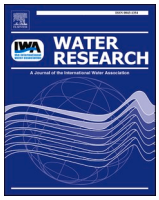
Water Research 267 (2024) 122546

Contents lists available at ScienceDirect

Water Research

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



Review

Application of remote sensing technology in water quality monitoring: From traditional approaches to artificial intelligence

Yuan Sun [a,1], Denghui Wang [a,1], Lei Li [a,b,*], Rongsheng Ning [a], Shuili Yu [a], Naiyun Gao [a]

[a] State Key Laboratory of Pollution Control and Resource Reuse, College of Environmental Science and Engineering, Tongji University, Shanghai, 200092, China
[b] Shanghai Institute of Pollution Control and Ecological Security, Shanghai 200092, China

ARTICLE INFO

Keywords:
Water quality monitoring
Remote sensing
Artificial intelligence
Inland water bodies
Machine learning

ABSTRACT

Quantitative estimation is a key and challenging issue in water quality monitoring. Remote sensing technology has increasingly demonstrated its potential to address these challenges. Remote sensing imagery, combined with retrieval algorithms such as empirical band ratio methods, analytical bio-optical models, and semi-empirical three-band models, enables efficient, large-scale, real-time acquisition of water quality distribution characteristics, overcoming the limitations of traditional monitoring methods. Furthermore, artificial intelligence (AI), with its powerful autonomous learning capabilities and ability to solve complex problems, can deal with the nonlinear relationships between different spectral bands' apparent optical properties and various water quality parameter concentrations. This review provides a comprehensive overview of remote sensing applications in retrieving concentrations of nine water quality parameters, ranging from traditional methods to AI-based approaches. These parameters include chlorophyll-a (Chl-a), phycocyanin (PC), total suspended matter (TSM), colored dissolved organic matter (CDOM) and five non-optically active constituents (NOACs). Finally, it discusses five major issues that need further research in the application of remote sensing technology and AI in water quality monitoring. This review aims to provide researchers and relevant management departments with a potential roadmap and information support for innovative exploration in automated and intelligent water quality remote sensing monitoring.

1. Introduction

Inland water bodies, including rivers, lakes, and reservoirs, play vital roles such as flood control, irrigation, and climate regulation, and they are important components of terrestrial ecosystems (Tian et al., 2023). These water bodies are sensitive to human activities and climate change. Under the context of global warming, external pollution due to excessive use of agricultural pesticides and fertilizers, and improper discharge of industrial wastewater, has led to severe pollution in inland water bodies. This includes organic and inorganic pollution and frequent occurrences of eutrophication and blue-green algae blooms (Ho et al., 2019). Water quality parameters are key indicators of the ecological environment of water bodies, reflecting the quality level and changing trends of the water environment. The decline in the water quality of inland water bodies threatens normal human production, life, and sustainable development (Fang et al., 2022). Therefore, monitoring the spatiotemporal changes in water quality has gradually attracted attention, making water quality monitoring an essential part of water environment management. It is crucial for the timely detection, tracing, and containment of water pollution (Niu et al., 2021).

Traditional water quality monitoring involves: a) on-site sample collection with lab analysis, which is accurate but time-consuming and costly; b) rapid on-site testing, which is efficient but measures fewer parameters with lower precision; and c) automatic monitoring stations, which provide continuous data but are expensive and have limited coverage. These methods make comprehensive spatiotemporal water quality monitoring challenging for regulatory agencies (Chawla et al., 2020).

Remote sensing provides cost-effective, large-scale water quality monitoring with high temporal coverage and reasonable accuracy. It captures spatial and dynamic changes, detects pollutants, and supports real-time monitoring. Although many satellite sensors, such as earth resources satellites like Landsat, SPOT, SeaWiFS, Terra and Aqua/MODIS, and ENVISAT/MERIS, and geostationary satellites as GOCI, are

* Corresponding author.
E-mail address: lilei@tongji.edu.cn (L. Li).
[1] Equal contributions.

https://doi.org/10.1016/j.watres.2024.122546
Received 19 July 2024; Received in revised form 13 September 2024; Accepted 28 September 2024
Available online 29 September 2024

designed for ocean or land observation, many studies have already applied them to inland water bodies. However, remote sensing of inland water bodies faces challenges such as optical complexity, satellite sensor limitations, and atmospheric correction accuracy. Thus, promoting the understanding and application of remote sensing for inland water quality monitoring is essential.

Previous research on water quality monitoring have focused on perfecting the retrieval of remote sensing reflectance to more accurately estimate OACs (Ouma et al., 2020), including chlorophyll-a (Chl-a) (Li et al., 2021), total suspended matter (TSM) (Fu et al., 2022), and colored dissolved organic matter (CDOM) (Xu et al., 2018). Efforts, such as improving the quality of atmospheric correction algorithms, cloud detection and glint removal algorithms (Duan et al., 2020; Sun et al., 2017), establishing the open data policy for accessing satellite images and providing customizable algorithms based on users' application needs, have greatly promoted the application of remote sensing technology in the field of water quality monitoring. Besides these, key water quality parameters related to eutrophication, such as total nitrogen (TN), total phosphorus (TP), dissolved oxygen (DO), chemical oxygen demand (COD), and biochemical oxygen demand (BOD), are not optically active. Despite this, numerous studies have explored the correlation between OACs and NOACs to estimate NOACs indirectly through model construction (Cai et al., 2023). However, these models are limited by the weak correlation between OACs and NOACs and the variability of this relationship across spatial and temporal scales, which hampers the generalizability of the algorithms (Cao et al., 2021). Traditional retrieval methods for OACs and NOACs—empirical, semi-empirical/semi-nalytical, and analytical—struggle with capturing nonlinear relationships, facing challenges in accuracy, generalizability, and complexity.

In recent years, advancements in computer science have made artificial intelligence (AI) a research hotspot, revitalizing fields such as natural language processing, facial recognition, medical imaging, pedestrian re-identification, and autonomous driving (Zhu et al., 2017). We are now in an era characterized by massive remote sensing data from satellites, drones, and other devices, with datasets that have varying imaging methods and spatiotemporal resolutions. AI has unique advantages in solving complex nonlinear problems due to its ability for continuous learning and model correction without a fixed framework. Beyond constructing water quality parameter retrieval models, AI also offers new possibilities for multi-source remote sensing data fusion, which is crucial for improving temporal, spatial, and spectral resolution, thereby providing higher-quality water quality monitoring data. The application of AI and remote sensing technology enables efficient, large-scale, and continuous water quality monitoring, providing strong support for eutrophication assessment (Li et al., 2023; Zhang et al., 2024).

Currently, few review articles address the synergistic application of remote sensing technology and AI in water quality monitoring. Most previous works have focused on remote sensing data and retrieval methods, briefly summarizing the application of AI-based methods in constructing water quality parameter retrieval models (Wasehun et al., 2024), or assessing only a few water quality parameters (Cao et al., 2022). This article comprehensively analyzes the retrieval of nine water quality parameters, including both OACs and NOACs. It reviews traditional retrieval models, including empirical, analytical, semi-empirical, and semi-analytical approaches, with a particular focus on recent advancements in Al-based retrieval techniques. The limitations and adaptability of AI are discussed as well. Additionally, it explores the challenges, future trends, and potential of the application of AI and remote sensing technology in inland water quality monitoring.

2. Data sources from sensors

Based on the type of platform on which they are situated, observing sensors can be divided into two major categories: airborne sensors and satellite sensors. The performance of sensors is fundamental to the application of remote sensing technology in water quality monitoring (refer to SI Section 1 for a detailed description of the data sources from sensors).

3. Preprocessing of remote sensing data

Remote sensing images are often limited by spatial resolution, radiometric resolution, and spectrum. Therefore, preprocessing is crucial. This includes radiometric calibration, atmospheric correction, water body extraction, and statistical analysis to determine the distribution characteristics of water quality parameters. Intelligent identification and extraction of water bodies is essential for long-term analytical studies that require accurate extraction of water bodies. Integrating multi-source remote sensing data fusion with AI enhances resolution in temporal, spatial, and spectral dimensions, providing robust data support for inland water quality monitoring.

3.1. Water body extraction

Water body extraction is based on the principle that water is significantly less reflective in the infrared channel than other land types. Intelligent iterations of water body extraction methods provide a more convenient means of studying various water bodies over long time series (refer to SI Section 2 for a detailed description of the methods).

3.2. Fusion of multi-source remote sensing data

Multi-source remote sensing data includes data recorded by different sensors, as well as data obtained by the same sensor across different spectral bands. The intelligent iteration of multi-source remote sensing data fusion techniques offers a viable research direction to mitigate the limitations of a single data source, overcome constraints in temporal, spatial, and spectral resolution, and improve the accuracy of water quality predictions (refer to SI Section 3 for a detailed description of the techniques).

4. Retrieval of water quality parameters

The principle of water quality monitoring via remote sensing links spectral information from water components to quality parameters, divided into OACs like Chl-a, TSM, and CDOM, and NOACs like DO, COD, BOD, and TN. OACs interact with light through absorption, refraction, and scattering, whereas NOACs lack significant optical properties and cannot be directly measured via spectral methods. Common retrieval methods include analytical/semi-analytical, empirical/semi-empirical, and AI-based methods.

The analytical/semi-analytical method, based on bio-optical models and water radiative transfer theory, is highly interpretable but requires optical parameters for different regions and seasons, which limits its practical use. The empirical method is simple, establishing statistical relationships between spectral data and water quality, but it lacks physical interpretability and generalizability across regions. The semi-empirical method combines analytical and empirical approaches, improving generalizability and interpretability, but still requires synchronous data for calibration and is limited to the retrieval of OACs. The AI-based method, handles complex, nonlinear data effectively, providing robust predictions, though it requires extensive data and face challenges in interpretability.

Most of current research focuses on OACs, but significant progress has been made in retrieving NOACs. Although the indirect retrieval of NOACs has limitations, studies on the estimation of NOACs still hold certain application scenarios and value. This article precisely investigates common methods and models for retrieving the concentrations of nine water quality parameters. Sections 4.1-4.5 respectively cover Chl-a, phycocyanin (PC), TSM, CDOM, and NOACs (including DO,

COD, BOD, TN, and TP).

4.1. Chlorophyll–a (Chl-a)

Although other indicators are also used to monitor algal blooms (Binding et al., 2021; Legleiter et al., 2022; Liu et al., 2022; Richard and Michelle, 2005), chlorophyll–a remains the most primary indicator for this purpose. Additionally, the concentration of Chl-a in water is an important indicator for assessing the degree of eutrophication. Effective monitoring of chlorophyll-a helps prevent further deterioration of water quality (Andrade et al., 2019).

Chlorophyll-a has absorption peaks in the 430–700 nm. Algal substances in the water show absorption peaks in the blue-violet band (around 440 nm) and the red band (around 675 nm). Thus, water bodies with high chlorophyll-a concentrations will exhibit troughs in their reflectance curves at these wavelengths due to the strong absorption of chlorophyll-a. Meanwhile, high Chl-a concentrations result in a significant reflectance peak around 700 nm in the near-infrared. An increase in the concentration of Chl-a leads to a decrease in the spectral response at short wavelengths, particularly in the blue band (George, 1997). Analyzing these spectral characteristics allows for accurate remote sensing retrieval of Chl-a concentrations in water bodies.

Traditional retrieval methods include empirical, semi-empirical/semi-analytical, and analytical methods. Table 1 provides a detailed summary of the retrieval methods for Chl-a in the water, including the application and accuracy of their models.

The analytical method originates from Gordon et al.'s bio-optical algorithm in 1975 (Gordon et al., 1975). Subsequent studies have made several improvements to this, such as Lee et al.'s Quasi-Analytical Algorithm (QAA) for type II water bodies (Lee et al., 2002) and the application of bio-optical models in inland waters. Giardino et al. (2007) obtained Chl-a concentration data in Lake Garda within 3 h of Hyperion satellite imaging using in-situ water samples based on the analytical method ISO 10260-E (1992). Based on optical parameters provided by Matthews (2011), they established a bio-optical model. They compared the subsurface irradiance reflectance R(0-, λ) measured using a PR-650 spectroradiometer with those simulated by the bio-optical model to verify the model's accuracy. With a sensitivity analysis applied to the bio-optical model, they selected Hyperion bands, specifically binning the 480–500 nm and 550–560 nm bands, to retrieve Chl-a concentration. The results were in good agreement with in situ point data concentrations measured in 8 pelagic stations, showing a correlation coefficient (r) of 0.77, a RMSE of 0.36 mg/m^3, and a bias of 0.12 mg/m^3 (relative bias 6%). This demonstrates that remote sensing can support relevant management applications for Lake Garda and other large subalpine lakes. While effective, these models require extensive inherent optical properties (IOPs) data, complicating their implementation and limiting their validation in complex waters.

The main models of the empirical method include the widely used two-band ratio model, which could reduce the impact of suspended matter, yellow substances, and atmospheric reflection. For instance, the ratio of near infrared (NIR) and red reflectance shows a strong correlation with Chl-a concentration (Gurlin et al., 2011), with common wavelengths around 700 nm and 670 nm (e.g., R_{708}/R_{665}(Gilerson et al., 2010)). Additionally, the normalized difference chlorophyll index (NDCI) model (Mishra and Mishra, 2012) and the single-band model (George, 1997) are used to retrieve the chlorophyll-a concentration in the water column. The single-band model typically selects wavelengths centered on the reflectance peaks or absorption valleys in the reflectance eigen-spectrum of Chl-a.

The empirical method has also undergone relevant iterations in the past decade (Allan et al., 2015). Beck et al. (2016) conducted a study in a temperate reservoir in southwestern Ohio using coincident VNIR CASI hyperspectral aircraft imagery and intensive coincident surface observations (44 densely overlapping surface observations acquired at 400-meter grid spacing within 1 h of image acquisition). They compared the performance of 10 existing models and 2 new models for retrieving Chl-a. The results showed that NDCI was the most widely applicable algorithm, performing well across all synthetic imaging systems, with 2BDA following. The performance of NDCI was best on CASI and the synthesized WorldView-2/-3, Sentinel-2, and MERIS satellites, with R^2 values ranging from 0.687 to 0.845. This confirms the strong potential of NDCI for routine Chl-a estimation in smaller inland water bodies using operational and future satellite systems.

The empirical method's retrieval models are simple but heavily reliant on actual sampling data, leading to regional limitations and a lack of physical basis. To address these issues, semi-empirical/semi-analytical models have been developed. Key models include the three-band model, the four-band model, APPLE model (El-Alem et al., 2012), and synthetic chlorophyll index (SCI) model (Shen et al., 2010). Gitelson et al. (2008) first developed the three-band model for Chl-a, later improved by Moses et al. (2009) and Gilerson et al. (2010) by adding reflectance at 753 nm to account for suspended particles' backscattering. Le et al. (2009) further improved it for turbid waters by adding a fourth band in the NIR (730–780 nm) to minimize interference from suspended particles and pure water absorption. The four-band model outperformed the three-band model in Taihu Lake. Semi-empirical/semi-analytical models aim to eliminate the influence of suspended matter, yellow substances, backscattering, and pure water on the Chl-a reflectance spectrum, thus maximizing the extraction of Chl-a concentration. Both empirical and semi-empirical/semi-analytical methods, however, still rely on field data, lacking strong physical significance.

The above models have limitations in retrieval accuracy, general applicability, fault tolerance, and computational complexity. With the advancement of technology, AI has become widely used in scientific fields. ML (machine learning), a branch of AI, avoids atmospheric correction errors common in empirical and semi-analytical methods and provides generalized models for retrieving Chl-a. ML algorithms such as neural networks (Chen et al., 2022), support vector machine (SVM) (Li et al., 2021), extreme gradient boosting (XGBoost) (Cao et al., 2020), and random forest (RF) (Fang et al., 2024) have been introduced for water quality parameter retrieval. These algorithms address nonlinear optimization problems in complex water bodies, enhancing inversion accuracy and generality (Park et al., 2015). For instance, Li et al. (2021) collected 273 samples from 45 typical lakes in China between 2017 and 2019, and applied an SVM model for Chl-a retrieval, achieving better results than linear regression (LR) and Catboost models. This suggests that SVM is effective for large-scale monitoring, especially at medium/low Chl-a levels. Pahlevan et al. (2020) proposed a mixed density network (MDN) model for Chl-a inversion with Sentinel-2 MSI and Sentinel-3 OLCI observations. The MDN, which learns covariances between target variables, improves performance by simultaneously retrieving Chl-a, total suspended solids (TSS), IOPs, and other parameters.

Current research on ML in ecological studies often relies on traditional algorithms, with limited use of advanced ones. Different water bodies have unique optical characteristics, making a single algorithm insufficient for accurate retrieval. To address this, an algorithm system based on water body classification has been developed. Neil et al. (2019) collected data from 185 water bodies and developed a dynamic ensemble algorithm based on the water body classification standards proposed by Spyrakos et al. (2018). This system considers spatiotemporal differences and optimizes parameters for each water body type, improving retrieval accuracy by 25%, with a correlation coefficient of 0.89 and an average absolute error of 0.18 mg/m^3. To address multicollinearity among feature bands, Zhang et al. (2022) proposed a feature bands selection strategy (FD-FI) based on knee-point detection and variance inflation factor (VIF) for retrieving Chl-a concentration. They also introduced stacking model fusion technology, combining nine ML algorithms to construct a MixModel, which showed superior generalization, stability, and sensitivity to extreme values of Chl-a compared to

Table 1
Retrieval methods, models and the accuracy for Chl-a.

Retrieval Methods	Retrieval Models	Equation/ Algorithm	Study Area	Chl-a Concentration (mg/m^3)	Sensor	R^2 (Number of Samples)	RMSE (mg/m^3)	Reference
The empirical method	Single-band model	$Chl-a = -41.63 * R_{560} + 6.91$	Six lakes and ten tarns in the English Lake District	1–70	Daedalus ATM	$r = 0.97$ ($N = 9$)	/	(George, 1997)
	Two-band ratio model	$Chl-a = 61.324 * [R_{665}^{-1} * R_{708}] - 37.94$	The Taganrog Bay and the Azov Sea, Russia	0.63–65.51	MERIS	0.97 ($N = 8$)	3.65	(Moses et al., 2009)
		$Chl-a = [35.75 * (R_{708}/R_{665}) - 19.30]^{1.124}$	Fremont State Lakes, USA	2–100	MERIS	0.96 ($N = 85$)	/	(Gilerson et al., 2010)
	Normalized difference chlorophyll index (NDCI) model	$Chl-a = 14.039 + 86.115 * \frac{[R_{708} - R_{665}]}{[R_{708} + R_{665}]} + 194.325 * \left(\frac{[R_{708} - R_{665}]}{[R_{708} + R_{665}]}\right)^2$	Chesapeake Bay, Delaware Bay, the river Mississippi Delta region, and the Mobile Bay, USA	1–60	MERIS	0.80 ($N = 20$)	1.89	(Mishra and Mishra, 2012)
The semi-analytical/ semi-empirical method	Three-band model (3BDM)	$Chl-a = 232.29 * (R_{665}^{-1} - R_{708}^{-1}) * R_{753} + 23.174$	The Taganrog Bay and the Azov Sea, Russia	0.63–65.51	MERIS	0.95 ($N = 8$)	5.02	(Moses et al., 2009)
		$Chl-a = \{113.36 * [(R_{665}^{-1} - R_{708}^{-1}) * R_{753}] + 16.45\}^{1.124}$	Fremont State Lakes, USA	2–100	MERIS	0.96 ($N = 85$)	/	(Gilerson et al., 2010)
		$Chl-a = 331.01 * [R_{684}^{-1} - R_{690}^{-1}] * R_{718} + 14.609$	The Pearl River Estuary, China	4.8–92.6	EO-1 Hyperion	0.95 ($N = 16$)	6.44	(Chen et al., 2011)
	Four-band model (4BDM)	$Chl-a = 0.0097[R_{662}^{-1} - R_{693}^{1}] * [R_{740}^{-1} - R_{705}^{-1}]^{-1} - 0.1268$	Taihu Lake, China	4–158	Field measurements	0.89 ($N = 80$)	9.74	(Le et al., 2009)
	APPLE model	$Chl-a = 8.5739 * Exp(28.176 * SI) SI = B_2 * [(B_1 * B_2) + (B_3 * B_2) * B_2]$	Missisquoi Bay of Lake Champlain, Lake Brome, Lake William and Lake Nairne, Canada	2.5–91,000	MERIS	0.93 ($N = 51$)	RMSEr=69%	(El-Alem et al., 2012)
	Synthetic chlorophyll index (SCI) model	$Chl-a_{spring} = 179378 * SCI^2 + 92.934 * SCI + 0.2736 Chl-a_{summer} = 550383 * SCI^2 + 2769 * SCI + 4.3866 SCI = H_{chl} - H_{\Delta} H_{chl} = (0.74 R_{681} + 0.26 R_{620}) - R_{665} H_{\Delta} = R_{620} - 0.5(R_{560} + R_{681})$	The Changjiang Estuary, China	Spr: 0.03–3.10 Sum: 0.88–31.50	MERIS	Spr: 0.72 ($N = 19$) Sum: 0.91 ($N = 21$)	Spr: 0.86 Sum: 2.87	(Shen et al., 2010)
The analytical method	Bio-optical model	Bio-optical algorithms	Lake Garda, Italy	1.30–2.16	Hyperion	0.59 ($N = 8$)	0.36	(Giardino et al., 2007)
	QAA-based model	Quasi-Analytical Algorithm	Waters around Baja California, USA	0.03–30	/	/	/	(Lee et al., 2002)
The AI-based method	Traditional machine learning models	Back propagation (BP) neural network	Middle and lower reaches of the Hanjiang River, China	2.774–41.837	MODIS	0.99 ($N = 5$)	3.1462	(Chen et al., 2022)
		Support vector machine (SVM)	45 typical lakes across China	0–120.99	Sentinel-2 MSI	0.88 ($N = 91$)	6.28	(Li et al., 2021)
		Extreme gradient boosting tree (XGBoost)	Lakes in the middle and lower reach of Yangtze River and a reach of Huai River, China	0.4–258.7	Landsat-8 OLI	0.79 ($N = 102$)	7.1	(Cao et al., 2020)
		Random forest (RF)	40 lakes across the Northeast China	0.45–781.42	MODIS	(a): 0.83 ($N = 177$) (b): 0.77 ($N = 665$) (c): 0.82 ($N = 73$)	(a): 2.19 (b): 4.57 (c): 6.31	(Fang et al., 2024)
	More advanced AI-based models	FD-FI-MixModel	Nansi Lake, China	10–70	Zhuhai-1 CMOS	0.8664 ($N = 99$)	5.7926	(Zhang et al., 2022)
		Dynamic ensemble algorithm	185 global inland and coastal aquatic systems	0–1000	MERIS	0.79 ($N = 2807$)	0.25	(Neil et al., 2019)

Note:
(1) In the equations, R_{665} refers to the remote sensing reflectance at a wavelength of 665 nanometers (nm). Similar expressions can be applied analogously.
(2) B_i represents the remote sensing reflectance of the ith band of the corresponding sensor.

other models.

4.2. Phycocyanin (PC)

For a long time, research on cyanobacterial blooms in inland waters has primarily used Chl-a concentration as an indicator of cyanobacterial biomass. However, Chl-a is present in all eukaryotic algae and aquatic vegetation. Phycocyanin (PC), the characteristic pigment of cyanobacteria, is unique to cyanobacteria and can accurately indicate their biomass. PC absorbs and transfers light energy, appears bright blue, and is an intracellular protein (Simis et al., 2012). Utilizing remote sensing to monitor PC concentration offers an effective, large-scale, rapid, and accurate strategy for monitoring cyanobacterial blooms.

Phycocyanin (PC) has a distinct absorption peak near 620 nm, providing a basis for remote sensing monitoring. However, several challenges limit the accuracy and application of PC retrieval. First, PC's absorption signal is weak, only 20% of that of Chl-a (Schalles and Yacobi, 2000). Higher PC concentrations yield stronger remote sensing signals, but in low-PC waters, especially oligotrophic ones, reliable estimation models are hard to establish (Liu et al., 2018). Additionally, PC signals are easily interfered with by other substances like Chl-a, CDOM and TSM in turbid waters (Liu et al., 2018). Researchers have developed empirical and semi-analytical methods to retrieve PC concentration based on absorption features between 615 and 630 nm (Ogashawara et al., 2013). Table 2 lists the classic remote sensing retrieval models for PC concentration that have been developed so far.

In the empirical method, two-band ratio models are commonly used to eliminate interference factors like atmospheric effects. Commonly used ratios are, for example, R_{710}/R_{620}(Hunter et al., 2008), R_{709}/R_{620}(Kwon et al., 2020). Based on Sentinel-2 MSI imagery data, Sòria-Perpinyà et al. (Sòria-Perpinyà et al., 2020) collected 76 georeferenced PC concentration (measured by 3D spectrofluorimetry) samples at 0.2 m depth from seven different points in the Albufera of Valencia (Eastern Iberian Peninsula) spanning two years. They selected the 21 samples for calibration and the remaining 55 for method validation. They successfully developed a two-band ratio retrieval model with the ratioR_{740}/R_{665}, achieving a validation accuracy with $R^2 = 0.775$, and RMSE% = 40%. This study provides a concise and efficient PC retrieval algorithm for Sentinel-2, but the algorithm still requires calibration before being applied to other lakes. The baseline model proposed by Dekker (1993) is another frequently used method. Castagna et al. (2020) developed the Orange Contra-band Algorithm model based on Dekker's baseline model to eliminate the effects of particle scattering. Recent studies by Sun et al. (2015) and Jin et al. (2017) constructed multiple band (or band ratio) linear regression models for waters, achieving high accuracy.

The empirical method does not consider the optical radiation principles of PC, thus having certain spatial limitations. So researchers have proposed the semi-analytical method. These methods include three-band models, four-band models, nested band ratio models, and derivative algorithm model (spectral shape algorithm-based model). Models of the semi-analytical method mainly aim to solve the issues of atmospheric correction, decomposition of PC absorption at 620 nm, and insufficient spectral resolution.

The nested band ratio model, developed by Simis et al. (2005), isolates PC absorption by removing Chl-a contributions, achieving high accuracy within specific study areas but larger errors elsewhere. Variants of this model have been widely used for PC retrieval in various regions (Chawira et al., 2013). The derivative model, or spectral shape algorithm (SS(λ)), was proposed by Wynne et al. (2008). Qi et al. (2014) modified the spectral shape algorithm, referred to as the PCI index, achieving good accuracy in Lake Taihu. Initially designed for Chl-a retrieval, the three-band and four-band models were adapted for PC by modifying the absorption bands. Hunter et al. (2010) introduced the three-band model for PC retrieval, achieving high accuracy. Le et al. (2011) adapted the four-band model for PC in Taihu Lake, achieving high accuracy (R^2=0.86, RMSE=6.8 $\mu g \cdot L^{-1}$). Liu et al. (2018) improved the four-band model as FBA_PC by separating the absorption spectra of yellow substances and other phytoplankton pigments from that of PC. Using PC concentration data from seven water bodies in China, the United States, and the Netherlands, along with in situ spectral data simulating the MERIS and Sentinel-3 OLCI bands, they calibrated and validated the model with 215 samples respectively. The model achieved satisfactory validation results with $R^2 = 0.730$ and RMSE= 27.691 $mg \cdot m^{-3}$. It performed better than the band ratio model, the three-band model, and the PCI index model, demonstrating robustness when applied to a wider range of water bodies.

The complex optical absorption and scattering characteristics of substances in water bodies, which defy simple linear relationships, have driven the trend toward the AI-based method for retrieving PC. ML models that have been widely used, including genetic algorithm (GA) (Song et al., 2012a), SVM (Sun et al., 2013), artificial neural network (ANN) (Park et al., 2017), RF (Beal et al., 2024). Further optimization of these shallow ML models can address problems such as overfitting, high dimensional data processing and local minima. Liu et al. (2023) addressed the problem of local minima in the BP neural network (BPNN) by optimizing BPNN using a particle swarm optimization algorithm (PSO). They demonstrated that PSO-BPNN outperformed traditional BPNN and support vector machine regression (SVR) in monitoring low-concentration PC.

Deep learning, a powerful form of ML, excels with complex, nonlinear, and redundant datasets. Pyo et al. (2019) constructed the Point-centered regression convolutional neural network (PRCNN) model based on hyperspectral images to accurately retrieve PC concentration, demonstrating that convolutional neural network (CNN) regression has the potential for high-precision detection and quantification of cyanobacteria. Yim et al. (2020) enhanced deep neural network (DNN) feature learning with stacked autoencoders (SAE-DNN), outperforming DNNs and band ratio models. Pyo et al. (2020) further showcased deep learning's potential with SAE for advanced feature extraction. However, these models rely heavily on mathematical analysis and lack mechanistic studies related to the IOPs and apparent optical properties (AOPs) of water bodies.

4.3. Total suspended matter (TSM)

Suspended matter in inland water bodies includes inorganic substances (insoluble sediments, clay, minerals) and organic substances (phytoplankton, plankton, plant and animal remains) (Schartau et al., 2019). Changes in the concentration of suspended matter affect underwater light distribution and aquatic plant light absorption (Dekker et al., 2002). Thus, timely and accurate monitoring of suspended matter dynamics is crucial for effective water management and protection.

Changes in suspended matter concentration alter IOPs, which is the theoretical basis for constructing TSM retrieval models. Reflectance values at various wavelengths rise with increasing suspended matter concentration. There is a reflection peak near 750 nm and an absorption peak near 950 nm. However, as the suspended matter concentration continues to increase, the peak reflectance values in the red and green bands become saturated, the reflection peak exhibits a red shift (i.e., shifts to longer wavelengths), and the absorption peak shows a blue shift (i.e., shifts to shorter wavelengths) (Dörnhöfer et al., 2018). Remote sensing of TSM primarily uses reflectance changes in the red and NIR bands. Traditional methods for retrieving TSM in inland water bodies primarily include two types: the empirical/semi-empirical method and the analytical/semi-analytical method. Table 3 provides a detailed summary of the retrieval methods for TSM in the water column and the application and accuracy of their models.

Based on the analytical method, Dekker et al. (2002) used a bio-optical model to retrieve TSM in the Frisian Lakes of the Netherlands. The analytical method, requiring many parameters and default conditions, become semi-analytical when some empirical

Table 2
Retrieval methods, models and the accuracy for PC.

Retrieval Methods	Retrieval Models	Equation/ Algorithm	Study Area	PC Concentration (μg/L)	Sensor	R^2 (Number of Samples)	RMSE (μg/L)	Reference
The empirical method	Two-band ratio model	$PC = 5.8281 * (R_{709}/R_{620}) - 2.7126$	The Daechung Reservoir, South Korea	0.39–803.01	Drone-based Hyperspectral	0.94 ($N = 92$)	/	(Kwon et al., 2020)
		$PC = \exp[2.6151 * (R_{740}/R_{665}) - 3.6369]$	Albufera of Valencia, Spain	10–1287.96	Sentinel-2 MSI	0.775 ($N = 55$)	40% (RMSE%)	(Sòria-Perpinyà et al., 2020)
	Baseline model	$PC = -24.6 + 13686(0.5 * (R(0^-)_{600} + R(0^-)_{648}) - R(0^-)_{624})$	Vecht lakes area, Netherlands	46–130	CASI	0.99 ($N = 10$)	2.34	(Dekker, 1993)
	Orange Contra-band Algorithm model	$PC \propto R_{rs}^{orange}$ $R_{rs}^{orange} = 2.2861(\pm 0.1303)R_{rs}^{pan} - 0.9467(\pm 0.0611)R_{rs}^{green} - 0.1989(\pm 0.0712)R_{rs}^{red}$	Two lakes in the Netherlands and seven lakes in Belgium	0.01–329.41	Landsat-8 OLI	/ ($N = 428$)	5.39% (RMSE%)	(Castagna et al., 2020)
	Multi-band model	$Log_{10}(PC) = K_0 + K_1 * B_1 + K_2 * B_2 + K_3 * B_3 + K_4 * B_4 + K_5(B_4/B_3) + K_6(B_4/B_2) + K_7(B_4/B_1) + K_8(B_3/B_2) + K_9(B_3/B_1) + K_{10}(B_2/B_1)$	Lake Dianchi, China	77.6–754.9	Landsat 4 TM Landsat 5 TM Landsat 7 ETM+ Landsat 8 OLI	>0.97 ($N = 14$)	RMSE<10%	(Sun et al., 2015)
		$PC = (89.1711 * B_5 - 262.292 * B_6 + 221.193 * B_7 - 1.7065 * B_8 - 26.1116 * B_9 + 0.363418) * 100\%$	Lake Taihu and Lake Chaohu, China	3.26–804.11	MERIS	0.72 ($N = 41$)	7.56	(Jin et al., 2017)
The semi-analytical method	Nested band ratio model	$PC = a(620)_{pc}/a_{pc}^*(620)$ $a_{pc}(620) = \{[R_{709}/R_{620}] * [a_w(709) + b_b]\} - b_b - a_w(620))\delta^{-1} - [\varepsilon * a_{chl}(665)]$	Missisquoi Bay, USA	4.1–105	Quick Bird MERIS	0.68 ($N = 16$)	/	(Wheeler et al., 2012)
	Three-band model	$PC = 1.58 + 0.984\left[R_{615}^{-1} - R_{600}^{-1}\right] * R_{725}$	Loch Leven and Esthwaite Water, UK	5.74–93.7	CASI-2 AISA	0.98 ($N = 15$)	3.98	(Hunter et al., 2010)
	Four-band model	$PC = 462.5 * \text{FBA\_PC} + 22.598\text{FBA\_PC} = \left[\frac{1}{R_{620}} - \frac{0.4}{R_{560}} - \frac{0.6}{R_{709}}\right] R_{754}$	12 inland waters in the United States, the Netherlands, and China	0.327–317.74	Field measurements	0.73 ($N = 215$)	27.691	(Liu et al., 2018)
	Spectral shape algorithm-based model	$PC = 3.87\exp[1154 * PCI(R_{rs})] PCI = R'_{620} - R_{620}R'_{620} = R_{560} + \frac{620 - 560}{665 - 560} * (R_{665} - R_{560})$	Lake Taihu, China	1–300	MERIS	0.79 ($N = 37$)	58% (RMSE%)	(Qi et al., 2014)
The AI-based method	Traditional machine learning models	artificial neural network(ANN)	Baekje Reservoir, South Korea	0.2–147	AisaFENIX	NSE=0.8 ($N = 39$)	/	(Park et al., 2017)
		Random forest (RF)	Lake Mendota, USA	0–5	Sentinel-2 MSI	0.69 ($N = 41$)	MAE=0.21	(Beal et al., 2024)
	Coupling machine learning models	PSO-BPNN	QinZhou Bay, China	0.799–4.855	Sentinel-2 MSI	0.70 ($N = 22$)	0.615	(Liu et al., 2023)
	Deep learning models	DNN	Waterbodies across the States of Illinois and Missouri, USA	0.1–9.34	Landsat 8 OLI Sentinel-2 MSI	0.82 ($N = 35$)	0.43	(Sagan et al., 2020)
		Point-centered regression CNN (PRCNN)	Baekje weir, South Korea	0.19—150.9	AISA	0.86 ($N = 36$)	9.39	(Pyo et al., 2019)
		SAE-DNN	Baekje Reservoir, South Korea	0.02—280.87	AISA	0.87 ($N = 61$)	14.45	(Yim et al., 2020)
		SAE-ANN	Baekje Reservoir, South Korea	0.19—146.9	AISA	0.83 ($N = 86$)	9.75	(Pyo et al., 2020)

Note: (1) In the equations,R_{665} refers to the remote sensing reflectance at a wavelength of 665 nanometers (nm). Similar expressions can be applied analogously. (2) B_i represents the remote sensing reflectance of the ith band of the corresponding sensor. (3) $R(0^-)_{600}$ represents the subsurface irradiance reflectance at 600 nm. Similar expressions can be applied analogously.

Table 3
Retrieval methods, models and the accuracy for TSM.

Retrieval Methods	Retrieval Models	Equation/ Algorithm	Study Area	TSM Concentration (mg/L)	Sensor	R^2 (Number of Samples)	RMSE (mg/L)	Reference
The empirical/semi-empirical method	Single-band model	$TSM = 1.6543 * \exp(0.0034 * Red)$	423 lakes across China	0.04–2426.7	Landsat TM/ ETM+/OLI	0.76 ($N = 212$)	21.4	(Wen et al., 2022)
		$TSM = 9.65 * \exp(58.81 * R_{645})$	Lake Taihu, China	13.9–301.3	MODIS	0.8 ($N = 150$)	14.0	(Shi et al., 2015)
		$TSM = 1.8035 * \exp(0.0039 * Red)$	Lakes across Inner Mongolia, China	1.2–860	Landsat TM/OLI	0.82 ($N = 33$)	8.23	(Du et al., 2021)
		$TSM_{OLI} = 6110.3 * NIR - 1.8242$ $TSM_{ETM^+/TM} = 4616.4 * NIR - 4.362$ $TSM_{MSS} = 4508.2 * NIR - 4.4551$	Dongting Lake, China	4–101	Landsat MSS; TM/ETM+; OLI	OLI:0.81 ETM+/ TM:0.82 MSS:0.82 ($N = 20$)	OLI:5.79 ETM+/ TM:7.01 MSS:12.21	(Zheng et al., 2015)
	Multi-band model	$TSM = \exp(15.4 * [R_{645} - R_{1240}]) + 1.99$	Lake Hongze, China	10–80	MODIS	0.55 ($N = 41$)	7.64	(Cao et al., 2017)
		$TSM = 7234.4\left(\frac{Green * Red}{Blue}\right)^2 - 146.46\frac{Green * Red}{Blue} + 8.0425$	22 water bodies in Songnen Plain, China	8.33–136	Landsat TM/ ETM+/OLI	0.66 ($N = 38$)	16.62	(Du et al., 2020)
		$TSM = -44.97 * R_{475} + 4.29 * R_{560} + 46.10 * R_{660} + 0.33$	Xin'anjing Reservoir, China	0.67–5.11	HJ-1A/B	0.85 ($N = 35$)	0.53	(Zhang et al., 2019)
The analytical/ semi-analytical method	Analytical model	Bio-optical algorithms	The Frisian lakes, Netherlands	3–245	Landsat TM	0.99 ($N = 22$)	/	(Dekker et al., 2002)
	Semi-analytical model	A two-step, IOP-based model	Hangzhou Bay and Lake Taihu, China	2.4–695.2	GOCI	0.83 ($N = 112$)	43.1	(Zhang et al., 2018)
The AI-based method	Traditional machine learning models	Multilayer back propagation neural network (MBPNN)	The Bohai Sea, Yellow Sea, and East China Sea, China	0.6–350	MODIS	0.91 ($N = 77$)	MRE=31.88%	(Chen et al., 2015)
		Random forest (RF)	423 lakes across China	0.04–2426.7	Landsat TM/ ETM+/OLI	0.81 ($N = 212$)	16.39	(Wen et al., 2022)
		Extreme gradient boosting (XGBoost)	San Francisco Bay, USA	2–120	Sentinel-2 MSI	0.77 ($N = 70$)	6.1	(Niroumand-Jadidi and Bovolo, 2022)
		Extreme learning machine (ELM)	Five lakes in Mexico	10–790	Landsat-8 OLI	0.91 ($N = 14$)	60.86	(Arias-Rodriguez et al., 2021)
	Coupling machine learning models	PLS-PSO-BPNN	Lower reaches of the Haihe River, China	6–34	Gaofen-2	0.92 ($N = 6$)	3.05	(Guo et al., 2022)
		LOOCV-XGBoost	Poyang Lake, China	3.6–99.6	Landsat-8 OLI, Sentinel-2 A/B	0.96 ($N = 12$)	MRE=12.79%	(Fu et al., 2022)
		GA-RF	Nansi Lake in North China	4–54.8	EO-1 Hyperion	0.98 ($N = 6$)	1.715	(Liu et al., 2021)

Note: In the equations, R_{665} refers to the remote sensing reflectance at a wavelength of 665 nanometers (nm). Similar expressions can be applied analogously.

parameters are used as approximations. Zhang et al. (2018) derived absorption and backscattering coefficients of suspended matter from the reflectance signals based on GOCI image data. They developed a two-step, IOP-based model to retrieve TSM in Taihu Lake and Hangzhou Bay, enhancing retrieval accuracy and applicability. The mechanistic parameters of the analytical/semi-analytical method that is more applicable, accurate and theoretically supported, still need to be further explored and optimized.

Remote sensing retrieval of total suspended matter (TSM) is mainly based on the empirical/semi-empirical method, including single-band models and multi-band models. Single-band models, generally use the red or NIR band to retrieve TSM concentration. Shi et al. (2015) found the 645 nm band in MODIS-Aqua data correlated well with TSM concentration in Taihu Lake. Thereby, an exponential model was constructed to estimate the TSM in Taihu Lake from 2003 to 2013, analyzing its annual and seasonal dynamics. Multi-band models, using statistical regression, can improve predictions. Du et al. (2020) collected 142 water samples (sampling depth of 1 m) from 22 water bodies in the Songnen Plain (China) during 2012 to 2015, and measured TSM concentrations using weighing method (Song et al., 2012b). They conducted a Pearson correlation analysis between various band combinations from Landsat TM/ETM+/OLI images and field-measured TSM concentrations, finding that the combination of Green*Red/Blue best correlated with TSM. Based on this finding, they established a power function model using TSM data from 104 samples and corresponding remote sensing data, achieving a validation accuracy with $R^2 = 0.66$ and an RMSE of 16.62 mg·L^{-1} ($N = 38$). Using this empirical model, they then retrieved the spatiotemporal variations in the annual average suspended matter concentration in the lakes of the Songnen Plain from 1984 to 2018. This study confirms the effectiveness of a multiple regression fitting model, using a combination of three bands as independent variables, in TSM retrieval.

Models of the Empirical/semi-empirical method, based on satellite data and water optical properties, often suffer from regional and seasonal specificity due to the spatial and temporal variability of optically active substances. These models, usually simple linear functions, struggle to capture complex relationships between water spectral characteristics and water color elements. Recently, ML models such as neural network (Chen et al., 2015), RF (Wen et al., 2022), XGBoost (Niroumand-Jadidi and Bovolo, 2022), and extreme learning machines (ELM) (Arias-Rodriguez et al., 2021) have been used for TSM retrieval. These models can detect both linear and nonlinear interactions, improving the identification of complex relationships (Chen et al., 2021). Milad et al. (Niroumand-Jadidi and Bovolo, 2022) developed an XGBoost model based on Sentinel-2 MSI images, showing high temporal robustness compared to the optimal band ratio analysis (OBRA) model. Leonardo F et al. (Arias-Rodriguez et al., 2021) proposed an ELM model with Landsat-8 OLI images to retrieve TSM concentration in Lake Cuitzeo, Mexico's second-largest lake, outperforming SVR and LR.

Researchers have also focused on optimizing traditional ML models. Guo et al. (2022) improved the BPNN model by combining partial least squares (PLS) with PSO, selecting wavelength factors that influence water quality parameters as input data for the neural network model. This method screened and reduced the dimensionality of large-band combination raw data, optimizing the BPNN model and effectively avoiding overfitting, enhancing retrieval accuracy from 9.50 mg/L to 4.04 mg/L. Fu et al. (2022) used multi-sensor images to assess a total of OACs and NOACs of Poyang Lake in China. They developed a leave-one-out cross validation (LOOCV) combined with GradientBoost (LOOCV-GradientBoost) model. The residuals of the TSM retrieval results were only half of those obtained by ML (XGB, CatBoost, and GradientBoost) and ensemble machine learning (SEL) models, indicating that LOOCV- GradientBoost was the best model for estimating a single water quality parameter in this study. Using Hyperion hyperspectral remote sensing data in Nansi Lake in North China, Liu et al. (2021) developed a RF model based on genetic algorithm optimization (GA_RF). The two parameters of optimal design are the number of features and the number of decision trees were optimized by the genetic algorithm. Six training-testing datasets (20 sample points for training and 6 for testing) were random selected. The validation results showed that the GA-RF model significantly outperformed seven other models, including LR, BPNN, KNN, and RF, with $R^2 = 0.98$, an RMSE of 1.715 mg·L^{-1}. This study provides useful methods and recommendations for the quantitative retrieval of TSM concentration in large shallow lakes.

4.4. Colored dissolved organic matter (CDOM)

Colored Dissolved Organic Matter (CDOM) consists of soluble, colored organic substances in water bodies, including humic substances, proteins, algal metabolic products, and other organic materials. It is a significant component and the largest reservoir of Dissolved Organic Carbon (DOC) in aquatic environments, appearing yellow and brown (Zhang et al., 2021). CDOM reduces sunlight penetration, and at high concentrations, its absorption overlaps with that of phytoplankton, potentially inhibiting their photosynthesis and disrupting aquatic ecosystems (Blough and Del Vecchio, 2002).

CDOM has unique spectral characteristics, with its absorption spectra showing good consistency across studies. Within the range of 250–700 nm, the spectral absorption coefficient of CDOM decreases exponentially with increasing wavelength. CDOM absorbs strongly in the blue and ultraviolet wavelengths and weakly in the range of 500–700 nm, with minimal absorption (approaching zero) beyond 700 nm.

In general, the parametric equation for the wavelength dependence of the absorption coefficient of CDOM in the visible band can be expressed as a negative exponential function (Carder et al., 1989; Zhang et al., 2007), as shown in Eq. (1):

$$a_{CDOM}(\lambda) = a_{CDOM}(\lambda_0)\exp[S_{CDOM}(\lambda_0 - \lambda)] \quad (1)$$

Where λ_0 is the wavelength of the selected reference band, $a_{CDOM}(\lambda)$ is the absorption coefficient of CDOM at any wavelength (m^{-1}), and S_{CDOM} is the slope parameter of the CDOM absorption spectrum. To measure the concentration of CDOM, it can be calculated directly or expressed as an absorption coefficient. Due to the complexity of CDOM composition, most studies use absorption coefficients to represent CDOM concentration. CDOM absorbs strongly in the ultraviolet and blue wavelength bands. The absorption coefficients at wavelengths of 355 nm ($a_{CDOM}(355)$), 400 nm ($a_{CDOM}(400)$), 420 nm ($a_{CDOM}(420)$), and 440 nm ($a_{CDOM}(440)$) are usually chosen to express the concentration of CDOM (Griffin et al., 2011; Zhu et al., 2014).Traditional methods for retrieving inland CDOM include the empirical/semi-empirical method and the semi-analytical method. Table 4 provides a detailed summary of the retrieval methods for CDOM in the water column and the application and accuracy of their models.

The widely used Quasi-Analytical Algorithm (QAA) model proposed by Lee et al. (2002), which belongs to the semi-analytical method, retrieves the total absorption coefficients of CDOM and non-algal species. Improved QAA-based models have since been applied to various water bodies. For instance, Zhu et al. (2011) developed the QAA-E model for better separation of CDOM and non-algae absorption coefficients. Dong et al. (2013) proposed a method with three wavelength absorption coefficients (412, 443, and 490 nm). Wang et al. (2017) accumulated 144 matchup samples based on GOCI images and nine cruise campaigns conducted in the Changjiang estuarine and coastal waters during 2011 and 2015, where CDOM water samples were obtained by filtering with a 0.22 μm polycarbonate membrane under low vacuum conditions immediately after sampling. They developed a new model (QAA_cj), adjusting several submodule coefficients within the QAA model by combining QAA_v6 (Zhongping, 2014) and QAA_CDOM (Zhu and Yu, 2013). This model can separate $a_{CDOM}(443)$ (the absorption coefficient of CDOM) from adg (CDOM and non-pigmented particles absorption

Table 4
Retrieval methods, models and the accuracy for CDOM.

Retrieval Methods	Retrieval Models	Equation/ Algorithm	Study Area	$a_{CDOM}(\lambda)$ or Concentration	Sensor	R^2 (Number of Samples)	RMSE (m^{-1}) or (mg/L)	Reference
The empirical/ semi-empirical method	Multi-band model	$a_{CDOM}(420) = 2.8091 * (R_{560}/R_{665})^{-2.341}$	46 lakes in Sweden	$a_{CDOM}(420) =$ 0.14–12.24 m^{-1}	Sentinel-2 MSI	0.65 ($N = 41$)	3.48	(Al-Kharusi et al., 2020)
		$a_{CDOM}(440) = 2.70\left(\frac{R_{650}}{R_{480}}\right)^2 - 6.14\frac{R_{650}}{R_{480}} + 4.19$	Barra Bonita Reservoir, Brazil	$a_{CDOM}(440) =$ 0.644–1.413 m^{-1}	Landsat-8 OLI	0.7 ($N = 19$)	10.65% (RMSE)	(Alcântara et al., 2016)
		$a_{CDOM}(440) = 40.75 * \exp[-2.463 * (B_3/B_4)]$	Lake Huron, China	$a_{CDOM}(440) =$ 0.11–8.46 m^{-1}	Landsat-8 OLI	0.95 ($N = 15$)	0.504	(Chen et al., 2017)
		Clean: $a_{CDOM}(355) = 0.6346 * (B_5/B_2) + 1.0811$ Turbid:$a_{CDOM}(355) = 0.417 * (B_7/B_8) + 1.3039$	Poyang Lake, China	$a_{CDOM}(355) =$ 1.70–3.33 m^{-1}	Sentinel-2A MSI	Clean:0.70 ($N = 15$) Turbid:0.73 ($N = 46$)	Clean:0.24 Turbid:0.19	(Xu et al., 2018)
The semi-analytical method	QAA-based model	QAA-E model	The Mississippi and Atchafalaya river plumes and the northern Gulf of Mexico, USA	$a_{CDOM}(440) =$ 0.192–7.068 m^{-1}	EO-1 Hyperion	0.81 ($N = 498$)	0.458	(Zhu et al., 2011)
		QAA-3R model	The Taiwan Strait and the South China Sea, China	$a_{CDOM}(412) =$ 0.008–1.613 m^{-1}	MODIS	0.67 ($N = 104$)	0.222	(Dong et al., 2013)
		QAA_cj model	The Changjiang estuarine and coastal waters, China	$a_{CDOM}(443) =$ 0.029–0.65 m^{-1}	GOCI	0.90 ($N = 43$)	0.07	(Wang et al., 2017)
The AI-based method	Traditional machine learning models	Backpropagation (BP) neural network	Lakes in the upper reaches of the Huai River, the middle and lower reaches of the Yangtze River and the Yunnan-Guizhou Plateau, China	$a_{CDOM}(254) =$ 2.64–34.04 m^{-1}	Landsat 8 OLI	0.75 ($N = 408$)	3.66	(Sun et al., 2021)
		Random forest (RF)	Baekje Reservoir, South Korea	$a_{CDOM}(355) =$ 2.1–11.0 m^{-1}	AisaFENIX hyperspectral	0.85 ($N = 22$)	0.70	(Kim et al., 2022)
		Extreme gradient boosting (XGBoost)	The Pearl River Estuary, China	$a_{CDOM}(290) =$ 0–12 m^{-1}	Landsat-8 OLI	0.9 ($N = 18$)	0.37	(Huang et al., 2023)
	Coupling machine learning models	PCA-extremely randomized trees (ET)	The river Elbe, Germany	Concentration = 0–46 ppbQS	Cubert UHD 285	0.95 ($N = 562$)	0.48	(Keller et al., 2018)
		GA-XGBoost	45 lakes in Estonia	Concentration = 0.85–81 mg/L	Sentinel-2 MSI	0.94 ($N = 21$)	3.77	(Toming et al., 2024)

Note:
(1) In the equations,R_{665} refers to the remote sensing reflectance at a wavelength of 665 nanometers (nm). Similar expressions can be applied analogously.
(2) B_i represents the remote sensing reflectance of the ith band of the corresponding sensor.

coefficient). Validation results showed that QAA_cj achieved high accuracy in estimating $a_{CDOM}(443)$, with an R^2 of 0.90 and an RMSE of 0.07 m^{-1} ($N = 43$). This study proposes an algorithm that is more suitable for retrieving CDOM in highly turbid waters. The semi-analytical method offers enhanced model generalizability and accuracy due to its clear physical parameters. However, it requires precise measurement of optical properties. Additionally, the Quasi-Analytical Algorithm is limited by the spectral channels available on some satellite remote sensing platforms.

The empirical/semi-empirical method establishes statistical relationships between CDOM concentrations and measured surface water spectral data, focusing on multi-band models. These models often use the ratio between green and red bands due to low blue spectral reflectance from high CDOM absorption and interference from particulates and phytoplankton (Alcântara et al., 2016; Kutser et al., 2016). Chen et al. (2017) estimated $a_{CDOM}(440)$ in the Saginaw River and the Kakagon River with the band ratio B_3/B_4 from the Landsat-8. Based on Sentinel-2 remote sensing imagery. Xu et al. (2018) found that in both clear water (TSM $<$ 10 mg/L) and turbid water (TSM $\geq$ 10 mg/L), multi-band models constructed with R_{689}/R_{497} and R_{767}/R_{826} provided the best retrieval results for $a_{CDOM}(355)$. Changes in boundary conditions can limit their applicability, and complex water environments may present challenges due to collinearity issues among different water elements.

With the rapid development of computer technology, the AI-based method for water quality parameter retrieval is gaining popularity. Unlike traditional empirical algorithms, ML algorithms like random forest regression (RFR), kernel ridge regression (KRR), Gaussian process

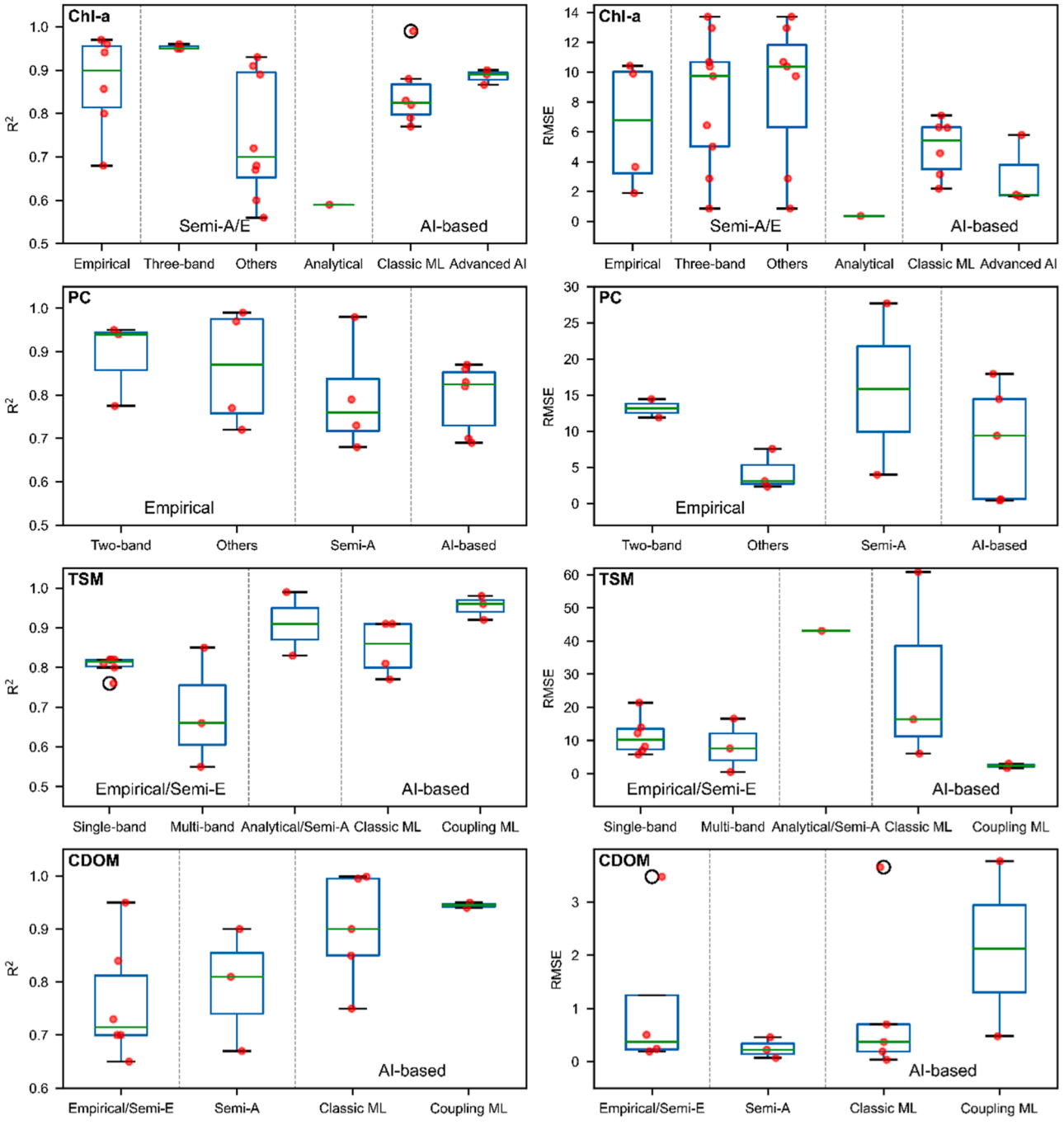


Fig. 1. Boxplot showing the R^2 and RMSE of the retrieval models for Chl-a, PC, TSM, and CDOM presented in the existing literature.

regression and SVR capture complex relationships in the data (Sun et al., 2021). Kim et al. (2022) collected 108 data points in the Baekje Reservoir based on seven sampling events for airborne hyperspectral imagery and CDOM absorption coefficient data from 350 nm to 440 nm over two years (2016–2017), which were divided into 80% and 20% for training and test datasets, respectively. They developed a RF model to retrieve CDOM and evaluated the best combination of input wavelength bands (475, 497, and 660 nm) that best represents the CDOM absorption coefficient at 355 nm. The RF model exhibited the best performance for CDOM estimation, with an R^2 of 0.85 and an RMSE of 0.70 m^{-1}. This study demonstrated the capability of RF models based on hyperspectral imagery in understanding CDOM in optically complex inland waters. Huang et al. (2023) developed an XGBoost model based on Landsat-8 OLI data to retrieve the CDOM absorption coefficient $a_{CDOM}(290)$, which outperformed SVM, RF, multilayer perceptron (MLP), and CNN models.

However, single ML methods can suffer from overfitting, high-dimensional data processing issues, and local minima, affecting accuracy and robustness. To address this, researchers are optimizing traditional models. Keller et al. (2018) combined principal component analysis (PCA) with extremely randomized trees based on hyperspectral data, successfully retrieving CDOM concentrations by making more sophisticated adjustments to the hyperparameters. Toming et al. (2024) constructed a model combining GA with XGBoost based on Sentinel-2 MSI images to estimate 16 water quality parameters, including CDOM concentration. This model was validated for its effectiveness.

This review statistically analyzes the retrieval models of Chl-a, PC, TSM, and CDOM presented in some existing literature, as well as summarizes the corresponding advantages and disadvantages. Among them: a) The analytical/semi-analytical method has a rigorous physical basis and versatility, but the complexity of the model is high and the acquisition of IOPs is demanding. b) The empirical/semi-empirical method uses linear regression between spectral radiance values and in-situ measurements. It is simple and low in computational effort but has poor generalizability due to sensitivity to changes in water composition. c) AI-based methods combine the benefits of the empirical/semi-empirical method, overcoming multicollinearity issues. Improved ML models avoid overfitting and local minima problems, while deep learning models excel with high-dimensional data. However, their intrinsic mechanisms need further exploration.

Fig. 1 shows the box plots of accuracy statistics based on the results. For Chl-a estimation, the AI-based method yields an R^2 comparable to the semi-empirical/semi-analytical method, slightly lower than the empirical method; the performance of RMSE is relatively strong, showing lower variability compared to the other methods. For PC estimation, the fitting performance of the AI-based method is comparable to the semi-analytical method, slightly lower than the empirical method, though RMSE is somewhat higher than the empirical method. For TSM estimation, the AI-based method outperforms both the empirical/semi-empirical method and the analytical/semi-analytical method, achieving higher R^2 and relatively lower RMSE. For CDOM estimation, R^2 of the AI-based method performs exceptionally well, far exceeding the empirical/semi-empirical method and the analytical/semi-analytical method, though the RMSE of the AI-based method is occasionally higher. Overall, considering that AI-based models are applied to retrieve water quality parameters across a large spatial range of various water bodies, and the concentration range is usually wide, it may lead to slightly weaker fitting performance and occasionally higher RMSE for some water quality parameters.

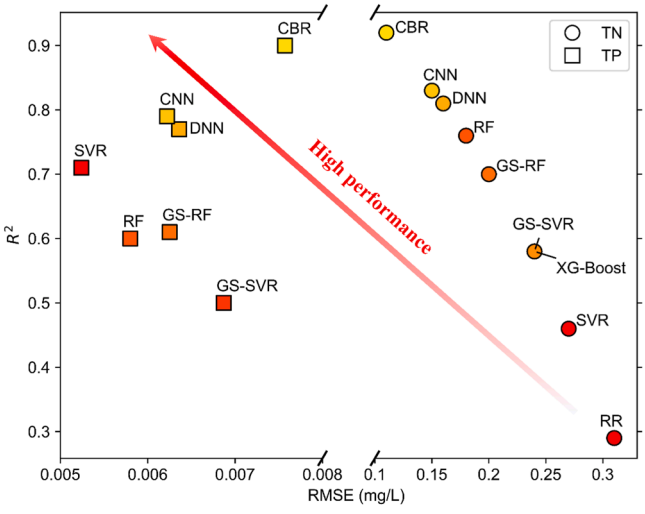


Fig. 2. The accuracy comparisons of each machine learning model.

4.5. *Dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), total nitrogen (TN), and total phosphorus (TP)*

Current research on NOACs is limited. NOACs have a weak correlation with the spectrum, making it challenging to estimate NOACs directly from spectral data. However, many studies have attempted to retrieve NOACs using indirect methods. Indirect methods first use spectral reflectance to retrieve OACs like Chl-a, TSM, and CDOM. Based on the correlations between OACs and NOACs, NOACs can then be indirectly estimated using traditional models (such as empirical/semi-empirical or semi-analytical models) (Cai et al., 2023; Chen and Quan, 2012). Additionally, in specific water bodies, NOACs may be correlated with multiple OAC parameters, resulting in a correlation with several bands in the spectral curve. AI-based models (Arias-Rodriguez et al., 2023; Niu et al., 2021; Zhang et al., 2020) are particularly adept at extracting hidden features related to OACs from complex spectral curves, allowing for the development of more accurate NOACs retrieval models, even without establishing a precise correlation between NOACs and OACs. For instance, Chen et al. (2024) used high-resolution UAV multispectral images and ground monitoring data for a typical rural stream to retrieve TN and TP using nine machine learning models. The results (Fig. 2) showed that the best model (Catboost regression, CBR) achieved a retrieval accuracy with an R^2 above 0.9. However, the retrieval accuracy of both traditional models and AI-based models depends on the robustness of the relationship between NOACs and OACs, which is environmentally dependent and difficult to generalize to other water bodies (Cao et al., 2021; Zang et al., 2011). Still, NOACs retrieval holds certain reference value and application potential for water quality monitoring.

5. Conclusion and prospects

Driven by rising environmental awareness and concerns over water safety, water quality monitoring technology has rapidly advanced. AI-based methods for water quality retrieval have gained prominence. This paper provides a comprehensive review of the retrieval methods for nine water quality parameters. It explores recent AI advancements in preprocessing, particularly in fusion techniques of multi-source remote sensing data, optimizing the use of open-source satellite data. To maximize the synergistic application of AI and remote sensing technologies in water quality monitoring of inland water bodies, more work needs to be done in the following five areas.

(a) The quality of remote sensing data is influenced by sensor resolution and atmospheric correction models, and research on AI-based preprocessing technologies is expected to significantly advance intelligent water quality monitoring. Multi-source remote sensing data fusion can address the limitations of single data sources in spatial and spectral resolution, enhancing image

quality and water quality prediction accuracy. The author believes that the development of advanced sensors as well as fusion techniques to improve temporal, spatial, and spectral resolutions is crucial for reducing the impact of data source quality on monitoring accuracy.

The atmospheric path radiance in the signals received by remote sensing significantly exceeds the reflectance of the water body, complicating compositional analysis. Effective atmospheric correction, removing effects from molecules, aerosols, and cloud particles, is crucial for accurate monitoring. Current methods, assuming zero reflectance in the near-infrared band, often fail for inland waters with high suspended matter and Chl-a concentrations. In the author's opinion, research should focus on the coupling mechanism and decoupling methods between water and atmospheric aerosol, and develop more atmospheric correction methods for inland water bodies to enhance the portability between sensors.

(b) Enriching the optical mechanisms of water quality retrieval models and integrating them with advanced artificial intelligence algorithms to improve model interpretability. This approach has already been attempted in the literature (Li et al., 2023) and is meaningful for remote sensing applications in the field of water quality monitoring. Traditional ML algorithms face issues like overfitting, high dimensionality, slow convergence, and local optima. Consequently, deep learning and coupled ML models have gained more attention. They often lack explainable physical mechanisms and struggle to integrate IOPs, limiting model accuracy and generalization.

Optical characteristics of inland water bodies vary by season and region, reducing model transferability. In the author's opinion, more efforts should be made to utilize large amounts of in-situ data and more advanced AI algorithms for water quality estimation. This includes enhancing the analysis of IOPs across a wider range of water bodies and enriching the physical mechanisms of AI-based water quality retrieval models. It is necessary to develop a universal algorithm that is not constrained by season or region, applicable to various water bodies, and capable of constructing transferable retrieval models.

(c) Exploring more water quality parameters with remote sensing technology and AI, strengthening the monitoring of special types of water bodies, and expanding to larger scales remain key focuses of current intelligent water quality monitoring. Although extensive research on OACs exists, special water bodies, such as black-odorous and eutrophic waters, require multiple parameters for accurate characterization due to their complex spectral properties, resulting in slow progress. Therefore, using deep learning methods to retrieve multiple OACs outputs from a single remote sensing reflectance input will be necessary in the future (Zhang et al., 2024). Meanwhile, with the increasing impact of climate change, global-scale data and research are becoming increasingly needed. Global remote sensing products, such as Forel-Ule Index (FUI) (Wang et al., 2021) and CIE (International Commission on Illumination)-based algal bloom detection (Hou et al., 2022), should be vigorously developed to support the SDG Goal 6: Clean Water and Sanitation.

Furthermore, NOACs remain insufficiently explored. The authors believe that, in addition to advancing NOACs retrieval for individual water bodies, emphasis should also be placed on data sharing and the establishment of large-scale datasets. Based on these datasets, artificial intelligence algorithms can be employed to classify water bodies with similar spectral characteristics and OACs, followed by efforts to build corresponding retrieval models for each category, thereby expanding the application scope of NOACs retrieval. At the same time, the long-term databases accumulated through online monitoring technologies should be fully utilized, and combined with AI algorithms, to explore more retrieval algorithms for NOACs, such as ammonia nitrogen, DOC, and others.

(d) The effective use of big data, cloud technology, and remote sensing platforms integrated with Large Language Models (LLMs) is the future trend for inland water quality monitoring. This meets the need for dense time series at high spatial resolution by leveraging multi-source remote sensing data, balancing extensive image preprocessing with real-time processing requirements. Cloud computing platforms like Google Earth Engine (GEE) and Amazon Web Services (AWS) offer high-end computing resources at low costs. This allows users to focus on earth science discoveries rather than data processing. New remote sensing platforms integrated with LLMs, like Alibaba DAMO Academy's AI Earth and SenseTime's SenseEarth, enable interactive dialogue with natural language, leading to more intelligent data processing.

The author suggests utilizing cloud computing infrastructures like GEE and AWS for large-scale water quality monitoring based on deep learning, or developing new cloud computing platforms that can directly use advanced intelligent algorithms, as well as remote sensing platforms integrated with LLMs. This will enable accurate, objective, dynamic, and rapid intelligent monitoring of inland water quality and trend prediction.

(e) Combining water quality remote sensing retrieval models with hydrodynamic or hydrological models to obtain spatiotemporally continuous water quality distributions provides a basis for scientific management and quantitative decision-making for water environment management. A single data source and model cannot monitor aquatic systems in a fine-grained, high-precision, comprehensive, and uninterrupted manner. Integrating water quality observation data with hydrodynamic or hydrological data into numerical simulation models through data assimilation can address this issue. This approach, originally applied in weather forecasting, includes optimal interpolation, variational algorithms, Kalman filtering, and particle filtering.

The author suggests utilizing data assimilation to combine different water environment information, optimizing water quality model accuracy. This allows dynamic simulation of inland water body runoff, pollution impact ranges, and water environment evolution, enhancing the ecological significance of remote sensing monitoring and promoting intelligent monitoring of inland water bodies.

CRediT authorship contribution statement

Yuan Sun: Writing – original draft, Visualization, Validation, Formal analysis, Data curation, Conceptualization. **Denghui Wang:** Writing – review & editing, Visualization, Software, Formal analysis, Conceptualization. **Lei Li:** Writing – review & editing, Resources, Project administration, Funding acquisition, Conceptualization. **Rongsheng Ning:** Writing – review & editing, Validation, Software. **Shuili Yu:** Writing – review & editing, Funding acquisition. **Naiyun Gao:** Writing – review & editing, Funding acquisition.

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 work was supported by the National Major Science and Technology Project of China (No. 2017ZX07502003-03) and the National Natural Science Foundation of China (No. 51678420).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2024.122546.

References

Al-Kharusi, E.S., Tenenbaum, D.E., Abdi, A.M., Kutser, T., Karlsson, J., Bergström, A.-K., Berggren, M., 2020. Large-scale retrieval of coloured dissolved organic matter in northern lakes using Sentinel-2 data. Remote Sens. (Basel) 12 (1), 157.

Alcântara, E., Bernardo, N., Watanabe, F., Rodrigues, T., Rotta, L., Carmo, A., Shimabukuro, M., Gonçalves, S., Imai, N., 2016. Estimating the CDOM absorption coefficient in tropical inland waters using OLI/Landsat-8 images. Remote Sens. Lett. 7 (7), 661–670.

Allan, M.G., Hamilton, D.P., Hicks, B., Brabyn, L., 2015. Empirical and semi-analytical chlorophyll a algorithms for multi-temporal monitoring of New Zealand lakes using Landsat. Environ. Monit. Assess. 187 (6), 364.

Andrade, C., Alcântara, E., Bernardo, N., Kampel, M., 2019. An assessment of semi-analytical models based on the absorption coefficient in retrieving the chlorophyll-a concentration from a reservoir. Adv. Space Res. 63 (7), 2175–2188.

Arias-Rodriguez, L.F., Duan, Z., Díaz-Torres, J.d.J., Basilio Hazas, M., Huang, J., Kumar, B.U., Tuo, Y., Disse, M., 2021. Integration of remote sensing and Mexican water quality monitoring system using an extreme learning machine. Sensors 21 (12), 4118.

Arias-Rodriguez, L.F., Tüzün, U.F., Duan, Z., Huang, J., Tuo, Y., Disse, M., 2023. Global water quality of inland waters with harmonized landsat-8 and Sentinel-2 using cloud-computed machine learning. Remote Sens. (Basel) 15 (5), 1390.

Beal, M.R.W., Özdoğan, M., Block, P.J., 2024. A machine learning and remote sensing-based model for algae pigment and dissolved oxygen retrieval on a small inland lake. Water Resour. Res. 60 (3), e2023WR035744.

Beck, R., Zhan, S., Liu, H., Tong, S., Yang, B., Xu, M., Ye, Z., Huang, Y., Shu, S., Wu, Q., Wang, S., Berling, K., Murray, A., Emery, E., Reif, M., Harwood, J., Young, J., Nietch, C., Macke, D., Martin, M., Stillings, G., Stump, R., Su, H., 2016. Comparison of satellite reflectance algorithms for estimating chlorophyll-a in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations. Remote Sens. Environ. 178, 15–30.

Binding, C.E., Pizzolato, L., Zeng, C., 2021. EOLakeWatch; delivering a comprehensive suite of remote sensing algal bloom indices for enhanced monitoring of Canadian eutrophic lakes. Ecol. Indic. 121, 106999.

Blough, N.V., Del Vecchio, R., 2002. Chromophoric DOM in the coastal environment. In: Hansell, D.A., Carlson, C.A. (Eds.), Biogeochemistry of Marine Dissolved Organic Matter. Academic Press, San Diego, pp. 509–546. https://doi.org/10.1016/B978-012323841-2/50012-9.

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. Sci. Total Environ. 856, 158869.

Cao, L., Zhang, D., Guo, Q., Zhan, J., 2021. Inversion of water quality parameter BOD5 based on hyperspectral remotely sensed data in Qinghai Lake. In: IEEE International Geoscience and Remote Sensing Symposium IGARSS. Brussels, Belgium, pp. 5036–5039. https://doi.org/10.1109/IGARSS47720.2021.9553783.

Cao, Q., Yu, G., Qiao, Z., 2022. Application and recent progress of inland water monitoring using remote sensing techniques. Environ. Monit. Assess. 195 (1), 125.

Cao, Z., Duan, H., Feng, L., Ma, R., Xue, K., 2017. Climate- and human-induced changes in suspended particulate matter over Lake Hongze on short and long timescales. Remote Sens. Environ. 192, 98–113.

Cao, Z., Ma, R., Duan, H., 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.

Carder, K.L., Steward, R.G., Harvey, G.R., Ortner, P.B., 1989. Marine humic and fulvic acids: their effects on remote sensing of ocean chlorophyll. Limnol. Oceanogr. 34 (1), 68–81.

Castagna, A., Simis, S., Dierssen, H., Vanhellemont, Q., Sabbe, K., Vyverman, W., 2020. Extending landsat 8: retrieval of an orange contra-band for inland water quality applications. Remote Sens. (Basel) 12 (4), 637.

Chawira, M., Dube, T., Gumindoga, W., 2013. Remote sensing based water quality monitoring in Chivero and Manyame lakes of Zimbabwe. Phys. Chem. Earth Parts A/B/C 66, 38–44.

Chawla, I., Karthikeyan, L., Mishra, A.K., 2020. A review of remote sensing applications for water security: quantity, quality, and extremes. J. Hydrol. 585, 124826.

Chen, B., Mu, X., Chen, P., Wang, B., Choi, J., Park, H., Xu, S., Wu, Y., Yang, H., 2021. Machine learning-based inversion of water quality parameters in typical reach of the urban river by UAV multispectral data. Ecol. Indic. 133, 108434.

Chen, J., Quan, W., 2012. Using Landsat/TM imagery to estimate nitrogen and phosphorus concentration in Taihu Lake, China. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5 (1), 273–280.

Chen, J., Quan, W., Cui, T., Song, Q., 2015. Estimation of total suspended matter concentration from MODIS data using a neural network model in the China eastern coastal zone. Estuarine Coastal Shelf Sci. 155, 104–113.

Chen, J., Zhu, W.N., Tian, Y.Q., Yu, Q., 2017. Estimation of colored dissolved organic matter from landsat-8 imagery for complex inland water: case study of Lake Huron. IEEE Trans. Geosci. Remote Sens. 55 (4), 2201–2212.

Chen, S., Fang, L., Li, H., Chen, W., Huang, W., 2011. Evaluation of a three-band model for estimating chlorophyll-a concentration in tidal reaches of the Pearl River Estuary, China. ISPRS J. Photogramm. Remote Sens. 66 (3), 356–364.

Chen, Y., Yao, K., Zhu, B., Gao, Z., Xu, J., Li, Y., Hu, Y., Lin, F., Zhang, X., 2024. Water quality inversion of a typical rural small river in Southeastern China based on UAV multispectral imagery: a comparison of multiple machine learning algorithms. Water (Basel) 16 (4), 553.

Chen, Z., Dou, M., Xia, R., Li, G., Shen, L., 2022. Spatiotemporal evolution of chlorophyll-a concentration from MODIS data inversion in the middle and lower reaches of the Hanjiang River, China. Environ. Sci. Pollut. Res. 29 (25), 38143–38160.

Dekker, A.G., 1993. Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing- Vrije Universiteit Amsterdam (vu.nl). Free UniversitPrint, Amsterdam. ISBNs: 9090062343.

Dekker, A.G., Vos, R.J., Peters, S.W.M., 2002. Analytical algorithms for lake water TSM estimation for retrospective analyses of TM and SPOT sensor data. Int. J. Remote Sens. 23 (1), 15–35.

Dong, Q., Shang, S., Lee, Z., 2013. An algorithm to retrieve absorption coefficient of chromophoric dissolved organic matter from ocean color. Remote Sens. Environ. 128, 259–267.

Dörnhöfer, K., Scholze, J., Stelzer, K., Oppelt, N., 2018. Water colour analysis of Lake Kummerow using time series of remote sensing and in situ data. PFG J. Photogramm. Remote Sens. Geoinf. Sci. 86 (2), 103–120.

Du, Y., Song, K., Liu, G., Wen, Z., Fang, C., Shang, Y., Zhao, F., Wang, Q., Du, J., Zhang, B., 2020. Quantifying total suspended matter (TSM) in waters using Landsat images during 1984–2018 across the Songnen Plain, Northeast China. J. Environ. Manage. 262, 110334.

Du, Y., Song, K., Wang, Q., Liu, G., Wen, Z., Shang, Y., Lyu, L., Du, J., Li, S., Tao, H., Zhang, B., Wang, X., 2021. Using Remote Sensing to Understand the Total Suspended Matter Dynamics in Lakes Across Inner Mongolia. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, 7478–7488.

Duan, P.H., Lai, J.B., Kang, J., Kang, X.D., Ghamisi, P., Li, S.T., 2020. Texture-aware total variation-based removal of sun glint in hyperspectral images. ISPRS J. Photogramm. Remote Sens. 166, 359–372.

El-Alem, A., Chokmani, K., Laurion, I., El-Adlouni, S.E., 2012. Comparative analysis of four models to estimate chlorophyll-a concentration in case-2 waters using MODerate Resolution Imaging Spectroradiometer (MODIS) imagery. Remote Sens. (Basel) 4 (8), 2373–2400.

Fang, C., Song, C., Wen, Z., Liu, G., Wang, X., Li, S., Shang, Y., Tao, H., Lyu, L., Song, K., 2024. A novel chlorophyll-a retrieval model based on suspended particulate matter classification and different machine learning. Environ. Res. 240, 117430.

Fang, C., Song, K., Paerl, H.W., Jacinthe, P.-A., Wen, Z., Liu, G., Tao, H., Xu, X., Kutser, T., Wang, Z., Duan, H., Shi, K., Shang, Y., Lyu, L., Li, S., Yang, Q., Lyu, D., Mao, D., Zhang, B., Cheng, S., Lyu, Y., 2022. Global divergent trends of algal blooms detected by satellite during 1982–2018. Glob. Chang. Biol. 28 (7), 2327–2340.

Fu, B., Lao, Z., Liang, Y., Sun, J., He, X., Deng, T., He, W., Fan, D., Gao, E., Hou, Q., 2022. Evaluating optically and non-optically active water quality and its response relationship to hydro-meteorology using multi-source data in Poyang Lake, China. Ecol. Indic. 145, 109675.

George, D.G., 1997. The airborne remote sensing of phytoplankton chlorophyll in the lakes and tarns of the English Lake District. Int. J. Remote Sens. 18 (9), 1961–1975.

Giardino, C., Brando, V.E., Dekker, A.G., Strömbeck, N., Candiani, G., 2007. Assessment of water quality in Lake Garda (Italy) using Hyperion. Remote Sens. Environ. 109 (2), 183–195.

Gilerson, A.A., Gitelson, A.A., Zhou, J., Gurlin, D., Moses, W., Ioannou, I., Ahmed, S.A., 2010. Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. Opt. Express 18 (23), 24109–24125.

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.

Gordon, H.R., Brown, O.B., Jacobs, M.M., 1975. Computed relationships between the inherent and apparent optical properties of a flat homogeneous ocean. Appl. Opt. 14 (2), 417–427.

Griffin, C.G., Frey, K.E., Rogan, J., Holmes, R.M., 2011. Spatial and interannual variability of dissolved organic matter in the Kolyma River, East Siberia, observed using satellite imagery. J. Geophys. Res. Biogeosci. 116 (G3).

Guo, Q., Wu, H., Jin, H., Yang, G., Wu, X., 2022. Remote sensing inversion of suspended matter concentration using a neural network model optimized by the partial least squares and particle swarm optimization algorithms. Sustainability 14 (4), 2221.

Gurlin, D., Gitelson, A.A., Moses, W.J., 2011. Remote estimation of chl-a concentration in turbid productive waters — Return to a simple two-band NIR-red model? Remote Sens. Environ. 115 (12), 3479–3490.

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.

Hou, X., Feng, L., Dai, Y., Hu, C., Gibson, L., Tang, J., Lee, Z., Wang, Y., Cai, X., Liu, J., Zheng, Y., Zheng, C., 2022. Global mapping reveals increase in lacustrine algal blooms over the past decade. Nat. Geosci. 15 (2), 130–134.

Huang, Y., Pan, J., Devlin, A.T., 2023. Enhanced estimate of chromophoric dissolved organic matter using machine learning algorithms from landsat-8 OLI data in the Pearl River estuary. Remote Sens. (Basel) 15 (8), 1963.

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 (11), 2705–2718.
Hunter, P.D., Tyler, A.N., Willby, N.J., Gilvear, D.J., 2008. The spatial dynamics of vertical migration by Microcystis aeruginosa in a eutrophic shallow lake: a case study using high spatial resolution time-series airborne remote sensing. Limnol. Oceanogr. 53 (6), 2391–2406.
Jin, Q., Lyu, H., Shi, L., Miao, S., Wu, Z., Li, Y., Wang, Q., 2017. Developing a two-step method for retrieving cyanobacteria abundance from inland eutrophic lakes using MERIS data. Ecol. Indic. 81, 543–554.
Keller, S., Maier, P.M., Riese, F.M., Norra, S., Holbach, A., Börsig, N., Wilhelms, A., Moldaenke, C., Zaake, A., Hinz, S., 2018. Hyperspectral data and machine learning for estimating CDOM, chlorophyll a, diatoms, green algae and turbidity. Int. J. Environ. Res. Public Health 15 (9), 1881.
Kim, J., Jang, W., Hwi Kim, J., Lee, J., Hwa Cho, K., Lee, Y.-G., Chon, K., Park, S., Pyo, J., Park, Y., Kim, S., 2022. Application of airborne hyperspectral imagery to retrieve spatiotemporal CDOM distribution using machine learning in a reservoir. Int. J. Appl. Earth Obs. Geoinformation 114, 103053.
Kutser, T., Paavel, B., Verpoorter, C., Ligi, M., Soomets, T., Toming, K., Casal, G., 2016. Remote sensing of black lakes and using 810 nm reflectance peak for retrieving water quality parameters of optically complex waters. Remote Sens (Basel) 8 (6), 497.
Kwon, Y.S., Pyo, J., Kwon, Y.-H., Duan, H., Cho, K.H., Park, Y., 2020. Drone-based hyperspectral remote sensing of cyanobacteria using vertical cumulative pigment concentration in a deep reservoir. Remote Sens. Environ. 236, 111517.
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.
Le, C., Li, Y., Zha, Y., Wang, Q., Zhang, H., Yin, B., 2011. Remote sensing of phycocyanin pigment in highly turbid inland waters in Lake Taihu, China. Int. J. Remote Sens. 32 (23), 8253–8269.
Lee, Z., Carder, K.L., Arnone, R.A., 2002. Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. Appl. Opt. 41 (27), 5755–5772.
Legleiter, C.J., King, T.V., Carpenter, K.D., Hall, N.C., Mumford, A.C., Slonecker, T., Graham, J.L., Stengel, V.G., Simon, N., Rosen, B.H., 2022. Spectral mixture analysis for surveillance of harmful algal blooms (SMASH): a field-, laboratory-, and satellite-based approach to identifying cyanobacteria genera from remotely sensed data. Remote Sens. Environ. 279, 113089.
Li, S., Song, K., Wang, S., Liu, G., Wen, Z., Shang, Y., Lyu, L., Chen, F., Xu, S., Tao, H., Du, Y., Fang, C., Mu, G., 2021. Quantification of chlorophyll-a in typical lakes across China using Sentinel-2 MSI imagery with machine learning algorithm. Sci. Total Environ. 778, 146271.
Li, X., Yang, Y., Ishizaka, J., Li, X., 2023. Global estimation of phytoplankton pigment concentrations from satellite data using a deep-learning-based model. Remote Sens. Environ. 294, 113628.
Liu, G., Simis, S.G.H., Li, L., Wang, Q., Li, Y., Song, K., Lyu, H., Zheng, Z., Shi, K., 2018. A four-band semi-analytical model for estimating phycocyanin in inland waters from simulated MERIS and OLCI data. IEEE Trans. Geosci. Remote Sens. 56 (3), 1374–1385.
Liu, S., Glamore, W., Tamburic, B., Morrow, A., Johnson, F., 2022. Remote sensing to detect harmful algal blooms in inland waterbodies. Sci. Total Environ. 851, 158096.
Liu, X., Zhang, Z., Jiang, T., Li, X., Li, Y., 2021. Evaluation of the effectiveness of multiple machine learning methods in remote sensing quantitative retrieval of suspended matter concentrations: a case study of Nansi Lake in North China. J. Spectrosc. 2021, 5957376.
Liu, Y., Yao, H., Chen, H., Wang, M., Huang, Z., Zhong, W., 2023. Early warning of red tide of Phaeocystis globosa based on phycocyanin concentration retrieval in Qinzhou Bay, China. Appl. Sci. 13 (20), 11449.
Matthews, M.W., 2011. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. Int. J. Remote Sens. 32 (21), 6855–6899. https://doi.org/10.1080/01431161.2010.512947.
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.
Moses, W.J., Gitelson, A.A., Berdnikov, S., Povazhnyy, V., 2009. Satellite estimation of chlorophyll-a concentration using the red and NIR bands of MERIS—The Azov Sea case study. IEEE Geosci. Remote Sens. Lett. 6 (4), 845–849.
Neil, C., Spyrakos, 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.
Niroumand-Jadidi, M., Bovolo, F., 2022. Extreme Gradient Boosting Machine Learning for Total Suspended Matter (TSM) Retrieval from Sentinel-2 Imagery. SPIE.
Niu, C., Tan, K., Jia, X., Wang, X., 2021. Deep learning based regression for optically inactive inland water quality parameter estimation using airborne hyperspectral imagery. Environ. Pollut. 286, 117534.
Ogashawara, I., Mishra, D.R., Mishra, S., Curtarelli, M.P., Stech, J.L., 2013. A performance review of reflectance based algorithms for predicting phycocyanin concentrations in inland waters. Remote Sens. (Basel) 5 (10), 4774–4798.
Ouma, Y.O., Noor, K., Herbert, K., 2020. Modelling reservoir chlorophyll-a, TSS, and turbidity using Sentinel-2A MSI and landsat-8 OLI satellite sensors with empirical multivariate regression. J. Sens. 2020 (1), 8858408.
Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M.K., Ondrusek, M., Oppelt, N., Stumpf, R., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: a machine-learning approach. Remote Sens. Environ. 240, 111604.
Park, Y., Cho, K.H., Park, J., Cha, S.M., Kim, J.H., 2015. Development of early-warning protocol for predicting chlorophyll-a concentration using machine learning models in freshwater and estuarine reservoirs, Korea. Sci. Total Environ. 502, 31–41.
Park, Y., Pyo, J., Kwon, Y.S., Cha, Y., Lee, H., Kang, T., Cho, K.H., 2017. Evaluating physico-chemical influences on cyanobacterial blooms using hyperspectral images in inland water, Korea. Water Res. 126, 319–328.
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.
Pyo, J., Duan, H., Ligaray, M., Kim, M., Baek, S., Kwon, Y.S., Lee, H., Kang, T., Kim, K., Cha, Y., Cho, K.H., 2020. An integrative remote sensing application of stacked autoencoder for atmospheric correction and cyanobacteria estimation using hyperspectral imagery. Remote Sens. 12 (7), 1073.
Qi, L., Hu, C., Duan, H., Cannizzaro, J., Ma, R., 2014. A novel MERIS algorithm to derive cyanobacterial phycocyanin pigment concentrations in a eutrophic lake: theoretical basis and practical considerations. Remote Sens. Environ. 154, 298–317.
Richard, P.S., Michelle, C.T., 2005. Use of remote sensing in monitoring and forecasting of harmful algal blooms. In: Proceedings Volume 5885, Remote Sensing of the Coastal Oceanic Environment, p. 58850I. https://doi.org/10.1117/12.614376.
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-Science Rev. 205, 103187.
Schalles, J., Yacobi, Y., 2000. Remote detection and seasonal patterns of phycocyanin, carotenoid and chlorophyll pigments in eutrophic waters. Archiv. Hydrobiol. Special Issues Adv. Limnol. 55, 153–168.
Schartau, M., Riethmüller, R., Flöser, G., van Beusekom, J.E.E., Krasemann, H., Hofmeister, R., Wirtz, K., 2019. On the separation between inorganic and organic fractions of suspended matter in a marine coastal environment. Prog. Oceanogr. 171, 231–250.
Shen, F., Zhou, Y.-X., Li, D.-J., Zhu, W.-J., Suhyb Salama, M., 2010. Medium resolution imaging spectrometer (MERIS) estimation of chlorophyll-a concentration in the turbid sediment-laden waters of the Changjiang (Yangtze) Estuary. Int. J. Remote Sens. 31 (17–18), 4635–4650.
Shi, K., Zhang, Y., Zhu, G., Liu, X., Zhou, Y., Xu, H., Qin, B., Liu, G., Li, Y., 2015. Long-term remote monitoring of total suspended matter concentration in Lake Taihu using 250m MODIS-Aqua data. Remote Sens. Environ. 164, 43–56.
Simis, S.G.H., Huot, Y., Babin, M., Seppälä, J., Metsamaa, L., 2012. Optimization of variable fluorescence measurements of phytoplankton communities with cyanobacteria. Photosyn. Res. 112 (1), 13–30.
Simis, S.G.H., Peters, S.W.M., Gons, H.J., 2005. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. Limnol. Oceanogr. 50 (1), 237–245.
Song, K., Li, L., Tedesco, L.P., Li, S., Clercin, N.A., Hall, B.E., Li, Z., Shi, K., 2012a. Hyperspectral determination of eutrophication for a water supply source via genetic algorithm–partial least squares (GA–PLS) modeling. Sci. Total Environ. 426, 220–232.
Song, K., Li, L., Wang, Z., Liu, D., Zhang, B., Xu, J., Du, J., Li, L., Li, S., Wang, Y., 2012b. Retrieval of total suspended matter (TSM) and chlorophyll-a (Chl-a) concentration from remote-sensing data for drinking water resources. Environ. Monit. Assess. 184 (3), 1449–1470.
Sòria-Perpinyà, X., Vicente, E., Urrego, P., Pereira-Sandoval, M., Ruíz-Verdú, A., Delegido, J., Soria, J.M., Moreno, J., 2020. Remote sensing of cyanobacterial blooms in a hypertrophic lagoon (Albufera of València, Eastern Iberian Peninsula) using multitemporal Sentinel-2 images. Sci. Total Environ. 698, 134305.
Spyrakos, E., O'Donnell, R., Hunter, P.D., Miller, C., Scott, M., Simis, S.G.H., Neil, C., Barbosa, C.C.F., Binding, C.E., Bradt, S., Bresciani, M., Dall'Olmo, G., Giardino, C., Gitelson, A.A., Kutser, T., Li, L., Matsushita, B., Martinez-Vicente, V., Matthews, M. W., Ogashawara, I., Ruiz-Verdú, A., Schalles, J.F., Tebbs, E., Zhang, Y., Tyler, A.N., 2018. Optical types of inland and coastal waters. Limnol. Oceanogr. 63 (2), 846–870.
Sun, D., Hu, C., Qiu, Z., Shi, K., 2015. Estimating phycocyanin pigment concentration in productive inland waters using Landsat measurements: a case study in Lake Dianchi. Opt. Express 23 (3), 3055–3074.
Sun, D., Li, Y., Wang, Q., Gao, J., Le, C., Huang, C., Gong, S., 2013. Hyperspectral remote sensing of the pigment C-phycocyanin in turbid inland waters, based on optical classification. IEEE Trans. Geosci. Remote Sens. 51 (7), 3871–3884.
Sun, L., Zhou, X.Y., Wang, R.L., Wei, J., Yang, Y.K., Wang, Q., 2017. A comparison of the cloud detection results between the UDTCDA mask and MOD35 cloud products. In: Fort Worth. Ieee, TX, pp. 25–28.
Sun, X., Zhang, Y., Zhang, Y., Shi, K., Zhou, Y., Li, N., 2021. Machine learning algorithms for Chromophoric Dissolved Organic Matter (CDOM) estimation based on landsat 8 images. Remote Sens. (Basel) 13 (18), 3560.
Tian, S., Guo, H., Xu, W., Zhu, X., Wang, B., Zeng, Q., Mai, Y., Huang, J.J., 2023. Remote sensing retrieval of inland water quality parameters using Sentinel-2 and multiple machine learning algorithms. Environ. Sci. Pollut. Res. 30 (7), 18617–18630.
Toming, K., Liu, H., Soomets, T., Uuemaa, E., Nõges, T., Kutser, T., 2024. Estimation of the biogeochemical and physical properties of lakes based on remote sensing and artificial intelligence applications. Remote Sens. 16 (3), 464.
Wang, S., Li, J., Zhang, W., Cao, C., Zhang, F., Shen, Q., Zhang, X., Zhang, B., 2021. A dataset of remote-sensed Forel-Ule Index for global inland waters during 2000–2018. Sci. Data 8 (1), 26.
Wang, Y., Shen, F., Sokoletsky, L., Sun, X., 2017. Validation and calibration of QAA algorithm for CDOM absorption retrieval in the Changjiang (Yangtze) estuarine and coastal waters. Remote Sens. (Basel) 9 (11), 1192.

Wasehun, E.T., Hashemi Beni, L., Di Vittorio, C.A., 2024. UAV and satellite remote sensing for inland water quality assessments: a literature review. Environ. Monit. Assess. 196 (3), 277.
Wen, Z., Wang, Q., Liu, G., Jacinthe, P.-A., Wang, X., Lyu, L., Tao, H., Ma, Y., Duan, H., Shang, Y., Zhang, B., Du, Y., Du, J., Li, S., Cheng, S., Song, K., 2022. Remote sensing of total suspended matter concentration in lakes across China using Landsat images and Google Earth Engine. ISPRS J. Photogramm. Remote Sens. 187, 61–78.
Wheeler, S.M., Morrissey, L.A., Levine, S.N., Livingston, G.P., Vincent, W.F., 2012. Mapping cyanobacterial blooms in Lake Champlain's Missisquoi Bay using QuickBird and MERIS satellite data. J. Great Lakes Res. 38, 68–75.
Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Warner, R.A., Tester, P.A., Dyble, J., Fahnenstiel, G.L., 2008. Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. Int. J. Remote Sens. 29 (12), 3665–3672.
Xu, J., Fang, C., Gao, D., Zhang, H., Gao, C., Xu, Z., Wang, Y., 2018. Optical models for remote sensing of chromophoric dissolved organic matter (CDOM) absorption in Poyang Lake. ISPRS J. Photogramm. Remote Sens. 142, 124–136.
Yim, I., Shin, J., Lee, H., Park, S., Nam, G., Kang, T., Cho, K.H., Cha, Y., 2020. Deep learning-based retrieval of cyanobacteria pigment in inland water for in-situ and airborne hyperspectral data. Ecol. Indic. 110, 105879.
Zang, C., Huang, S., Wu, M., Du, S., Scholz, M., Gao, F., Lin, C., Guo, Y., Dong, Y., 2011. Comparison of relationships between pH, dissolved oxygen and chlorophyll a for aquaculture and non-aquaculture waters. Water Air Soil Pollut. 219 (1), 157–174.
Zhang, D., Shi, K., Wang, W., Wang, X., Zhang, Y., Qin, B., Zhu, M., Dong, B., Zhang, Y., 2024. An optical mechanism-based deep learning approach for deriving water trophic state of China's lakes from Landsat images. Water Res. 252, 121181.
Zhang, J., Fu, P., Meng, F., Yang, X., Xu, J., Cui, Y., 2022. Estimation algorithm for chlorophyll-a concentrations in water from hyperspectral images based on feature derivation and ensemble learning. Ecol. Inform. 71, 101783.
Zhang, Y., Qin, B., Zhu, G., Zhang, L., Yang, L., 2007. Chromophoric dissolved organic matter (CDOM) absorption characteristics in relation to fluorescence in Lake Taihu, China, a large shallow subtropical lake. Hydrobiologia 581 (1), 43–52.
Zhang, Y., Shi, K., Zhang, Y., Moreno-Madrinan, M.J., Li, Y., Li, N., 2018. A semi-analytical model for estimating total suspended matter in highly turbid waters. Opt. Express 26 (26), 34094–34112.
Zhang, Y., Shi, K., Zhang, Y., Moreno-Madriñán, M.J., Zhu, G., Zhou, Y., Yao, X., 2019. Long-term change of total suspended matter in a deep-valley reservoir with HJ-1A/B: implications for reservoir management. Environ. Sci. Pollut. Res. 26 (3), 3041–3054.
Zhang, Y., Wu, L., Ren, H., Deng, L., Zhang, P., 2020. Retrieval of water quality parameters from hyperspectral images using hybrid bayesian probabilistic neural network. Remote Sen. 12 (10), 1567. https://doi.org/10.3390/rs12101567.
Zhang, Y., Zhou, L., Zhou, Y., Zhang, L., Yao, X., Shi, K., Jeppesen, E., Yu, Q., Zhu, W., 2021. Chromophoric dissolved organic matter in inland waters: present knowledge and future challenges. Sci. Total Environ. 759, 143550.
Zheng, Z., Li, Y., Guo, Y., Xu, Y., Liu, G., Du, C., 2015. Landsat-based long-term monitoring of total suspended matter concentration pattern change in the wet season for Dongting Lake, China. Remote Sens. 7 (10), 13975–13999.
Zhongping, L. 2014. Update of the Quasi-Analytical Algorithm (QAA_v6). Available online: http://www.ioccg.org/groups/Software_OCA/QAA_v6_2014209.pdf (accessed on 19 November 2017).
Zhu, W., Yu, Q., 2013. Inversion of chromophoric dissolved organic matter from EO-1 hyperion imagery for turbid estuarine and coastal waters. IEEE Trans. Geosci. Remote Sens. 51 (6), 3286–3298.
Zhu, W., Yu, Q., Tian, Y.Q., Becker, B.L., Zheng, T., Carrick, H.J., 2014. An assessment of remote sensing algorithms for colored dissolved organic matter in complex freshwater environments. Remote Sens. Environ. 140, 766–778.
Zhu, W., Yu, Q., Tian, Y.Q., Chen, R.F., Gardner, G.B., 2011. Estimation of chromophoric dissolved organic matter in the Mississippi and Atchafalaya river plume regions using above-surface hyperspectral remote sensing. J. Geophys. Res. Oceans 116 (C2).
Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag. 5 (4), 8–36.