inland waters to persistently expand the GRSS database and improve machine learning algorithms by adding physical mechanisms. In addition, establishing a regional algorithm library for different types of inland water according to their optical properties should be considered. In this way, the spectra obtained by GRSS will first be matched to the most similar category based on their characteristics and then incorporated with the regional algorithm to estimate Chla with higher accuracy.

4.3. Initial applications and implications for environmental management

The fundamental product presented in the GRSS is a platform capable of displaying the real-time dynamics of Chla with a high frequency of 20 s. To demonstrate the role and significance of the GRSS in monitoring phytoplankton blooms, we set up a GRSS hyperspectral sensor in Lake Taihu to capture the dynamics of phytoplankton blooms. As Lake Taihu is a large eutrophic shallow lake, its water surface conditions are sensitive to short-term environmental changes, especially phytoplankton blooms, which change rapidly with wind and temperature. Two phytoplankton blooms that occurred in August 2021 in Lake Taihu were clearly captured with the GRSS (Fig. 6).

On August 11, 2021, the Chla at the study site remained low, at $\sim 20 \mu\text{g/L}$, from 8:30–14:00, except for a small increase at approximately 9:00 (Fig. 6). An apparent increase in Chla began at 14:20, reflecting the initiation of a diminutive bloom. The Chla peaked at $150 \mu\text{g/L}$ on that day at 15:10. Then, Chla continued to decline and finally dropped to the background level of $\sim 20 \mu\text{g/L}$ at 16:50. On August 25, 2021, an intense phytoplankton bloom occurred. From 7:50–9:00, the Chla fluctuated slightly between 40 and $50 \mu\text{g/L}$. Then, it manifested itself similar to the bloom event observed on August 11, 2021. The Chla began to increase at 9:00 and reached a peak of $\sim 200 \mu\text{g/L}$ at 11:20, representing a rapid accumulation of phytoplankton. It increased abruptly to $70 \mu\text{g/L}$ at 12:00. Subsequently, Chla began to increase rapidly. The increase in Chla continued until 15:00, and then Chla reached $275 \mu\text{g/L}$. In the last four hours of observations, Chla was maintained at a high value of approximately $250\text{--}275 \mu\text{g/L}$ with small fluctuations, and it reached the highest value of the day of $325 \mu\text{g/L}$ at approximately 18:00.

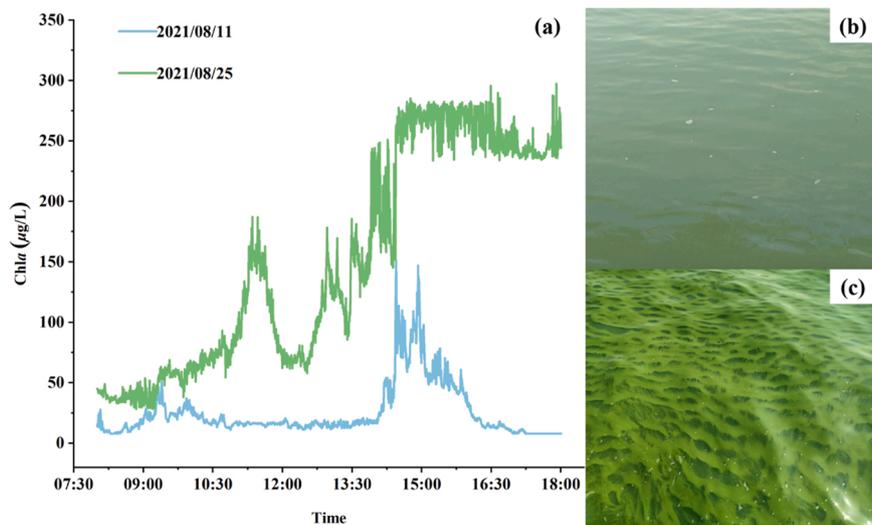


Fig. 6. Two phytoplankton blooms captured by the GRSS. (a) The blue curve represents the change in Chla observed by the GRSS on August 11, 2021, while the green curve represents the Chla on August 25, 2021, and photographs of the water condition at the highest Chla value on August 11, 2021 (b) and August 25, 2021 (c) are shown.

From the perspective of dramatic changes in Chla, the observation site experienced a rapid bloom process in just a few hours, which is too short for a complete cycle of phytoplankton reproduction. Therefore, the rapid changes in the phytoplankton bloom reflected by the monitored Chla indicate that the distribution of cyanobacteria was greatly affected by wind. Intraday variability of phytoplankton blooms has been documented with Geostationary Ocean Color Imager (GOCI) data in previous studies, but it is still difficult to finely characterize the dynamics over shorter time scales (Huang et al., 2015; Qi et al., 2018). The capture of similar rapid and dynamic changes in phytoplankton blooms plays an important role in deep understandings of phytoplankton bloom formation mechanism, their forecast, early warning and management. Such rapid shifts in Chla over a period of hours underscore the value of high-frequency monitoring. However, low-frequency *in situ* observations and satellite remote sensing monitoring miss many key processes of phytoplankton bloom dynamics, inhibiting our deep understanding of phytoplankton bloom formation mechanisms and controlling strategies. Therefore, installing multiple GRSS hyperspectral probes at important observation points can provide evidence for the processes of surface environmental changes and thus provide strong evidence for mechanistic studies.

The GRSS proposed in this study can well compensate for the deficiencies of existing phytoplankton bloom monitoring methods and is proven to be capable of monitoring real-time Chla accurately at high frequencies, thus effectively reflecting the dynamic process of phytoplankton blooms. As a result, it can serve as a foundation and prerequisite for management agencies to make prudent decisions when implementing phytoplankton bloom management strategies. In a broader perspective, the GRSS can also be extended to other carriers, such as aircraft, ships and unmanned aerial vehicles, to more flexibly monitor phytoplankton blooms. In this way, phytoplankton bloom monitoring will no longer be fixed points, and linear monitoring can be formed along the routes of ships or aircrafts. In addition, a phytoplankton monitoring network can be formed when multiple GRSSs work simultaneously, thus greatly improving the efficiency and accuracy of phytoplankton bloom dynamic monitoring. Therefore, we recommend that future inland water monitoring plans include the GRSS to aid in the interpretation of the spatial and temporal patterns of phytoplankton blooms.

5. Conclusion

In this study, we propose a novel system named the GRSS for the accurate, real-time acquisition of Chla in inland waters with a high frequency. Empirical, semi-analytical, and machine learning algorithms (BP, SVR, and RFR) were developed to retrieve Chla from the GRSS based on 481 pairs of sampled data covering different optical properties. The RFR algorithm is the best algorithm for inverting Chla with the GRSS ($R^2 \geq 0.95$). Compared with *in situ* sampling, automatic monitoring stations and satellite remote sensing, the GRSS can continuously provide high-frequency data and conserve labor and material resources. The hyperspectral camera of the GRSS is not polluted by substances in water; thus, its operation and maintenance costs are low. The proposed system has an accurate radiation signal and a high signal-to-noise ratio. With its high temporal and spatial resolutions, it can be applied to the monitoring of small streams and ponds. Two distinct bloom processes were successfully captured by the GRSS in Lake Taihu, proving its importance in phytoplankton blooms dynamics monitoring. Algorithm optimization including more synchronous data is needed to expand the application in inland waters in future research. Further insight and a deep understanding of phytoplankton bloom dynamics can be gained by combining the GRSS with satellite remote sensing to provide a comprehensive analysis of the increasing risk of harmful algal blooms. Therefore, the GRSS can serve as a significant component of a comprehensive aquatic environment monitoring system with many application prospects in the future.

CRedit authorship contribution statement

Weijia Wang: Investigation, Conceptualization, Data curation, Methodology, Writing - original draft. **Shi Kun:** Supervision, Methodology and Editing. **Yibo Zhang:** Methodology, Software, Validation. **Na Li:** Data curation, Methodology, Validation. **Xiao Sun:** Investigation, Conceptualization. **Dong Zhang:** Software, Validation. **Yunlin Zhang:** Data curation and Editing. **Boqiang Qin:** Conceptualization, Validation and Editing. **Guangwei Zhu:** Writing - review and editing.

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.

Data availability

Data will be made available on request.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (41922005), the Scientific Instrument Developing Project of the Chinese Academy of Sciences (YJKYYQ20200071), the Tibetan Plateau Scientific Expedition and Research Program, (2019QZKK0202), NIGLAS foundation (E1SL002), and the Water Resource Science and Technology Project in Jiangsu Province (2020057). We would like to thank all of the partners for their participation in field sample collection and experimental analysis.

Environmental implications

It is very difficult to accurately capture harmful algal blooms using the traditional cross-section monitoring and satellite remote sensing due to their rapid migration. In this study, we propose a novel ground-based remote sensing system (GRSS) that can monitor real-time harmful algal blooms in inland waters with a high frequency. The high-frequency, high-accuracy and real-time data acquisition capabilities of the GRSS

can compensate for the insufficiencies of *in situ* sampling, automatic observation stations and satellite observations. The GRSS can provide new insights into phytoplankton bloom dynamics and thus improve our ability of harmful algal blooms prediction, early warning and management.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jhazmat.2022.129623.

References

- APHA (1995) Standard methods for the examination of water and waste water, 19th edn. NewYork, USA.
- Bartosiewicz, M., Przytulska, A., Deshpande, B.N., Antoniadis, D., Cortes, A., MacIntyre, S., Lehmann, M.F., Laurion, I., 2019. Effects of climate change and episodic heat events on cyanobacteria in a eutrophic polymictic lake. *Sci. Total Environ.* 693, 133414.
- Bazi, Y., Melgani, F., 2007. Semisupervised PSO-SVM regression for biophysical parameter estimation. *IEEE Trans. Geosci. Remote Sens.* 45 (6), 1887–1895.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.U., 2006. Complex networks: structure and dynamics. *Phys. Rep. -Rev. Sect. Phys. Lett.* 424 (4–5), 175–308.
- Bonanea, M., Claudia Rodriguez, M., Pinotti, L., Ferrero, S., 2015. Using multi-temporal Landsat imagery and linear mixed models for assessing water quality parameters in Rio Tercero reservoir (Argentina). *Remote Sens. Environ.* 158, 28–41.
- Bowling, L.C., Zamyadi, A., Henderson, R.K., 2016. Assessment of in situ fluorometry to measure cyanobacterial presence in water bodies with diverse cyanobacterial populations. *Water Res.* 105, 22–33.
- Breiman, L., 2001. Random forests. *Mach. Learn* 45 (1), 5–32.
- Bricker, S.B., Longstaff, B., Dennison, W., Jones, A., Boicourt, K., Wicks, C., Woerner, J., 2008. Effects of nutrient enrichment in the nation's estuaries: a decade of change. *Harmful Algae* 8 (1), 21–32.
- Cao, Z.G., Ma, R.H., Duan, H.T., Pahlevan, N., Melack, J., Shen, M., Xue, K., 2020. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. *Remote Sens. Environ.* 248, 111974.
- Cao, Z.G., Ma, R.H., Melack, J.M., Duan, H.T., Liu, M., Kutser, T., Xue, K., Shen, M., Qi, T. C., Yuan, H.L., 2022. Landsat observations of chlorophyll-a variations in Lake Taihu from 1984 to 2019. *Int. J. Appl. Earth Obs. Geoinf.* 106, 102642.
- Carmichael, W.W., 2001. Health effects of toxin-producing cyanobacteria: "The CyanoHABs". *Hum. Ecol. Risk Assess.* 7 (5), 1393–1407.
- Chang, C.C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. *ACM T Intell. Syst. Tec.* 2 (3), 1–27.
- Chavez, P.S., 1996. Image-based atmospheric corrections revisited and improved. *Photogramm. Eng. Remote Sens.* 62 (9), 1025–1036.
- Cogliati, S., Rossini, M., Julitta, T., Meroni, M., Schickling, A., Burkart, A., Pinto, F., Rascher, U., Colombo, R., 2015. Continuous and long-term measurements of reflectance and sun-induced chlorophyll fluorescence by using novel automated field spectroscopy systems. *Remote Sens. Environ.* 164, 270–281.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20 (3), 273–297.
- Dall'Olmo, G., Gitelson, A.A., 2005. Effect of bio-optical parameter variability on the remote estimation of chlorophyll-a concentration in turbid productive waters: experimental results. *Appl. Opt.* 44 (3), 412–422.
- 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 (9), 2255–2266.
- Du, P., Samat, A., Waske, B., Liu, S., Li, Z., 2015. Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS J. Photogramm. Remote Sens.* 105, 38–53.
- Duan, H.T., Ma, R.H., Hu, C.M., 2012. Evaluation of remote sensing algorithms for cyanobacterial pigment retrievals during spring bloom formation in several lakes of East China. *Remote Sens. Environ.* 126, 126–135.
- Duan, H.T., Ma, R.H., Zhang, Y.Z., Loisel, S.A., Xu, J.P., Zhao, C.L., Zhou, L., Shang, L. L., 2010. A new three-band algorithm for estimating chlorophyll concentrations in turbid inland lakes. *Environ. Res. Lett.* 5 (4), 044009.
- Ebina, J., Tsutsui, T., Shirai, T., 1983. Simultaneous determination of total nitrogen and total phosphorus in water using peroxodisulfate oxidation. *Water Res.* 17 (12), 1721–1726.
- Geetha, S., Gouthami, S., 2017. Internet of things enabled real time water quality monitoring system. *Smart Water* 2, 1.
- Gitelson, A.A., Dall'Olmo, G., Moses, W., Rundquist, D.C., Barrow, T., Fisher, T.R., Gurlin, D., Holz, J., 2008. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: validation. *Remote Sens. Environ.* 112 (9), 3582–3593.
- Gitelson, A.A., Schalles, J.F., Hladik, C.M., 2007. Remote chlorophyll-a retrieval in turbid, productive estuaries: Chesapeake Bay case study. *Remote Sens. Environ.* 109 (4), 464–472.
- Gons, H.J., Auer, M.T., Effler, S.W., 2008. MERIS satellite chlorophyll mapping of oligotrophic and eutrophic waters in the Laurentian Great Lakes. *Remote Sens. Environ.* 112 (11), 4098–4106.

- Gons, H.J., Rijkeboer, M., Ruddick, K.G., 2002. A chlorophyll-retrieval algorithm for satellite imagery (Medium Resolution Imaging Spectrometer) of inland and coastal waters. *J. Plankton Res.* 24 (9), 947–951.
- Gons, H.J., Rijkeboer, M., Ruddick, K.G., 2005. Effect of a waveband shift on chlorophyll retrieval from MERIS imagery of inland and coastal waters. *J. Plankton Res.* 27 (1), 125–127.
- Gower, J., King, S., Borstad, G., Brown, L., 2005. Detection of intense plankton blooms using the 709 nm band of the MERIS imaging spectrometer. *Int. J. Remote Sens.* 26 (9), 2005–2012.
- Haeder, D.-P., Banaszak, A.T., Villafane, V.E., Narvarte, M.A., Gonzalez, R.A., Walter Helbling, E., 2020. Anthropogenic pollution of aquatic ecosystems: Emerging problems with global implications. *Sci. Total Environ.* 713, 136586.
- 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.
- Hu, C., 2009. A novel ocean color index to detect floating algae in the global oceans. *Remote Sens. Environ.* 113 (10), 2118–2129.
- Huang, C.C., Shi, K., Yang, H., Li, Y.M., Zhu, A.X., Sun, D.Y., Xu, L.J., Zou, J., Chen, X., 2015. Satellite observation of hourly dynamic characteristics of algae with Geostationary Ocean Color Imager (GOCI) data in Lake Taihu. *Remote Sens. Environ.* 159, 278–287.
- Huber, S., Tagesson, T., Fensholt, R., 2014. An automated field spectrometer system for studying VIS, NIR and SWIR anisotropy for semi-arid savanna. *Remote Sens. Environ.* 152, 547–556.
- Huisman, J., Codd, G.A., Paerl, H.W., Ibelings, B.W., Verspagen, J.M.H., Visser, P.M., 2018. Cyanobacterial blooms. *Nat. Rev. Microbiol.* 16 (8), 471–483.
- Iverson, L.R., Prasad, A.M., Matthews, S.N., Peters, M., 2008. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *Ecol. Manag.* 254 (3), 390–406.
- Johnk, K.D., Huisman, J., Sharples, J., Sommeijer, B., Visser, P.M., Stroom, J.M., 2008. Summer heatwaves promote blooms of harmful cyanobacteria. *Glob. Change Biol.* 14 (3), 495–512.
- Kahru, M., Brotas, V., Manzano-Sarabia, M., Mitchell, B.G., 2011. Are phytoplankton blooms occurring earlier in the Arctic? *Glob. Change Biol.* 17 (4), 1733–1739.
- Keiner, L.E., Yan, X.H., 1998. A neural network model for estimating sea surface chlorophyll and sediments from thematic mapper imagery. *Remote Sens. Environ.* 66 (2), 153–165.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60 (6), 84–90.
- Kutser, T., 2004. Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. *Limnol. Oceanogr.* 49 (6), 2179–2189.
- Le, C., Li, Y., Zha, Y., Sun, D., Huang, C., Lu, H., 2009. A four-band semi-analytical model for estimating chlorophyll a in highly turbid lakes: the case of Taihu Lake, China. *Remote Sens. Environ.* 113 (6), 1175–1182.
- Lee, Z.P., Carder, K.L., Mobley, C.D., Steward, R.G., Patch, J.S., 1998. Hyperspectral remote sensing for shallow waters. I. A semianalytical model. *Appl. Opt.* 37 (27), 6329–6338.
- Li, N., Zhang, Y., Shi, K., Zhang, Y., Sun, X., Wang, W., Huang, X., 2022. Monitoring water transparency, total suspended matter and the beam attenuation coefficient in inland water using innovative ground-based proximal sensing technology. *J. Environ. Manag.* 306, 114477.
- Liu, G., Li, L., Song, K., Li, Y., Lyu, H., Wen, Z., Fang, C., Bi, S., Sun, X., Wang, Z., Cao, Z., Shang, Y., Yu, G., Zheng, Z., Huang, C., Xu, Y., Shi, K., 2020. An OLCI-based algorithm for semi-empirically partitioning absorption coefficient and estimating chlorophyll a concentration in various turbid case-2 waters. *Remote Sens. Environ.* 239, 11648.
- Liu, W., Zhang, Q., Liu, G., 2012. Influences of watershed landscape composition and configuration on lake-water quality in the Yangtze River basin of China. *Hydro. Process.* 26 (4), 570–578.
- Matthews, M.W., Bernard, S., Robertson, L., 2012. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sens. Environ.* 124, 637–652.
- Matthews, M.W., Bernard, S., Winter, K., 2010. Remote sensing of cyanobacteria-dominant algal blooms and water quality parameters in Zeekoevlei, a small hypertrophic lake, using MERIS. *Remote Sens. Environ.* 114 (9), 2070–2087.
- 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 Sens. Environ.* 117, 394–406.
- Mittenzwey, K.-H., Ullrich, S., Gitelson, A.A., Kondratiev, K.Y., 1992. Determination of chlorophyll a of inland waters on the basis of spectral reflectance. *Limnol. Oceanogr.* 37, 147–149.
- Mouw, C.B., Greb, S., Aurin, D., DiGiacomo, P.M., Lee, Z., Twardowski, M., Binding, C., Hu, C., Ma, R., Moore, T., Moses, W., Craig, S.E., 2015. Aquatic color radiometry remote sensing of coastal and inland waters: challenges and recommendations for future satellite missions. *Remote Sens. Environ.* 160, 15–30.
- Mutanga, O., Adam, E., Cho, M.A., 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *Int. J. Appl. Earth Obs. Geoinf.* 18, 399–406.
- Nasirudin, M.A., Za'bah, U.N., Sidek, O., 2011. Fresh water real-time monitoring system based on Wireless Sensor Network and GSM. 2011 IEEE Conf. Open Syst. 354–357.
- Neil, C., Spyros, E., Hunter, P.D., Tyler, A.N., 2019. A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types. *Remote Sens. Environ.* 229, 159–178.
- Odiini, J., Adam, E., Ngubane, Z., Mutanga, O., Slotow, R., 2014. Comparison between WorldView-2 and SPOT-5 images in mapping the bracken fern using the random forest algorithm. *J. Appl. Remote Sens.* 8 (1), 083527.
- O'Reilly, J.E., Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M., McClain, C., 1998. Ocean color chlorophyll algorithms for SeaWiFS. *J. Geophys. Res. -Oceans* 103 (C11), 24937–24953.
- Paerl, H.W., Barnard, M.A., 2020. Mitigating the global expansion of harmful cyanobacterial blooms: moving targets in a human- and climatically-altered world. *Harmful Algae* 96, 101845.
- Paerl, H.W., Huisman, J., 2008. Climate - blooms like it hot. *Science* 320 (5872), 57–58.
- Paerl, H.W., Otten, T.G., 2013. Blooms bite the hand that feeds them. *Science* 342 (6157), 433–434.
- Pahlevan, N., Mangin, A., Balasubramanian, S.V., Smith, B., Alikas, K., Arai, K., Barbosa, C., Belanger, S., Binding, C., Bresciani, M., Giardino, C., Gurlin, D., Fan, Y., Harmel, T., Hunter, P., Ishikawa, J., Kratzer, S., Lehmann, M.K., Ligi, M., Ma, R., Martin-Lauzer, F.-R., Olmanson, L., Oppelt, N., Pan, Y., Peters, S., Reynaud, N., de Carvalho, L.A.S., Simis, S., Spyros, E., Steinmetz, F., Stelzer, K., Sterckx, S., Tormos, T., Tyler, A., Vanhellemont, Q., Warren, M., 2021. ACIX-Aqua: a global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters. *Remote Sens. Environ.* 258, 112366.
- Pahlevan, N., Smith, B., Alikas, K., Anstee, J., Barbosa, C., Binding, C., Bresciani, M., Cremella, B., Giardino, C., Gurlin, D., Fernandez, V., Jamet, C., Kangro, K., Lehmann, M.K., Loisel, H., Matsushita, B., Ha, N., Olmanson, L., Potvin, G., Simis, S. G.H., VanderWoude, A., Vantrepotte, V., Ruiz-Verdu, A., 2022. Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. *Remote Sens. Environ.* 270.
- Pal, M., Foody, G.M., 2010. Feature selection for classification of hyperspectral data by SVM. *IEEE Trans. Geosci. Remote Sens.* 48 (5), 2297–2307.
- Palmer, S.C.J., Hunter, P.D., Lankester, T., Hubbard, S., Spyros, E., Tyler, A.N., Presing, M., Horvath, H., Lamb, A., Balzter, H., Toth, V.R., 2015. Validation of Envisat MERIS algorithms for chlorophyll retrieval in a large, turbid and optically-complex shallow lake. *Remote Sens. Environ.* 157, 158–169.
- Pyo, J., Duan, H., Baek, S., Kim, M.S., Jeon, T., Kwon, Y.S., Lee, H., Cho, K.H., 2019. A convolutional neural network regression for quantifying cyanobacteria using hyperspectral imagery. *Remote Sens. Environ.* 233, 111350.
- Qi, L., Hu, C.M., Visser, P.M., Ma, R.H., 2018. Diurnal changes of cyanobacteria blooms in Taihu Lake as derived from GOCI observations. *Limnol. Oceanogr.* 63 (4), 1711–1726.
- Qin, B., Li, W., Zhu, G., Zhang, Y., Wu, T., Gao, G., 2015. Cyanobacterial bloom management through integrated monitoring and forecasting in large shallow eutrophic Lake Taihu (China). *J. Hazard. Mater.* 287, 356–363.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323 (6088), 533–536.
- Sagan, V., Peterson, K.T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B.A., Maalouf, S., Adams, C., 2020. Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. *Earth Sci. Rev.* 205, 103187.
- Shi, K., Zhang, Y., Liu, X., Wang, M., Qin, B., 2014. Remote sensing of diffuse attenuation coefficient of photosynthetically active radiation in Lake Taihu using MERIS data. *Remote Sens. Environ.* 140, 365–377.
- Shi, K., Zhang, Y., Xu, H., Zhu, G., Qin, B., Huang, C., Liu, X., Zhou, Y., Lv, H., 2015. Long-term satellite observations of microcystin concentrations in Lake Taihu during cyanobacterial bloom periods. *Environ. Sci. Technol.* 49 (11), 6448–6456.
- Shi, K., Zhang, Y., Zhang, Y., Li, N., Qin, B., Zhu, G., Zhou, Y., 2019. Phenology of phytoplankton blooms in a trophic lake observed from long-term MODIS data. *Environ. Sci. Technol.* 53 (5), 2324–2331.
- Song, K.S., Li, L., Tedesco, L.P., Li, S., Duan, H.T., Liu, D.W., Hall, B.E., Du, J., Li, Z.C., Shi, K., Zhao, Y., 2013. Remote estimation of chlorophyll-a in turbid inland waters: three-band model versus GA-PLS model. *Remote Sens. Environ.* 136, 342–357.
- Sun, D., Li, Y., Wang, Q., 2009. A unified model for remotely estimating chlorophyll a in Lake Taihu, China, based on SVM and in situ hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* 47 (8), 2957–2965.
- Sun, X., Zhang, Y., Shi, K., Zhang, Y., Li, N., Wang, W., Huang, X., Qin, B., 2022. Monitoring water quality using proximal remote sensing technology. *Sci. Total Environ.* 803, 149805.
- Suykens, J.A.K., Vandewalle, J., De Moor, B., 2001. Optimal control by least squares support vector machines. *Neural Netw.* 14 (1), 23–35.
- Taranu, Z.E., Zurawell, R.W., Pick, F., Gregory-Eaves, I., 2012. Predicting cyanobacterial dynamics in the face of global change: the importance of scale and environmental context. *Glob. Change Biol.* 18 (12), 3477–3490.
- Tebbs, E.J., Remedios, J.J., Harper, D.M., 2013. Remote sensing of chlorophyll-a as a measure of cyanobacterial biomass in Lake Bogoria, a hypertrophic, saline-alkaline, flamingo lake, using Landsat ETM. *Remote Sens. Environ.* 135, 92–106.
- Topouzelis, K., Pysillos, A., 2012. Oil spill feature selection and classification using decision tree forest on SAR image data. *ISPRS J. Photogramm. Remote Sens.* 68, 135–143.
- Van Beusekom, J.E.E., Mengedocht, D., Augustin, C.B., Schilling, M., Boersma, M., 2009. Phytoplankton, protozooplankton and nutrient dynamics in the Bornholm Basin (Baltic Sea) in 2002–2003 during the German GLOBEC Project. *Int. J. Earth Sci.* 98 (2), 251–260.
- Vapnik, V. 1998. *The Support Vector method of function estimation.* Norwell: Kluwer Academic Publishers.
- Wang, L., Zeng, Y., Chen, T., 2015. Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. *Expert Syst. Appl.* 42 (2), 855–863.
- Wang, M., 2002. Evaluate method and classification standard on lake eutrophication. *Environ. Monit. China* 18 (5), 47–49.

- Wang, M., Shi, W., Tang, J., 2011. Water property monitoring and assessment for China's inland Lake Taihu from MODIS-Aqua measurements. *Remote Sens. Environ.* 115 (3), 841–854.
- Wang, S., Zhang, X., Chen, N., Wang, W., 2022. Classifying diurnal changes of cyanobacterial blooms in Lake Taihu to identify hot patterns, seasons and hotspots based on hourly GOCI observations. *J. Environ. Manag.* 310, 114782.
- Warren, M.A., Simis, S.G.H., Martinez-Vicente, V., Poser, K., Bresciani, M., Alikas, K., Spyarakos, E., Giardino, C., Ansper, A., 2019. Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. *Remote Sens. Environ.* 225, 267–289.
- Xu, H., Qin, B., Paerl, H.W., Peng, K., Zhang, Q., Zhu, G., Zhang, Y., 2021. Environmental controls of harmful cyanobacterial blooms in Chinese inland waters. *Harmful Algae* 110, 102190.
- Zapadka, T., Ostrowska, M., Stoltmann, D., Krężel, A., 2020. A satellite system for monitoring the radiation budget at the Baltic Sea surface. *Remote Sens. Environ.* 240, 111683.
- Zeng, S., Li, Y., Lyu, H., Xu, J., Dong, X., Wang, R., Yang, Z., Li, J., 2020. Mapping spatio-temporal dynamics of main water parameters and understanding their relationships with driving factors using GF-1 images in a clear reservoir. *Environ. Sci. Pollut. Res.* 27 (27), 33929–33950.
- Zhang, Y., Shi, K., Liu, J., Deng, J., Qin, B., Zhu, G., Zhou, Y., 2016a. Meteorological and hydrological conditions driving the formation and disappearance of black blooms, an ecological disaster phenomena of eutrophication and algal blooms. *Sci. Total Environ.* 569, 1517–1529.
- Zhang, Y., Shi, K., Liu, X., Zhou, Y., Qin, B., 2014. Lake topography and wind waves determining seasonal-spatial dynamics of total suspended matter in turbid Lake Taihu, China: assessment using long-term high-resolution meris data. *Plos One* 9 (5), e98055.
- Zhang, Y., Zhang, Y., Shi, K., Zha, Y., Zhou, Y., Liu, M., 2016b. A Landsat 8 OLI-based, semianalytical model for estimating the total suspended matter concentration in the slightly turbid Xin'anjiang Reservoir (China). *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 9 (1), 398–413.
- Zhang, Y.C., Hu, M.Q., Shi, K., Zhang, M., Han, T., Lai, L., Zhan, P.F., 2021. Sensitivity of phytoplankton to climatic factors in a large shallow lake revealed by column-integrated algal biomass from long-term satellite observations. *Water Res.* 207, 117786.
- Zhang, Y.L., Liu, X.H., Qin, B.Q., Shi, K., Deng, J.M., Zhou, Y.Q., 2016c. Aquatic vegetation in response to increased eutrophication and degraded light climate in Eastern Lake Taihu: Implications for lake ecological restoration. *Sci. Rep.* -Uk 6 (1), 1–12.
- Zhou, Y.Q., Jeppesen, E., Zhang, Y.L., Shi, K., Liu, X.H., Zhu, G.W., 2016. Dissolved organic matter fluorescence at wavelength 275/342 nm as a key indicator for detection of point-source contamination in a large Chinese drinking water lake. *Chemosphere* 144, 503–509.
- Zhu, M.Y., Paerl, H.W., Zhu, G.W., Wu, T.F., Li, W., Shi, K., Zhao, L.L., Zhang, Y.L., Qin, B.Q., Caruso, A.M., 2014. The role of tropical cyclones in stimulating cyanobacterial (*Microcystis* spp.) blooms in hypertrophic Lake Taihu, China. *Harmful Algae* 39, 310–321.