Remote sensing inversion of suspended particulate matter in the estuary of the Pinglu Canal in China based on machine learning algorithms

Jinying Mo[1,2], Yichao Tian[1,2,3,4,5]*, Jiale Wang[1], Qiang Zhang[1], Yali Zhang[1], Jin Tao[1] and Junliang Lin[1]

[1]School of Resources and Environment, Beibu Gulf University, Qinzhou, China, [2]School of Resources, Environment and Materials, Guangxi University, Nanning, China, [3]Guangxi Key Laboratory of Marine Environmental Change and Disaster in Beibu Gulf, Qinzhou, China, [4]Gulf Ocean Development Research Center, Beibu Gulf University, Qinzhou, China, [5]Key Laboratory of Marine Geographic Information Resources Development and Utilization in the Beibu Gulf, Beibu Gulf University, Qinzhou, China

Introduction: Suspended particulate matter (SPM) is a critical indicator of water quality and has a significant impact on the nearshore ecological environment. Consequently, the quantitative evaluation of SPM concentrations is essential for managing nearshore environments and planning marine resources.

Methods: This study utilized Sentinel-2's single band and water index variables to develop a remote sensing inversion model for oceanic SPM in the estuary of the Pinglu Canal in China. Six machine learning algorithms were employed: K-nearest neighbor regression (KNNR), AdaBoost regression (ABR), random forest (RF), gradient boosting regression (GBR), extreme gradient boosting regression (XGBR), and light generalized boosted regression (LGBM). The model with the optimal performance was then selected for further analysis. This research applied the established model to investigate the spatial-temporal dynamics of SPM from 2021 to 2023.

Results: The findings indicated that (1) the XGBR algorithm exhibited superior performance (R^2 = 0.9042, RMSE = 3.0258 mg/L), with LGBM (R^2 =0.8258, RMSE = 4.0813 mg/L) and GBR (R^2 = 0.823, RMSE = 4.3477 mg/L) also demonstrating effective fitting. However, the ABR, RF, and KNNR algorithms produced less satisfactory fitting results. (2) Additionally, the study revealed that the combination of input variables in the XGBR algorithm was more accurate than single-variable inputs. (3) The contribution of single-band variables to the XGBR algorithm surpassed that of water index variables, with B12, B4, and B11 emerging as the top three influential variables in the model. (4) The annual SPM concentration in the study area exhibited an overall increasing trend, while its spatial distribution generally decreased from the estuary toward the Maowei Sea and Qinzhou Bay.

Discussion: The combination of Sentinel-2 data and XGBR model has shown good performance in retrieving SPM concentration, providing a new method and approach for large-scale estimation of SPM concentration.

1 Introduction

The suspended particulate matter (SPM) concentration serves as a pivotal parameter in the evaluation of water quality and plays a significant role in coastal ecosystems (Chen et al., 2015; Kratzer et al., 2020). The concentration of SPM directly correlates with the optical properties of water, including transparency, turbidity, and color. These properties, in turn, influence the distribution of underwater light fields, imposing limitations on the growth and reproductive processes of primary productivity, such as phytoplankton (Li et al., 2019; Zhang et al., 2021; Wang et al., 2022a; Wang et al., 2022). Concurrently, SPM serves as a crucial medium for the transport and transformation of waterborne pollutants, facilitating the cycling of nutrients, organic contaminants, and heavy metals (Cheng et al., 2019; Liu et al., 2019). Consequently, changes in SPM concentration may have adverse effects on nearshore ecosystems. The spatial distribution characteristics of SPM plays a pivotal role in the analysis of erosion and sedimentation processes in estuarine and coastal regions, serving as the primary influencing factor in topographic and geomorphic evolution (Ji et al., 2018; Ma et al., 2024). Thus, the quantitative assessment of spatiotemporal dynamic changes in SPM concentration holds great significance for marine engineering construction, nearshore marine environmental management, and marine resource planning.

Traditionally, SPM monitoring involves field sampling and subsequent laboratory analysis. This method provides precise SPM concentration information at specific locations. However, this approach demands considerable human resources, material inputs, and financial investments, and its monitoring scope is limited, hampering the acquisition of spatial distribution trends in SPM variations (Li et al., 2006). In light of advancements in remote sensing satellite technology, coupled with the rapid development of image processing technology, remote sensing technology has brought us a novel monitoring approach (Yang et al., 2022; Li et al., 2023; Li et al., 2024b; Saha and Pal, 2024). Compared to traditional methods, remote sensing offers distinct advantages, such as streamlined data acquisition, well-established processing techniques, and the capacity for large-scale monitoring. Consequently, this technique has become a crucial technical tool in the field of offshore water quality monitoring and has been applied extensively in water monitoring (Guang et al., 2007;

González Vilas et al., 2024). Numerous scholars have conducted extensive research on SPM remote sensing quantitative inversion methods employing various remote sensing data sources. Ocean color satellites such as Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) (Ramaswamy et al., 2004), Medium Resolution Imaging Spectrometer (MERIS) (Moore et al., 1997), Moderate Resolution Imaging Spectroradiometer (MODIS) (Miller and Mckee, 2004), and Geostationary Ocean Color Imager (GOCI) (Liu et al., 2013) have successfully analyzed the SPM distribution in coastal waters. Nevertheless, their utility in smaller-scale areas such as estuaries, rivers, and lakes is constrained due to limited spatial resolution (Yan et al., 2024). Several medium-resolution remote sensing satellites designed for land applications have been used to study SPM concentrations, and notable progress has been made. Noteworthy examples include Williamson and Grabau (1974) investigation using measured data and Landsat series data and Wang et al. (2022b) successful mapping of total suspended matter concentrations in major river mouths using Landsat5 TM and Landsat8 OLI images. Additionally, Lu et al. (2019) established a suspended matter inversion model for Donghu Lake using 48 Landsat satellite images spanning from 1973 to 2018, thereby analyzing the long-term trends in total suspended solids concentration.

In comparison to medium-resolution satellites such as the Landsat series, high-resolution satellites have the potential to construct more refined SPM models owing to their superior spatial resolution. High-spectral-resolution satellites such as EO-1 (Yan et al., 2006; Liu et al., 2021), HJ-1A/B (Xiao et al., 2012; Xing et al., 2014; Yu et al., 2020), and Zhuhai 1 (Yin et al., 2021) have all been applied in the field of SPM. However, these high-resolution satellites often fail to simultaneously meet the demands for both high spatial and temporal resolution while missing key spectral bands essential for SPM inversion. Compared to optical sensors, Synthetic Aperture Radar (SAR), Polarimetric SAR, and Unmanned Aerial Vehicle (UAV) LiDAR exhibit higher spatial resolution when acquiring image data (Lu et al., 2024; Tang et al., 2023; Li et al., 2024a). However, monitoring large areas requires substantial human and material resources for these sensors, thus posing challenges to their widespread adoption in practical applications. With a spatial resolution of up to 10 m, a revisit period of 5 days, and a spectral range covering visible light, near-infrared, and shortwave infrared, Sentinel-2 presents promising prospects for

SPM inversion. Pahlevan et al. (2017) employed Sentinel-2 data for the remote sensing inversion of total suspended solids, achieving high-quality inversion products. Li et al. (2021) conducted a comparative analysis of commonly used satellites, including MODIS, Landsat-8, Sentinel-2, and HJ-1B, finding Sentinel-2 to be the most suitable for monitoring total suspended matter in the Yangtze River mainstream. Moreover, Li et al. (2023) used remote sensing image data from Landsat-8, Sentinel-2, and GF-1 to invert suspended solid concentrations in the Three Gorges Reservoir area of the Yangtze River and Changshou Lake and concluded that Sentinel-2 exhibited the highest accuracy. These studies collectively demonstrate that the utilization of Sentinel-2 data can yield a robust model for SPM inversion. Furthermore, a comparative analysis of different remote sensing data sources indicates that Sentinel-2 data excel in SPM inversion, underscoring its considerable advantages and potential for application in SPM inversion. However, further validation of its effectiveness in this specific study area is warranted.

Currently, various methods, including analytical methods (Dekker and Peters, 1993), semianalytical methods (Jiang et al., 2023), and empirical methods (Meng et al., 2011), are employed for SPM remote sensing inversion. The analytical approach relies on bio-optical and radiative transfer models to simulate the absorption and backscattering coefficients in relation to remote sensing data and water quality parameters (Pan and Ma, 2008). This method, while accurate, necessitates measuring the inherent optical properties of water components, resulting in complex algorithms with limited applications (Lee et al., 1994). Semianalytical methods combine spectral characteristics with statistical models to achieve water quality parameter inversion (Koponen et al., 2002). Shun et al. (2019); Jiang et al. (2021), and Jiang et al. (2023) have successfully estimated the suspended sediment concentration in water bodies using semi-analytical methods. However, the construction of these models requires the measurement of multiple optical parameters of water bodies, making the models relatively complex. Empirical methods, on the other hand, base SPM inversion on measured SPM data and remote sensing data (Chen et al., 2015). These methods are simpler than traditional methods and have become the primary approach for SPM inversion. Numerous scholars have successfully utilized empirical methods for SPM remote sensing monitoring in inland and marine waters, yielding favorable results. The established empirical models include single-band models (Chen et al., 2015; Gao et al., 2019), band ratio models (Wirabumi et al., 2021; Zhong et al., 2022), and multiple regression models (Molkov et al., 2019). However, for optically complex regions such as estuaries and nearshore waters, linear models may not be suitable when applied in isolation. The ongoing integration of computer and remote sensing technology has introduced machine learning algorithms as a promising direction for SPM inversion due to their ability to handle complex nonlinearities. Traditional machine learning algorithms, such as Support Vector Regression (SVR), Random Forest (RF), and K-Nearest Neighbor Regression (KNNR), have been widely applied in the field of water quality remote sensing inversion and have been successfully used in estimating SPM concentrations in most cases (Chen et al., 2019; Hu et al., 2020; Kolluru and Surya, 2022; Liu et al., 2021; Wang et al., 2022b). These research findings

demonstrate the significant potential of machine learning algorithms in SPM inversion. In recent years, thanks to the rapid development of computer technology, an increasing number of novel machine learning algorithms have been introduced into the inversion study of SPM concentrations in water bodies (Balasubramanian et al., 2020; Duan et al., 2022; Kolluru and Surya, 2022; Pahlevan et al., 2022). Gradient boosting algorithms, which enhance model accuracy by transforming weak learners into strong learners, encompass various models such as ABR, GBR, XGBR, and LGBM. As demonstrated by studies conducted by Chen et al. (2022); Duan et al. (2022), and Wen et al. (2024), these algorithms exhibit a reduced sensitivity to the quality of training data, particularly when applied to SPM inversion, resulting in favorable prediction outcomes. However, further in-depth research is still needed to determine whether these new gradient boosting algorithms outperform traditional machine learning methods (such as KNNR and RF) in prediction performance and which algorithm is more suitable for SPM inversion. Additionally, numerous scholars conduct research on SPM inversion using machine learning algorithms, their focus primarily lies in areas such as rivers, estuaries, and coastal seas. Despite the fact that scholars such as Sipelgas et al. (2006); Feng et al. (2014), and Song et al. (2018) have utilized traditional empirical regression models to analyze SPM in engineering construction areas and achieved promising results, studies attempting to apply machine learning inversion methods to observe SPM variations in engineering construction, especially in artificial canal construction, are exceedingly scarce. Consequently, how effective are different machine learning algorithms in inverting SPM in coastal engineering construction areas, particularly in the estuary regions of canal construction? Can they efficiently monitor the trends in SPM concentration changes? These questions necessitate further exploration and validation.
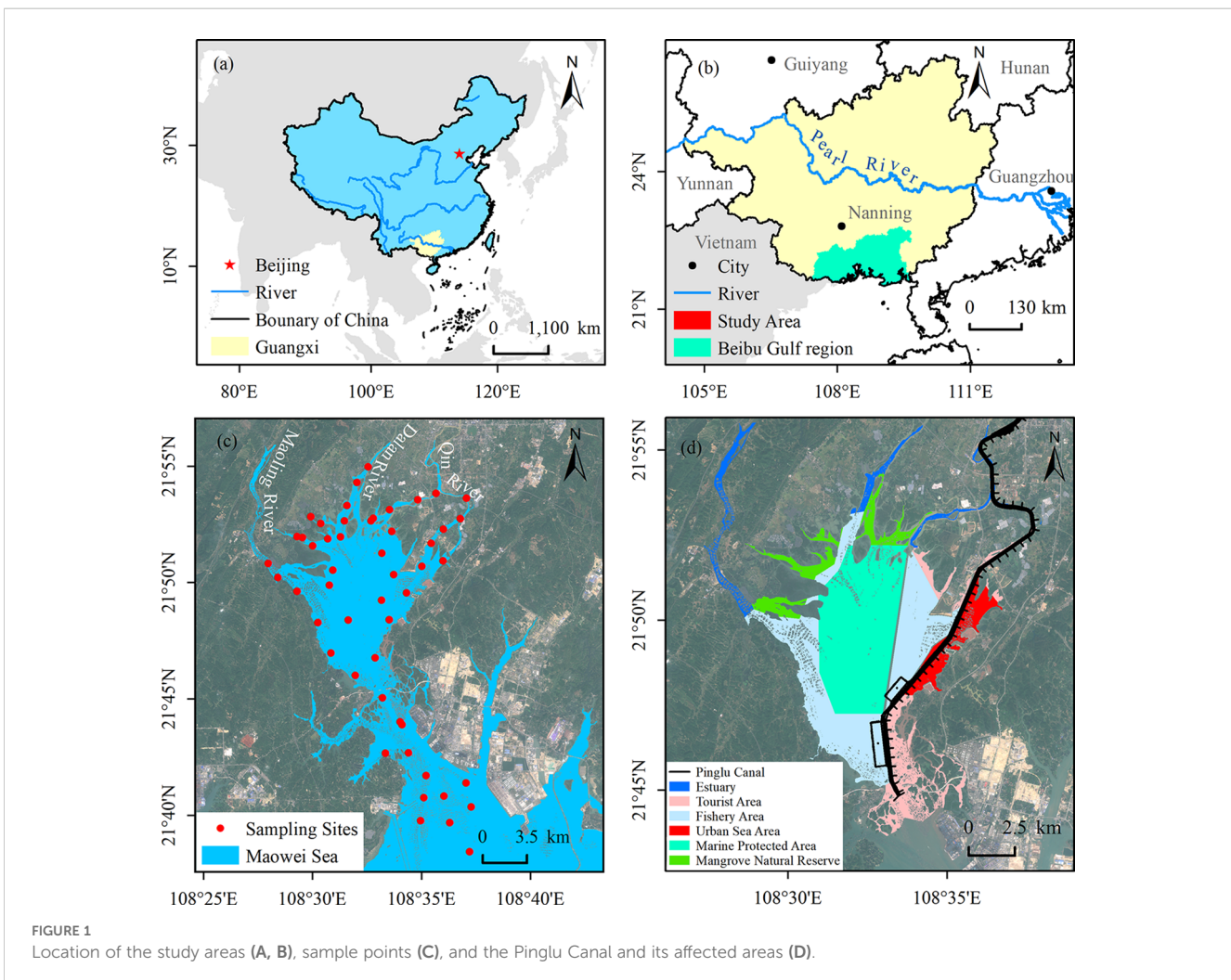
The Maowei Sea's nearshore area is a crucial mangrove conservation region. This area boasts abundant marine resources and biodiversity and serves as a prominent natural habitat for near-river oysters in southern China. It also supports marine aquaculture products such as green crabs, groupers, and sea bass. The Maowei Sea is a multifunctional semienclosed bay that integrates mangrove conservation, aquaculture, tourism, and port activities. In recent years, the Maowei Sea's marine economy has improved, with the construction of the Pinglu Canal serving as a key development. The Pinglu Canal project originates at Pingtang River Estuary in Hengzhou City, Nanning, and extends south along the mainstream of Qin River through Qinzhou City into Maowei Sea in the Beibu Gulf, with a total length of approximately 140 km. It serves as a waterway connecting rivers to the sea. Specifically, the Pinglu Canal involves two segments of the estuary waterway within the study area, namely the urban segment of Qin River and the offshore segment of the estuary. The primary channel construction activity in this area during the project involves dredging, which is carried out using dredgers to excavate and dredge the underwater channel. During construction, activities such as channel dredging, sediment transfer, and marine waste disposal will generate suspended solids. As the concentration of suspended solids increases, it may adversely affect organisms in the surrounding

ecological environment, including mangroves, oysters, and green crabs. Therefore, monitoring of SPM in the study area is of particular importance. However, study on changes in SPM concentration in this region, especially those resulting from the construction of the Pinglu Canal, is currently extremely rare and scarce. However, study focusing on the changes in SPM concentration in this region, particularly those induced by the construction of the Pinglu Canal, remains extremely scarce. Meanwhile, no scholars have yet attempted to obtain the distribution of SPM concentrations in this area using machine learning methods. Based on this, the present study aims to evaluate the accuracy of different algorithms in estimating SPM concentrations using four novel machine learning algorithms and two traditional ones. Based on this evaluation, the optimal machine learning model is selected to estimate SPM concentrations in the study area. This paper has three main research objectives: (1) to explore the applicability of different novel machine learning algorithms and traditional machine learning algorithms in SPM inversion; (2) to analyze the impact of different input variables on the accuracy of the inversion model; and (3) to understand the spatio-temporal trends of SPM concentrations under the influence of the construction of the Pinglu Canal project.

2 Materials and methods

2.1 Study area

The Maowei Sea is located in the northern part of the Beibu Gulf of Guangxi (108°28′E-108°37′E, 21°48′N-21°55′N), with a total area of approximately 135 km^2. It measures approximately 12.6 km in width from east to west and approximately 18 km in depth from north to south (Figure 1). The region is situated in a subtropical monsoon zone influenced by the southeast monsoon and experiences an average annual precipitation of 1658 mm. There are distinct seasonal variations characterized by dry and wet periods: from May to September, more than 70% of the annual rainfall occurs, while the rainfall is low from October to April of the following year (Lu et al., 2022). The study area consists primarily of three rivers that flow into the sea, namely, the Qin River, Dalan River, and Maoling River. Annually, these rivers discharge significant amounts of fresh water and sediment directly into the bay. In 2021, China explicitly proposed a plan to advance the construction of the Pinglu Canal.The Pinglu Canal starts at the mouth of the Pingtang River in Nanning, passes through Qinzhou city along the Qin River, traverses the Maowei Sea, and then enters the Beibu Gulf (Figure 1D).

FIGURE 1
Location of the study areas **(A, B)**, sample points **(C)**, and the Pinglu Canal and its affected areas **(D)**.

2.2 Data sources

2.2.1 *In situ* measurements

In November 4, 2023, a field survey was conducted, with 51 sampling points established at the mouth of the Qin River, Dalan River, Maoling River, Maowei Sea area, and Qinzhou Port, as depicted in Figure 1C. The weather during the field sampling was clear, with calm water surfaces. The collection process of water samples strictly adhered to the Specification for Offshore Environmental Monitoring (HJ 442-2008) and the Technical Specification for Offshore Environmental Monitoring Part 3: Offshore Seawater Quality Monitoring (HJ 442.3-2020). Seawater was collected at a depth of 0.5 m below the surface, then sealed and transported to the laboratory for water quality testing to determine the concentration of Suspended Particulate Matter (SPM). According to the Specification for Marine Monitoring - Part 4: Seawater Analysis (GB 17378.4-2007), the gravimetric method was employed to detect SPM. This involved taking a 2 L water sample and passing it through a 0.45 μm microporous filter membrane. After drying at a constant temperature, the weight of SPM retained on the filter membrane was measured using an analytical balance. The concentration of SPM in the water was then calculated based on the volume of the water sample. Additionally, each sampling point was recorded with longitudinal and latitudinal coordinates using a GPS positioning device.

2.2.2 Remote sensing data

Sentinel-2 is a high-resolution multispectral satellite launched by the European Space Agency (ESA) that consists of 2A and 2B satellites. The Multi-Spectral Instrument (MSI) sensor carried by the Sentinel-2 satellite is capable of capturing images in 13 different spectral bands, covering a range from visible light to shortwave infrared. These images have varying spatial resolutions. Table 1 highlights the main characteristics of Sentinel-2 images. Compared

to Landsat-8, Sentinel-2 has superior spatial and temporal resolutions, making it advantageous for monitoring terrestrial ecology, inland rivers, and coastal ecological environments (Cao et al., 2023). For this study, Sentinel-2 images from the European Aviation Copernicus Data Center (https://dataspace.copernicus.eu/browser/) were utilized, specifically the L2A level products. Compared to the L1C class, Sentinel-2 L2A data have undergone preprocessing, including geometric correction, radiometric correction, and atmospheric correction.

Due to the influence of cloud cover and access time, it was difficult to synchronize the field sampling time with the imagery. Therefore, in this study, we obtained an image from November 1, 2023, which was close to the SPM sampling time and had a cloud cover below 10%, for spatial matching. Additionally, the field sampling was conducted during the dry season in Maowei Sea, while the Pinglu Canal construction officially commenced in 2022. To investigate the impact of the Pinglu Canal construction project on SPM concentrations, we selected eight satellite images from the dry seasons of Maowei Sea between 2021 and 2023, with cloud cover below 10% and ensuring that the study area was not obscured by clouds. These images cover three critical periods: before the commencement of the Pinglu Canal construction, the initial phase of construction, and the full-scale construction. The specific usage of these satellite images is detailed in Table 2. SNAP software was used to resample L2 A-level products, convert Sentinel-2 bands to a resolution of 10 m using the nearest neighbor method, and extract single-band information, based on the positional information of the sampling points, obtained their image spectra (Figure 2).

2.3 Methods

2.3.1 Research technique

This study utilized Sentinel-2 image band data and field-measured SPM data to estimate the SPM concentration in the ocean at the mouth of the Pinglu Canal. Six machine learning methods were employed to determine the spatiotemporal

TABLE 1 Sentinel-2 spectral variables list.

Band	Introduction of Bands	Central Wavelength (μm)	Spatial Resolution (m)
B1	Coastal aerosol	0.443	60
B2	Blue	0.490	10
B3	Green	0.560	10
B4	Red	0.665	10
B5	Vegetation red edge	0.705	20
B6	Vegetation red edge	0.740	20
B7	Vegetation red edge	0.783	20
B8	NIR	0.842	10
B8a	Narrow NIR	0.865	20
B9	Water vapor	0.945	60
B10	SWIR-Cirrus	1.375	60
B11	SWIR	1.610	20
B12	SWIR	2.190	20

TABLE 2 Utilization of sentinel-2 data for analyzing the spatio-temporal distribution of SPM.

Number	Image Level	Date
1	L2A	2021.11.26
2	L2A	2021.12.06
3	L2A	2022.12.31
4	L2A	2023.01.20
5	L2A	2023.01.25
6*	L2A	2023.11.01
7	L2A	2023.11.21
8	L2A	2023.11.26

*Indicating its simultaneous use for matching with ground samples and establishing a model.
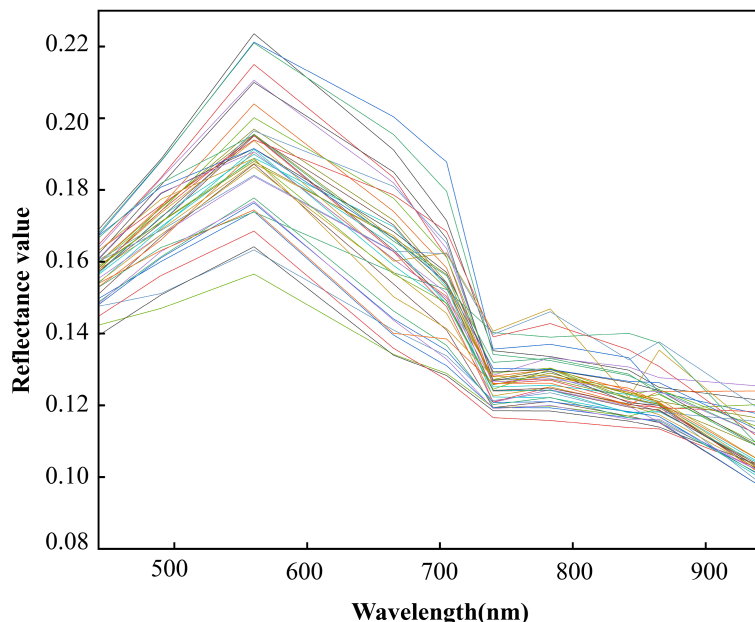
FIGURE 2
presents the remote sensing reflectance (Rrs) spectra of the sampling points obtained using Sentinel-2 imagery.

distribution characteristics of the SPM. The specific technical process of this study is illustrated in Figure 3 and is as follows:

1. Preprocessing of Sentinel-2 data and determination of SPM concentrations via the gravimetric method.
2. Extraction of feature variables from Sentinel-2 data, including single-band variables and water index variables, and variable importance analysis.
3. The SPM test dataset was divided randomly into 80% for training and 20% for testing. Six machine models were evaluated using R^2 and RMSE to select the most suitable machine learning model.
4. An SPM concentration distribution map was created using the best machine learning algorithm model.

2.3.2 Model variable selection

Research has demonstrated that multi-band indices exhibit greater sensitivity than single-band indices (Gao et al., 2020). To improve the accuracy of the inversion model, a water index sensitive to SPM was incorporated in addition to the single band. This study employed two sets of variables and their combinations for analysis: the 12 single bands extracted from Sentinel-2 (excluding the cirrus band B10) and the 7 water body indices calculated using the band operation tool in ENVI software (Table 3). The water body indices included the normalized difference water index (NDWI), modified normalized difference water index (MNDWI), enhanced water index (EWI), revised normalized difference water index (RNDWI), normalized difference turbidity index (NDTI), simple ratio (SR), and simple ratio water color (SRWC).

2.3.3 Methods

To assess the differences in performance among various machine learning algorithms, this study selected six different algorithms, namely, KNNR, ABR, RF, GBR, XGBR, and LGBM, for estimating SPM concentration.

2.3.3.1 K-nearest neighbor regression

The KNNR algorithm is a nonparametric algorithm that operates based on distance measurements. It is renowned for its simplicity in understanding and implementing. The regression principle of KNNR mirrors that of classification (Altman, 1992). However, KNNR has been extended to the field of regression based on the classification foundation. When conducting regression analysis using this algorithm, it becomes essential to locate the k nearest neighbors of a sample. The sample attributes are then assigned by averaging the attributes of those neighbors, with different weights assigned based on the degree of influence of each neighbor on the sample (Zhang and Zhou, 2007). The accuracy of KNNR is dependent on the number of nearest neighbors, k. If k is set too small, the prediction result will be sensitive to noise. Conversely, setting k too large will result in the nearest neighbor list containing an excessive number of nonapproximate samples, leading to accuracy difficulties and significant errors (Jia et al., 2014).

2.3.3.2 AdaBoost regression

ABR is an iterative algorithm that builds strong learners by integrating weak classifiers. During training, the method continuously updates the distribution weights of the samples and iterates to improve the prediction output probability (Lu et al.,
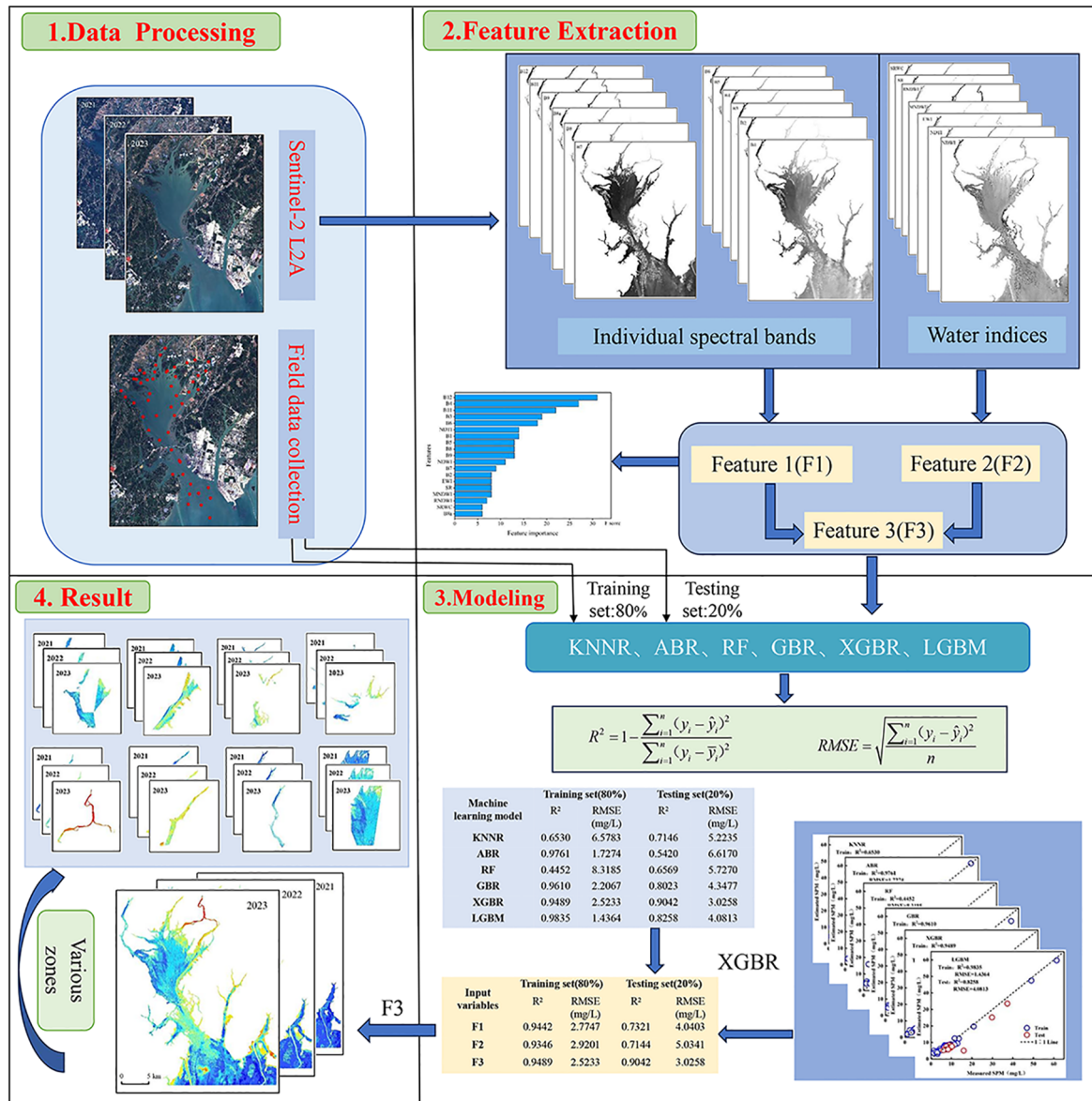
FIGURE 3
Research technical route for estimation of SPM.

2015). The main process involves iteratively selecting a training set from the samples to train the model and assigning weights. In each subsequent training iteration, a greater weight is given to unclassified samples. After each iteration, ABR assigns weights to weak classifiers based on an optimized decision tree classifier—higher weights correspond to better classification performance (Liang et al., 2013).

2.3.3.3 Random forest

The RF algorithm was proposed by Breiman (2001) in 2001. This algorithm employs decision trees as base learners and employs the bagging method to randomly sample them, generating multiple subsets for constructing a regression tree and obtaining the final result (Cheng et al., 2023). RF randomly sample and return the original sample set, reusing some data while others are left unused.

The unused data, known as out-of-bag data, are employed to evaluate the performance of the RF algorithm models. The RF algorithm uses out-of-bag data estimation methods to calculate the error generated by the model, and the importance of each input factor is determined by the estimation error.

2.3.3.4 Gradient boosting regression

The GBR is the regression form of GBDT. GBR initially constructs a regression tree with equal weights based on the original data. This approach estimates different data points during each training session, which effectively reduces overfitting. The model progressively minimizes the loss function and expedites convergence, achieving local or global optimal solutions. Through continuous iteration, the predictions are combined and averaged to obtain the final solution (Friedman, 2001).

TABLE 3 Water indices details.

Indices	Features	Formulas	References
NDWI	Normalized Difference Water Index	(B3−B8)/(B3+B8)	McFeeters (1996)
MNDWI	Modified Normalized Difference Water Index	(B3 − B11)/(B3+ B11)	Xu (2006)
EWI	Enhanced Water Index	(B3-B8-B11)/(B3+B8+B11)	Luo (2015)
RNDWI	Revised Normalized Difference Water Index	(B11 − B4)/(B11+ B4)	Luo (2015)
NDTI	Normalized Difference Turbidity Index	(B4-B3)/(B4+B3)	Lacaux et al. (2007)
SR	Simple Ratio	B4/B8	Birth and McVey (1968)
SRWC	Simple Ratio Water Color	B4/B2	Zarco-Tejada and Ustin (2001)

2.3.3.5 Extreme gradient boosting regression

XGBR is a novel gradient boosting decision tree model that was proposed by Chen and Guestrin (2016) in 2016. It combines the cause-based decision tree (CBDT) and gradient boosting machine (GBM) methods, resulting in enhanced problem-solving capabilities. XGBRs have gained widespread adoption across various industries. Its approach involves utilizing Taylor expansion for more efficient function minimization. Additionally, regularization terms are incorporated to improve the model's generalization ability and prevent overfitting.

2.3.3.6 Light generalized boosted regression

LGBM is an enhanced model based on GBDT proposed by Microsoft in 2017 (Zhou et al., 2019b). This algorithm excels in handling massive amounts of data processing and offers several advantages, including short running time, low memory consumption, and high accuracy. The underlying mechanism of the LGBM algorithm involves the linear combination of m weak regression trees to form a strong regression tree. During the generation process of each tree, a random selection of samples and data features is employed for training, ensuring tree diversity. The algorithm also incorporates leaf segmentation, leading to the generation of more complex trees (Gu et al., 2020). Consequently, LGBM may experience issues of overcomplexity and overfitting during operation. Therefore, it is

crucial to reduce model complexity and prevent overfitting by limiting the tree depth and number of leaves.

2.3.4 Model establishment and evaluation

The remote sensing inversion of SPM concentrations involves three steps: training set construction, model training, and model inversion. In this study, 80% of the samples (40) were utilized to train the model, while 20% of the samples (11) were used for testing purposes. When constructing different machine learning model parameters, initial values are first set according to the characteristics of each model. Subsequently, a grid search method is utilized to optimize the hyperparameters. Through hyperparameter optimization, the optimal parameters for different machine learning algorithms are obtained. Based on these optimal parameters, various machine learning models are constructed. The optimal parameters for the six machine learning models are presented in Table 4. Initially, 19 variables (12 single-band variables and 7 water index variables) were input into six machine learning models for training. The optimal prediction model was subsequently determined based on the coefficient of determination (R^2) and root mean square error (RMSE). To evaluate the impact of different variable characteristics on SPM modeling, different variable combinations were established. Through comparisons of different variable combinations, the optimal input variable was identified to enhance the accuracy of the SPM concentration estimation.

The performance of the model was evaluated using the R^2 coefficient and RMSE. R^2 represents the degree of fit between the predicted and measured values. Higher R^2 values indicate better fit and greater accuracy. On the other hand, the RMSE gauges the deviation between the predicted and measured values, with lower deviations indicating higher model accuracy (Tan et al., 2023). The formula for the RMSE is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (1)$$

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

where y_i is the observed value of the water quality parameter concentration at the sampling point, $\bar{y}_i$ is the average value of the water quality parameter concentration at the sampling point, and $\hat{y}_i$

TABLE 4 Hyperparameters of different machine learning models in this study.

Machine learning model	learning_rate epsilon	min_samples_leaf min_child_weight	Gamma random_state	max_depth max_features	n_estimators n_neighbors
KNNR	NA	NA	NA	NA	5
ABR	0.01	NA	0	NA	50
RF	NA	5	0	15	100
GBR	0.10	1	NA	1	100
XGBR	0.10	3	0.1	5	150
LGBM	0.05	20	NA	2	100

is the predicted value of the water quality parameter concentration at the sampling point.

3 Results

3.1 Modeling results, assessment and comparison

Figure 4 and Table 5 illustrate the accuracy evaluation results of the SPM inversion using the six machine learning algorithm models. The six machine learning models utilized 19 input variables, including 12 single-band variables and 7 water body indices. By comparing the R^2 and RMSE values, it becomes apparent that the XGBR model exhibits the best fit among the six machine learning models, with R^2 values above 0.9042 in both the training and testing stages. The LGBM was the second most common model for fitting

(R^2 = 0.8258, RMSE=4.0813 mg/L). The GBR also demonstrated a good fitting effect (R^2 = 0.8023, RMSE=4.3477 mg/L). Although the ABR model exhibited a high R^2 (R^2 = 0.9761) during the training phase, the R^2 (R^2 = 0.5420) during the testing phase fell short of expectations. Additionally, RF and KNNR performed poorly, with R^2 values of 0.6569 and 0.7146, respectively.

This study employs the XGBR machine learning model to test the accuracy of three different input variables: single-band variables, water index variables, and all variables. As indicated in Table 6, both the R^2 values of the single band variable (F1) and the water index variable (F2) exceeded 0.7144. Among them, the single-band input variable (R^2 = 0.7321) slightly outperforms the water index variable (R^2 = 0.7144), indicating that both the single-band variable and the water index variable can be employed to predict the SPM concentration in this study area. By integrating the single band index and water index (F3) into the XGBR model, the model exhibited the strongest fit and enhanced the prediction accuracy
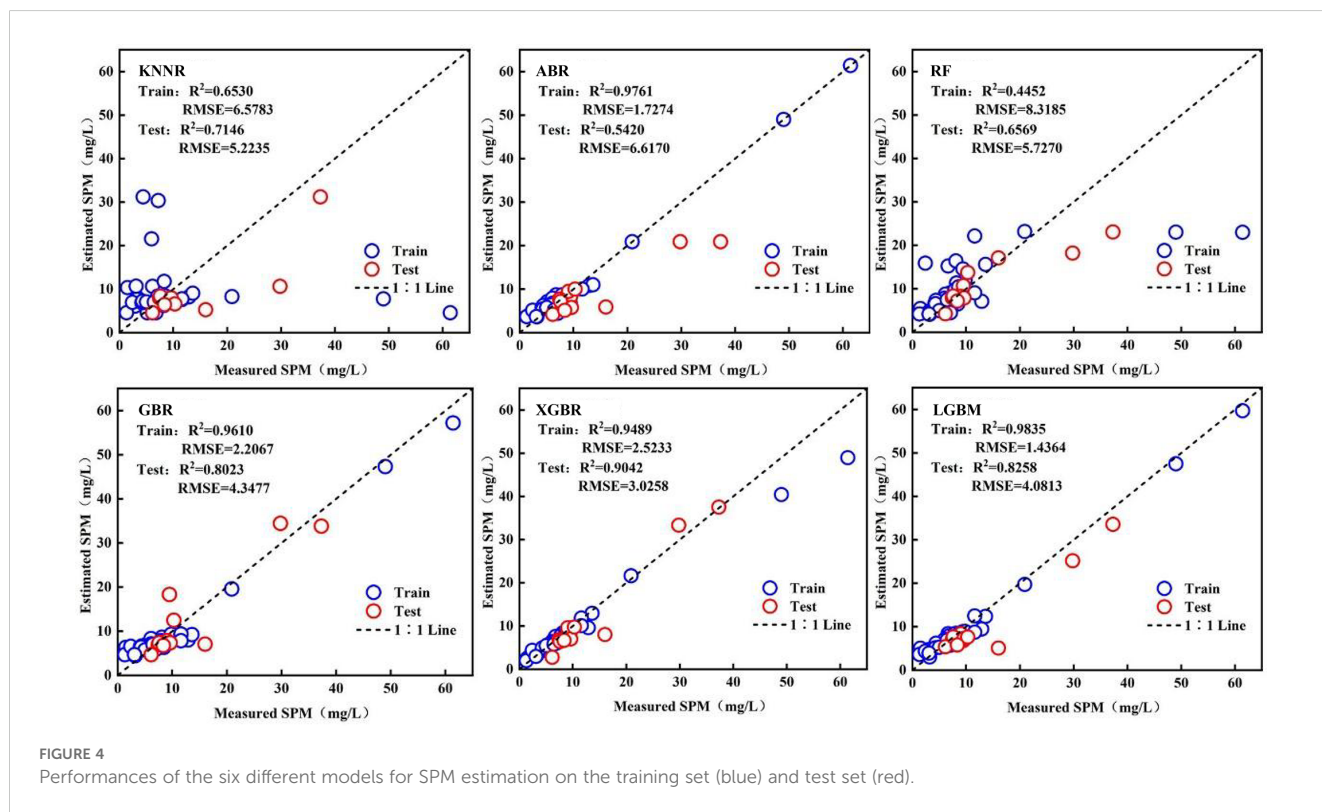

FIGURE 4
Performances of the six different models for SPM estimation on the training set (blue) and test set (red).

TABLE 5 Accuracy comparison results of different models.

Machine learning model	Training set (80%)		Testing set (20%)	
	R^2	RMSE (mg/L)	R^2	RMSE (mg/L)
K-nearest neighbor regression (KNNR)	0.6530	6.5783	0.7146	5.2235
Adaboost regression (ABR)	0.9761	1.7274	0.5420	6.6170
Random forest (RF)	0.4452	8.3185	0.6569	5.7270
Gradient boosting regression (GBR)	0.9610	2.2067	0.8023	4.3477
Extreme gradient boosting regression (XGBR)	0.9489	2.5233	0.9042	3.0258
Light generalized boosted regression (LGBM)	0.9835	1.4364	0.8258	4.0813

TABLE 6 Performance of the XGBR model using different numbers of variables.

Input variables	Training set (80%)		Testing set (20%)	
	R^2	RMSE (mg/L)	R^2	RMSE (mg/L)
Individual spectral bands (F1)	0.9442	2.7747	0.7321	4.0403
Water indices (F2)	0.9346	2.9201	0.7144	5.0341
All indices (F3)	0.9489	2.5233	0.9042	3.0258

of the XGBR model for the SPM concentration in the study area (R^2 = 0.9042, RMSE=3.0258 mg/L).

3.2 Variable importance

The XGBR algorithm possesses a unique capability to discern the significance of input variables, where a higher importance of a variable translates into a more pronounced contribution to the inversion of SPM concentration. In this study, we incorporated all input variables into the XGBR model to thoroughly analyze and evaluate the importance of each feature variable during the SPM concentration inversion process. The features are sorted in descending order according to the intensity of importance, with their contribution represented by the F score. The results, displayed in Figure 5, reveal that the 5 variables with the highest contributions to the SPM concentration inversion model are all single-band variables: B12, B4, B11, B3, and B6, which span visible light and the NIR to the SWIR. The subsequent variable of importance is the NDTI, which denotes the normalized turbidity index. Notably, among the top ten variables in terms of importance, nine are single-band variables, indicating their substantial contribution to the SPM concentration inversion model.

3.3 Spatio-temporal evolution of the SPM concentration

For comparative analysis of the trends in SPM concentrations, we divided the eight images obtained between 2021 and 2023 into three groups based on dry season characteristics. Then, using the SPM inversion model constructed based on the XGBR algorithm, we calculated the pixel averages for Sentinel-2 during the dry seasons of 2021, 2022, and 2023, respectively, and plotted the average SPM concentration distribution map for the dry seasons in the study area from 2021 to 2023 (Figure 6). As shown in Figure 6, the SPM concentration in the study area ranges from 1 to 50 mg/L, with most areas having concentrations below 10 mg/L. The average SPM concentration during the dry season increased from 3.06 mg/L in 2021 to 3.96 mg/L in 2022, and further rose to 4.85 mg/L in 2023, showing a significant upward trend overall. In terms of spatial distribution, the SPM concentration generally decreases gradually from the inner bay to the outer bay. The study results indicate that in 2021, the SPM concentration in most areas was below 3 mg/L, which was relatively low overall; specifically, the SPM concentration in the inner sea of Maowei Sea was slightly higher than that in the outer bay, while higher concentrations were observed in the water bodies near oyster
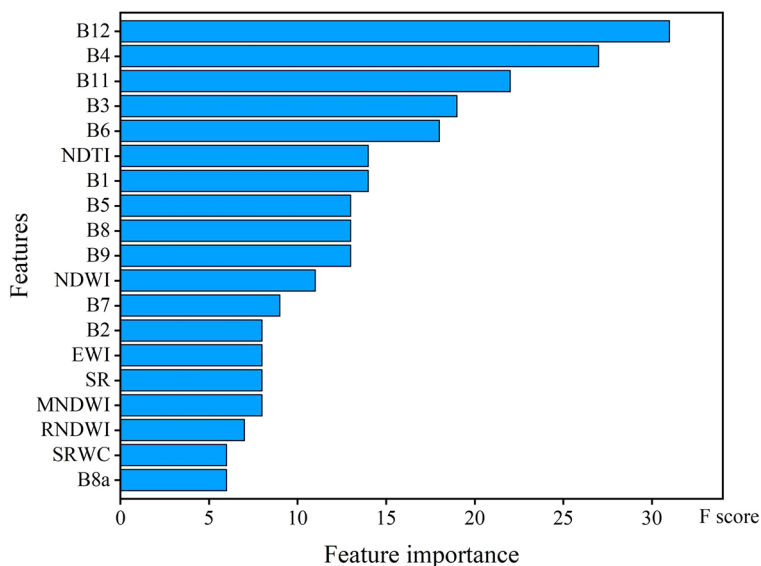

FIGURE 5
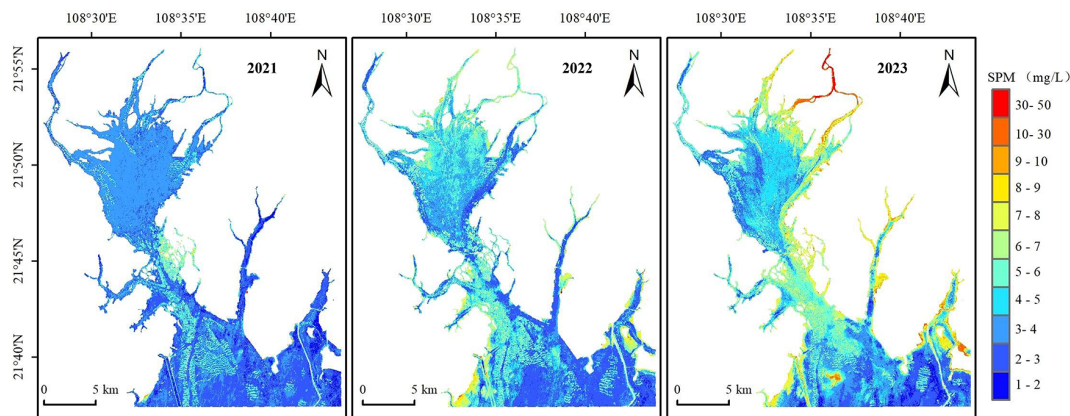Order of selected variable importance of SPM in this study area.

FIGURE 6
Spatio-temporal distribution of SPM concentration.

rafts. By 2022, the general trend of SPM concentration was a decrease from north to south, with higher concentrations observed at the mouths of the Qin River and Dalan River, as well as along the coastal areas of Fangchenggang Industrial Zone, basically maintained within the range of 5 to 6 mg/L. In 2023, the SPM concentration fluctuated significantly, with concentrations in nearshore areas higher than those in offshore areas; particularly, the SPM concentration at the mouth of the Qin River increased significantly, generally exceeding 10 mg/L, and even reaching above 30 mg/L in upstream areas of the river mouth.

To further investigate the changes in SPM concentration across the different zones, we selected three estuary areas of the Maowei Sea estuary and five areas affected by the Pinglu Canal for analysis. These included the Qin River estuary, Dalan River estuary, Maoling River estuary, marine protected area, mangrove natural area, fishing area, tourist area, and urban sea area (Figure 1D). The SPM concentration in each partition was calculated as the average value of all the effective grid data in that partition, as shown in Figure 6. Figure 7 illustrates the trend of the annual SPM concentration changes in each zone. From 2021 to 2023, the SPM concentration in all zones, except for the marine protected area, exhibited an upward trend. Among the subregions, the Qin River Estuary experienced the largest change in SPM concentration from 2022 to 2023, with an increase from 3.55 mg/L in 2021 to 26.71 mg/L in 2023, representing an annual increase of nearly 20 mg/L.

4 Discussion

4.1 Selection and importance analysis of the feature parameters

With the advancement of multispectral remote sensing satellites, numerous scholars have developed models to estimate SPM concentrations based on band information from satellite data sources such as MODIS (Feng et al., 2014), GOCI (Chu et al., 2022), and Landsat-8 (Wang et al., 2018). However, relying solely on input variables of the same category may not capture all necessary information, resulting in suboptimal model performance (Chen et al., 2022). For instance, Maniyar et al. (2023) developed a water body index based on the Sentinel-2 band to estimate the total suspended sediment concentration in the Belize coastal lagoon, achieving an accuracy of $R^2 = 0.82$. Wu et al. (2023) employed 13 bands of Sentinel-2 images as input variables in a machine learning model but achieved less satisfactory results ($R^2 = 0.534$). The accuracy of models utilizing a single category of input variables is lower than that of the results of this study (Table 6). When adding various water indices to the machine learning model on the basis of adding a single band, the uncertainty caused by using one index is avoided, and the predictive ability of the model is improved (Gao et al., 2022). Fang et al. (2019) opted to include exponential indicators in their modeling based on a single band and discovered that the collective involvement of multiple variables enhanced the model's accuracy, confirming our findings.

Based on the importance results of the input variables in this study (Figure 5), it is evident that the single band variable is the most sensitive to the inversion of the SPM concentration. Among the top ten important variables, nine are single-band variables, with only one being the water index variable. The five contributing variables are B12, B4, B11, B3, and B6, indicating that the shortwave infrared reflectance (SWIR), red reflectance (R), green band (G), and near-infrared band (NIR) of the water in this study are sensitive to the SPM concentration. E. Knaeps et al. (2015) established an inversion model through the SWIR band and successfully generated a map of the total suspended matter concentration. Novoa et al. (2017); Din et al. (2017), and Qing et al. (2017) also noted that using the SWIR band can, to some extent, help obtain the distribution characteristics of SPM concentrations. These studies further demonstrated the usefulness of the SWIR band for inverting SPM concentrations, supporting the conclusions of this study. Yin et al. (2022); Zhang et al. (2023), and Yang et al. (2023) used red light, green light, and near-infrared bands to establish SPM inversion models and found that the models had good accuracy, which is consistent with the results of this study. In most cases, the emission peaks in the measured spectra are located in the G, R, and NIR bands and are strongly correlated with SPM. Therefore,
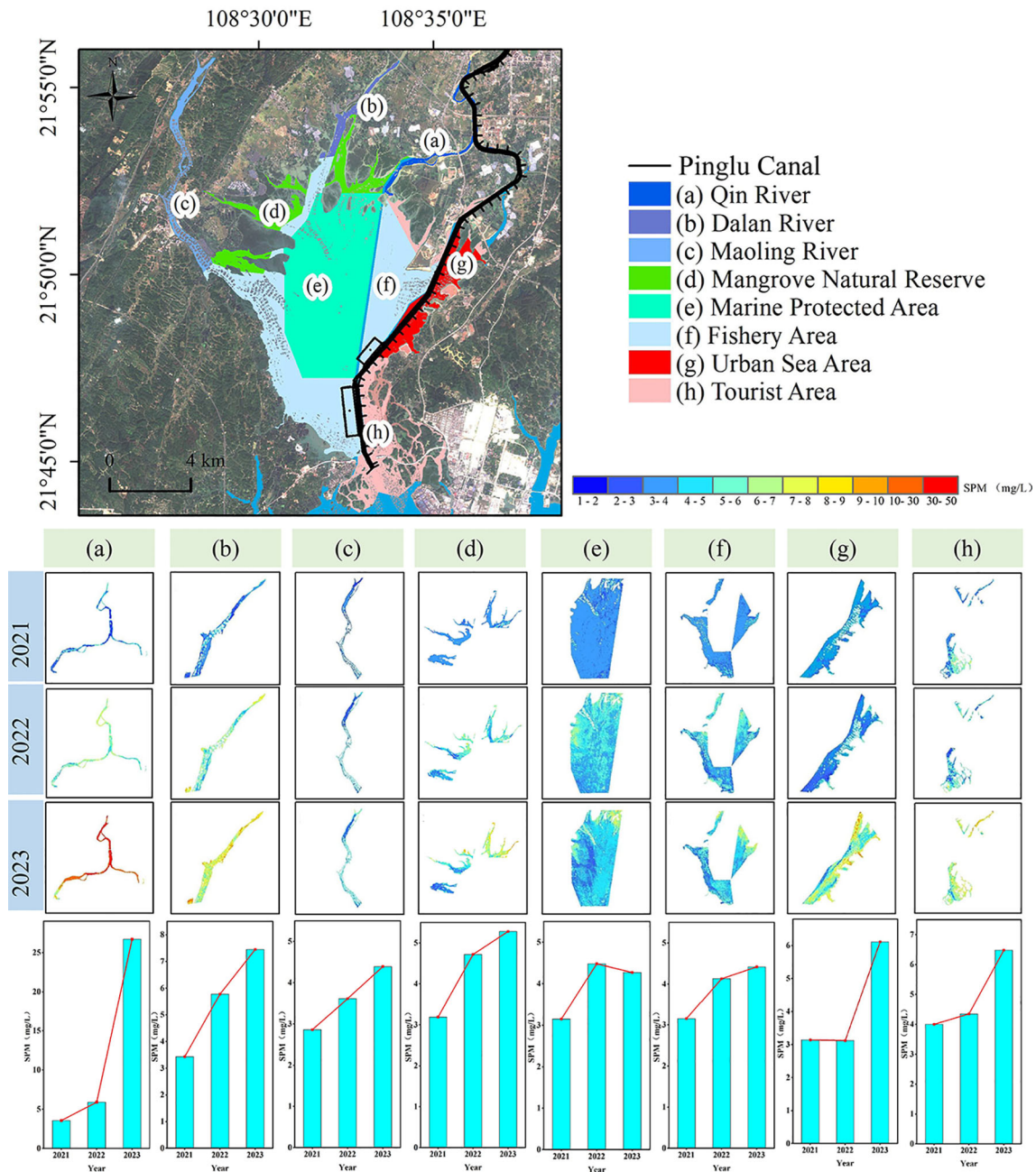
FIGURE 7
Trend chart of SPM concentration changes in various zones from 2021 to 2023.

these bands are more sensitive to changes in SPM concentration and have a good ability to predict SPM concentrations in the study area. Additionally, the NDTI in the water index contributed to the inversion of the SPM concentration (Figure 5). Using the water index as an input variable improved the accuracy of the model (Table 6). NDTI is an index that utilizes the reflectance of R and G. As turbidity increases, the reflectance of R increases, while the reflectance of G decreases, allowing for qualitative determination of turbidity levels (Virtanen et al., 2020). This process significantly contributed to the model inversion, possibly due to the close correlation between turbidity and SPM, with higher turbidity indicating higher SPM concentrations (Chen et al., 2023).

Sankaran et al. (2023) used the spectral bands of Sentinel-2 data and combined them with water indices such as the NDTI to confirm the ability of these indices to plot SPM concentrations in bay waters.

4.2 Selection of the machine learning model

Currently, although many studies utilize linear models to forecast SPM concentrations (Chen et al., 2015; Gao et al., 2019), linear models are ineffective at predicting SPM concentrations in

areas heavily impacted by human activities, such as estuaries, rivers, and lakes (Chen et al., 2022). With the advancement of computers, machine learning algorithms have been employed in the inversion of SPM concentrations in lakes and nearshore waters due to their ability to handle complex nonlinearities (Liu et al., 2021; Wang et al., 2022b). Different machine learning algorithms and input variables have varying effects on SPM concentration inversion (Maniyar et al., 2023). In this study, we employ six machine learning methods to estimate SPM concentrations based on single-band and water index variables. Based on the inversion results (as depicted in Figure 4 and Table 5), the XGBR model demonstrated the best fitting effect during inversion (R^2 = 0.9042, RMSE=2.0258 mg/L). The conclusion that XGBR outperforms other algorithms, such as RF, support vector machines, XGBR, and deep neural networks, in the global inversion of coastal and inland water SPM, as reported by Cao et al. (2022), is consistent with our findings. Additionally, from Figure 4 and Table 5, it can be observed that the XGBR model exhibits the least change in accuracy when evaluating model performance using the test set, suggesting its strong robustness. The XGBR model is an enhancement of the traditional GBDT algorithm, and it has undergone extensive optimization during the specific training process, resulting in higher accuracy, improved flexibility, and superior performance (Chen and Guestrin, 2016). The R^2 of this model surpasses that of Ding et al. (2022), who estimated the accuracy of suspended matter inversion in the Maowei Sea and its estuary using a cubic polynomial regression algorithm (R^2 = 0.88). This indicates that the XGBR model is more suitable for total suspended matter inversion in our study area. According to Table 6, compared to a single variable combination, combining single-band and exponential-band variables can enhance the accuracy of SPM remote sensing inversion and yield superior results, which aligns with the research findings of Saberioon et al. (2020) and Fang et al. (2019) The results of this study demonstrate the high performance of XGBR in estimating SPM concentration, and this model holds significant potential for application, providing a new research direction for SPM concentration estimation.

4.3 Analysis of the spatiotemporal evolution of the SPM concentration

Numerous studies have demonstrated that the SPM concentration is influenced by factors such as runoff flow (Gong et al., 2017), tidal currents (Zhang et al., 2015), and human activities (Zhou et al., 2021). In this study, the XGBR model was used to analyze the data, which revealed a consistent year-to-year increase in SPM concentration, particularly near the mouth of the Qin River. Moreover, compared with those of other areas, urban sea area and tourist area intersected by the Pinglu Canal Channel exhibited greater increases in SPM concentrations. These findings suggest that the construction of the Pinglu Canal may have contributed to the increase in SPM concentration. The Pinglu Canal commenced full-scale construction in August 2023, with dredging being the primary construction method in the urban segment of Qin River and the offshore segment of the estuary. During the construction period, grab

dredgers, chain dredgers, and cutter-suction dredgers were mainly used for dredging and debris removal operations. Studies have shown that activities such as channel dredging, sediment transfer, and hydraulic fill overflow can increase the concentration of SPM in surrounding water bodies (Xu et al., 2010). For the study area, the volume of dredging, excavation, and hydraulic fill operations involved in the construction of the Pinglu Canal is expected to exceed 3000000 m^3. This large-scale engineering activity has led to a significant increase in SPM concentrations in the study area, especially in the three key areas crossed by the Pinglu Canal—the Qin River estuary, tourist area, and urban sea area. Mangrove wetland ecosystems are among the sensitive areas affected by the construction activities of the Pinglu Canal. As clearly shown in Figures 6, 7, the SPM concentrations in the mangrove nature reserve, particularly in the mangrove growth areas around the estuaries of the Qin River and the Dalan River in Maowei Sea, have also increased. Fortunately, the concentration values in these areas basically remain below 9 mg/L. This phenomenon indicates that effective control of SPM diffusion was implemented during the construction of the Pinglu Canal channel, resulting in relatively minor negative impacts on the mangrove ecosystem. The Chinese government attaches great importance to the protection of coastal ecological environments. The construction of the Pinglu Canal project must comply with relevant environmental protection laws, regulations, and policy requirements, and coordinate with relevant plans such as watershed ecological protection plans, channel plans, or overall port plans. According to the requirements of the "Approval Principles for Environmental Impact Assessment Documents of Waterway Construction Projects," if activities such as dredging, hydraulic fill, and mud dumping have adverse effects on water quality, specific control measures for SPM must be clearly proposed for each construction link. During the construction process of this project, anti-pollution curtains were installed to effectively control the diffusion of SPM. Additionally, studies by Capello et al. (2013); He et al. (2013), and Chen et al. (2014) have shown that the impact of dredging-induced SPM diffusion is limited in range and diminishes with distance. The zone experiencing a significant increase in SPM concentration in this study corresponds to the Qinzhou estuary where Pinglu Canal dredging is taking place. As the distance from the construction site increases, the magnitude of the increase in SPM concentration decreases, ultimately leading to a negligible impact, which aligns with the findings of previous researchers.

4.4 Uncertainty of the model evaluation results and future research directions

The results obtained from the application of the XGBR model for estimating SPM concentrations reveal certain spatial distribution characteristics. Specifically, the SPM concentration exhibited a decreasing trend as one moved from the estuary to the inner sea of the Maowei Sea and finally to the outer bay. These findings differ slightly from the results obtained by Ding et al. (2022), who used optical images and microwave data to analyze the Maowei Sea. One potential reason for the discrepancy between these two sets of results could be the recent construction of the Pinglu Canal. This

construction has resulted in an increase in the SPM concentration in the Qin River and Dalan River. Furthermore, based on the XGBR inversion model, this study revealed that the SPM concentration along the coast of the Maowei Sea is mostly less than 10 mg/L, which is lower than the SPM concentration in the coastal waters of Guangxi reported by Li et al. (2020) via empirical models (20 mg/L). It is important to note that the differences in remote sensing inversion methods, sampling periods and collection locations could also contribute to the variations observed in the results. Hence, it is evident that the evaluation results of the model are influenced not only by machine algorithms but also by external factors. Figure 4A demonstrates that the XGBR model underestimates the prediction of higher concentrations of SPM. This suggests that the applicability of the XGBR model for inverting SPM at different concentrations may vary, in accordance with the research findings of Maniyar et al. (2023). Moreover, the prediction accuracy of the XGBR model in this study (R^2 = 0.9042) is lower than that of Si (2022) research, which utilized unmanned aerial vehicle hyperspectral data in combination with the XGBR machine algorithm (R^2 = 0.93). This indicates that different spatial resolutions can impact the accuracy of the XGBR model. Additionally, the limited sampling point data used in this study, along with the potential inconsistency between the sampling time and image acquisition time, undermines the accuracy of SPM estimation. Future research should focus on improving the following areas: (1) increasing the number of SPM samples to provide more authentic and reliable data; (2) collecting *in situ* spectral data to reduce uncertainty resulting from inconsistent sampling times and image acquisition times; (3) integrating Sentinel-2 data with higher resolution image data to minimize errors attributed to traditional spatial resampling methods; and (4) fine-tuning and optimizing the XGBR algorithm to enhance prediction accuracy.

5 Conclusions

This study examined the accuracy of six machine learning algorithms, namely, KNNR, ABR, RF, GBR, XGBR, and LGBM, in estimating the concentration of SPM in the Pinglu Canal estuary ocean. This study also analyzed the spatiotemporal distribution characteristics of the SPM concentration.

1. Among the six machine learning models, the XGBR model demonstrated the highest accuracy in estimating SPM concentration, with an R^2 value of 0.9042 and an RMSE of 3.0258 mg/L. The LGBM (R^2 = 0.8258, RMSE=4.0813 mg/L) and the GBR (R^2 = 0.8023, RMSE=4.3477 mg/L) models also exhibited reasonably good accuracy. However, the ABR, RF, and KNNR models perform poorly in terms of fitting effects.
2. Furthermore, incorporating both F1 and F2 as input variables into the XGBR model significantly improved its fitting ability. The R^2 value increases from 0.7321 for F1 and 0.7144 for F2 to 0.9042 when utilizing both F1 and F2 (F3).
3. Regarding the contribution of the input variables in the XGBR model, F1 demonstrated a relatively high importance.

The SWIR and R variables rank highest in terms of importance. Conversely, the contribution of F2 was relatively small, with the NDTI being the most influential.

4. The inversion results of the XGBR model reveal that the SPM concentration in the study area exhibited an upward trend from 2021 to 2023. The spatial distribution pattern indicates a decrease in concentration when the estuary moves to the inner sea of the Maowei Sea to the outer bay. Specifically, the SPM concentration increased from 3.06 mg/L to 4.85 mg/L over the course of 2021 to 2023. Notably, the mouth of the Qin River experienced the largest change, with an increase of nearly 20 mg/L.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

JM: Writing – original draft, Visualization, Validation, Formal analysis, Data curation. YT: Writing – review & editing, Resources, Methodology, Investigation, Conceptualization. JW: Writing – review & editing, Visualization, Investigation, Data curation. QZ: Writing – review & editing, Software, Methodology. YZ: Writing – review & editing, Investigation, Data curation. JT: Writing – review & editing, Investigation. JL: Writing – review & editing, Investigation.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric re-gression. *Am. Stat.* 46, 175–185. doi: 10.1080/00031305.1992.10475879

Balasubramanian, S. V., Pahlevan, N., Smith, B., Binding, C., Schalles, J., Loisel, H., et al. (2020). Robust algorithm for estimating total suspended solids (TSS) in inland and nearshore coastal waters. *Remote Sens. Environ.* 246, 111768. doi: 10.1016/j.rse.2020.111768

Birth, G. S., and McVey, G. R. (1968). Measuring the color of growing turf with a reflectance spectrophotometer. *Agron. J.* 60, 640–643. doi: 10.2134/agronj1968.00021962006000060016x

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Cao, X., Zhang, J., Meng, H. B., Lai, Y. Q., and Xu, M. F. (2023). Remote sensing inversion of water quality parameters in the Yellow River Delta. *Ecol. Indic.* 155, 110914. doi: 10.1016/j.ecolind.2023.110914

Cao, Z. G., Shen, M., Kutser, T., Liu, M., Qi, T. C., Ma, J. G., et al. (2022). What water color parameters could be mapped using MODIS land reflectance products: A global evaluation over coastal and inland waters. *Earth Sci. Rev.* 232, 104154. doi: 10.1016/j.earscirev.2022.104154

Capello, M., Cutroneo, L., Ferranti, M. P., Castellano, M., Povero, P., Budillon, G., et al. (2013). Mathematical simulation of the suspended solids diffusion during dredging operations on the continental shelf off the coast of Lazio (Central Tyrrhenian Sea, Italy). *Ocean Eng.* 72, 140–148. doi: 10.1016/j.oceaneng.2013.06.008

Chen, D. D., Chen, Y. Z., Feng, X. F., and Wu, S. (2022). Retrieving suspended matter concentration in rivers based on hyperparameter optimized CatBoost algorithm. *J. Geo-Inf. Sci.* 24, 780–791. doi: 10.12082/dqxkx.2022.210446

Chen, J., Jiang, C. B., Zhang, S. H., and Hu, B. A. (2014). Suspended solids diffusion induced by bucket dredger. *J. Transport Sci. Eng.* 30, 55–59. doi: 10.3969/j.issn.1674-599X.2014.02.010

Chen, S., Hu, C., Barnes, B. B., Xie, Y., Lin, G., and Qiu, Z. (2019). Improving ocean color data coverage through machine learning. *Remote Sens. Environ.* 222, 286–302. doi: 10.1016/j.rse.2018.12.023

Chen, S. S., Han, L. S., Chen, X. Z., Li, D., Sun, L., and Li, Y. (2015). Estimating wide range total suspended solids concentrations from MODIS 250-m imageries: An improved method. *ISPRS J. Phtogramm. Remote Sens.* 99, 58–69. doi: 10.1016/j.isprsjprs.2014.10.006

Chen, T., and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* San Francisco, CA, USA, 13-17, August. 785–794. doi: 10.1145/2939672.2939785

Chen, X. J., Liu, H. X., Zhang, D. Q., Yuan, S. M., Lin, C., and Pang, C. G. (2023). Remote sensing inversion model of seawater particle size based on multispectral data. *Mar. Sci.* 47, 54–68. doi: 10.11759/hykx20220822002

Cheng, Q., Zhou, W., Zhang, J., Shi, L., Xie, Y., and Li, X. (2019). Spatial variations of arsenic and heavy metal pollutants before and after the water-sediment regulation in the wetland sediments of the Yellow River Estuary, China. *Mar. pollut. Bull.* 145, 138–147. doi: 10.1016/j.marpolbul.2019.05.032

Cheng, W. Q., Yuan, D. B., Xiong, P., Liu, H. L., Wang, Y., Xu, Q., et al. (2023). Construction and evaluation of city water quality index prediction model based on multiple machine learning algorithms. *Acta Sci. Circumstantiae.* 43, 144–152. doi: 10.13671/j.hjkxxb.2023.0182

Chu, Y. H., Wu, W. J., Li, P., and Chen, S. L. (2022). Temporal and spatial dynamics of suspended sediment and its driving mechanism in the Yellow River Estuary. *Haiyang Xuebao.* 44, 150–1635. doi: 10.12284/hyxb2022059

Dekker, A. G., and Peters, S. W. M. (1993). The use of the Thematic Mapper for the analysis of eutrophic lakes: a case study in the Netherlands. *Int. J. Remote. sens.* 14, 799–822. doi: 10.1080/01431169308904379

Din, E. S. E., Zhang, Y., and Suliman, A. (2017). Mapping concentrations of surface water quality parameters using a novel remote sensing and artificial intelligence framework. *Int. J. Remote Sens.* 38, 1023–1042. doi: 10.1080/01431161.2016.1275056

Ding, B., Li, W., and Hu, K. (2022). Inversion of total suspended matter concentration in Maowei Sea and its estuary, Southwest China using contemporaneous optical data and GF SAR data. *Remote Sens. Natural Resour.* 34, 10–17. doi: 10.6046/zrzyyg.2021094

Duan, P., Zhang, F., Liu, C., Tan, M. L., Shi, J., Wang, W., et al. (2022). High-resolution planetscope imagery and machine learning for estimating suspended particulate matter in the Ebinur Lake, Xinjiang, China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens* 16, 1019–1032. doi: 10.1109/jstars.2022.3233113

Fang, X. R., Wen, Z. F., Chen, J. L., Wu, S. J., Huang, Y. Y., and Ma, M. H. (2019). Remote sensing estimation of suspended sediment concentration based on Random Forest regression model. *Natl. Remote Sens. Bulletin.* 23, 756–772. doi: 10.11834/jrs.20197498

Feng, L., Hu, C., and Chen, X. (2014). Influence of the Three Gorges Dam on total suspended matters in the Yangtze Estuary and its adjacent coastal waters: Observations from MODIS. *Remote Sens. Environ.* 140, 779–788. doi: 10.1016/j.rse.2013.10.002

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Ann.Stat* 29, 1189–1232. doi: 10.1214/AOS%2F1013203451

Gao, Y. N., Gao, J. F., Yin, H. B., Liu, C. S., Xia, T., Wang, J., et al. (2022). Total phosphorus and nitrogen dynamics and influencing factors in Dongting Lake using Landsat data. *Remote Sens.* 14, 5648–5648. doi: 10.3390/RS14225648

Gao, L., Wang, X. F., Johnson, B. A., Tian, Q. J., Wang, Y., Verrelst, J., et al. (2020). Remote sensing algorithms for estimation of fractional vegetation cover using pure vegetation index values: A review. *ISPRS J. Photogramm. Remote Sens.* 159, 364–377. doi: 10.1016/j.isprsjprs.2019.11.018

Gao, C., Xu, J., Gao, D., Wang, L. L., and Wang, Y. Q. (2019). Retrieval of concentration of total suspended matter from GF- 1 satellite and field measured spectral data during flood period in Poyang Lake. *Remote Sens. Natural Resour.* 31, 101–109. doi: 10.6046/gtzyyg.2019.01.14

Gong, S. B., Gao, A. G., Lin, J. J., Zhu, X. X., Zhang, Y. P., and Hou, Y. T. (2017). Temporal-spatial distribution and its influencing factors of suspended particulate matters in Minjiang Lower reaches and estuary. *J. Earth Sci. Environ.* 39, 826–836. doi: 10.3969/j.issn.1672-6561.2017.06.012

González Vilas, L., Brando, V. E., Di Cicco, A., Colella, S., D'Alimonte, D., Kajiyama, T., et al. (2024). Assessment of ocean color atmospheric correction methods and development of a regional ocean color operational dataset for the Baltic Sea based on Sentinel-3 OLCI. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1256990

Gu, Z. J., Wu, D., Zhao, C. D., and Zhou, J. X. (2020). Resampling and boosting techniques for balanced traffic classification. *Comput. Eng. Applications.* 56, 86–91. doi: 10.3778/j.issn.1002-8331.1811-0323

Guang, J., Wei, Y. C., Huang, J. D., Li, Y. M., Wen, J. G., and Guo, J. P. (2007). Study on seasonal remote sensing estimation model of suspended solids in Taihu lake. *J. Lake Sci.* 03, 241–249. doi: 10.3321/j.issn:1003-5427.2007.03.003

He, D. H., He, Q. Y., Wu, G. R., Jiang, Z. S., and Hong, Y. D. (2013). Probing on the dispersion character of dredged material after dumping in sea area around Cangnan. *Ocean Engineering.* 31, 101–106. doi: 10.16483/j.issn.1005-9865.2013.03.015

Hu, C., Lian, F., and Qi, G. (2020). A machine learning approach to estimate surface chlorophyll a concentrations in global oceans from satellite measurements. *IEEE Trans. Geosci. Remote Sens.* 59, 4590–4607. doi: 10.1109/tgrs.2020.3016473

Ji, H., Chen, S., Pan, S., Xu, C., Jiang, C., and Fan, Y. (2018). Morphological variability of the active Yellow River mouth under the new regime of riverine delivery. *J. Hydrol.* 564, 329–341. doi: 10.1016/j.jhydrol.2018.07.014

Jia, Q. N., Ma, L., and He, J. F. (2014). Research on twice supervised learning algorithm applied for clinical survival time prediction. *J. Front. Comput. Sci. Technol.* 8, 1391–1399. doi: 10.3778/j.issn.1673-9418.1406019

Jiang, D., Matsushita, B., Pahlevan, N., Gurlin, D., Fichot, C. G., Harringmeyer, J., et al. (2023). Estimating the concentration of total suspended solids in inland and coastal waters from Sentinel-2 MSI: A semi-analytical approach. *ISPRS J. Photogramm. Remote Sens.* 204, 362–377. doi: 10.1016/j.isprsjprs.2023.09.020

Jiang, D. L., Matsushita, B., Pahlevan, N., Gurlin, D., Lehmann, M. K., Fichot, C. G., et al. (2021). Remotely estimating total suspended solids concen-tration in clear to extremely turbid waters using a novel semi-analytical method. *Remote Sens. Environ.* 258, 112386. doi: 10.1016/j.rse.2021.112386

Knaeps, E., Ruddick, K. G., Doxaran, D., Dogliotti, A. I., NeChad, B., Raymaekers, D., et al. (2015). A SWIR based algorithm to retrieve total suspended matter in extremely turbid waters. *Remote Sens. Environ.* 168, 66–79. doi: 10.1016/j.rse.2015.06.022

Kolluru, S., and Surya, P. T. (2022). Modeling ocean surface chlorophyll-a concentration from ocean color remote sensing reflectance in global waters using machine learning. *Sci. Total Environ.* 844, 157191. doi: 10.1016/j.scitotenv.2022.157191

Koponen, S., Pulliainen, J., Kalliob, K., and Hallikainen, M. (2002). Lake water quality classification with airborne hyperspectral spectrometer and simulated MERIS data. *Remote Sens. Environ.* 79, 51–59. doi: 10.1016/S0034-4257(01)00238-3

Kratzer, S., Kyryliuk, D., and Brockmann, C. (2020). Inorganic suspended matter as an indicator of terrestrial influence in Baltic Sea coastal areas — algorithm development and validation, and ecological relevance. *Remote Sens. Environ.* 237, 111609. doi: 10.1016/j.rse.2019.111609

Lacaux, J., Tourre, Y., Vignolles, C., Ndione, J., and Lafaye, M. (2007). Classification of ponds from high-spatial resolution remote sensing: application to Rift Valley Fever epidemics in Senegal. *Remote Sens. Environ.* 106, 66–74. doi: 10.1016/j.rse.2006.07.012

Lee, Z. P., Carder, K. L., Hawes, S. H., Steward, R. G., Peacock, T. G., and Davis, C. O. (1994). A model for interpretation of hyperspectral remote-sensing reflectance. *Appl. Opt.* 33, 5721–5732. doi: 10.1364/ao.33.005721

Li, H., Fan, L., Xu, W. X., Wang, L. H., Li, J. H., and Cui, L. L. (2023). Remote sensing monitoring of suspended solids concentration in the Three Gorges Reservoir Area based on Multi-source satellite data. *Resour. Environ. Yangtze Basin* 32, 611–625. doi: 10.11870/cjlyzyyhj202303015

Li, Y., Guo, Y. L., Cheng, C. M., Zhang, Y. B., Hu, Y. D., Xia, Z., et al. (2019). Remote estimation of total suspended matter concentration in the Hangzhou Bay based on OLCI and its water color product applicability analysis. *Acta Oceanol. Sin.* 41, 156–169. doi: 10.3969/j.issn.0253–4193.2019.09.015

Li, J. H., Huang, C. C., Cha, Y., Wang, C., Shang, N. N., and Hao, W. Y. (2021). Spatial variation characteristics and remote sensing retrieval of total suspended matter in surface water of the Yangtze River. *Environ. Sci.* 42, 5239–5249. doi: 10.13227/j.hjkx.202103245

Li, L., Lv, M., Jia, Z., Jin, Q., Liu, M., Chen, L., et al. (2023). An effective infrared and visible image fusion approach via rolling guidance filtering and gradient saliency map. *Remote Sens.* 15, 2486. doi: 10.3390/rs15102486

Li, L., Ma, H., Zhang, X., Zhao, X., Lv, M., and Jia, Z. (2024a). Synthetic aperture radar image change detection based on principal component analysis and two-level clustering. *Remote Sens.* 16, 1861. doi: 10.3390/rs16111861

Li, J., Qiao, L. L., Le, D. C., Xue, W. J., Yang, H. D., Wang, Y. Z., et al. (2020). Surficial distribution of suspended sediment in Beibu Gulf of the South China. *Sea. Mar. Geol. Quat. Geol.* 40, 10–18. doi: 10.16562/j.cnki.0256-1492.2019021301

Li, L., Shi, Y., Lv, M., Jia, Z., Liu, M., Zhao, X., et al. (2024b). Infrared and visible image fusion via sparse representation and guided filtering in laplacian pyramid domain. *Remote Sens.* 16, 3804. doi: 10.3390/rs16203804

Li, H. L., Zhang, Y., and Jiang, J. (2006). Study on the inversion model for the suspended sediment consentration in remote sensing technology. *Adv. Water Sci.* 17, 242–245. doi: 10.14042/j.cnki.32.1309.2006.02.015

Liang, D., Guan, Q. S., Huang, W. J., Huang, L. S., and Yang, G. J. (2013). Remote sensing inversion of leaf area index based on support vector machine regression in winter wheat. *Trans. Chin. Soc Agric. Eng.* 29, 117–123. doi: 10.3969/j.issn.1002-6819.2013.07.015

Liu, X., Liu, D., Wang, Y., Shi, Y., Wang, Y., and Sun, X. (2019). Temporal and spatial variations and impact factors of nutrients in Bohai Bay, China. *Mar. pollut. Bull.* 140, 549–562. doi: 10.1016/j.marpolbul.2019.02.011

Liu, M., Shen, F., Ge, J. Z., and Kong, Y. Z. (2013). Diurnal variation of suspended sediment concentration in Hangzhou Bay from geostationary satellite observation and its hydrodynamic analysis. *J. Sediment Res.* 1, 7–13. doi: 10.16239/j.cnki.0468155x.2013.01.003

Liu, X. Y., Zhang, Z., Jiang, T., Li, X. H., and Li, Y. 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.* 17, 5957376. doi: 10.1155/2021/5957376

Lu, J. N., Hu, H. P., and Bai, Y. P. (2015). Generalized radial basis function neural network based on an im-proved dynamic particle swarm optimization and AdaBoost algorithm. *Neurocomputing* 152, 305–315. doi: 10.1016/j.neucom.2014.10.065

Lu, D. S., Li, J., and Filippi, A. (2019). Analysis of total suspended solids concentration in water bodies of East Lake based on long time series Landsat imagery. *J. Wuhan Univ.* 52, 854–861. doi: 10.14188/j.1671-8844.2019-10-002

Lu, L., Luo, J., Xin, Y., Xu, Y., Sun, Z., Duan, H., et al. (2024). A novel strategy for estimating biomass of submerged aquatic vegetation in lake integrating UAV and Sentinel data. *Sci. Total Environ.* 912, 169404. doi: 10.1016/j.scitotenv.2023.169404

Lu, D. L., Zhang, D., Zhu, W. J., Dan, S. F., Yang, B., Kang, Z. J., et al. (2022). Sources and long-term variation characteristics of dissolved nutrients in Maowei Sea, Beibu Gulf, China. *J. Hydrol.* 615, 128576. doi: 10.1016/j.jhydrol.2022.128576

Luo, C. L. (2015). Comparative study on extracting water area of Abihu Lake based on water body index. *Sci. Technol. Innovation Herald.* 12, 34–35. doi: 10.3969/j.issn.1674-098X.2015.24.015

Ma, M., Porz, L., Schrum, C., and Zhang, W. (2024). Physical mechanisms, dynamics and interconnections of multiple estuarine turbidity maximum in the Pearl River estuary. *Front. Mar. Sci.* 11. doi: 10.3389/fmars.2024.1385382

Maniyar, C. B., Rudresh, M., Callejas, I. A., Osborn, K., Lee, C. M., Jay, J., et al. (2023). Spatio-temporal dynamics of total suspended sediments in the Belize Coastal Lagoon. *Remote Sens.* 15, 5625. doi: 10.3390/rs15235625

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* 17, 1425–1432. doi: 10.1080/01431169608948714

Meng, L., Qu, F. Z., and Bi, X. L. (2011). A review of retrieval algorithms for suspended sediment concentration by remote sensing. *J. Zhejiang Ocean Univ. Nat. Sci.* 30, 443–449. doi: 10.3969/j.issn.1008-830X.2011.05.014

Miller, R. L., and Mckee, B. A. (2004). Using MODIS Terra 250 m imagery to map concentrations of total suspended matter in coastal waters. *Remote Sens Environ.* 93, 259–266. doi: 10.1016/j.rse.2004.07.012

Molkov, A. A., Fedorov, S., Pelevin, V. V., and Korchemkina, E. N. (2019). Regional models for high-resolution retrieval of chlorophyll a and TSM Concentrations in the Gorky Reservoir by Sentinel-2 imagery. *Remote Sens.* 11, 1215. doi: 10.3390/rs11101215

Moore, K. A., Wetzel, R. L., and Orth, R. J. (1997). Seasonal pulses of turbidity and their relations to eelgrass (Zostera marina L.) survival in an estuary. *J. Exp. Mar. Biol. Ecol.* 215, 115–134. doi: 10.1016/S0022-0981(96)02774-8

Novoa, S., Doxaran, D., Ody, A., Vanhellemont, Q., Lafon, V., Lubac, B., et al. (2017). Atmospheric corrections and multi-conditional algorithm for multi-sensor remote sensing of suspended particulate matter in low-to-high turbidity levels coastal waters. *Remote Sens.* 9, 61. doi: 10.3390/rs9010061

Pahlevan, N., Sarkar, S., Franz, B. A., Balasubramanian, S. V., and He, J. (2017). Sentinel-2 MultiSpectral Instrument (MSI) data processing for aquatic science applications: Demonstrations and validations. *Remote Sens. Environ.* 201, 47–56. doi: 10.1016/j.rse.2017.08.033

Pahlevan, N., Smith, B., Alikas, K., Anstee, J., Barbosa, C., Binding, C., et al. (2022). Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. *Remote Sens. Environ.* 270, 112860. doi: 10.1016/j.rse.2021.112860

Pan, D. L., and Ma, R. H. (2008). Some key problems in remote sensing of lake water quality. *J. Lake Sci.* 20, 139–144. doi: 10.18307/2008.0201

Qing, S., Bao, Y. H., and Hao, Y. L. (2017). Atmospheric correction of Landsat-8 OLI data for Wuliangsuhai Lake based on SWIR bands. *Infrared* 38, 21–30. doi: 10.3969/j.issn.1672-8785.2017.03.005

Ramaswamy, V., Rao, P. S., Rao, K. H., Thwin, S., Rao, N. S., and Raiker, V. (2004). Tidal influence on suspended sediment distribution and dispersal in the Northern Andaman Sea and Gulf of Martaban. *Mar. Geol.* 208, 33–42. doi: 10.1016/j.margeo.2004.04.019

Saberioon, M., Brom, J., Nedbal, V., Soucek, P., and Cisař, P. (2020). Chlorophyll-a and total suspended solids retrieval and mapping using Sentinel-2A and machine learning for inland waters. *Ecol. Indic.* 113, 106236. doi: 10.1016/j.ecolind.2020.106236

Saha, A., and Pal, S. C. (2024). Application of machine learning and emerging remote sensing techniques in hydrology: A state-of-the-art review and current research trends. *J. Hydrol.* 632, 130907. doi: 10.1016/j.jhydrol.2024.130907

Sankaran, R., Al-Khayat, J. A., Aravinth, J., Chatting, M. E., Sadooni, F. N., and Al-Kuwari, H. A. (2023). Retrieval of suspended sediment concentration (SSC) in the Arabian Gulf water of arid region by Sentinel-2 data. *Sci. Total Environ.* 904, 166875. doi: 10.1016/j.scitotenv.2023.166875

Shun, B. R., Qing, S., and Hao, Y. L. (2019). Remote sensing inversion of suspended sediment concentration in the Yellow River estuary based on semi-analytical method. *Mar. Sci.* 43, 17–27. doi: 10.11759/hykx20190414002

Si, W. (2022). *Research on UAV-borne hyperspectral imagery for re-trieval water quality parameters by machine learning algo-rithms* (Wuhan: Hubei University).

Sipelgas, L., Raudsepp, U., and Kõuts, T. (2006). Operational monitoring of suspended matter distribution using MODIS images and numerical modelling. *Adv. Space Res.* 38, 2182–2188. doi: 10.1016/j.asr.2006.03.011

Song, Q., Wang, N., and Wu, N. (2018). Monitoring of suspended particulate matter diffusion during reclamation construction based on numerical model and satellite remote sensing-Taking the Dalian Offshore Airport as the background. *Mar. Sci. Bull.* 37, 201–208. doi: 10.11840/j.issn.1001-6392.2018.02.011

Tan, Z., Ren, J., Li, S. D., Li, W., Zhang, R., and Sun, T. G. (2023). Inversion of nutrient concentrations using machine learning and influencing factors in Minjiang River. *Water* 15, 1398. doi: 10.3390/w15071398

Tang, Y., Pan, Y., Zhang, L., Yi, H., Gu, Y., and Sun, W. (2023). Efficient monitoring of total suspended matter in urban water based on UAV multi-spectral Images. *Water Resour. Manage.* 37, 2143–2160. doi: 10.1007/s11269-023-03484-2

Virtanen, O., Constantinidou, E., and Tyystjarvi, E. (2020). Chlorophyll does not reflect green light – how to correct a misconception. *J. Biol. Educ.* 56, 552–559. doi: 10.1080/00219266.2020.1858930

Wang, X., Bai, J., Yan, J., Cui, B., and Shao, D. (2022a). How turbidity mediates the combined effects of nutrient enrichment and herbivory on seagrass ecosystems. *Front. Mar. Sci.* 9, 787041. doi: 10.3389/fmars.2022.787041

Wang, C. Y., Li, W. J., Chen, S. S., Li, D., Wang, D. N., and Liu, J. (2018). The spatial and temporal variation of total suspended solid concentration in Pearl River Estuary during 1987-2015 based on remote sensing. *Sci. Total Environ.* 618, 1125–1138. doi: 10.1016/j.scitotenv.2017.09.196

Wang, L., Wang, X., Wang, X. X., Meng, Q. H., Ma, Y. J., and Chen, Y. L. (2022). Retrieval of suspended particulate matter concentration from Sentinel-3 OLCI image in the Coastal Waters of Qinhuangdao. *China Environ. Sci.* 42, 3867–3875. doi: 10.19674/j.cnki.issn1000-6923.20220314.003

Wang, X., Wen, Z. D., Liu, G., Tao, H., and Song, K. S. (2022b). Remote estimates of total suspended matter in China's main estuaries using Landsat images and a weight

random forest model. *ISPRS.J. Phtogramm. Remote Sens.* 183, 94–110. doi: 10.1016/j.isprsjprs.2021.11.001

Wen, Z., Wang, Q., Ma, Y., Jacinthe, P. A., Liu, G., Li, S., et al. (2024). Remote estimates of suspended particulate matter in global lakes using machine learning models. *Int. Soil Water Conserv. Res.* 12, 200–216. doi: 10.1016/j.iswcr.2023.07.002

Williamson, A. N., and Grabau, W. E. (1974). Sediment concentration mapping in tidal estuaries. In *NASA Goddard Space Flight Center 3rd ERTS-1 Symposium*, Volume 1, Section B (Paper-M5). NASA Goddard Space Flight Center.

Wirabumi, P., Muhammad, K., and Pramaditya, W. (2021). Determining effective water depth for total suspended solids (TSS) mapping using PlanetScope imagery. *Int. J. Remote Sens.* 42, 5784–5810. doi: 10.1080/01431161.2021.1931538

Wu, C. H., Fu, X. L., Li, H. H., and Hu, H. (2023). Study on inversion of suspended matter in Wuliansu Lake based on M-GA-BP. *Water Resour. Power.* 41, 49–52. doi: 10.20040/j.cnki.1000-7709

Xiao, Y. F., Zhao, W. J., and Zhu, L. (2012). Quantitative retrieval model of suspended sediment concentration in estuary based on HJ-1 CCD image. *Mar. Sci.* 36, 59–63.

Xing, Q. G., Lou, M. J., Tian, L. Q., Yu, D. F., Braga, F., Tosi, L., et al. (2014). Quasi-simultaneous measurements of suspended sediments concentration (SSC) of very turbid waters at the Yellow River Estuary with the multi-spectral HJ-1 Imageries and *in-situ* sampling. *Ocean Remote Sens. Monit. Space* 9261, 170–174. doi: 10.1117/12.2068930

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* 27, 3025–3033. doi: 10.1080/01431160600589179

Xu, Z. L., Li, M., Gao, Q., and Chen, H. (2010). Analysis on key factors of influence of Yangshan Project on marine environment. *Mar. Environ. Sci.* 29, 617–622 + 635. doi: 10.3969/j.issn.1007-6336.2010.05.001

Yan, F. L., Wang, S. X., Zhou, Y., Xiao, Q., Zhu, L. Y., Wang, L. T., et al. (2006). Monitoring the water qualty of Taihu lake by using Hyperion hyperspectral data. *J. Infrared Millim. Waves* 25, 460–464. doi: 10.3321/j.issn:1001-9014.2006.06.015

Yan, R., Zhang, X., Bi, W., Wang, N., Zhao, Y., Bi, L., et al. (2024). Extraction and analysis of the sea ice parameter dataset of the Bohai Sea from 2011 to 2021 based on GOCI. *Front. Mar. Sci.* 11. doi: 10.3389/fmars.2024.1364889

Yang, H., Kong, J., Hu, H., Du, Y., Gao, M., and Chen, F. (2022). A review of remote sensing for water quality retrieval: progress and challenges. *Remote Sens.* 14, 1770. doi: 10.3390/rs14081770

Yang, C. X., Li, Y., Yang, J., and Shu, S. J. (2023). Remote sensing inversion and regularity analysis of suspended sediment in Pearl River Estuary based on machine

learning model. *Bull. Surveying Mapping* 09, 117–123. doi: 10.13474/j.cnki.11-2246.2023.0275

Yin, Z. Y., Li, J. S., Fan, H. S., Gao, M., and Xie, Y. (2021). Preliminary study on water quality parameter inversion for the Yuqiao reservoir based on Zhuhai-1 Hyperspectral satellite data. *Spectrosc. Spectral Anal.* 41, 494–498. doi: 10.3964/j.issn.1000-0593(2021)02-0494-05

Yin, Z. Y., Li, J. S., Liu, Y., Zhang, F. F., Wang, S. L., Xie, Y., et al. (2022). Decline of suspended particulate matter concentrations in Lake Taihu from 1984 to 2020: observations from Landsat TM and OLI. *Opt. Express* 30, 22572–22589. doi: 10.1364/OE.454814

Yu, Z. X., Xu, P., Luo, W. X., and Zhang, C. (2020). A study on the suspended sediment concentration in Dianchi Lake using HJ-1A hyperspectral data. *J. Southwest For. Univ.* 40, 94–104. doi: 10.11929/j.swfu.201902056

Zarco-Tejada, P. J., and Ustin, S. L. (2001). Modeling canopy water content for carbon estimates from MODIS data at land EOS validation sites. In: *International Geoscience and Remote Sensing Sympo sium 2001, IGARSS'01.* New York, USA, 342–344. doi: 10.1109/IGARSS.2001.976152

Zhang, E. F., Chen, S. L., Gu, G. C., Yang, H. F., and Wang, R. S. (2015). Temporal and spatial variations in suspended sediment concentration and transport in the North Branch of the Yangtze Estuary. *Haiyang Xuebao.* 37, 138–151. doi: 10.3969/j.issn.0253-4193.2015.09.014

Zhang, X., Huang, J., Chen, J. J., and Zhao, Y. F. (2023). Remote sensing monitoring of total suspended solids concentration in Jiaozhou Bay based on multi-source data. *Ecol. Indic.* 154, 110513. doi: 10.1016/j.ecolind.2023.110513

Zhang, J., Li, F., Lv, Q., Wang, Y., Yu, J., Gao, Y., et al. (2021). Impact of the water–sediment regulation scheme on the phytoplankton community in the Yellow River estuary. *J. Clean. Prod.* 294, 126291. doi: 10.1016/j.jclepro.2021.126291

Zhang, M. L., and Zhou, Z. H. (2007). ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recogn.* 40, 2038–2048. doi: 10.1016/j.patcog.2006.12.019

Zhong, S. K., Lv, H., Yang, Z. Q., Li, Y. Y., Xu, J. F., and Miao, S. (2022). Remote sensing estimation method of organic suspended matter concentration in inland lakes based on Sentinel-3 OLCI data. *Natl. Remote Sens. Bulletin.* 26, 155–167. doi: 10.11834/jrs.20221266

Zhou, T., Yang, J., Zhou, Q., Tan, B., Zhou, Y., Jian, X., et al. (2019b). Power system transient stability assessment method based on modified LightGBM. *Power Syst. Technol.* 43, 1931–1940. doi: 10.13335/j.1000-3673.pst.2019.0085

Zhou, D., Yu, M., Yu, J. B., Li, Y. Z., Guan, B., Wang, X. H., et al. (2021). Impacts of inland pollution input on coastal water quality of the Bohai Sea. *Sci. Total Environ.* 765, 142691. doi: 10.1016/j.scitotenv.2020.142691