High-resolution ocean color reconstruction and analysis focusing on Kd490 via machine learning model integration of MODIS and Sentinel-2 (MSI)

Yulin Yang[1,2,3], Ziyao Wang[1,3], Peng Chen[4], Xue Shen[1,2,5]*, Wei Kong[1,2,3,5], Genghua Huang[1,2,3,5]* and Rong Shu[1,2,3,5]

[1]Key Laboratory of Space Active Opto-electronics Technology, Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China, [2]Shanghai Branch, Hefei National Laboratory, Shanghai, China, [3]University of Chinese Academy of Sciences, Beijing, China, [4]State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, State Oceanic Administration, Hangzhou, China, [5]Shanghai Research Center for Quantum Sciences, Shanghai, China

Oceanic water quality monitoring is essential for environmental protection, resource management, and ecosystem vitality. Optical remote sensing from space plays a pivotal role in global surveillance of oceanic water quality. However, the spatial resolution of current ocean color data products falls short of scrutinizing intricate small-scale marine features. This study introduces a hybrid model that fuses MODIS (Moderate Resolution lmaging Spectroradiometer) ocean color products with Sentinel-2 's remote sensing reflectance data to generate high-resolution ocean color imagery, specifically investigating the diffuse attenuation coefficient at a wavelength of 490 nm (Kd490). To address the intricacies of coastal environments, we propose two complementary strategies to improve the accuracy of inversion. The first strategy leverages MODIS ocean color products alongside a geographic segmentation model to perform distinct inversions for separate marine zones, enhancing spatial resolution and specificity in coastal regions. The second strategy bolsters model interpretability during training by integrating predictions from conventional physical models into a Random Forest-based Regression Ensemble (RFRE) model. This study focuses on the coastal regions surrounding the Beibu Gulf, near Hainan Island in China. Our findings exhibit a strong concordance with MODIS products, achieving a monthly average coefficient of determination (R^2) of 0.90, peaking at 0.97, and sustaining a monthly average root-mean-square error (RMSE) of less than 0.02. These results substantiate the model's efficacy. Moreover, the annual trend analysis and localized assessment of the reconstructed Kd490 offer nuanced insights that surpass MODIS data, establishing a robust foundation for high-resolution water quality monitoring in coastal zones.

1 Introduction

Oceanic waters have recently experienced persistent deterioration in water quality due to human activities in neighboring inland and coastal areas. Monitoring oceanic environments is crucial for assessing ecosystem health and supporting local fisheries economies, forming the foundation for sustainable development (Marzano et al., 2020). However, obtaining uniformly distributed time-series data across the entire study area poses a challenge for traditional field measurement methods. In this regard, leveraging remote sensing technology can effectively overcome these challenges and enable more comprehensive and consistent water quality monitoring.

Spaceborne optical remote sensing technology complements traditional *in-situ* sampling methods by providing large-scale, high-resolution images (with resolutions of tens of meters or finer) covering large areas with revisit times of a few days (Blondeau-Patissier et al., 2014; Matthews, 2011). Ocean remote sensing relies on measuring the spectral characteristics of radiation emitted or reflected by its components (Mobley, 1994; Manzo et al., 2018; Braga et al., 2016). In estimating chlorophyll-a (Chl-a) on centration, a series of algorithms including OC3 (Ocean Chlorophyll 3), OC4 (Ocean Chlorophyll 4), and those proposed by O'Reilly et al. have been widely adopted (O'Reilly, 2000). However, the global ocean color products provided by MODIS are primarily targeted toward open ocean waters with a spatial resolution of only 1 km. The coarse resolution is detrimental to depicting fine oceanic structures, hindering the study of small-scale oceanic activities. Additionally, coastal water bodies exhibit more complex compositions, influenced by colored dissolved organic matter (Zhu and Yu, 2013), phytoplankton (Asim et al., 2021), and suspended sediments (Zhang et al., 2016). The complexity in the composition of water bodies leads to rich textural features in reflectance images. As a result, the remote sensing of these optically complex water bodies presents greater challenges and typically requires instruments with higher spatial resolution to provide more detailed information (Wang et al., 2009; Kremezi and Karathanassi, 2020; Mouw et al., 2015; Yang et al., 2011). To address this challenge, satellite sensors with higher resolution, such as those on Landsat 8, have emerged, offering a moderate spatial resolution of 30 m. For example, Prasetyo et al. successfully utilized this sensor to estimate the diffuse attenuation coefficient at a wavelength of 490nm (Kd490) in the waters surrounding Bangka Island (Jaelani et al., 2016). Similarly, in experiments conducted in Eastern Moreton Bay, Australia, Jaelani et al. successfully estimated the total suspended matter (TSM) and Chl-a concentrations in the water column (Rodrigues et al., 2020).

Sentinel-2 satellites have recently become instrumental in ocean monitoring (Hedley et al., 2018; Pahlevan et al., 2017), whose spatial resolution is up to 10m. It provides high-resolution optical imagery, (with spatial resolution ranging from 10 to 60 meters depending on the visible spectral bands) and achieves systematic global coverage every 2-3 days in mid-latitude regions (Harmel et al., 2018). Many studies have used Sentinel-2 to successfully perform ocean color remote sensing tasks. For instance, Shundan Dong et al. conducted research in the coastal waters of Hong Kong, comparing the inversion results of Landsat-8 and Sentinel-2 based on multiple

models, demonstrating the feasibility of using both remote sensing datasets for Chl-a concentration inversion in Hong Kong's nearshore waters (Dong et al., 2021). Gernez et al. successfully applied a simulated Sentinel 2 MSI dataset to retrieve CSPM values in turbid estuary waters (Gernez et al., 2015).

There have also been studies on the construction of high-resolution ocean color data. Kremezi et al. improved marine monitoring by integrating Sentinel 3 and Sentinel 2 image data. They generated Chl-a and TSM maps to assess the fused image, yielding favorable results (Kremezi and Karathanassi, 2020). Pan Duan et al. aimed to develop and evaluate the performance of two image reconstruction strategies, namely spatiotemporal fusion reflectance image inversion SPM and SPM spatiotemporal fusion, based on the measured SPM concentration data with Sentinel-2 and Sentinel-3 (Duan et al., 2023).

In the realm of ocean color inversion methods, empirical models, while straightforward, suffer from a lack of spectral feature consideration. This limitation restricts their applicability to regions beyond their initial calibration area, consequently affecting prediction accuracy in the current region (Cao et al., 2020). Conversely, bio-optical models, characterized by their complexity and stability, encounter challenges associated with the acquisition of inherent optical properties, thereby constraining their widespread adoption (Song et al., 2014; Giardino et al., 2007).

In recent years, machine learning has been widely used to retrieve water quality parameters, including random forest (RF) (Maier et al., 2018; Chen et al., 2019), Support Vector Regression (SVR) (Camps-Valls et al., 2006), and artificial neural network (ANN) (Pahlevan et al., 2020; Gonzalez Vilas et al., 2011). Frank S. Marzano et al. proposed using an empirical and model-based framework to design regression and neural network (NN) models for the inversion of Chl-a and TSM along the coasts of the Adriatic and Tyrrhenian Seas in Italy. Comparisons with *in situ* data demonstrated the effectiveness of using Sentinel-2 for ocean color inversion (Marzano et al., 2020). Asim et al. developed techniques for ocean Chl-a remote sensing in the Barents Sea, improving monitoring through *in situ* data integration and machine learning, resulting in enhanced Chl-a estimation accuracy (Asim et al., 2021). Chen et al. evaluated various methods and found that the random forest-based regression ensemble (RFRE) model performed best in remotely estimating surface pCO2 in the Gulf of Mexico, providing a robust estimation framework for ocean carbon flux and acidification research (Chen et al., 2019).

The preceding studies relied on *in-situ* measurement results for validation, which resulted in a limited number of samples available for model training and testing. In contrast, this paper adopts a methodology that replaces *in-situ* measurement results with filtered MODIS standard ocean color products. While Sentinel-2 provides high spatial resolution data essential for detailed coastal analysis, there is currently no universal model. MODIS can complement Sentinel-2 by providing lower-resolution but frequent observations and a mature ocean color data product. Leveraging Sentinel-2's multispectral remote sensing imagery, a model combining machine learning and empirical formulas is constructed to re-invert and obtain high-resolution ocean color data for the study area.

In this paper, we employed RFRE as the machine learning method. A notable advantage of the machine learning-based RFRE

approach is the capability to approximate the nonlinear relationship between predictor variables (i.e., reflectance values from different spectral bands in remote sensing) and the target variable (i.e., Kd490) without the need for explicit knowledge of their functional dependency (Lunetta et al., 2004).

The coastal waters of the Beibu Gulf, adjacent to Hainan Island, China (spanning from 18.8°N, 108.1°E to 19.8°N, 109.1°E), have been designated as the research area. It is a semi-enclosed and shallow marine environment bordered by the southern coast of China and the northeastern coast of Vietnam. With an average depth of about 50m, it is influenced by complex hydrodynamic processes, including monsoons, tides, and currents. The gulf is rich in biodiversity and serves as a vital fishing ground, supporting various fish species, crustaceans, and mollusks. Nutrient input from rivers contributes to its high primary productivity. Accurately understanding the spatiotemporal characteristics of marine ocean color parameters in this region holds significant scientific and strategic importance (Zhuang et al., 2010). Kd490 is chosen as our research target, which quantifies the reduction in the intensity of light at a wavelength of 490 nm due to scattering and absorption by water and its constituents. This coefficient is important in oceanography and marine science for understanding light penetration in water, which affects photosynthesis, ocean color, and the overall health of aquatic ecosystems (Tomlinson et al., 2019). Utilizing data collected by both sensors throughout the months in 2023, Kd490 standard ocean products from NASA's MODIS are employed to train and test the inversion model with Sentinel-2 data. The test results have shown promising outcomes. With the proposed model, we reconstruct Kd490 images at a 10m resolution for the coastal waters of Hainan Island. Subsequently, a trend analysis of Kd490 is conducted based on these reconstructed images.

2 Data

This section introduces the Sentinel-2 satellite imagery and MODIS ocean products used in this study.

2.1 Sentinel-2 imagery data set

The Sentinel-2 mission under the Copernicus program, driven by the European Space Agency (ESA), aims to acquire high-precision imagery data of land and coastal areas to support the monitoring of vegetation, soil, and coastal waters (Pahlevan et al., 2017; Liu et al., 2017). The full mission specification of the twin satellites flying in the same orbit but phased at 180°, is designed to give a high revisit frequency of 5 days at the Equator. The mission consists of two identical satellites, Sentinel-2A and Sentinel-2B, both of which are equipped with Multi-Spectral Imagers (MSI).

The MSI measures the Earth's reflected radiance in 13 spectral bands from VNIR to SWIR with an orbital swath width of 290 kilometers. Spatial resolution varies are shown in Table 1: 1) Four bands (B2-4, B8) for visible and near-infrared spectra at a resolution of 10 meters; 2) Six bands (B5-7, B8A) at 20-meter resolution for red-edge and SWIR for vegetation detection, and (B11, B12) for snow, ice, and cloud discrimination; 3) Three bands (B1, B9, B10) at a resolution of 60 meters targeting coastal aerosols, water vapor, and cirrus detection, respectively.

For this study, 24 Sentinel-2 satellite images were selected from the Copernicus Open Access Hub (https://browser.dataspace.copernicus.eu) and categorized by month. These Level-2A images, obtained in the free SAFE format and projected in UTM/WGS84, underwent official atmospheric

TABLE 1 Overview of the MSI spectral bands with their spatial resolutions (From Sentinel-2 MSI technical guide).

	S2A		S2B		
Band Number	Central wave-length (nm)	Bandwidth (nm)	Central wave-length (nm)	Bandwidth (nm)	Spatial resolution (m)
1	442.7	20	442.3	20	60
2	492.7	65	492.3	65	10
3	559.8	35	558.9	35	10
4	664.6	30	664.9	31	10
5	704.1	14	703.8	15	20
6	740.5	14	739.1	13	20
7	782.8	19	779.7	19	20
8	832.8	105	832.9	104	10
8A	864.7	21	864.0	21	20
9	945.1	19	943.2	20	60
10	1373.5	29	1376.9	29	60
11	1613.7	90	1610.4	94	20
12	2202.4	174	2185.7	184	20

(From ESA Sentinel-2 MSI Technical Guide: https://sentiwiki.copernicus.eu/web/s2-mission#MSI-Instrument).

correction by ESA, eliminating the need for further correction during subsequent processing. To ensure data quality, images with cloud cover exceeding 15% were excluded, resulting in a final selection of 24 pairs of high-quality images for model training and reconstruction.

2.2 MODIS Kd490 (marine level-3 products) data set

MODIS, onboard NASA's Aqua and Terra satellites, plays a vital role in monitoring Earth's surface parameters, including ocean color. Kd490, which characterizes the diffuse attenuation of light in the water column, is particularly significant in ocean color remote sensing as it reflects water transparency and particle concentration (Rodrigues et al., 2020). To provide reference and validation results for Sentinel-2 inversion, data from NASA's website (https://oceandata.sci.gsfc.nasa.gov) were accessed to obtain Kd490 products covering the same year in the vicinity of Hainan Island, with a spatial resolution of 1 km. Two images were selected each month to match with Sentinel-2. Due to potential cloud cover and data gaps in MODIS images, two images were selected each month to match the temporal resolution of Sentinel-2 data. A maximum time difference of 2 days was allowed between corresponding Sentinel-2 and MODIS images to ensure consistency. Images with extensive missing data were manually excluded, resulting in a final selection of 24 pairs for analysis.

3 Method

This section provides an overview of the methodology employed in the study. It commences with the sequential preprocessing, filtering, and matching of pairs of single-day images for each month. Subsequently, the resulting datasets are recombined by month for model training and testing. Following this, the model framework, training process, testing results, and improvements are detailed.

3.1 Data preprocessing, filtering, and matching

Before developing the model, we first conducted preprocessing steps on the original Sentinel-2 imagery and MODIS Kd490 ocean color products, including spatial resolution matching, dataset selection, and geographic location alignment to facilitate subsequent work, as shown in Figure 1A.

The inversion algorithm primarily utilizes four Sentinel-2 bands: ultra blue, blue, green, and red, with different spatial resolutions. To ensure consistency, the coastal band of each Sentinel-2 image is upsampled to a 10m resolution to match the

other bands. Subsequently, due to the large data volume and resolution mismatch with MODIS, all Sentinel-2 images are downsampled to a uniform resolution of 1 km.

During dataset filtering, the corresponding MODIS Kd490 data within the coverage area of Sentinel-2 images are initially selected due to the limited coverage of Sentinel-2 images. Due to having previewed the approximate range of valid values for MODIS Kd490 marine products, to ensure the validity and accuracy of the data, the range of Kd490 values is restricted to between 0 and 1 m-1, thereby excluding potential outliers. This filtering process results in the creation of a MODIS dataset, referred to as set A, containing geographic coordinates and corresponding Kd490 values.

To improve geographic coordinate matching accuracy, a proximity-based matching method is employed. By comparing the distance at corresponding locations on Sentinel-2 and MODIS images, matching is determined. Given that Sentinel-2 images have already been downsampled to 1 km spatial resolution, and each degree of latitude is approximately 111 km, a spatial resolution of 1 km corresponds to a difference of 0.01° in latitude and longitude. Therefore, a distance threshold of 0.01° is set for matching.

To be specific, first, iterate through each pixel of the Sentinel-2 image. Then, for each pixel's corresponding geographical coordinates, iterate through and calculate the Haversine distance with the geographical coordinates in set A. This calculation considers Earth's curvature.

The Haversine distance between two points with specific latitude and longitude coordinates (φ_1, λ_1) and (φ_2, λ_2) is calculated as follows:

$$a = \sin^2\left(\frac{\Delta\varphi}{2}\right) + cos(\varphi_1) \cdot cos(\varphi_2) \cdot \sin^2\left(\frac{\Delta\lambda}{2}\right) \tag{1}$$

$$c = 2 \cdot \text{atan2}(\sqrt{a}, \sqrt{1-a}) \tag{2}$$

$$d = R \cdot c \tag{3}$$

where (φ_1, λ_1) and are the latitudes of the two points in radians, $\Delta\varphi = \varphi_2 - \varphi_1$ is the difference in latitudes, $\Delta\lambda = \lambda_2 - \lambda_1$ is the difference in longitudes, and R is the Earth's radius (mean radius = 6371 km).

For each pixel in Sentinel-2, if the minimum distance to the coordinates in Set A is less than the predefined threshold of 0.01°, it indicates that the pixel has a geographic location match in the preprocessed MODIS data. In such cases, the Sentinel data corresponding to this minimum distance is selected as the input variable, while the associated MODIS data is assigned as the label.

After completing the preprocessing steps, two datasets are obtained, representing input variables (the sampling points include the remote sensing reflectance of Sentinel-2 bands 1-4 and their corresponding latitude and longitude) and target labels (the Kd490 values of sampling points given by MODIS), forming the basis for subsequent model training and analysis.
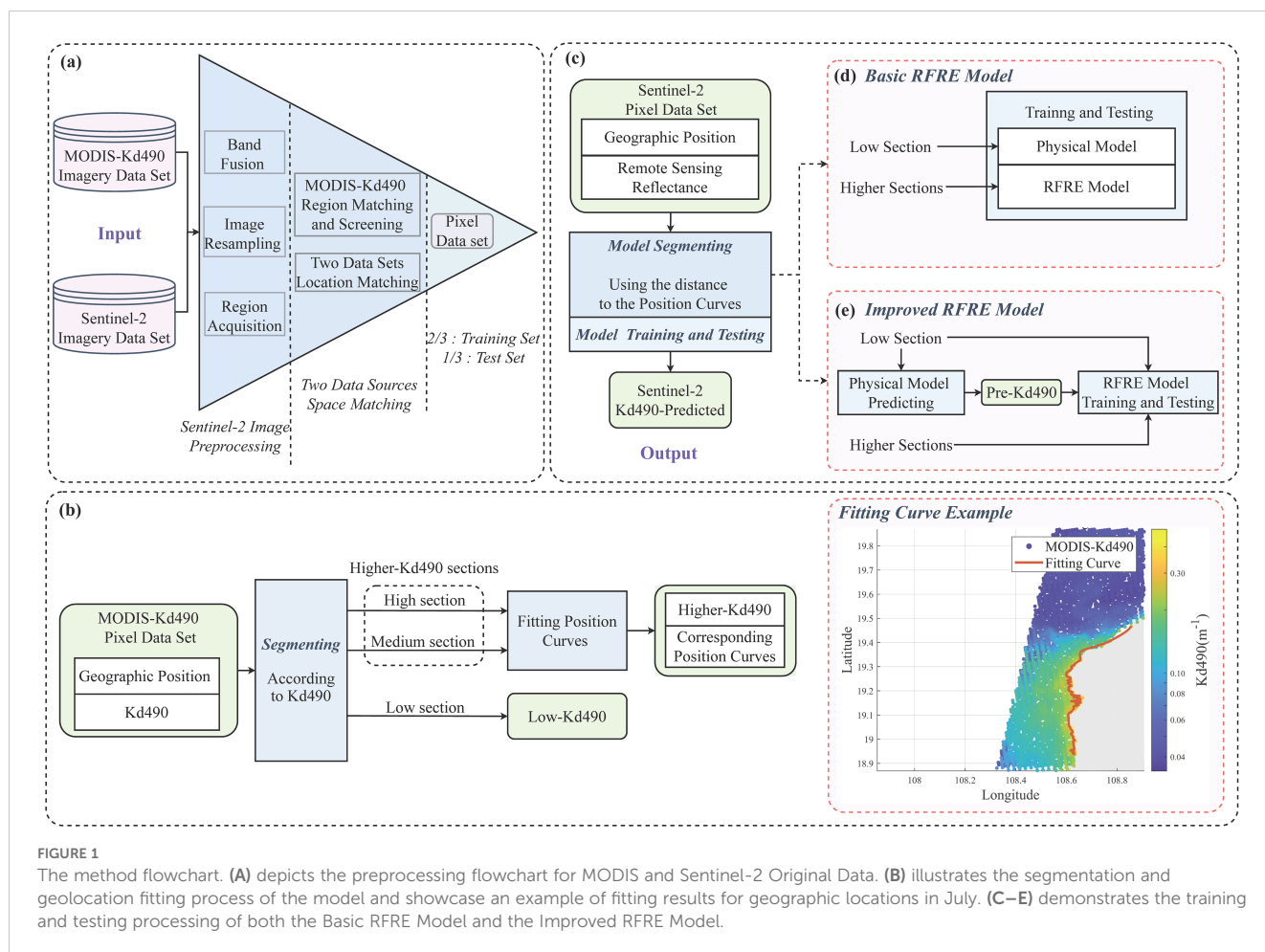
FIGURE 1
The method flowchart. **(A)** depicts the preprocessing flowchart for MODIS and Sentinel-2 Original Data. **(B)** illustrates the segmentation and geolocation fitting process of the model and showcase an example of fitting results for geographic locations in July. **(C−E)** demonstrates the training and testing processing of both the Basic RFRE Model and the Improved RFRE Model.

3.2 Geolocation segmentation of MODIS Kd490

Coastal and transitional waters often categorized as Case-II water types, encompassing algal blooms and natural or anthropogenic turbidity plumes, typically pose challenges for a single model approach to achieve satisfactory results, following the definition provided by Morel and Prieur (Morel and Prieur, 1977). Using MODIS data, we can obtain a rough estimate of Kd490 for the area. We found that the Kd490 values in the nearshore waters exhibit a pronounced tendency to higher values, which can probably be attributed to elevated concentrations of suspended particles, such as sediments and organic matter, which are common in coastal areas. Therefore, this study adopts a model segmentation approach that integrates Kd490 reference value from MODIS and geographic location information. The operational steps are outlined as follows:

1. Regional Partitioning: Data points with Kd490 values exceeding a specific threshold are labeled, and the boundaries of these regions are delineated using smooth curve fitting techniques.
2. Model Segmentation: For regions with lower Kd490 ground truth values, indicating clearer and simpler water

compositions, empirical algorithms remain applicable. Here, commonly used blue-green band ratio models for oceanic water bodies, collectively referred to as physical models, are primarily employed (see Empirical Formula in Table 2) (Blondeau-Patissier et al., 2014; O'Reilly, 2000; Sauer et al., 2012). Through multiple tests, the best-performing models are selected.

In contrast, for regions with higher Kd490 truth values due to water complexity, a machine learning model is utilized.

3. Model Training: During the inversion process, the choice of model is determined based on the proximity of Sentinel-2 data coordinates to the boundaries of the fitting curves. Data points closer to these boundaries undergo inversion using machine learning model, while others employ physical models. In the fitting process, Sentinel-2 remote sensing data serve as independent variables, with corresponding MODIS Kd490 values as dependent variables. These methods aim to enhance the model's adaptability to different water types, thereby improving the accuracy and reliability of ocean color parameter inversion.

The segmentation and geolocation fitting process of the model is illustrated in Figure 1B, along with an example of fitting results

TABLE 2 Physical model and RFRE model setting.

	Model Form	Expression
Empirical Formula (Based on Blue-Green Ratio Method)	Exponential Function	$Kd490 = t_1 e^{(t_2 \cdot \max(b_2, b_1))/b_3} + t_3$
		$Kd490 = t_1 e^{(t_2 \cdot b_2 + t_3 \cdot b_1)/b_3} + t_4$
		$Kd490 = t_1 e^{(t_2 \cdot b_2)/b_3} + t_3$
	Polynomial Function	$Kd490 = t_1 (b_2/b_3)^2 + t_2$
		$Kd490 = t_1 (b_2/b_3)^3 + t_2$
	Power Function	$Kd490 = t_1 (b_2/b_3)^{t_2} + t_3$
	Logarithmic Function	$Kd490 = t_1 \log(b_2/b_3) + t_2$
RFRE Model Setting	Physical Model	$Kd490 = t_1 e^{(t_2 \cdot \max(b_2, b_1))/b_3} + t_3$
	Matching Threshold	$D = 5 \times 10^{-5}$
	Sections breakpoints	0.25, 0.5
	Learning Cycles	25
	Minimum Leaf Size	8

for July geographic locations. The red solid line is the curve fitted by screening the higher value of MODIS-Kd490, while the scattered point represents the Kd490 value of the sample points. The Kd490 around the fitted curve is significantly higher, indicating that the segmentation of the model based on geographical location is effective. In addition, the curve is consistent with the lower part of the coastline, which can be preliminarily analyzed that the coastal water body in the south of this region is turbidity in summer, so it is not suitable to use a simple physical model such as blue-green ratio method for ocean color inversion.

3.3 Basic RFRE model and seasonal fixed model

Initially, Kd490 values are divided into three intervals using 0.25 and 0.5 as breakpoints. For MODIS-provided Kd490 falling within these intervals, separate position curves are fitted to determine the appropriate model for training. Based on the distance between Sentinel-2 data points and the fitted curves, data points are assigned to the corresponding Kd490 interval model. Geographic coordinates are used to evaluate the fit of each data point to the curves, with assignments made based on a matching threshold.

For sample points near the boundaries of the fitting curves, the RFRE model was chosen for this study after evaluating several machine learning models mentioned in the introduction. RFRE is a widely used ensemble learning algorithm that enhances model accuracy and robustness by aggregating predictions from multiple decision trees. In regression tasks, RFRE helps mitigate the overfitting commonly associated with individual regression trees by employing bagging, which improves the model's generalization capability (James et al., 2013).

During training, each regression tree in the RFRE model is independently grown on bootstrapped samples of the training data, which in this case consist of Rrs values from Sentinel-2 bands B1 to B4. At each decision split, a random subset of these Rrs values is selected as predictors, reducing correlations between trees and increasing their independence, ultimately boosting overall model performance. The bootstrap sampling technique also enables out-of-bag (OOB) estimation, allowing for the assessment of predictive performance using observations not included in the bootstrap samples.

Two key parameters define the RFRE model's structure: the minimum leaf size and the number of learning cycles (i.e., the number of regression trees). The minimum leaf size determines the smallest number of data samples required in each node of a regression tree, affecting both the depth of the tree and the way it splits. The number of learning cycles refers to the total number of regression trees in the ensemble. After optimizing these parameters, the Basic RFRE model was implemented to establish the predictive relationship between the Rrs values and Kd490, the target variable. Figure 1C depicts the overall process, while Figure 1D specifically details the Basic RFRE Model utilized within this framework.

After multiple rounds of cross-validation, we have selected the optimal configuration of the model's key parameters (see RFRE Model Setting in Table 2). The chosen configuration for the physical model is $y = t_1 e^{(t_2 \cdot \max(b_2, b_1))/b_3} + t_3$, the matching threshold is $D = 5 \times 10^{-5}$. Using 50 trees in the regression random forest and set the threshold accordingly. The RFRE optimal values for the minimum leaf size and number of learning cycles were determined to be 8 and 25, respectively. With these settings, the predictive accuracy of the RFRE model becomes stable, and the RFRE model has been developed for predicting Kd490.

Using random sampling, 2/3 of all data points are selected as the training set for training the Sentinel-2 Kd490 inversion model each month. The remaining 1/3 of data points are reserved as the test set to evaluate the accuracy and generalization ability of the model, and to adjust the model parameters appropriately.

We computed the overall coefficient of determination (R^2), the mean bias (Bias), and the root mean square error (RMSE) of the model on the test set for each month. These metrics serve to gauge the efficacy of the model across different time intervals and its overall performance.

The coefficient of determination, R^2, represents the proportion of the variance in the dependent variable that is explained by the model. R^2 values range from 0 to 1, with values closer to 1 indicating better model fitting:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (4)$$

Bias indicates the average difference between the predicted values and the true values, thereby offering a measure of model accuracy:

$$Bias = \frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y_i}) \quad (5)$$

RMSE, the square root of the mean squared differences between the predicted values and the true values, provides an assessment of

the overall prediction error of the model. Smaller bias and RMSE values indicate that the model's predictions are closer to the true values, signaling better performance:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \qquad (6)$$

Where y_i represents the observed value, $\hat{y}_i$ denotes the predicted value, $\bar{y}$ stands for the mean of the observed values, and n indicates the sample size, SS_{res} refers to the sum of squared residuals, which represents the total variance of the dependent variable that is not explained by the regression model. SS_{tot} stands for the total sum of squares, which represents the total variance of the dependent variable.

Moreover, an extensive examination of the physical model's fitting outcomes reveals a noticeable trend in coefficients across various seasons. Utilizing this insight, the coefficients within each season are weighted by their corresponding R^2 values to derive universal coefficients representative of each quarter. A similar approach is adopted for the RFRE model, where parameters are fixed within each quarter to better capture the seasonal data dynamics.

Consequently, this study develops a Seasonal RFRE model with predefined parameter settings, ensuring model stability across different seasons. By simultaneously employing the physical model and RFRE for joint inversion, a Seasonal Fixed Model is established, eliminating the need for recurrent model retraining during Sentinel inversion operations. The model's primary strength lies in its ability to perform inversions directly following model stabilization for each season, all without the need for retraining. In addition, because the parameters of the model are few and fixed, the application is very simple. However, it's important to note that while its fixed coefficients may compromise predictive accuracy to some extent, this trade-off is outweighed by its efficiency, making it particularly suitable for scenarios where

speed and simplicity are paramount, such as real-time decision-making processes or large-scale data analysis tasks.

3.4 Improved RFRE model

Without accounting for the seasonal aspect of the model and instead fitting and updating the model using MODIS images from nearby dates on a case-by-case basis, the prediction results from the physical model for the low Kd490 segment (Pre-Kd490) are integrated into the input of the RFRE model. At this stage, all segments derived from the geographically fitted curves are trained using the RFRE model, as shown in Equation 7. This approach allows the RFRE model to leverage not only raw data but also the predictive information generated by the physical model, thereby improving its accuracy and robustness. Figure 1C shows the entire process, while Figure 1E demonstrates the application of the Improved RFRE Model for data inversion.

$$\text{Kd490} = f_{\text{RFRE}}(b1, b2, b3, b4, \text{Pre-Kd490}) \qquad (7)$$

We also performed a significance analysis of the inputs to the Improved RFRE model, and the monthly average results are shown in Figures 2A, B), representing the importance ranking of high and low segments based on geographical location segmentation respectively. As can be seen from the figure, the low segmentation still mainly depends on the prediction results of the physical model (Pre-Kd490), the proportion reached 77.8%, which is far more important than other inputs, indicating that the blue-green ratio method is still valid. In addition to the physical model, the high segment is more dependent on the ultra blue wave segment (b1), the proportion is 43.0%, and other inputs also occupy a certain proportion. It further indicates that when the water body is complex, it cannot simply rely on the physical model
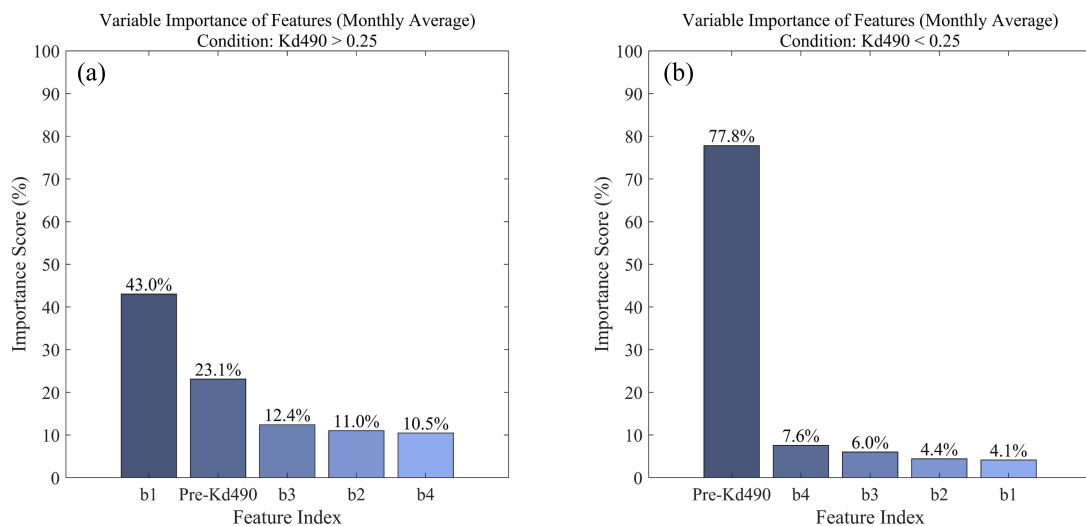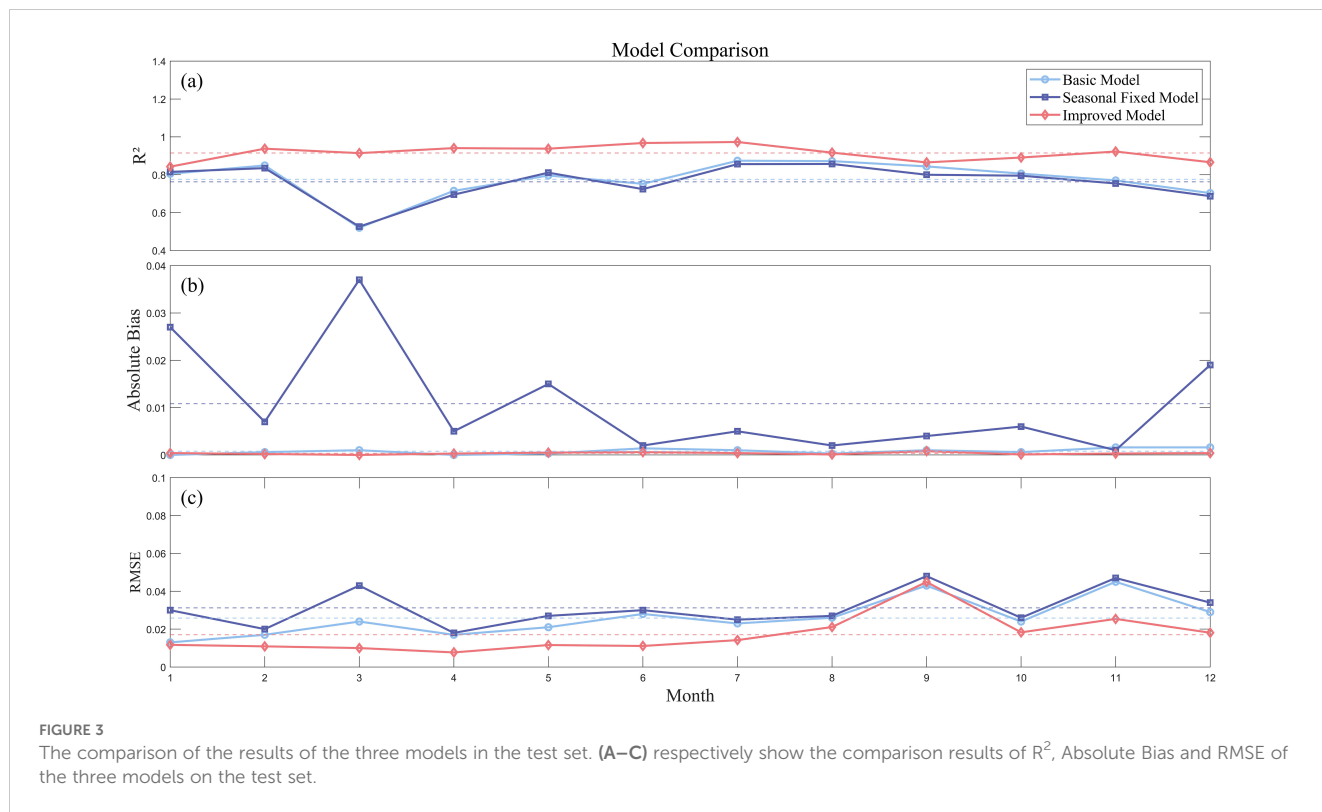

FIGURE 2
The significance analysis of the inputs to the improved RFRE model. **(A, B)** represent the importance ranking of high and low segments based on geographical location segmentation respectively.

FIGURE 3
The comparison of the results of the three models in the test set. **(A–C)** respectively show the comparison results of R^2, Absolute Bias and RMSE of the three models on the test set.

for training, nor can it only consider the information of the blue and green bands.

4 Results

4.1 Comparations and analysis of model performances

In this section, R^2, Bias and RMSE are used to evaluate the testing results of the three models, as shown in Figure 3.

The R^2 of Basic RFRE Model values mostly hover around 0.8, indicating a strong linear correlation between the model's predictions and the corresponding Kd490 values from MODIS products. Despite the lower R^2 value in March, both bias and RMSE remain low, indicating minimal discrepancies between the model predictions and actual values for that month. Overall, the model demonstrates robust performance. Additionally, it's worth noting that the geographic segmentation model for the selected MODIS dataset in March is null, indicating that the Kd490 values predominantly fall within the low range of 0 to 0.25. Consequently, the inversion mainly relies on the physical model, specifically the blue-green ratio method. The lower R^2 values also indicate that results obtained solely through various empirical models based on the blue-green ratio method may not be fully applicable to the more complex conditions of coastal waters. This suggests that the simplistic fitting from such empirical models may not adequately capture the nuances of coastal water bodies. When we fixed model, the R^2 and RMSE decreased slightly compared with the Basic RFRE

Model, but the absolute Bias was quite different, indicating that the accuracy of the seasonal fixed model decreased more.

The Improved RFRE Model demonstrates remarkable advancements in both data fitting and predictive accuracy. Through the integration of the physical model, further refinement is achieved, enhancing the accuracy and robustness of the prediction results. The monthly average R^2 on the test set reaches an impressive 0.91, with the highest value reaching an exceptional 0.97, highlighting the model's exceptional ability to capture variance and elucidate the variability of the observed data. Furthermore, the noteworthy improvement in R^2 values is evident, with increases ranging from 0.02 to 0.39 over the twelve months, emphasizing the significant enhancement achieved through the adoption of the Improved RFRE Model. Moreover, the model maintains consistently low average RMSE values, persistently remaining below the 0.02 threshold. This emphasizes the minimal discrepancies between predicted and observed values, indicating the model's remarkable accuracy in estimating target variables.

It is also noteworthy that the most significant improvement was observed in March, highlighting that even in relatively clean water environments, the Improved RFRE Model brings about substantial optimizations compared to solely relying on empirical formulas.

Due to the superior overall fitting performance of the Improved RFRE Model, its results were utilized for subsequent validation and high-resolution Kd490 reconstruction tasks. Figures 4A, B illustrates the density plots of the test results for July and November respectively using the Improved RFRE Model, along with the corresponding geographical locations marked shown in Figures 4C, D. In the figure, red hollow circles denote predicted high Kd490 locations.
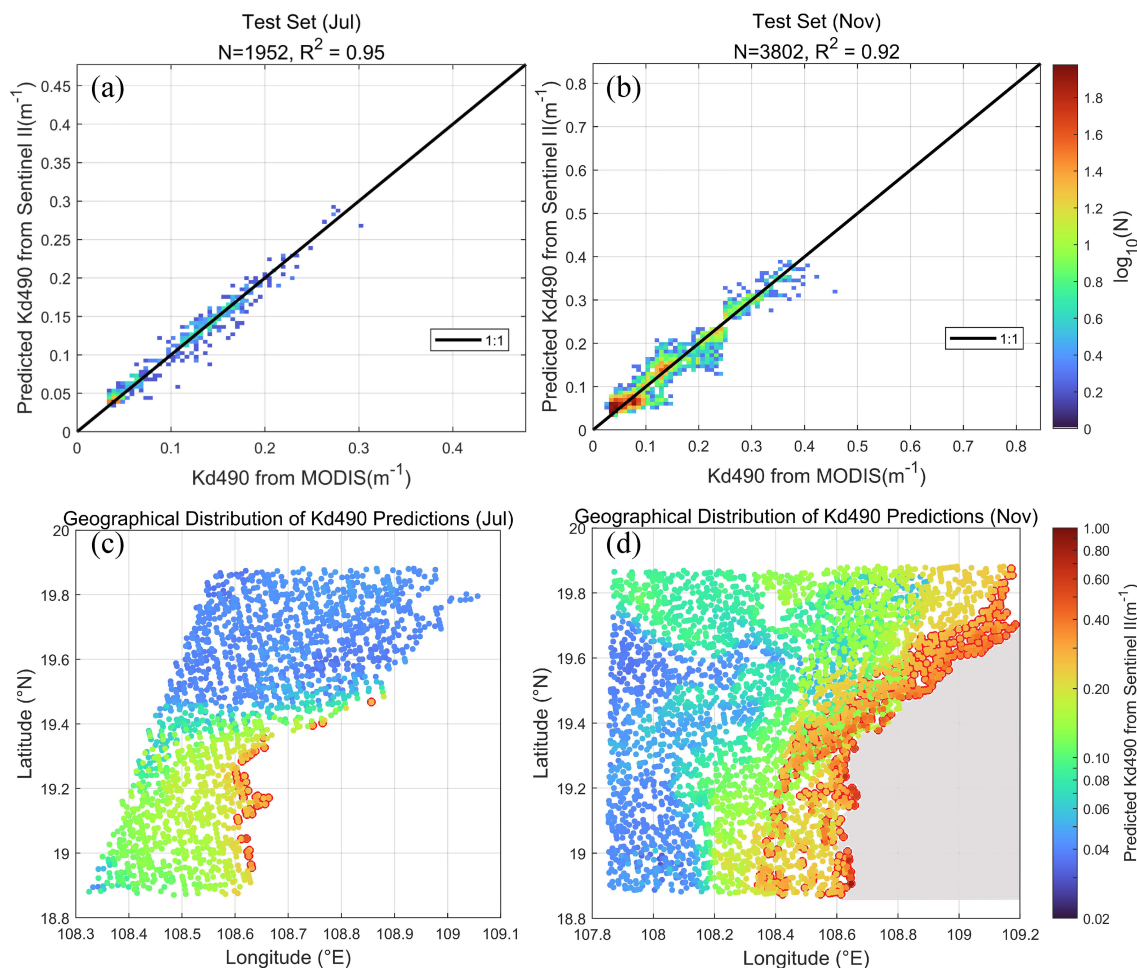
FIGURE 4
The testing results and corresponding geographical locations for selected months of the year 2023 are depicted below. **(A, B)** illustrate the model's testing outcomes for July and September, respectively, while **(C, D)** display their corresponding Kd490 scatter plots.

The test results reveal a striking alignment between the predictions generated by the Improved RFRE model, which incorporates geographical segmentation and integrates insights from the physical model, and the Kd490 ocean color products provided by MODIS. It is worth noting that the model excels in accurately predicting high Kd490 values. This demonstrates the effectiveness of integrating MODIS ocean color products for geographic segmentation, showcasing its exceptional fitting capability in addressing the complexities of coastal waters. Additionally, it highlights the model's high reliability in capturing and predicting changes in ocean color products.
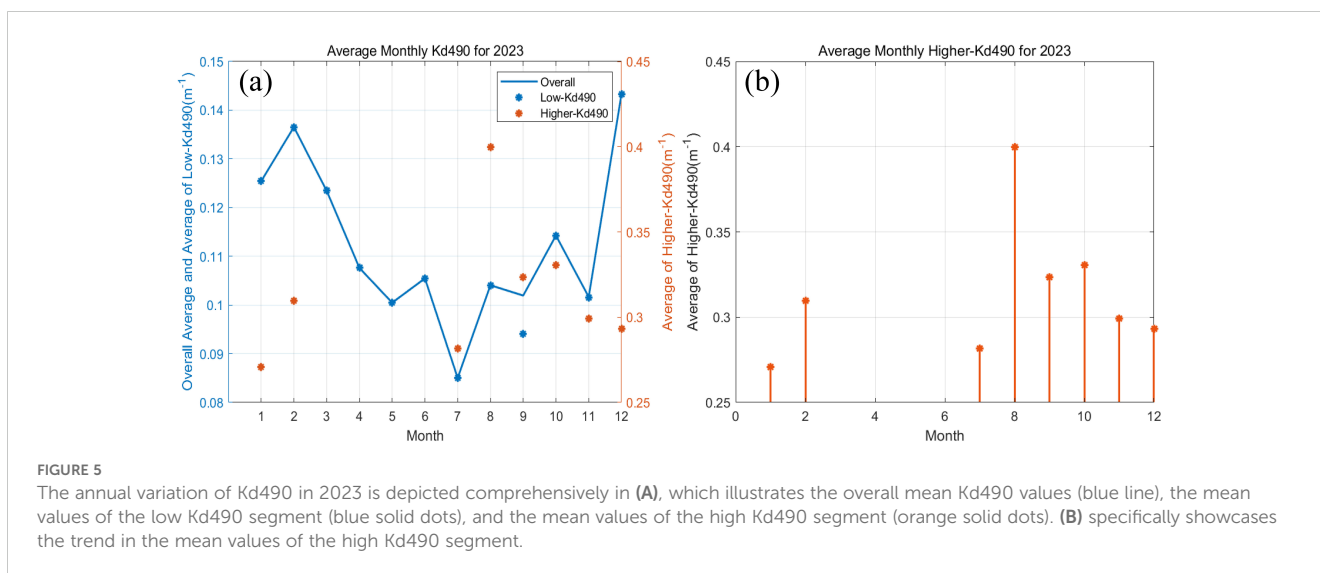
4.2 Analysis of annual variations in predicted Kd490 trends

By training the Improved RFRE Model monthly, Kd490 predictions for single days are eventually reconstructed. In this section, the reconstructed results will be evaluated and analyzed.

Firstly, a thorough analysis of the reconstructed dataset is conducted, including meticulous calculations of the monthly average

Kd490 values. Additionally, the analysis encompasses computations for the average Kd490 values in both high and low segments. These insightful research findings are vividly illustrated in Figure 5.

From Figure 5A, it's evident that the mean Kd490 generally peaks in winter and declines in other seasons. This indicates a probable increase in suspended particles, phytoplankton, or other factors influencing light propagation during the winter months. Such variations could signify an adaptive response known as light acclimation, where phytoplankton modulate their Chl-a content to cope with reduced light intensity during the weaker daylight periods of winter (Tiarasani et al., 2023). Looking at the trend of lower Kd490 mean values (indicated by blue solid dots), it closely overlaps with the overall mean, indicating lower Kd490 values in most areas of the study region. It also indicates that most areas of this water body are relatively clean. It also indicates that most areas of this water body are relatively clean. As for the trend of higher Kd490 mean values (from Figure 5B), there is a lack of data from March to June, suggesting lower Kd490 values during spring, indicating clearer water during this period. In summer, however, there is a higher value, especially with a significant peak in August, but the overall mean Kd490 remains low. This could indicate that during

FIGURE 5
The annual variation of Kd490 in 2023 is depicted comprehensively in **(A)**, which illustrates the overall mean Kd490 values (blue line), the mean values of the low Kd490 segment (blue solid dots), and the mean values of the high Kd490 segment (orange solid dots). **(B)** specifically showcases the trend in the mean values of the high Kd490 segment.

the summer, there are noticeable localized oceanic activities in the area, such as eddies and surface winds (Guo et al., 2017; He et al., 2016; Fang et al., 2006; Yu et al., 2019). These factors can also affect the growth environment of phytoplankton, further leading to seasonal variations in ocean color parameters. For example, Mesoscale eddies can alter biological productivity by changing nutrient distribution in the euphotic zone with their complex horizontal or vertical motion (Chen et al., 2011; Siegel et al., 2011). Wind-induced mixing can enhance Chl-a levels by entraining additional nutrients into the upper ocean in most tropical and subtropical areas (Kahru et al., 2010).

4.3 High-resolution predicted Kd490 reconstruction

We further compare the high-resolution reconstruction results based on Sentinel-2 with MODIS Kd490 products. Additionally, to corroborate the conclusion of locally elevated summer Kd490 values in the annual trend, we select three sets of summer reconstruction results for demonstration, as shown in Figure 6.

Based on the comparison of the images in Figure 6, it is evident that the reconstructed results are consistent with MODIS over the overall large-scale range. This indicates that the method proposed in this paper is effective in the coastal waters near Hainan Island. Furthermore, owing to the 10m spatial resolution of Sentinel satellite imagery, the reconstructed results offer more intricate details on small-scale ocean features that are less distinguishable in MODIS imagery with its 1km resolution. It is important to note that simply upsampling MODIS Ocean color products to match the spatial resolution of Sentinel-2 will not produce comparable results, as upsampling only interpolates new pixels without adding new information. So the proposed model enhances image resolution while compensating for data gaps in MODIS caused by cloud cover. This facilitates more refined studies of oceanic activities and enables precise monitoring of ocean color parameters. In addition, locally

elevated Kd490 values are indeed observed during the summer, primarily concentrated in the southern maritime areas. This may be related to seasonal biological activities, oceanic dynamic adjustments, marine currents, and mixing processes (Xiu and Chai, 2011; Liu et al., 2008).

In summary, these findings collectively indicate that the method proposed in this paper is crucial for in-depth analysis and understanding of the ocean color conditions in nearshore areas, especially for conducting small-scale studies. It should be noted that since MODIS Kd490 ocean color products are generally more suitable for open ocean waters, future research may need to integrate more *in-situ* data, especially for more detailed sampling and monitoring in coastal areas.

It is also worth noting that the original Sentinel-2 imagery did not undergo complex preprocessing. In future work, we will incorporate BRDF correction into the preprocessing of Sentinel-2 Rrs data, as variations in viewing and sun angles introduce significant anisotropy effects (Zhang et al., 2018). This will enhance the model's precision and robustness, especially in complex coastal environments. Furthermore, this study focuses on experimental verification in the Hainan Island region, and more case studies are needed to validate the applicability and generalizability of the method to other regions. Future research could select other representative geographic areas to further verify the applicability of the method in different aquatic environments and expand the model's scope of application.

5 Conclusion

This work proposes a model that integrates MODIS ocean color products with Sentinel-2 remote sensing reflectance images to ultimately reconstruct high-resolution ocean color data. Due to the complexity of coastal waters, the accuracy of ocean color inversion is improved through two methods:
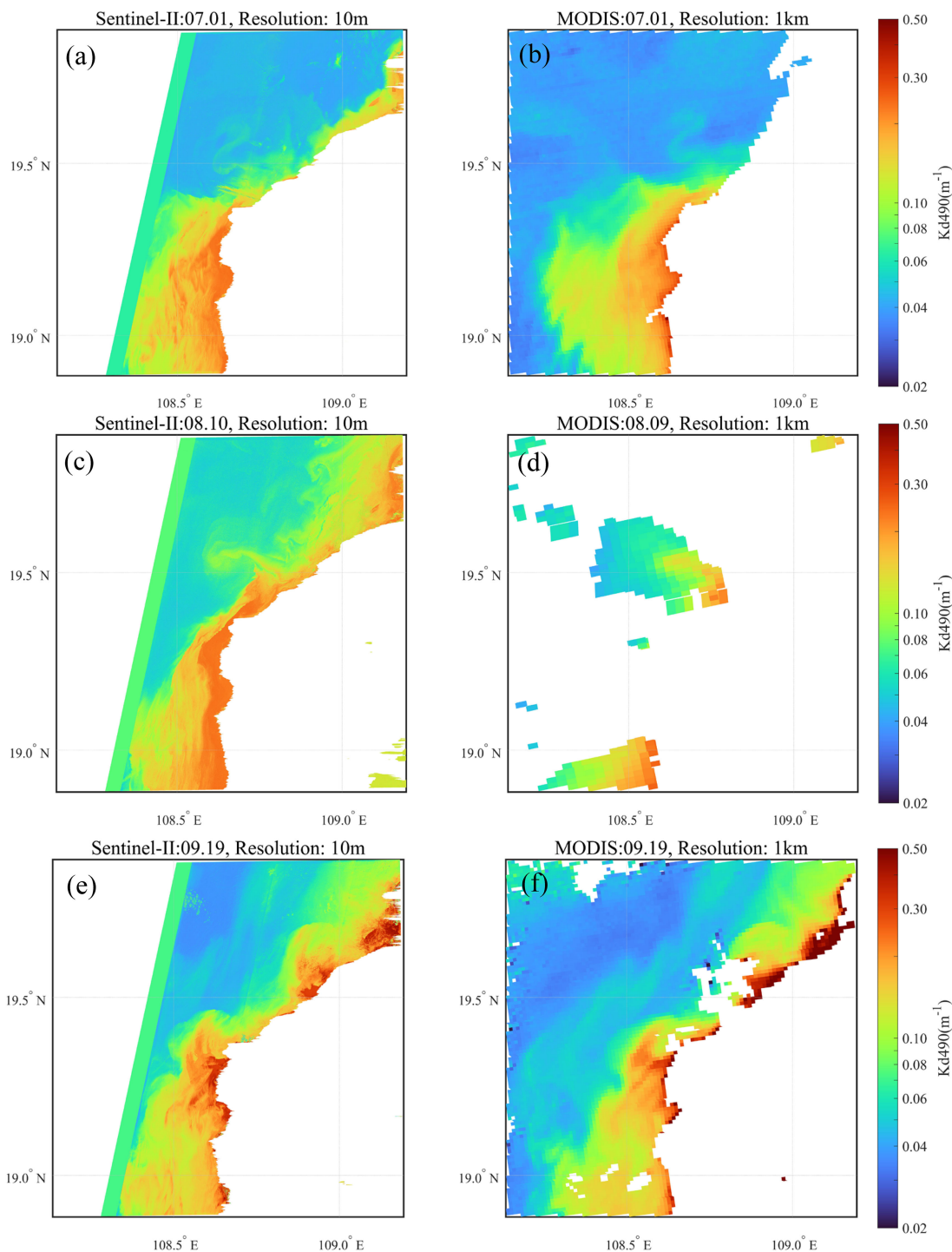
FIGURE 6
The comparison between the forecast results from Sentinel-2 in summer 2023, specifically **(A)** July, **(C)** August, **(E)** September, and the MODIS Kd490 product results for **(B)** July, **(D)** August, **(F)** September.

1. Using MODIS ocean color products (specifically Kd490) as a basis, combined with a geographic segmentation model, different marine areas are inverted separately.

2. Introducing the preliminary prediction results of an empirical physical model based on the blue-green ratio method into the input end of an RFRE model, enhancing the interpretability of the model learning process.

The method was tested in the coastal waters around the Beibu Gulf, near Hainan Island, and its performance was assessed using the overall monthly average coefficient of determination (R^2), mean bias (Bias), and root mean square error (RMSE) on the test set. The results indicate excellent performance of the model in terms of data fitting and prediction accuracy. The monthly average coefficient of determination (R^2) on the test set reached 0.90, with a maximum

of 0.97, and the monthly average root mean square error (RMSE) was less than 0.02.

We also demonstrate and analyze the annual variation of Kd490, comparing the 10m resolution Kd490 results obtained from Sentinel-2 inversion with MODIS products. The results show overall consistency between the reconstructed results and MODIS data, with Sentinel-2 providing more detailed information due to its higher resolution. This demonstrates the effectiveness of the proposed model in the research area. In addition, the method proposed in this paper can be further applied to other ocean color parameters and different marine areas in the future, providing valuable references for conducting finer monitoring of marine color parameters and studying small-scale oceanic activities.

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

YY: Writing – original draft, Methodology, Investigation, Conceptualization. ZW: Writing – review & editing, Methodology, Conceptualization. PC: Writing – review & editing, Methodology, Conceptualization. XS: Writing – review & editing, Methodology, Conceptualization. WK: Writing – review & editing, Supervision. GH: Writing – review & editing, Project administration. RS: Writing – review & editing, Project administration.

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

Asim, M., Brekke, C., Mahmood, A., Eltoft, T., and Reigstad, M. (2021). Improving chlorophyll-A estimation from sentinel-2 (MSI) in the barents sea using machine learning. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 14, 5529–5549. doi: 10.1109/JSTARS.2021.3074975

Blondeau-Patissier, D., Gower, J. F. R., Dekker, A. G., Phinn, S. R., and Brando, V. E. (2014). A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. *Prog. Oceanography* 123, 123–144. doi: 10.1016/j.pocean.2013.12.008

Braga, F., Zaggia, L., Bellafiore, D., Bresciani, M., Giardino, C., Lorenzetti, G., et al. (2016). Mapping turbidity patterns in the Po river prodelta using multi-temporal landsat 8 imagery. *Estuarine Coast. Shelf Sci.* 198. doi: 10.1016/j.ecss.2016.11.003

Camps-Valls, G., Bruzzone, L., Rojo-Álvarez, J. L., and Melgani, F. (2006). Robust support vector regression for biophysical variable estimation from remotely sensed images. *Geosci. Remote Sens. Letters IEEE* 3, 339–343. doi: 10.1109/LGRS.2006.871748

Cao, Z., Ma, R., Duan, H., Pahlevan, N., Melack, J., Shen, M., et al. (2020). A machine learning approach to estimate chlorophyll-a from landsat-8 measurements in inland lakes. *Remote Sens. Environ.* 248, 111974. doi: 10.1016/j.rse.2020.111974

Chen, G., Hou, Y., and Chu, X. (2011). Mesoscale eddies in the South China Sea: mean properties, spatiotemporal variability, and impact on thermohaline structure. *J. Geophysical Research: Oceans* 116. doi: 10.1029/2010JC006716

Chen, S., Hu, C., Barnes, B. B., Wanninkhof, R., Cai, W.-J., Barbero, L., et al. (2019). A machine learning approach to estimate surface ocean *p*CO2 from satellite measurements. *Remote Sens. Environ.* 228, 203–226. doi: 10.1016/j.rse.2019.04.019

Dong, S., He, H., Fu, B., Fan, D., and Wang, T. (2021). Remote sensing retrieval of chlorophyll-a concentration in the coastal waters of Hong Kong based on landsat-8 OLI and sentinel-2 MSI sensors. *IOP Conf. Series: Earth Environ. Sci.* 671, 0120335. doi: 10.1088/1755-1315/671/1/012033

Duan, P., Zhang, F., Jim, C. Y., Tan, M. L., Cai, Y., Shi, J., et al. (2023). Reconstruction of sentinel images for suspended particulate matter monitoring in arid regions. *Remote Sens.* 15, 872. doi: 10.3390/rs15040872

Fang, G., Chen, H., Wei, Z., Wang, Y., Wang, X., and Li, C. (2006). Trends and interannual variability of the South China Sea surface winds, surface height, and surface temperature in the recent decade. *J. Geophysical Res.* 111. doi: 10.1029/2005JC003276

Gernez, P., Lafon, V., Lerouxel, A., Curti, C., Lubac, B., Cerisier, S., et al. (2015). Toward sentinel-2 high resolution remote sensing of suspended particulate matter in very turbid waters: SPOT4 (Take5) experiment in the loire and gironde estuaries. *Remote Sens.* 7, 9507. doi: 10.3390/rs70809507

Giardino, C., Brando, V. E., Dekker, A. G., Strömbeck, N., and Candiani, G. (2007). Assessment of water quality in Lake Garda (Italy) using hyperion. *Remote Sens. Environ.* 109, 183–955. doi: 10.1016/j.rse.2006.12.017

Gonzalez Vilas, L., Spyrakos, E., and Torres, J. (2011). Neural network estimation of chlorophyll a from MERIS full resolution data for the coastal waters of galician rias (NW Spain). *Remote Sens. Environ.* 115, 524–535. doi: 10.1016/j.rse.2010.09.021

Guo, M., Xiu, P., Li, S., Chai, F., Xue, H., Zhou, K., et al. (2017). Seasonal variability and mechanisms regulating chlorophyll distribution in mesoscale eddies in the South China Sea. *J. Geophysical Research: Oceans* 122, 5329–5347. doi: 10.1002/2016JC012670

Harmel, T., Chami, M., Tormos, T., Reynaud, N., and Danis, P.-A. (2018). Sunglint correction of the multi-spectral instrument (MSI)-SENTINEL-2 imagery over inland and sea waters from SWIR bands. *Remote Sens. Environ.* 204, 308–321. doi: 10.1016/j.rse.2017.10.022

He, Q., Zhan, H., and Li, Z. (2016). Eddy effects on surface chlorophyll in the Northern South China Sea: mechanism investigation and temporal variability analysis. *Deep Sea Res. Part I: Oceanographic Res. Papers* 112. doi: 10.1016/j.dsr.2016.03.004

Hedley, J., Roelfsema, C., Brando, V., Giardino, C., Kutser, T., Phinn, S., et al. (2018). Coral reef applications of sentinel-2: coverage, characteristics, bathymetry and benthic mapping with comparison to landsat 8. *Remote Sens. Environ.* 216, 598–614. doi: 10.1016/j.rse.2018.07.014

Jaelani, L., Limehuwey, R., Kurniadin, N., Pamungkas, A., Koenhardono, E., and Sulisetyono, A. (2016). Estimation of TSS and Chl-a concentration from landsat 8-OLI: the effect of atmosphere and retrieval algorithm. *IPTEK J. Technol. Sci.* 27, 16–23. doi: 10.12962/j20882033.v27i1.1217

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). "Tree-based methods," in *An Introduction to Statistical Learning: With Applications in R*. Eds. G. James, D. Witten, T. Hastie and R. Tibshirani (Springer, New York, NY), 303–335. doi: 10.1007/978-1-4614-7138-7_8

Kahru, M., Gille, S. T., Murtugudde, R., Strutton, P. G., Manzano-Sarabia, M., Wang, H., et al. (2010). Global correlations between winds and ocean chlorophyll. *J. Geophysical Research: Oceans* 115. doi: 10.1029/2010JC006500

Kremezi, M., and Karathanassi, V. (2020). Data fusion for increasing monitoring capabilities of sentinel optical data in marine environment. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 13, 4809–4815. doi: 10.1109/JSTARS.2020.3018050

Liu, H., Li, Q., Shi, T., Hu, S., Wu, G., and Zhou, Q. (2017). Application of sentinel 2 MSI images to retrieve suspended particulate matter concentrations in poyang lake. *Remote Sens.* 9, 761. doi: 10.3390/rs9070761

Liu, Q., Kaneko, A., and Su, J. (2008). Recent progress in studies of the South China Sea circulation. *J. Oceanography* 64, 753–762. doi: 10.1007/s10872-008-0063-8

Lunetta, K. L., Hayward, L.B., Segal, J., and Van Eerdewegh, P. (2004). Screening large-scale association study data: exploiting interactions with random forests. *BMC Genet.* 5, 32. doi: 10.1186/1471-2156-5-32

Maier, P., Keller, S., and Hinz, S. (2018). Estimation of chlorophyll a, diatoms and green algae based on hyperspectral data with machine learning approaches.

Manzo, C., Braga, F., Zaggia, L., Brando, V., Giardino, C., Bresciani, M., et al. (2018). Spatio-temporal analysis of prodelta dynamics by means of new satellite generation: the case of Po river by landsat-8 data. *Int. J. Appl. Earth Observation Geoinformation* 66, 210–225. doi: 10.1016/j.jag.2017.11.012

Marzano, F., Iacobelli, M., Orlandi, M., and Cimini, D. (2020). Coastal water remote sensing from sentinel-2 satellite data using physical, statistical, and neural network retrieval approach. *IEEE Trans. Geosci. Remote Sens.*, 1–14. doi: 10.1109/TGRS.2020.2980941

Matthews, M. (2011). A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. *Int. J. Remote Sens. - Int. J. Remote SENS* 32, 1–45. doi: 10.1080/01431161.2010.512947

Mobley, C. (1994). *Light and Water: Radiative Transfer in Natural Waters* (Academic Press).

Morel, A., and Prieur, L. (1977). Analysis of variations in ocean color1. *Limnology Oceanography* 22, 709–722. doi: 10.4319/lo.1977.22.4.0709

Mouw, C., Greb, S., Aurin, D., DiGiacomo, P., Lee, Z., Twardowski, M., et al. (2015). Aquatic color radiometry remote sensing of coastal and inland waters: challenges and recommendations for future satellite missions. *Remote Sens. Environ.* 160, 15–30. doi: 10.1016/j.rse.2015.02.001

O'Reilly, J. (2000). Ocean color chlorophyll a algorithms for SeaWiFS, OC2, and OC4: version 4. *SeaWiFS Postlaunch Calibration Validation Analyses* 11, 9–23.

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., Schalles, J., Binding, C., Cao, Z., Ma, R., et al. (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. doi: 10.1016/j.rse.2019.111604

Rodrigues, G., Potes, M., Costa, M. J., Novais, M. H., Penha, A. M., Salgado, R., et al. (2020). Temporal and spatial variations of secchi depth and diffuse attenuation coefficient from sentinel-2 MSI over a large reservoir. *Remote Sens.* 12, 768. doi: 10.3390/rs12050768

Sauer, M., Roesler, C., Werdell, J., and Barnard, A. (2012). Under the hood of satellite empirical chlorophyll a algorithms: revealing the dependencies of maximum band ratio algorithms on inherent optical properties. *Optics Express* 20, 20920–20933. doi: 10.1364/OE.20.020920

Siegel, D. A., Peterson, P., McGillicuddy, D. J., Maritorena, S., and Nelson, N. B. (2011). Bio-optical footprints created by mesoscale eddies in the Sargasso Sea: EDDY FOOTPRINTS IN THE SARGASSO SEA. *Geophysical Res. Lett.* 38. doi: 10.1029/2011GL047660

Song, K., Li, L., Li, S., Tedesco, L. P., Duan, H., Li, Z., et al. (2014). Using partial least squares-artificial neural network for inversion of inland water chlorophyll-a. *Geosci. Remote Sensing IEEE Trans. On* 52, 1502–1517. doi: 10.1109/TGRS.2013.2251888

Tiarasani, A., Siregar, V., and Lumban-Gaol, J. (2023). Estimation of diffuse coefficient attenuation (K d 490) using sentinel 2A in Panggang Island and its surrounding water. *IOP Conf. Series: Earth Environ. Sci.* 1251, 12029. doi: 10.1088/1755-1315/1251/1/012029

Tomlinson, M. C., Stumpf, R. P., and Vogel, R. L. (2019). Approximation of diffuse attenuation, Kd, for MODIS high-resolution bands. *Remote Sens. Lett.* 10, 178–855. doi: 10.1080/2150704X.2018.1536301

Wang, M., Son, S., and Harding, L. W. (2009). Retrieval of diffuse attenuation coefficient in the chesapeake bay and turbid ocean regions for satellite ocean color applications. *J. Geophysical Res.* 114, C10011. doi: 10.1029/2009JC005286

Xiu, P., and Chai, F. (2011). Modeled biogeochemical responses to mesoscale eddies in the South China Sea. *J. Geophysical Research: Oceans* 116. doi: 10.1029/2010JC006800

Yang, W., Matsushita, B., Chen, J., and Fukushima, T. (2011). Estimating constituent concentrations in case II waters from MERIS satellite data by semi-analytical model optimizing and look-up tables. *Remote Sens. Environ.* 115, 1247–1259. doi: 10.1016/j.rse.2011.01.007

Yu, Y., Xing, X., Liu, H., Yuan, Y., Wang, Y., and Chai, F. (2019). The variability of chlorophyll-a and its relationship with dynamic factors in the basin of the South China Sea. *J. Mar. Syst.* 200, 103230. doi: 10.1016/j.jmarsys.2019.103230

Zhang, H. K., Roy, D. P., Yan, L., Li, Z., Huang, H., Vermote, E., et al. (2018). Characterization of sentinel-2A and landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences. *Remote Sens. Environ.* 215, 482–494. doi: 10.1016/j.rse.2018.04.031

Zhang, Y., Zhang, Y., Shi, K., Zha, Y., Zhou, Y., and Liu, M. (2016). A landsat 8 OLI-based, semianalytical model for estimating the total suspended matter concentration in the slightly turbid xin'anjiang reservoir (China). *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 9, 398–413. doi: 10.1109/JSTARS.2015.2509469

Zhu, W., and 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, 3286–3298. doi: 10.1109/TGRS.2012.2224117

Zhuang, W., Xie, S.-P., Wang, D., Taguchi, B., Aiki, H., and Sasaki, H. (2010). Intraseasonal variability in sea surface height over the South China Sea. *J. Geophysical Research: Oceans* 115. doi: 10.1029/2009JC005647