

Article

Integrating Multi-Source Environmental Variables with Sentinel-3 OLCI Imagery for Interpretable Retrieval of Eutrophication Parameters in Bohai

Qiguang Xing ^{1,2}, Zhen Sun ¹, Jingping Xu ¹, Siwen Gao ¹, Shanwei Liu ²  and Yanlong Chen ^{1,*}

¹ National Marine Environmental Monitoring Center, Ministry of Ecology Environment, Dalian 116023, China; s23160023@s.upc.edu.cn (Q.X.); zsun@nmemc.org.cn (Z.S.); xjp.pp@126.com (J.X.); siwengao95@163.com (S.G.)

² College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China; 20110052@upc.edu.cn

* Correspondence: ylchen@nmemc.org.cn

Highlights

What are the main findings?

- Interpretable models were developed based on optimal combination of spectral environmental features to estimate DIN and SRP in the Bohai Sea.
- Annual average concentrations of both DIN and SRP in the Bohai Sea have remained relatively stable, exhibiting a slight downward trend.

What are the implications of the main findings?

- To address the issue of weak optical responses for non-optically active water quality parameters, retrieval models that integrate spectral and environmental characteristics can effectively improve model performance. The proposed strategy for optimizing band combinations further enhances inversion accuracy.
- This method provides accurate and effective support for remote sensing monitoring of marine eutrophication parameters and water quality assessment.

Abstract

Satellite remote sensing technology, characterized by high spatiotemporal resolution and long-term continuous observations, has become a critical tool for water quality monitoring. Excessive inputs of nutrients, particularly nitrogen and phosphorus, into marine environments can induce eutrophication and a range of associated ecological problems. Dissolved Inorganic Nitrogen (DIN) and Soluble Reactive Phosphorus (SRP) are key parameters governing water quality. However, their inherently weak spectral response poses a persistent challenge for the construction of accurate satellite-based inversion models. Incorporating environmental variables into model input features alongside spectral reflectance represents a promising approach to improving inversion performance. Taking the Bohai Sea as a case study, multi-source environmental variables were integrated with remote sensing spectral features, and the optimal feature combination and corresponding model were identified using Bayesian optimization and an iterative feature-importance screening strategy. Bayes-CatBoost demonstrated superior performance in estimating DIN and SRP, achieving R² values of 0.82 and 0.68, respectively. After incorporating multi-source environmental variables, inversion errors decreased by 17.41% for DIN and 16.79% for SRP relative to models based solely on multispectral remote sensing data. Application of the proposed model to Sentinel-3 OLCI imagery enabled the generation of daily DIN and SRP distributions for the Bohai Sea. The results indicate that from 2018 to 2023, DIN and SRP concentrations in the



Academic Editor: Hatim Sharif

Received: 12 March 2026

Revised: 7 May 2026

Accepted: 23 May 2026

Published: 2 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Bohai Sea exhibited an overall stable but slightly declining trend with pronounced seasonal variability, accompanied by a gradual improvement in water quality.

Keywords: marine water quality inversion; eutrophication parameters; multi-source environmental variables; Bayesian-optimized CatBoost; Bohai Sea

1. Introduction

The rapid economic development of coastal regions, together with population growth, urban expansion, and intensified industrialization, has directly or indirectly resulted in excessive discharges of nutrients, particularly nitrogen and phosphorus, into coastal waters [1]. Excessive nutrient inputs into marine environments lead to water pollution and subsequently trigger eutrophication, frequent algal blooms, and hypoxic conditions. These processes pose serious threats to marine biodiversity and human health [2]. Therefore, timely and accurate monitoring of marine water quality, particularly the estimation of key parameters such as inorganic nitrogen and reactive phosphorus, is of paramount importance. Traditional fixed-point sampling methods are labor-intensive and costly, and their limited spatial coverage makes it difficult to comprehensively characterize the water quality of large water bodies [3,4]. Satellite remote sensing technology, characterized by short observation cycles, extensive spatial coverage, and strong retrospective data capabilities, has become an indispensable tool for large-scale water quality monitoring [5].

In the context of satellite-based monitoring, understanding the responses of water quality parameters to spectral characteristics is fundamental for remote sensing water quality monitoring [6]. Optically active parameters, such as chlorophyll-a and total suspended matter, exhibit distinct responses to specific spectral bands, enabling development of empirical or statistical regression models that link these parameters to spectral reflectance [7–9]. In contrast, non-optically active parameters such as DIN and SRP exhibit inherently weak spectral responses, making their variations difficult to detect directly from satellite-derived spectrum [10,11]. Traditionally, these parameters have been indirectly estimated by exploiting their correlations with optically active parameters. For instance, elevated nitrogen and phosphorus concentrations can promote algal proliferation, which increases chlorophyll-a levels and modifies optical properties [12,13]. Based on spectral information derived from optically active parameters, regression models can be developed to estimate optically insensitive indicators. However, substantial variability in optical properties among different water bodies limits the transferability of such models. Moreover, conventional statistical approaches often suffer from limited predictive accuracy [14,15]. Machine learning (ML) techniques offer significant advantages in capturing complex nonlinear relationships between non-optically active parameters and spectral information [16–18]. Consequently, ML-based approaches have emerged as the preferred choice for developing inversion models for non-optically active water quality parameters [15,19].

Relevant studies demonstrate that incorporating natural environmental drivers, such as hydrological and meteorological factors, can enhance the inversion capability and stability of water quality models. Different environmental factors exert distinct influences on non-optically active water quality parameters [2]. Fluctuations in water levels affect the transport and degradation of pollutants, thereby inducing changes in water quality [20,21]. Rainfall enhances runoff generation and convergence processes, thereby influencing nutrient levels in water bodies [22]. Appropriate sea surface temperatures accelerate processes such as nitrogen fixation and nitrification, thereby influencing the nitrogen cycle [23]. Ref. [23] incorporated ERA5 meteorological data, effectively mitigating model overfitting

while maintaining high predictive accuracy. Ref. [24] constructed a spectral–environmental model by integrating average temperature and cumulative precipitation during the seven days prior with UAV-derived spectral data, achieving significant reductions in prediction errors for total phosphorus, ammonia nitrogen, and chemical oxygen demand compared with conventional spectral-based ML models. Overall, incorporating natural environmental factors as feature variables has become a widely adopted strategy for constructing inversion models of non-optically active water quality parameters. However, most existing studies select spectral and environmental features solely based on statistical correlations with target variables. This feature selection strategy generally overlooks interactions between spectral data and natural environmental factors, thereby limiting the model’s ability to achieve an optimal combination of features for best performance.

To address this issue, three ML models were constructed based on Bayesian optimization, integrating multi-source environmental variables with spectral features as model inputs. Feature importance rankings derived from the model were used to iteratively remove the least significant features during training, thereby identifying the optimal feature combination for each parameter. Retrieval models for DIN and SRP were developed using Sentinel-3 OLCI multispectral data (400–885 nm) combined with environmental variables influencing marine water quality, and applied to characterize the spatiotemporal distributions of DIN and SRP in the Bohai Sea. The main contributions of this study are summarized as follows:

- (1) An iterative feature selection strategy based on model-derived feature importance ranking was proposed to identify the optimal combination of spectral and environmental variables for non-optical parameter retrieval.
- (2) SHAP value analysis revealed the contribution mechanisms of different environmental factors to the distribution of DIN and SRP.
- (3) A daily-resolution spatiotemporal distribution dataset for DIN and SRP in the Bohai Sea (2018–2023) was constructed and the distribution patterns of DIN and SRP were quantitatively analyzed, and the water quality of the Bohai Sea was assessed.

2. Materials

2.1. Study Area

The Bohai Sea (Figure 1) is a semi-enclosed, shallow marginal sea located at the northeastern margin of mainland China, covering approximately 77,000 km² with an average depth of about 18 m [25–27]. As China’s only inland sea, it is bordered by Liaodong Bay, Bohai Bay and Laizhou Bay, and connects to the northern Yellow Sea through the Bohai Strait, with a total coastline of approximately 3800 km [28]. The shore is covered by densely populated and highly industrialized regions where strong river inputs, together with intensive aquaculture, urban runoff, and industrial effluents, have led to elevated levels of DIN and SRP, making the region highly susceptible to eutrophication and harmful algal blooms [29–31]. High concentrations of suspended particulate matter (SPM), colored dissolved organic matter (CDOM), and phytoplankton interact nonlinearly with nutrient dynamics in the Bohai Sea, which generates pronounced optical complexity [32]. The complexity poses a substantial challenge for satellite-based retrieval of water quality parameters, particularly non-optically active constituents such as DIN and SRP. Overcoming these challenges to achieve reliable nutrient monitoring is essential for marine ecological protection and water quality management in this region.

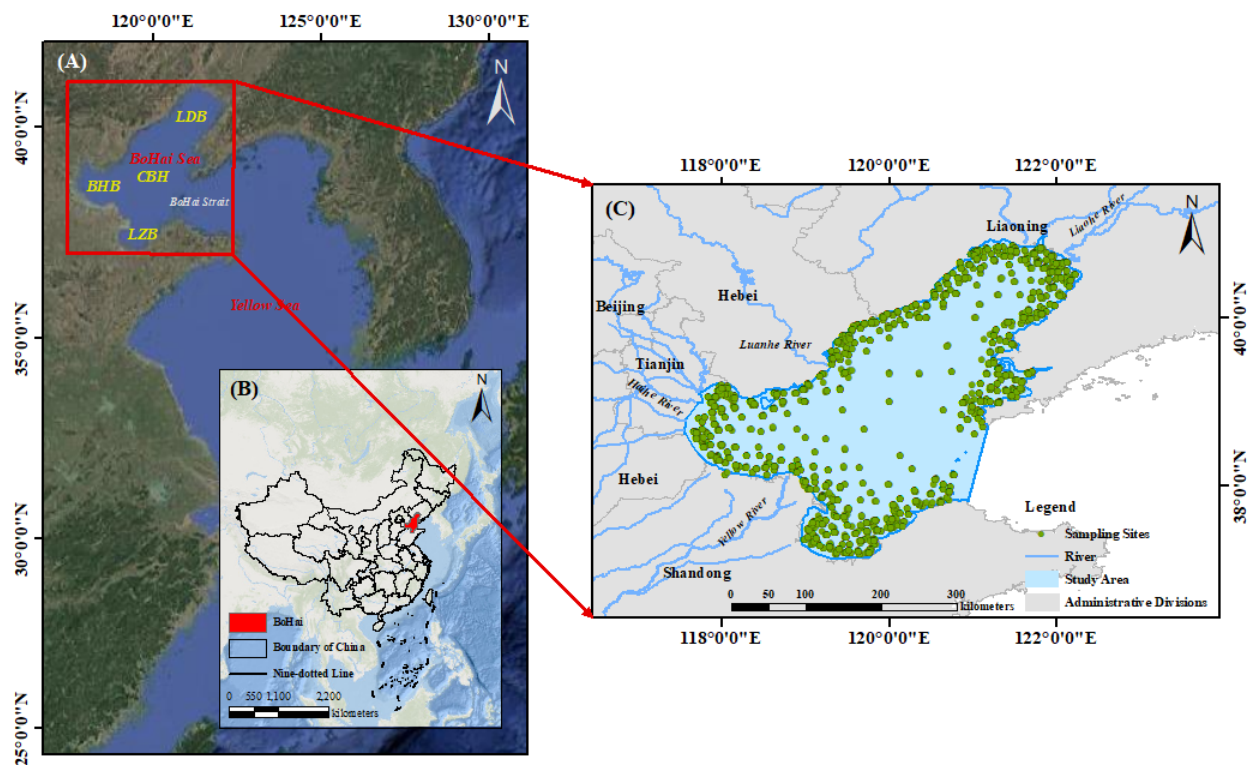


Figure 1. Satellite image shows the study area (**A**) and geographical location of it (**B**), (**C**) shows the spatial distribution of in situ water quality sampling sites.

2.2. Data and Preprocessing

2.2.1. In Situ Water Quality Dataset

In situ water quality monitoring dataset used in this study was acquired by the National Marine Environmental Monitoring Center (NMEMC), Ministry of Ecology and Environment, China, from 2018 to 2023. The sampling time for each year is shown in Table 1. Records with missing information on station coordinates, sampling time, or measured DIN and SRP concentrations were excluded. DIN and SRP were measured and analyzed in accordance with China's Seawater Quality Standards (GB 3097-1997) (Table 2). Figure 1C illustrates the spatial distribution of the in situ water quality monitoring stations used in this study (green dots). The histograms of the frequency distributions of the DIN and SRP data used in this study is shown in text.

Table 1. Sampling sites and time for in situ measurements.

Year	Number of Station Sites	Sampling Time (Month)
2018	322	2, 3, 5, 6, 7, 8, 10, 11, 12
2019	181	3, 5, 6, 7, 8, 10, 11, 12
2020	252	3, 5, 7, 8, 10, 11
2021	237	2, 3, 4, 5, 7, 8, 10, 11
2022	237	4, 5, 7, 8, 10, 11
2023	237	4, 5, 7, 8, 10, 11

Table 2. In situ data statistics for DIN and SRP. Statistical parameters include the minimum, maximum, mean, standard deviation, and coefficient of variation (CV).

	Min	Max	Mean	Standard Deviation	CV (%)
DIN	0.003	5.401	0.246	0.271	110.160
SRP	0.0003	0.2230	0.0088	0.0103	117.7690

2.2.2. Satellite Remote Sensing Images

In this study, the Sentinel-3 Ocean and Land Color Imager (OLCI) Level-1 product (OL_1_EFR) was used as the primary satellite data source. After excluding images affected by extensive cloud and cloud-shadow contamination and applying quality control procedures, a total of 612 daily remote sensing images were retained for the period 2018–2023. Atmospheric correction was performed using the Dark Spectrum Fitting (DSF) algorithm implemented in the ACOLITE processor [33]. During the generation of Level-2 products, 5 bands of the 21 Sentinel-3 OLCI bands are used as auxiliary inputs for atmospheric correction. Water-leaving reflectance from the first 15 spectral bands within the visible–near-infrared range (400–900 nm) was extracted and used as spectral input for model construction. In situ measurements of DIN and SRP were matched with atmospheric-corrected Sentinel-3 products using a 3×3 pixels spatial window and a temporal window of ± 1 day. This procedure resulted in 1010 SRP and 1027 DIN matched observations, defined as valid water-reflectance pairs with quasi-synchronous satellite imagery.

2.2.3. Environmental Driving Data

The Copernicus Marine Environment Monitoring Service (CMEMS) is the marine component of the European Union’s Earth observation program (<https://marine.copernicus.eu>, accessed on 13 July 2025). It provides global and regional information on physical oceanography, sea ice, and biogeochemical variables. GLORYS12V1 is CMEMS’s global ocean eddy-resolving reanalysis dataset, with a horizontal resolution of $1/12^\circ$ and 50 vertical layers. It provides daily data from 1993 onwards. In this study, sea water temperature (thetao), salinity (so), velocity of eastward and northward (uo and vo), and sea surface height above geoid (zos) from the surface layer of the Bohai Sea were used for the period 2018–2023.

Precipitation data (TD7pre) were obtained from NASA’s Global Precipitation Measurement (GPM) mission (<https://gpm.nasa.gov/data/IMERG>, accessed on 15 July 2025). Specifically, the GPM Integrated Multi-satellite Retrievals for GPM (IMERG) Final Precipitation L3 dataset (1-day, $0.1^\circ \times 0.1^\circ$, V07; GPM_3IMERGDF) was used. The dataset used in this study spans 2018–2023, with a spatial resolution of $0.1^\circ \times 0.1^\circ$ and a temporal resolution of 1 day.

First of all, we applied a mask to the CMEMS and GPM precipitation data to retain only the imagery within the Bohai Sea. Both CMEMS and GPM datasets were resampled (bilinear for them) to a spatial resolution of 300 m to match Sentinel-3 OLCI data. Corresponding pixel values were matched to latitude and longitude coordinates according to the in situ measurements.

3. Methods

As shown in Figure 2, the overall technical approach of this study involved the following parts: data preparation, model development, and the application and analysis of results. Data preparation primarily involves spatio-temporal matching of spectral data, environmental data, and field measurement locations to construct datasets of spectral and environmental features. Subsequently, the optimal feature combinations for each model are selected using an iterative method based on the importance ranking of model features. Model development involves training the models and optimizing them using Bayesian methods, while model evaluation is used to select the inversion models. Finally, the results are applied to Sentinel-3 OLCI imagery to derive water quality products, and the results are analyzed.

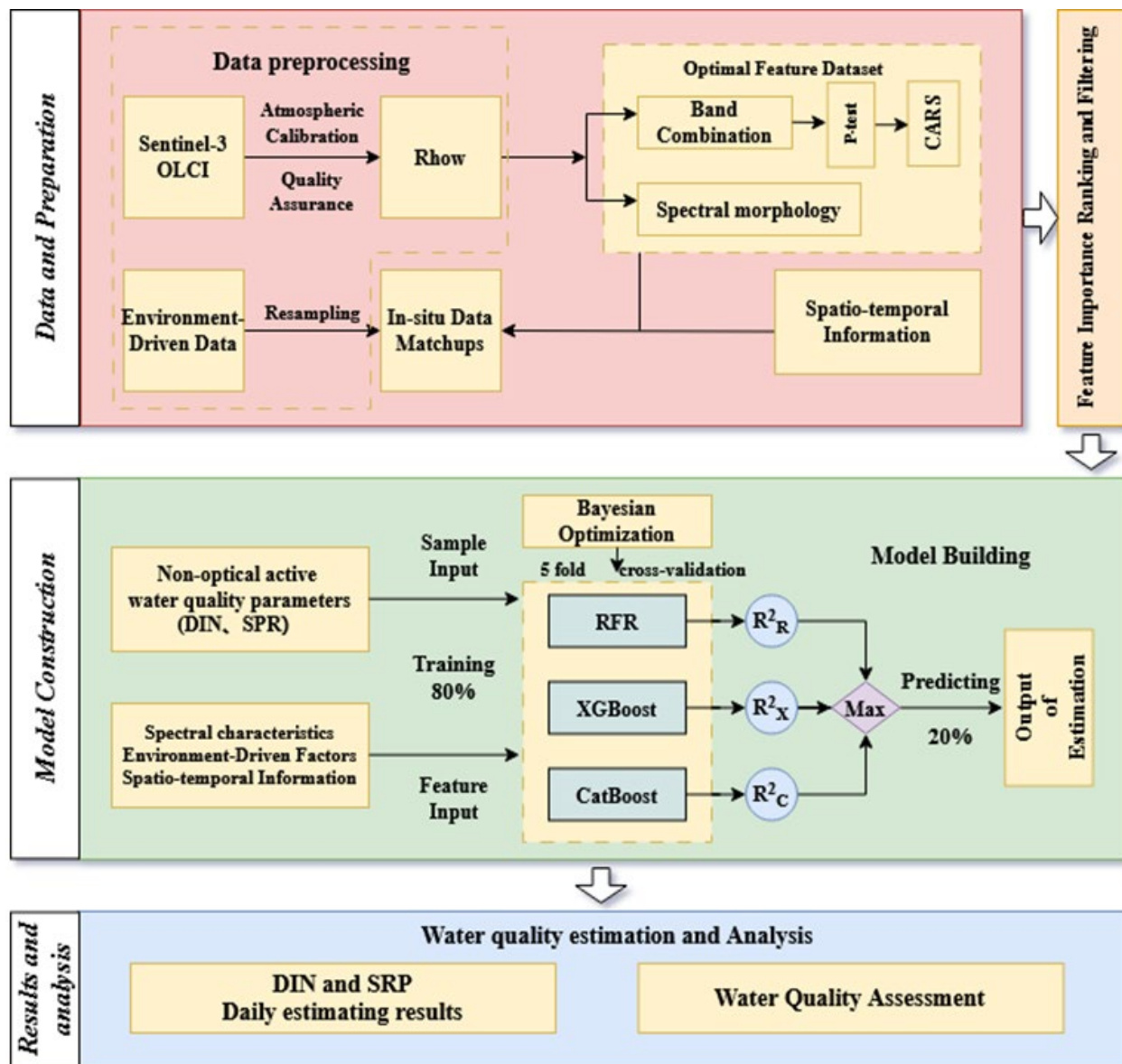


Figure 2. Flowchart of the proposed approach.

3.1. Extraction of Spectral Feature

3.1.1. Spectral Band Combination

Mathematical transformations were applied to individual bands and band combinations, including the Normalized Difference Index [34], Fractional Subtraction Difference [35], and Tri-band Index combinations [1] (Table 3). These operations included calculating the sum, difference, product, and ratio of any two bands; computing normalized indices and reciprocals for band pairs and performing three-band combinations. All possible band combinations were exhaustively enumerated, and their correlation coefficients with DIN and SRP were calculated, with significant features identified using *t*-tests ($p < 0.05$) [36,37]. Comprehensive band combinations can mitigate the inherent limitations of single-band water quality inversion and enhance feature representation [38]. To reduce feature redundancy and obtain optimal variables, the Competitive Adaptive Reweighted Sampling (CARS) algorithm was applied for selection of band combinations. CARS integrates Monte Carlo sampling with partial least squares regression coefficients to identify variable subsets that minimize the cross-validation SHapley Additive exPlanations [39,40]. Fifty iterations of CARS were conducted on significantly correlated bands, and the vari-

able set yielding the lowest 10-fold cross-validation RMSE was selected as the optimal feature dataset.

Table 3. Band Combination include two-band and three-band combinations, where N0(i) represent single band, N1 (i, j) to N9 (i, j) represent the name of these combinations. The boundary of i, j, k is from 1 to 16, which sequentially represent the band of spectra.

Combination	Expression	Combination Name	Source
Two-band combinations	$B_i + B_j, B_i - B_j, B_i \times B_j, B_i / B_j$	N1 (i, j)–N4 (i, j)	[38]
	$(B_i - B_j) / (B_i + B_j)$	N5 (i, j)	[34]
	$\frac{1}{B_i} + \frac{1}{B_j}, \frac{1}{B_i} - \frac{1}{B_j}$	N6 (i, j), N7 (i, j)	[35]
Three-band combinations	$\left[\frac{1}{B_i} - \frac{1}{B_j} \right] \times B_k$	N8 (i, j, k)	[1]
	$[B_j - B_i] - [B_k - B_i] \times \frac{\lambda_j - \lambda_i}{\lambda_k - \lambda_i}$	N9 (i, j, k)	

3.1.2. Extraction of Spectral Morphological Features

Spectral bands and band combinations typically capture the spectral characteristics of only one or a few bands closely associated with the target variable. The complex relationships between non-optically active water quality parameters and satellite spectral bands indicate that reliance solely on specific bands or band combinations is insufficient to capture deeper spectral characteristics [14]. Variations in these parameters can alter spectral curve morphology, including changes in elevation, undulation, and curvature. To capture such structural information, the simplified Multispectral Morphological Combination Feature (MSMCF) method [41] was adapted for the extraction of multispectral morphological features, which derived mean reflectance, gradient, curvature, and spectral similarity to class-averaged spectrum.

First, the average values of spectral reflectance, gradient, and curvature were calculated. The expression is as follows:

$$L_1 = \frac{1}{15} [1 \ 1 \ \cdots \ 1 \ 1]_{1 \times 15} \quad (1)$$

$$L_2 = \begin{cases} \begin{bmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}, x = 1, 15 \\ \begin{bmatrix} -1 & 1 \end{bmatrix}, x \in [2, 14] \end{cases} \quad (2)$$

$$L_3 = [-1 \ 0 \ 1], x \in [1, 15] \quad (3)$$

In the formula, L_1 was used to calculate the average spectral reflectance, whereas L_2 and L_3 were used to calculate the gradient and curvature, respectively. Here, x denotes the spectral index, with a step size of 1. The average gradient and average curvature were computed as the mean of the gradient and curvature values across all wavelengths.

Second, hierarchical clustering was employed to classify water bodies, grouping spectral curves into three categories (Figure 3). Finally, the cosine similarity between each spectrum and the average spectrum of its cluster was calculated. The cosine similarity was calculated using the following formula:

$$G_1 = \frac{F \times F_1}{|F| \times |F_1|} \quad (4)$$

$$G_2 = \frac{F \times F_2}{|F| \times |F_2|} \quad (5)$$

$$G_3 = \frac{F \times F_3}{|F| \times |F_3|} \quad (6)$$

where F denotes characteristic points on the spectral curve, while F_1 – F_3 represent the obtained from clustering.

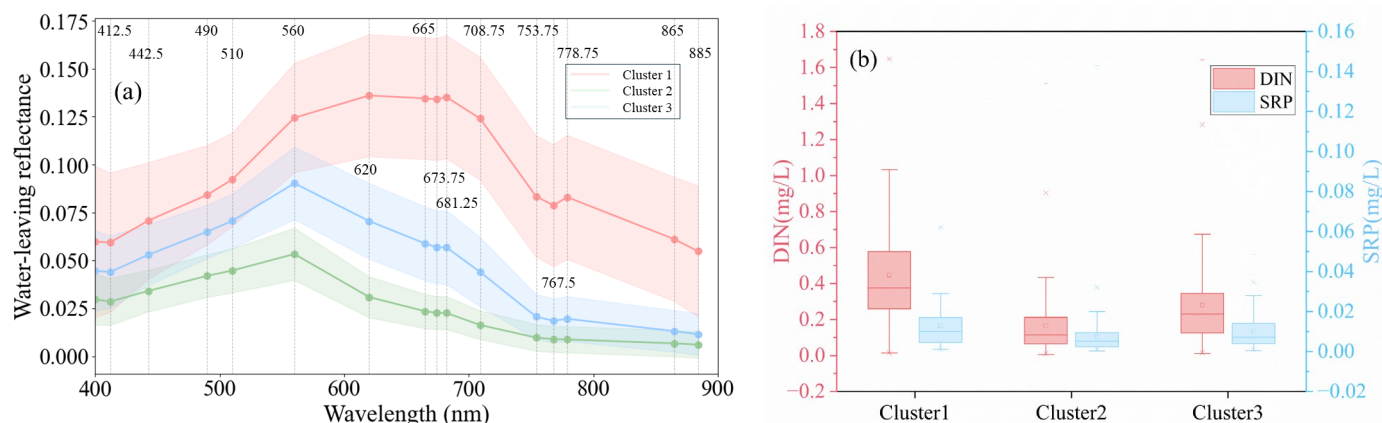


Figure 3. (a) Average spectral curve of each category. Light red, light green, and light blue represent the average spectral characteristics of the first, second, and third water body categories, respectively, where the shaded area represent the range of wavelengths for each spectral band after classification. (b) DIN and SRP concentration distribution of different clusters. Red represents DIN, and blue represents SRP.

3.2. Construction and Optimization of Model

3.2.1. Initial Feature Variable Input

Considering that dynamic changes in water quality parameters and remote sensing data are driven by inherent spatiotemporal heterogeneity in the environmental context [42–44]. Potential spatiotemporal variations in environmental variables were incorporated into the feature set [1,45]. Water quality parameters generally exhibit higher concentrations in estuarine and coastal regions than in open waters and show pronounced seasonal variability [46,47]. Spatial information was represented by geographic coordinates (latitude and longitude), while temporal variables include month of year (MOY), day of week (DOW), and day of year (DOY). Considering the cyclical nature of temporal variations in marine and environmental drivers [45,48], temporal features were expressed in aperiodic form. The periodic form was calculated using the following expression:

$$cT = 1 - \cos\left(\frac{t}{N_T} \times 2\pi\right) \quad (7)$$

where t represents the calendar month, the calendar weekday, and the calendar day, with N_T corresponding to 12, 7, and 365 respectively.

The spectral, environmental and spatiotemporal feature inputs for the inversion model was defined as Equation (8). We refer to the combination of bands after CARS feature filtering, spectral morphology, multi-source environmental and spatiotemporal features, as the initial feature input variables. The expression summarizes the specific categories and quantities of these initial input features used were in Tables 4 and S1.

$$NWQP = f(inputvariables) = f(C_{SC}, C_{ED}, C_{ST}) \quad (8)$$

where C denotes the feature variable category, C_{SC} , C_{ED} , C_{ST} represent, respectively, the factors of spectral characteristics, environmental data, spatiotemporal information, and f represents the nonlinear relationship between the input features and the water quality parameter to be inverted.

Table 4. Model input dataset for DIN and SRP, where 16 represents sixteen single band, 35 for DIN and 414 for SRP represent band combination, 6 in spectral characteristics represents six morphological features, 6 in spatio-temporal information represents time features, and 2 represents spatial features. NAWQPs represent non-optically active water quality parameters.

NAWQPs	Spectral Characteristics	Environmental Data	Spatio-Temporal Information
DIN	16 + 35 + 6	6	6 + 2
SRP	16 + 414 + 6	6	6 + 2

3.2.2. Model Building

To address performance differences among models and ensure accurate inversion of DIN and SRP, multiple inversion models were constructed for each parameter using Random Forest Regression (RFR), eXtreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost), with hyperparameters optimized via Bayesian optimization (Section 3.2.3). The initial feature dataset was input firstly (Section 3.2.1). Based on the feature importance rankings generated during each training iteration, the least important features were sequentially removed to identify the optimal feature combination for model inversion. After model training, the final inversion model was selected based on the performance of the three models on DIN and SRP. Using an R^2 -based voting mechanism, the model with the highest R^2 score was chosen as the inversion model for each parameter [24].

The dataset was randomly split into 80% for training and 20% for testing. To improve model generalization, five-fold cross-validation was applied to the training set, which was partitioned into five subsets, with four used for training and one for validation in each iteration. The mean R^2 across the five validation folds was used as the criterion for final feature selection.

3.2.3. Hyperparameter Optimization

The adjustment and optimization of model hyperparameters is critical for model performance [15]. Traditional methods, such as random search and grid search, for finding optimal hyperparameters are time-consuming and computationally intensive [49]. Bayesian optimization, which employs a Tree-structured Parzen Estimator (TPE) as a surrogate model, iteratively guides the search toward optimal hyperparameters based on previous evaluation results [15,50,51]. Accordingly, Bayesian optimization was adopted for hyperparameter tuning in this study, as formulated below:

$$I(x) = \max(0, y^* - y) \quad (9)$$

$$p(x|y) = \begin{cases} l(x) = p(x|y < y^*) \\ g(x) = p(x|y \geq y^*) \end{cases} \quad (10)$$

$$EI(x) = \mathbb{E}[I(x)] = \int_{-\infty}^{y^*} (y^* - y) p(y|x) dy = \int_{-\infty}^{y^*} (y^* - y) \frac{p(x|y)p(y)}{p(x)} dy \quad (11)$$

$$EI(x) = \frac{\gamma y^* l(x) - l(x) \int_{-\infty}^{y^*} p(y) dy}{\gamma l(x) + (1 - \gamma) g(x)} \propto \left(\gamma + \frac{g(x)}{l(x)} (1 - \gamma) \right)^{-1} \quad (12)$$

where $I(x)$ denotes the improvement function, x represent the hyperparameters set, y is the observed value, and y^* is the predefined quantization value, with the threshold set at the current observation corresponding to the $\gamma = 0.15$ quantile. The probability density functions $l(x)$ and $g(x)$ serve as TPE surrogate models, with expected improvement derived via maximum likelihood estimation serving as the acquisition function. Based on the threshold y^* , the two surrogate models were subsequently fitted. The candidate

points with the maximum $l(x)/g(x)$ was selected for objective evaluation. This process was iteratively repeated until convergence, yielding the optimal hyperparameter combination.

3.2.4. Model Evaluation

Four evaluation metrics—coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE)—were employed to assess the inversion models for DIN and SRP. The corresponding formulas are provided below.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (13)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (14)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (15)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (16)$$

where n represents the number of samples; y_i and $\hat{y}_i$ represent in situ measured and estimated values, respectively. $\bar{y}_i$ is the average value of the measured water quality concentration.

4. Results

4.1. Evaluation of Model Performance

Table 5 summarizes the regression performance of the RFR, XGBoost, and CatBoost models optimized using Bayesian optimization on the test dataset. Compared with the RFR and XGBoost models, CatBoost exhibited the highest accuracy and stability, particularly for SRP, and was therefore selected to construct the final inversion models for DIN and SRP. Figure 4 illustrates scatter plots of the test set results. Most data points are concentrated around the 1:1 line and display a relatively uniform distribution. The R^2 were 0.82 and 0.68, MAE were 0.06 and 0.003, RMSE were 0.099 mg/L and 0.004 mg/L, and MAPE were 29.34% and 40.89%, respectively. Table S2 presents the specific hyperparameter search space for each model.

Table 5. Accuracy evaluation of DIN and SRP retrieval models based on different Bayesian optimization machine learning algorithms (MAPE in % and RMSE in mg/L).

Model	DIN				SRP			
	R^2	MAE	RMSE	MAPE%	R^2	MAE	RMSE	MAPE%
Bayes-RFR	0.75	0.08	0.114	37.02	0.57	0.003	0.005	42.75
Bayes-XGBoost	0.80	0.07	0.093	37.27	0.57	0.003	0.005	39.51
Bayes-CatBoost	0.82	0.06	0.099	29.34	0.68	0.003	0.004	40.89

4.2. Model Consistency Verification

The model was applied to Sentinel-3 OLCI imagery to generate daily DIN and SRP maps for the Bohai Sea from 2018 to 2023 (Text S1). To account for the impacts of cloud cover and the proximity effect on satellite imagery [52], the Directed Interpolation of Missing Objects method was employed to reconstruct missing pixels [53,54]. Inversion results were validated using unmatched samples (due to missing pixels) as an independent dataset, which is not used in model training and validation. Figure 5 illustrates this process: satellite-

derived temporal trends and estimates agree well with in situ measurements, confirming the model's reliability.

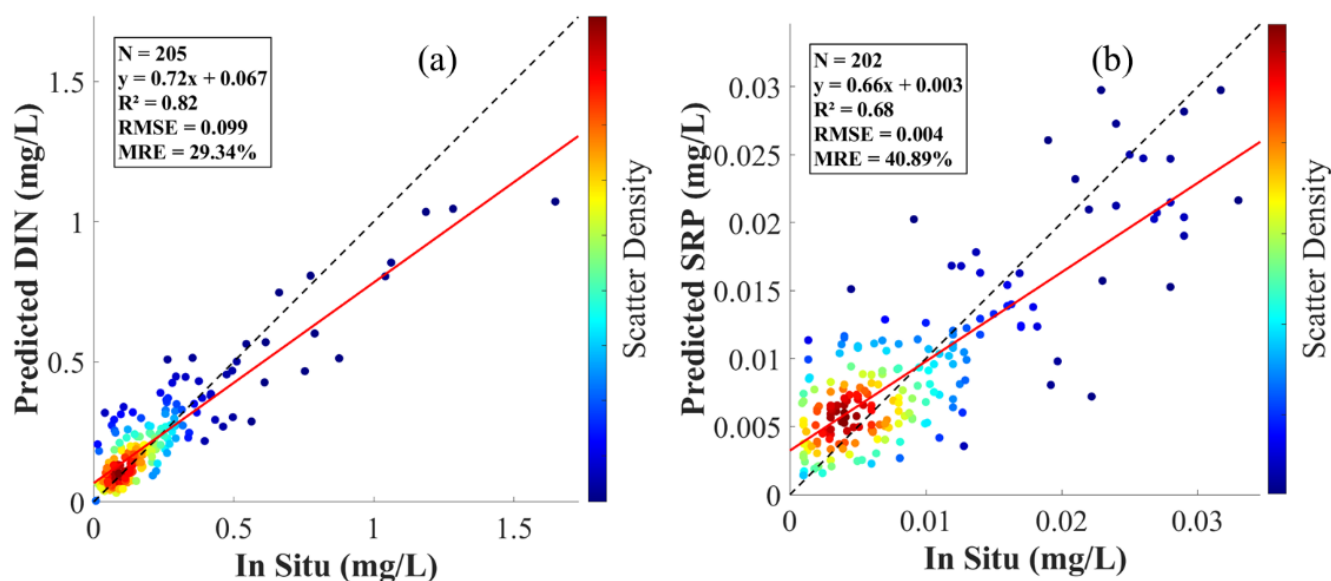


Figure 4. Prediction results of optimal model correlating with in situ value, which shows the performance on the test set of a Bayes-CatBoost model trained on the dataset using the optimal feature combination selected based on feature importance. Dotted line denotes the 1:1 line. Red line represents the results of the linear fit. The color scale encodes the normalized probability density of the samples, derived from a Gaussian kernel density estimation. (a,b) show the performance of the Bayes-CatBoost model on the DIN and SRP test set separately.

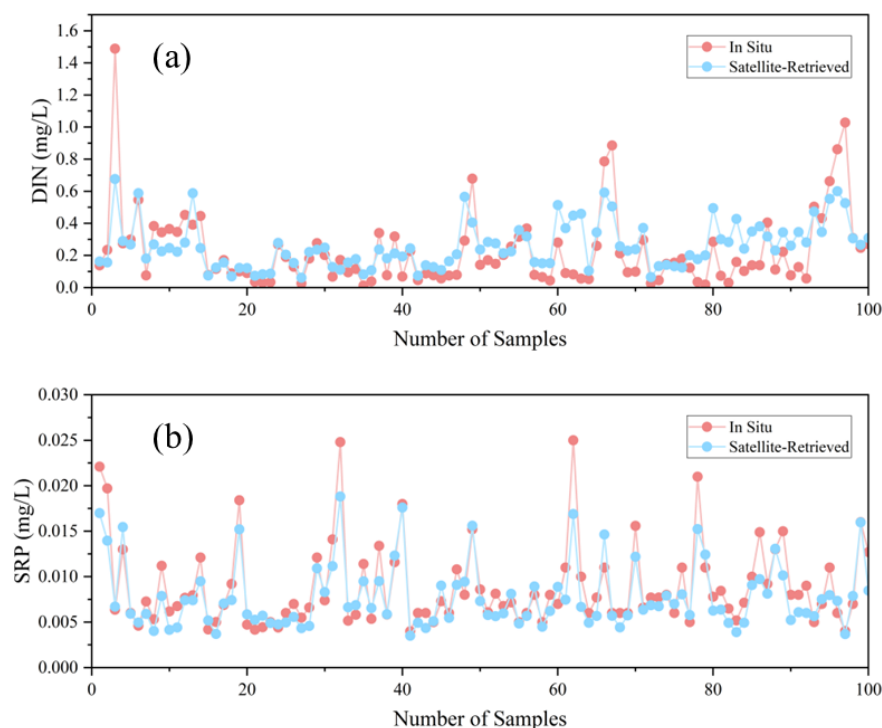


Figure 5. Model consistency evaluation results constructed using independent samples unused in the training and testing sets. (a) shows a comparison of actual value and predicted results about DIN, and (b) shows a comparison of actual value and predicted results about SRP.

4.3. Model Application

4.3.1. Annual Spatiotemporal Variation

Figure 6 illustrates the multi-year mean distributions and interannual variations of DIN and SRP, which used the method in Text S1. DIN concentrations are generally higher in coastal waters and decrease toward the central Bohai Sea, with high-value areas mainly located in Liaodong Bay and Laizhou Bay. The discharge of domestic sewage, agricultural runoff dominated by chemical fertilizers, and industrial wastewater transported by rivers has resulted in pronounced nutrient enrichment in coastal waters, especially in estuarine regions [55]. DIN concentrations at the Liao River estuary can exceed 0.8 mg/L. In contrast, low DIN levels, occur in offshore regions distant from major river mouths and densely populated areas, where dilution and exchange with the open ocean through the Bohai Strait dominate. The spatial distribution of SRP differs substantially from that of DIN, exhibiting a distinct north–south gradient with higher concentrations in the north. Elevated SRP levels were mainly observed in northern Liaodong Bay, the Dalian coastal area, and northern Bohai Bay, with maximum values reaching ~0.02 mg/L near the Liao and Hun River estuaries. Similar to DIN, SRP enrichment in nearshore waters is primarily driven by riverine inputs.

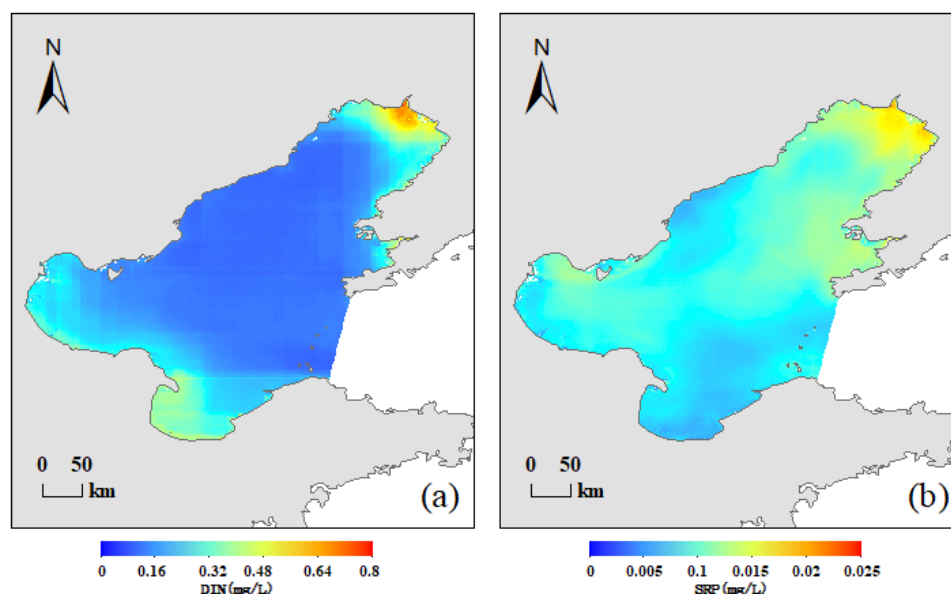


Figure 6. (a) shows the spatial distribution of DIN based on the average results over 6 years period 2018 to 2023. (b) shows the spatial distribution of SRP based on the average results over 6 years period 2018 to 2023.

Figure 7 illustrates the interannual variations in annual mean DIN and SRP concentrations in the Bohai Sea from 2018 to 2023. The interannual trends of DIN and SRP concentrations were further examined using linear regression and the MannKendall method. Both DIN and SRP exhibited weak decreasing trends over the study period, with estimated slopes of approximately $-0.0021 \text{ mg/L}^{-1} \cdot \text{year}^{-1}$ and $-0.00023 \text{ mg} \cdot \text{L}^{-1} \cdot \text{year}^{-1}$, respectively. Despite the overall downward trend, the annual average concentrations of DIN and SRP both rebounded to varying degrees in 2021. This may indicate that DIN and SRP levels in the Bohai Sea during that year increased due to environmental or human-induced disturbances.

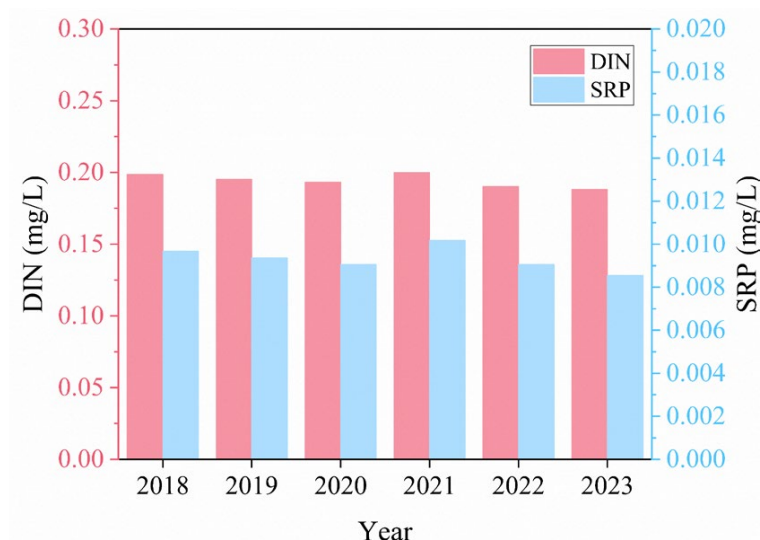


Figure 7. Annual DIN and SRP results for the Bohai Sea from 2018 to 2023. The annual average image of the Bohai Sea was calculated by averaging the statistics for each pixel, and then the average of the pixel concentrations in the annual average image for each year were calculated to obtain the annual average concentration.

4.3.2. Intra-Annual Spatiotemporal Variation

Seasonal statistics indicate that DIN and SRP in the Bohai Sea exhibit pronounced seasonal variability across the four seasons (Figures 8 and 9). Both DIN and SRP display consistent seasonal patterns, with minimum concentrations in summer and maximum concentrations in winter.

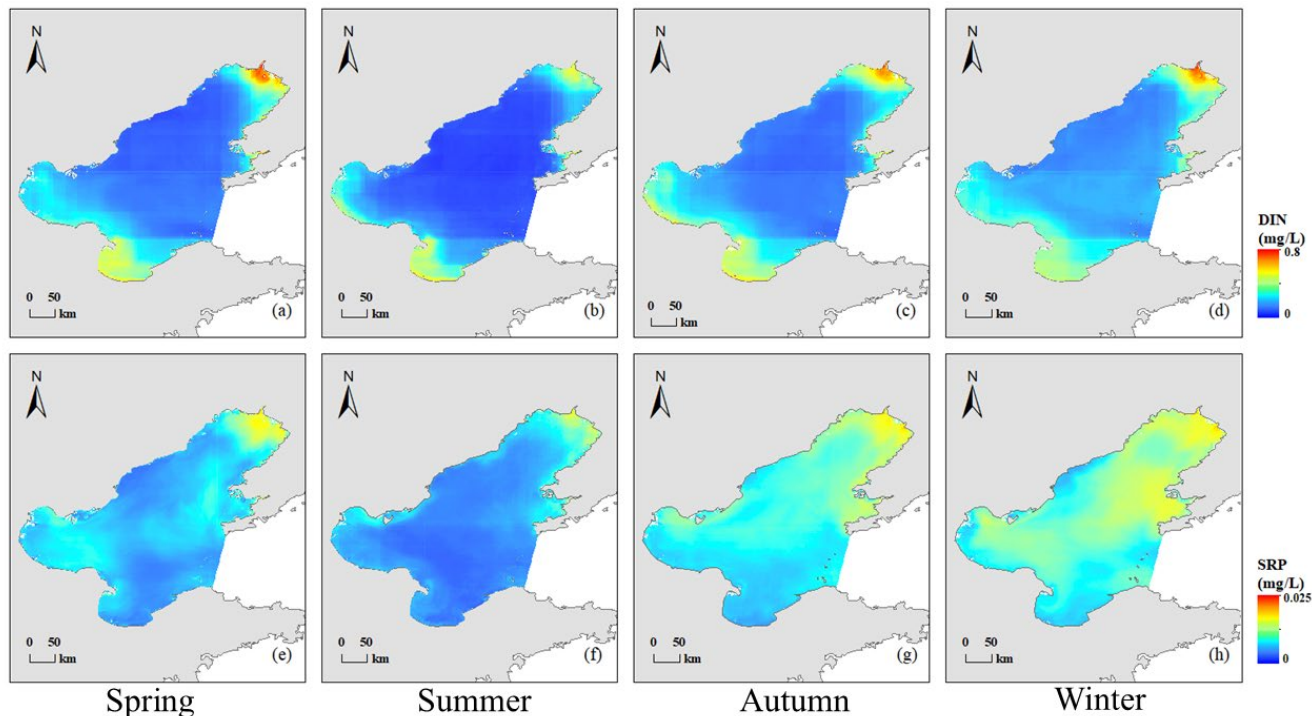


Figure 8. Seasonal distribution of DIN and SRP in the Bohai sea.

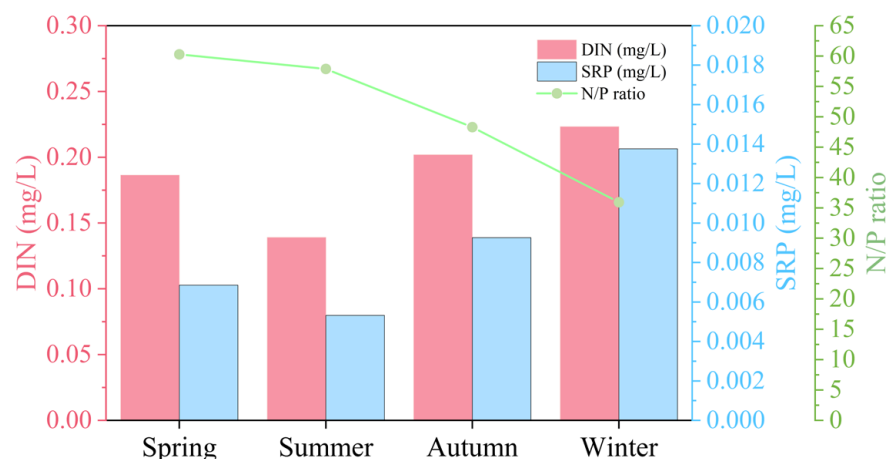


Figure 9. Seasonal concentrations of DIN and SRP in the Bohai sea, and the nitrogen-to-phosphorus ratios. The average of the daily concentrations recorded during each season between 2018 and 2023 as the average concentration for that season. Mean values were calculated from the observations and subjected to statistical significance testing ($p < 0.05$) to assess the reliability of differences.

In spring, DIN levels are generally elevated, particularly in estuarine and nearshore regions of northern Liaodong Bay and Laizhou Bay, reflecting strong terrestrial inputs from rivers and sewage outfalls and a rapid decrease toward offshore. During summer, DIN concentrations decline markedly across most areas, remaining relatively high only in localized coastal zones. This decrease is mainly attributed to enhanced phytoplankton uptake under warm conditions [56]. Notably, DIN levels along the Bohai Bay coast are higher in summer than in spring, likely due to increased precipitation and riverine nitrogen inputs during the wet season [57,58]. In autumn, DIN concentrations rebound as phytoplankton activity weakens, with high-value zones reappearing in northern Liaodong Bay and along the Laizhou Bay coast. In winter, DIN reaches its annual maximum and expands over most of the Bohai Sea, reflecting reduced biological uptake under low temperatures [59]. However, nearshore DIN levels in Bohai Bay and Laizhou Bay remain lower than in autumn, consistent with diminished precipitation and river runoff during the dry season.

The seasonal distribution of SRP exhibits pronounced differences. In spring, SRP concentrations are generally low, with high values mainly confined to northern Liaodong Bay. Summer shows a similar spatial pattern, but with slightly lower concentrations, as enhanced biological uptake offsets increased terrestrial phosphorus inputs. In autumn, SRP concentrations increase markedly, accompanied by a significant expansion of high-value zones. Persistent high-SRP zones occur in Liaodong Bay and adjacent waters, with elevated values also in northern Bohai Bay. SRP concentrations in the central Bohai Sea markedly exceed spring and summer levels. In winter, SRP reaches its annual maximum, with high-concentration areas further expanding across northern and eastern Liaodong Bay and along the coasts of Bohai Bay and Laizhou Bay.

DIN and SRP exhibit broadly consistent seasonal trends, whereas the mean N/P ratio (Text S2) remains relatively high during spring and summer. The N/P ratio reaches maximum in spring and then declines to a minimum in winter, indicating stronger seasonal variability in SRP than DIN. Seasonal mean N/P ratios in all four seasons exceed the Redfield ratio with a value of 16, highlighting the necessity for enhanced nitrogen input control in the Bohai Sea, especially during spring and summer.

Figure 10 illustrates the monthly variations in multi-year mean DIN and SRP concentrations in the Bohai Sea from 2018 to 2023 (Text S3). DIN concentrations range from 0.11 to 0.28 mg/L, while SRP varies from 0.004 to 0.016 mg/L. SRP exhibits stronger seasonal variability than DIN, reflecting greater biological regulation, with pronounced summer

depletion due to phytoplankton uptake and recovery during autumn and winter. DIN reaches its minimum in June–July and peaks in December–February, whereas SRP is lowest in June–July and attains its maximum in December.

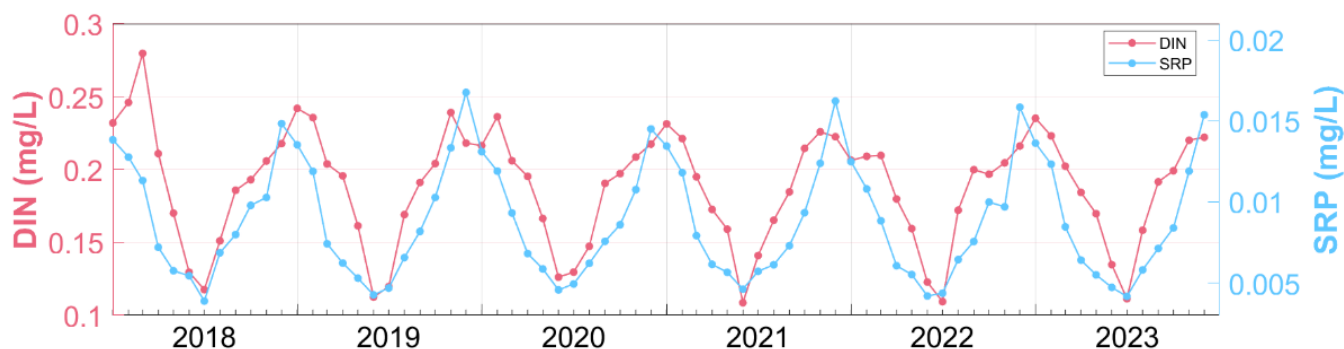


Figure 10. Monthly time series of DIN and SRP in the Bohai sea during 2018 to 2023.

4.3.3. Water Quality Assessment

Based on the National Seawater Quality Standards (NSWQS, GB3097-1997), DIN and SRP concentrations in the Bohai Sea were classified into five water quality grades (Figure 11). The details regarding the NSWQS are shown in Text S4 and Table S3. From 2018 to 2023, Case I consistently dominated the classifications for both parameters. For DIN, Case I accounted for 66.7–68.7% of observations, followed by Case II and Case III, with Case IV and Case V contributing marginally; the proportion of Case I showed slight fluctuations but an overall increasing trend. In contrast, SRP classifications were more uniform, with Case I accounting for 95.3–97.6% and Case II for only 2.4–4.3%, the latter declining after 2021. These trends suggest that the Bohai Comprehensive Management Action Plan implemented since 2018 has contributed to overall improvements in seawater quality.

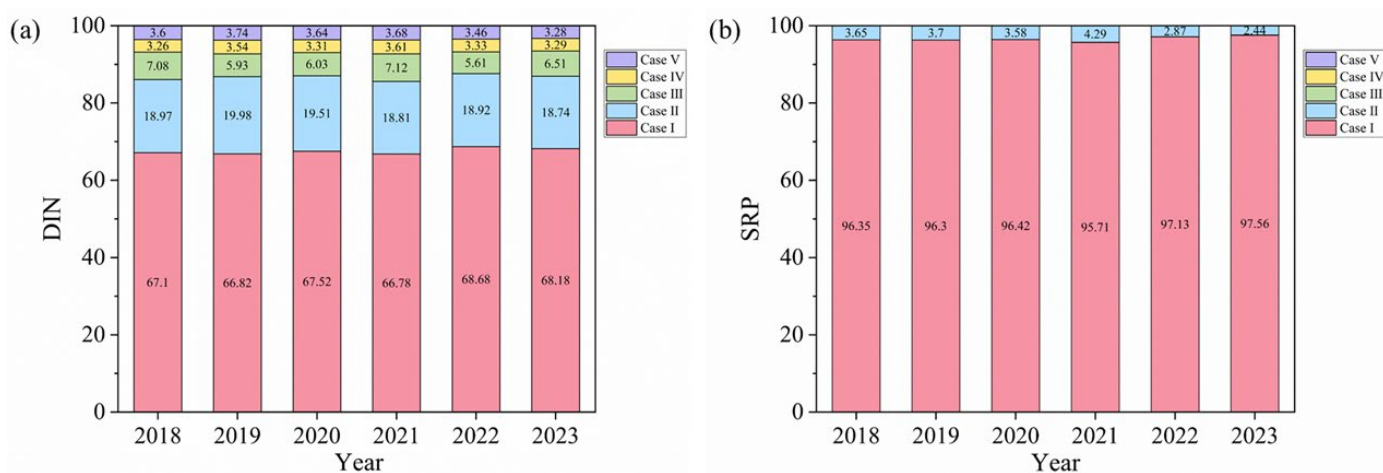


Figure 11. (a) shows the histograms of the water quality classification area in 2018 to 2023 based on DIN. (b) shows the histograms of the water quality classification area in 2018 to 2023 based on SRP. The annual average image was calculated by averaging the daily average image intensity for each year ($p < 0.05$) to classify water quality categories.

5. Discussion

5.1. Ablation Experiment

Different categories of feature combinations were input into the CatBoost model, and their inversion performance for DIN and SRP was evaluated after feature selection (Table 6). Model using only $Rrs(\lambda)$ showed relatively low accuracy. When spectral morphological

features were combined with $Rrs(\lambda)$, test-set MAPE decreased by 10.83% for DIN and 10.67% for SRP, with a slight increase in SRP R^2 , indicating that morphological features complement spectral information in capturing nonlinear relationships. As shown in rows 2–4 of Table 6, the accuracy of both DIN and SRP inversion models improved when spectral features were individually combined with environmental data or with spatiotemporal information. This suggests that nutrient concentrations in water bodies are influenced not only by optical properties but also by environmental factors, and their distributions exhibit pronounced spatiotemporal heterogeneity. Incorporating environmental factors effectively compensates for the limitations of spectral features in capturing water quality variations. The inclusion of spatiotemporal prior information further enables the model to capture patterns of background variability. By integrating spectral characteristics, environmental factors, and spatiotemporal elements as model inputs, the MAPE of the DIN and SRP inversion models decreased by 17.41% and 16.79%, respectively, accompanied by a substantial increase in R^2 .

Table 6. Accuracy evaluation of different feature inputs based on Bayes-CatBoost model. $Rrs(\lambda)$ represents the water-leaving reflectance of the sentinel-3 OLCI band in the visible spectrum. SC named as spectral characteristics represents the sum of spectral single-band data, band combinations and spectral morphological characteristics. ED named as environmental data, represents these environmental features input into the model. ST named as spatio-temporal information represents the information of space time and the cycle of time.

	Input Feature	DIN				SRP			
		R^2	MAE	RMSE	MAPE%	R^2	MAE	RMSE	MAPE%
1	$Rrs(\lambda)$	0.62	0.09	0.141	46.75	0.48	0.004	0.005	57.68
2	SC	0.62	0.09	0.140	35.92	0.49	0.004	0.006	47.01
3	SC&ED	0.74	0.08	0.114	34.93	0.57	0.004	0.005	46.15
4	SC&ST	0.77	0.07	0.109	39.17	0.64	0.003	0.005	48.40
5	ED&ST	0.81	0.06	0.096	32.02	0.46	0.004	0.005	48.85
6	SC&ED&ST	0.82	0.06	0.099	29.34	0.68	0.003	0.004	40.89

5.2. Interpretation of the Optimal Model Based on SHAP

The Figure 12 presents the SHapley Additive exPlanations (SHAP) values of individual features in the DIN and SRP models. The contributions of different feature types were further examined using SHAP values, and their physical interpretations were analyzed.

In the DIN model (Figure 12a,b), the SHAP values for latitude and longitude exhibit a wide distribution, with relatively high mean values in the global interpretation. This indicates that spatial distribution features provide strong explanatory power for DIN variations. Among the spectral features, the combination N8 (7,5,10) contributes most prominently to the model. Analysis of it and the other two band combinations suggests that the red and green bands play a significant role in the DIN inversion model. This suggests that DIN is indirectly represented in the spectral signal through its coupling with optically active substances, including chlorophyll, suspended matter, and CDOM. In the DIN model, SHAP values corresponding to high eigenvalue points are all negative, indicating that low salinity positively influences model predictions. Low-salinity zones typically occur in nearshore areas and estuaries, where terrestrial runoff dilutes seawater salinity while concurrently delivering substantial nutrient loads into the ocean [60]. Consequently, low-salinity areas generally correspond to elevated nutrient concentrations [61]. The SHAP values associated with precipitation are both positive and negative, yet its contribution to the model ranks relatively low. This indicates that precipitation does not directly increase DIN concentrations. Rather, precipitation likely influences DIN predictions indirectly by transporting nitrogen into water bodies via enhanced terrestrial runoff.

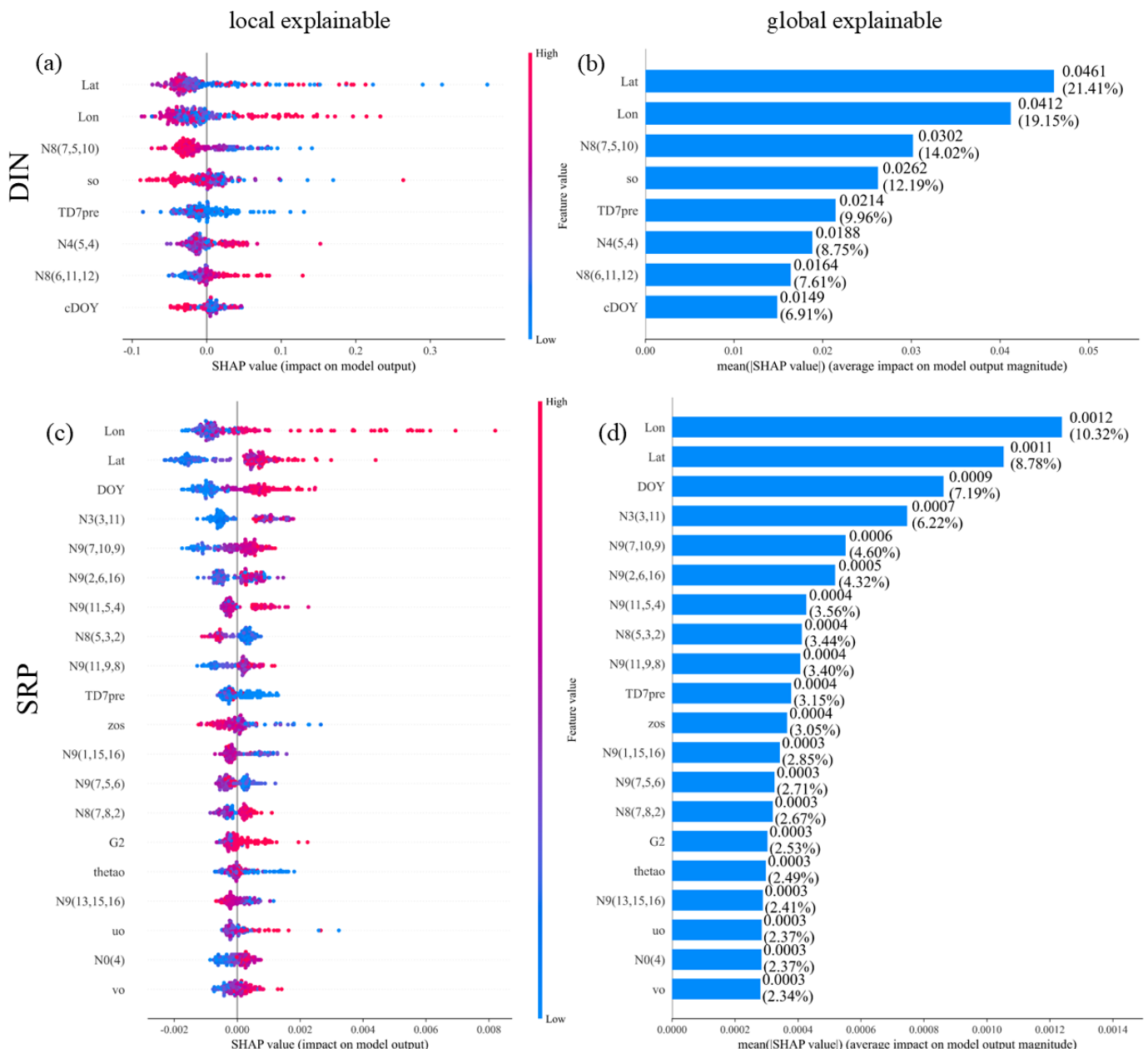


Figure 12. Interpretability results of features for Bayes-CatBoost model used on the retrieval of DIN and SRP by SHAP analysis. (a,c), respectively, show the local explainable results of DIN and SRP. (b,d), respectively, show the global explainable results of DIN and SRP. Higher values in the left column indicate stronger influence, with color representing feature magnitude (blue: low, red: high). In the right column, the mean absolute SHAP values represent contribution of features to the overall model.

In the SRP model (Figure 12c,d), latitude and longitude remain among the top-ranked features in the SHAP analysis, indicating that SRP concentrations are strongly influenced by spatial factors. DOY also exhibits a relatively high contribution in the SRP model, highlighting the significant influence of seasonal temporal factors. The SRP model employs a larger set of spectral features, combining multiple bands to span the entire spectral range. This reflects the complex nonlinear relationships between SRP and optically active substances in water, necessitating more extensive spectral information for accurate representation. Notably, the SHAP values for the spectral morphological parameter G2 (red dot) are predominantly positive, suggesting that more pronounced morphological features positively influence SRP concentrations. In addition, precipitation, sea surface height, sea surface

temperature, and water flow velocity all contribute to the SRP model. Taking ZOS as an example, ZOS showed a moderate strong correlation with SRP, and the SHAP values of red points were mostly less than zero. This indicates that higher ZOS values exert a stronger negative influence on SRP concentrations. In contrast, nearer to the coastline and estuaries, sea surface elevations are lower, corresponding to higher nutrient concentrations. Sea surface temperature and current velocity influence SRP retrieval by modulating plankton activity and altering nutrient spatial distributions, respectively [62,63].

Overall, amounts of nutrients, including DIN and SRP are discharged into the seawater, promoting the proliferation of phytoplankton, which affects the optical properties of the seawater. Riverine discharge typically carries high loads of nutrients into coastal waters while simultaneously reducing salinity. Therefore, low salinity often indicate the presence of the area of water enriched with DIN and SRP. This relationship is particularly significant in regions influenced by strong land and sea interactions, where salinity can effectively trace the spatial extent and intensity of terrestrial nutrient inputs. In this sense, salinity compensates for the weak spectral response of DIN and SRP by providing indirect but physically meaningful information about their sources. Temperature regulates phytoplankton growth rates and biogeochemical cycling processes, all of which influence nutrient concentrations. For instance, higher temperatures in summer enhance biological uptake of nutrients, while winter mixing can replenish surface nutrient levels. Precipitation is closely linked to the magnitude and timing of terrestrial nutrient inputs. Increased precipitation enhances river runoff and surface erosion, leading to higher fluxes of nitrogen and phosphorus into coastal waters. This process is often episodic and exhibits strong seasonal variability, especially in rainfall dominated period. Ocean currents and flow velocity represent the primary mechanisms governing the transport and dispersion of nutrients after their introduction into the marine environment. All of these parameters contribute to the retrieval of DIN and SRP depend on their direct or indirect effects on nutrients.

5.3. Comparison with Previous Studies

Due to differences in modeling approaches, models developed in previous studies exhibit substantial variability in their performance when estimating non-optical water quality parameters. In this study, spatiotemporal factors and multiple natural environmental variables are integrated with spectral data to construct inversion models. Feature selection is conducted using an iterative strategy based on feature importance ranking to identify the optimal feature combination which optimizes model performance. Ref. [14] proposed a novel zoning-based ensemble modeling strategy and achieved R^2 values of 0.56 for ammonia nitrogen, 0.54 for chemical oxygen demand, and 0.57 for dissolved oxygen. Ref. [64] developed six machine learning models optimized using intelligent optimization algorithms to retrieve water quality parameters, with the best-performing model achieving R^2 values of 0.545 for total phosphorus, 0.418 for total nitrogen, and 0.653 for dissolved oxygen. Ref. [38] employed multi-source data integration and advanced machine learning algorithms, achieving an R^2 value of 0.78 for total nitrogen in Poyang Lake. In contrast, the models developed in this study achieved R^2 values of 0.82 for nitrogen and 0.68 for phosphorus, respectively, indicating robust and competitive performance.

In terms of model applicability, we incorporated environmental variables influencing DIN and SRP concentrations in the Bohai Sea, including climatic factors such as precipitation and hydrological variables such as sea surface temperature and currents, thereby compensating for limited spectral information. In addition, integrating spatiotemporal factors enhanced the model's capacity to capture heterogeneity in nutrient dynamics with optically complex water. For model selection, multiple machine learning (ML) algorithms were evaluated and optimized using Bayesian optimization and cross-validation. This

approach ensured that the final inversion model achieved optimal performance while maintaining accuracy and robustness.

5.4. Limitations and Future Prospects

Regarding feature inputs, we did not restrict our analysis to spectral bands, band combinations, and various indices alone. We found that input features for the NAWQPs inversion model could additionally be derived from differences in spectral morphology [39]. By adapting the MSMCF for feature extraction across multiple spectral scales, we observed that its contribution to overall model performance enhancement was relatively limited. When constructing inversion models based solely on spectral features, the inclusion of spectral morphology parameters can significantly improve model performance in terms of MAPE. However, in the final integrated inversion model, the advantages of morphological features were not observed. Future research should aim to integrate water body morphological variations to systematically explore characteristics tailored to different water bodies, thereby further enhancing inversion accuracy and model capability.

Secondly, the availability of high-concentration training samples is limited. Among all training samples, only 34 exhibited DIN concentrations above 0.8 mg/L, and 46 exhibited SRP concentrations above 0.025 mg/L, corresponding to 3.31% and 4.57%, respectively. This scarcity directly results in underestimation of actual nutrient concentrations during model prediction at high levels. For instance, as shown in the validation results in Figure 5, when actual nutrient concentrations are excessively high, predicted values are generally underestimated relative to observations. The model demonstrates good performance in low-concentration regions, but underperforms in high-concentration areas, limiting its generalization and robustness [23].

Furthermore, during model development, incorporating spatiotemporal information may introduce spatial discontinuities in inversion outputs. This phenomenon was observed in the DIN inversion results shown in Figures 6a and 8a–d. Recently, the Geographical Extreme Gradient Boosting (G-XGBoost) model has been developed. This model generates localized submodels by incorporating spatial kernel functions and integrates them with the global model, thereby directly assigning spatial weights within the model framework [65,66]. Therefore, future studies may adopt G-XGBoost and Geographically Weighted Random Forest (GWRF) models [67,68]. These approaches address the issue by incorporating spatial information as model features, rather than directly including spatial coordinates as input variables.

6. Conclusions

In this study, we integrated Sentinel-3 OLCI remote sensing data, in situ measurements, multi-source environmental variables, and spatiotemporal information to develop RFR, XGBoost, and CatBoost models for the inversion of DIN and SRP concentrations. We analyzed the spectral differences arising from the alignment of in situ measurements with corresponding satellite pixels. Building on spectral characteristics, we further incorporated environmental drivers and spatiotemporal variables to enrich the feature set. By ranking features according to their importance, we investigated the influence of different feature combinations on model performance, ultimately determining the optimal input features for each model. Incorporation of spatiotemporal and environmental variables reduced model inversion errors by 17.41% and 16.79%, respectively. Bayesian optimization was applied to tune the hyperparameters of each model, resulting in Bayes-RFR, Bayes-XGBoost, and Bayes-CatBoost models. Evaluation of the three models demonstrates that Bayes-CatBoost achieves superior performance, with DIN inversion yielding $R^2 = 0.82$,

MAE = 0.06 mg/L, RMSE = 0.099 mg/L, and MAPE = 29.34%, and SRP inversion yielding $R^2 = 0.68$, MAE = 0.003 mg/L, RMSE = 0.004 mg/L, and MAPE = 40.89%.

The model was applied to Sentinel-3 OLCI imagery of the Bohai Sea to estimate DIN and SRP concentrations from 2018 to 2023 and to perform a spatiotemporal analysis. The results indicate that, in recent years, annual average concentrations of both DIN and SRP in the Bohai Sea have remained relatively stable, exhibiting a slight downward trend. Based on water quality classification according to the Sea Water Quality Standard (GB 3097-1997), DIN and SRP in the Bohai Sea are predominantly classified as Class I, with their relative proportions exhibiting a modest upward trend. This suggests that recent comprehensive management initiatives in the Bohai Sea have yielded measurable positive effects. Seasonally, both DIN and SRP exhibit minimum concentrations in summer and maximum concentrations in winter; notably, SRP shows greater seasonal variability, indicating a stronger influence of phytoplankton-mediated uptake and transformation processes. The N/P ratios in the Bohai Sea substantially exceed the Redfield ratio across all seasons, highlighting the continued need for targeted monitoring and management of nitrogen emissions.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs18152527/s1>, Text S1: Water quality images acquisition of daily DIN and SRP, Text S2: Mathematical operation of N/P ratio, Text S3: Analysis of Annual and Changes, Text S4: Water quality assessment; Figure S1: Histograms of the frequency distributions of the DIN and SRP data experiment; Table S1: Factors of model input. Description of SC, ED, ST's specific variables and explain these variables. SC, ED, ST represent respectively the factors of spectral characteristics, environmental data, spatiotemporal information. Table S2: Hyperparameter search space, Table S3: National Seawater Quality Standards (NSWQS, GB 3097-1997). Table S4: Evaluation of model performance on the test set following the exclusion of latitude and longitude. Table S5: The top 25 features and their importance scores based on the importance ranking of the initial feature inputs in the first iteration.

Author Contributions: Q.X.: conceptualization, methodology, software, formal analysis, investigation, visualization, writing—original draft preparation; Z.S.: writing—original draft preparation, writing—review and editing; J.X.: validation, writing—review and editing; S.G.: investigation and data curation; S.L.: writing—review and editing, supervision; Y.C.: writing—review and editing, supervision, project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data supporting the conclusion of this article are available after obtaining approval of the national marine environmental monitoring center, ministry of ecology environment. The data are not publicly available due to privacy reasons.

Acknowledgments: The authors gratefully acknowledge the colleagues from the remote sensing department of the national marine environmental monitoring center for their assistance and support. We also thank our colleagues at the center for their work in sample collection and laboratory analysis.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhu, L.D.; Cui, T.W.; Runa, A.; Pan, X.L.; Zhao, W.J.; Xiang, J.Z.; Cao, M.M. Robust remote sensing retrieval of key eutrophication indicators in coastal waters based on explainable machine learning. *ISPRS-J. Photogramm. Remote Sens.* **2024**, *211*, 262–280. [CrossRef]
2. Fu, B.; Lao, Z.; Liang, Y.; Sun, J.; He, X.; Deng, T.; He, W.; Fan, D.; Gao, E.; Hou, Q. Evaluating optically and non-optically active water quality and its response relationship to hydro-meteorology using multi-source data in Poyang Lake, China. *Ecol. Indic.* **2022**, *145*, 109675. [CrossRef]

3. Yang, H.; Kong, J.; Hu, H.; Du, Y.; Gao, M.; Chen, F. A Review of Remote Sensing for Water Quality Retrieval: Progress and Challenges. *Remote Sens.* **2022**, *14*, 1770. [[CrossRef](#)]
4. Zhou, Y.; Ma, J.; Zhang, Y.; Qin, B.; Jeppesen, E.; Shi, K.; Brookes, J.D.; Spencer, R.G.M.; Zhu, G.; Gao, G. Improving water quality in China: Environmental investment pays dividends. *Water Res.* **2017**, *118*, 152–159. [[CrossRef](#)] [[PubMed](#)]
5. Sun, Y.; Wang, D.; Li, L.; Ning, R.; Yu, S.; Gao, N. Application of remote sensing technology in water quality monitoring: From traditional approaches to artificial intelligence. *Water Res.* **2024**, *267*, 122546. [[CrossRef](#)] [[PubMed](#)]
6. Sun, W.; Du, Q. Hyperspectral Band Selection: A Review. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 118–139. [[CrossRef](#)]
7. Gaur, S.; Bhattacharjee, R.; Chander, S.; Ohri, A.; Srivastava, P.K. Globally validated non-unique inversion framework to estimate optically active water quality indicators using in situ and space-borne hyperspectral data sets. *Front. Environ. Sci. Eng.* **2025**, *19*, 10.
8. Hunter, P.D.; Tyler, A.N.; Carvalho, L.; Codd, G.A.; Maberly, S.C. Hyperspectral remote sensing of cyanobacterial pigments as indicators for cell populations and toxins in eutrophic lakes. *Remote Sens. Environ.* **2010**, *114*, 2705–2718. [[CrossRef](#)]
9. Olmanson, L.G.; Brezonik, P.L.; Bauer, M.E. Airborne hyperspectral remote sensing to assess spatial distribution of water quality characteristics in large rivers: The Mississippi River and its tributaries in Minnesota. *Remote Sens. Environ.* **2013**, *130*, 254–265. [[CrossRef](#)]
10. Gholizadeh, M.; Melesse, A.; Reddi, L. A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors* **2016**, *16*, 1298. [[CrossRef](#)] [[PubMed](#)]
11. Sagan, V.; Peterson, K.T.; Maimaitijiang, M.; Sidike, P.; Sloan, J.; Greeling, B.A.; Maalouf, S.; Adams, C. Monitoring inland water quality using remote sensing: Potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. *Earth-Sci. Rev.* **2020**, *205*, 103187. [[CrossRef](#)]
12. Niu, C.; Tan, K.; Jia, X.; Wang, X. Deep learning based regression for optically inactive inland water quality parameter estimation using airborne hyperspectral imagery. *Environ. Pollut.* **2021**, *286*, 117534. [[CrossRef](#)] [[PubMed](#)]
13. Tesfaye, A. Remote Sensing-Based Water Quality Parameters Retrieval Methods: A Review. *East Afr. J. Environ. Nat. Resour.* **2024**, *7*, 80–97. [[CrossRef](#)]
14. Cai, C.; Liu, L.; Wang, Z.; Pang, W.; Bai, C.; Zhang, H. Retrieval of non-optical active water quality parameters in complex Lake environments using a novel zoning-based ensemble modeling strategy. *Ecol. Indic.* **2025**, *176*, 113723. [[CrossRef](#)]
15. Cui, J.; Wu, S.; Dai, J.; Xue, W.; Zhang, Y.; You, J.; Lv, X.; Yang, X. Satellite retrievals of total phosphorus in Taihu Lake using Sentinel-2 images and an optimized XGBoost model. *Ecol. Indic.* **2025**, *175*, 113563. [[CrossRef](#)]
16. Arias-Rodriguez, L.F.; Tüzün, U.F.; Duan, Z.; Huang, J.; Tuo, Y.; Disse, M. Global Water Quality of Inland Waters with Harmonized Landsat-8 and Sentinel-2 Using Cloud-Computed Machine Learning. *Remote Sens.* **2023**, *15*, 1390. [[CrossRef](#)]
17. Toming, K.; Liu, H.; Soomets, T.; Uuema, E.; Nöges, T.; Kutser, T. Estimation of the Biogeochemical and Physical Properties of Lakes Based on Remote Sensing and Artificial Intelligence Applications. *Remote Sens.* **2024**, *16*, 464. [[CrossRef](#)]
18. Xiong, J.; Lin, C.; Cao, Z.; Hu, M.; Xue, K.; Chen, X.; Ma, R. Development of remote sensing algorithm for total phosphorus concentration in eutrophic lakes: Conventional or machine learning? *Water Res.* **2022**, *215*, 118213. [[CrossRef](#)] [[PubMed](#)]
19. Zhao, Y.; He, X.; Pan, S.; Bai, Y.; Wang, D.; Li, T.; Gong, F.; Zhang, X. Satellite retrievals of water quality for diverse inland waters from Sentinel-2 images: An example from Zhejiang Province, China. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *132*, 104048. [[CrossRef](#)]
20. Wang, H.; Zhou, Y.; Tang, Y.; Wu, M.; Deng, Y. Fluctuation of the Water Environmental Carrying Capacity in a Huge River-Connected Lake. *Int. J. Environ. Res. Public Health* **2015**, *12*, 3564–3578. [[CrossRef](#)] [[PubMed](#)]
21. White, M.S.; Xenopoulos, M.A.; Hogsden, K.; Metcalfe, R.A.; Dillon, P.J. Natural lake level fluctuation and associated concordance with water quality and aquatic communities within small lakes of the Laurentian Great Lakes region. *Hydrobiologia* **2008**, *613*, 21–31. [[CrossRef](#)]
22. Guo, H.W.; Huang, J.J.; Zhu, X.T.; Wang, B.; Tian, S.; Xu, W.; Mai, Y.Q. A generalized machine learning approach for dissolved oxygen estimation at multiple spatiotemporal scales using remote sensing. *Environ. Pollut.* **2021**, *288*, 117734. [[CrossRef](#)] [[PubMed](#)]
23. Zhang, Z.W.; Li, C.B.; Yang, P.; Xu, Z.H.; Yao, L.L.; Wang, Q.; Chen, G.J.; Tan, Q. Enhancing Remote Sensing Water Quality Inversion through Integration of Multisource Spatial Covariates: A Case Study of Hong Kong's Coastal Nutrient Concentrations. *Remote Sens.* **2024**, *16*, 3337. [[CrossRef](#)]
24. Zhou, X.; Liu, C.; Carrion, D.; Akbar, A.; Wang, H. Spectro-environmental factors integrated ensemble learning for urban river network water quality remote sensing. *Water Res.* **2024**, *267*, 122544. [[CrossRef](#)] [[PubMed](#)]
25. Maúre, E.d.R.; Terauchi, G.; Ishizaka, J.; Clinton, N.; DeWitt, M. Globally consistent assessment of coastal eutrophication. *Nat. Commun.* **2021**, *12*, 6142. [[CrossRef](#)] [[PubMed](#)]
26. Sun, D. Late Quaternary environmental change of Yellow River Basin: An organic geochemical record in Bohai Sea (North China). *Org. Geochem.* **2011**, *42*, 575–585. [[CrossRef](#)]
27. Zheng, L.-W.; Zhai, W.-D. Nutrient dynamics in the Bohai and North Yellow seas from seasonal to decadal scales: Unveiling Bohai Sea eutrophication mitigation in the 2010s. *Sci. Total Environ.* **2023**, *905*, 167417. [[CrossRef](#)] [[PubMed](#)]

28. Wei, Q.; Wang, B.; Yao, Q.; Xue, L.; Sun, J.; Xin, M.; Yu, Z. Spatiotemporal variations in the summer hypoxia in the Bohai Sea (China) and controlling mechanisms. *Mar. Pollut. Bull.* **2019**, *138*, 125–134. [[CrossRef](#)] [[PubMed](#)]
29. Liang, W. Nutrient changes in the Bohai Sea over the past two decades. *Sci. Total Environ.* **2023**, *903*, 166696. [[CrossRef](#)] [[PubMed](#)]
30. Wu, Z.; Yu, Z.; Song, X.; Yuan, Y.; Cao, X.; Liang, Y. Application of an integrated methodology for eutrophication assessment: A case study in the Bohai Sea. *Chin. J. Oceanol. Limnol.* **2013**, *31*, 1064–1078. [[CrossRef](#)]
31. Wang, X.; Cui, Z.; Guo, Q.; Han, X.; Wang, J. Distribution of nutrients and eutrophication assessment in the Bohai Sea of China. *Chin. J. Oceanol. Limnol.* **2009**, *27*, 177–183. [[CrossRef](#)]
32. Liu, H.; Lin, L.; Wang, Y.; Du, L.; Wang, S.; Zhou, P.; Yu, Y.; Gong, X.; Lu, X. Reconstruction of Monthly Surface Nutrient Concentrations in the Yellow and Bohai Seas from 2003–2019 Using Machine Learning. *Remote Sens.* **2022**, *14*, 5021. [[CrossRef](#)]
33. Vanhellemont, Q.; Ruddick, K. Atmospheric correction of Sentinel-3/OLCI data for mapping of suspended particulate matter and chlorophyll-a concentration in Belgian turbid coastal waters. *Remote Sens. Environ.* **2021**, *256*, 112284. [[CrossRef](#)]
34. Mishra, S.; Mishra, D.R. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sens. Environ.* **2012**, *117*, 394–406. [[CrossRef](#)]
35. Chen, B.; Chen, Y.; Chen, H. An Interpretable CatBoost Model Guided by Spectral Morphological Features for the Inversion of Coastal Water Quality Parameters. *Water* **2024**, *16*, 3615. [[CrossRef](#)]
36. Gao, C. Significance testing for cross correlation: A critical examination of correlations between ENSO and GRACE-derived terrestrial water storage variabilities. *Glob. Planet Change* **2024**, *241*, 104549. [[CrossRef](#)]
37. Wang, M.; Zhang, Z.; Hu, T.; Wang, G.; He, G.; Zhang, Z.; Li, H.; Wu, Z.; Liu, X. An Efficient Framework for Producing Landsat-Based Land Surface Temperature Data Using Google Earth Engine. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4689–4701. [[CrossRef](#)]
38. Xia, X.; Pan, J.; Pei, J. A new approach to estimate total nitrogen concentration in a seasonal lake based on multi-source data methodology. *Ecol. Inform.* **2024**, *83*, 102807. [[CrossRef](#)]
39. Fu, B.; Li, S.; Lao, Z.; Wei, Y.; Song, K.; Deng, T.; Wang, Y. A novel hierarchical approach to insight to spectral characteristics in surface water of karst wetlands and estimate its non-optically active parameters using field hyperspectral data. *Water Res.* **2024**, *257*, 121673. [[CrossRef](#)] [[PubMed](#)]
40. Li, H.; Liang, Y.; Xu, Q.; Cao, D. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Anal. Chim. Acta* **2009**, *648*, 77–84. [[CrossRef](#)] [[PubMed](#)]
41. Li, L.; Gu, M.; Gong, C.; Hu, Y.; Wang, X.; Yang, Z.; He, Z. An advanced remote sensing retrieval method for urban non-optically active water quality parameters: An example from Shanghai. *Sci. Total Environ.* **2023**, *880*, 163389. [[CrossRef](#)] [[PubMed](#)]
42. Chen, S.; Hu, C.; Barnes, B.B.; Wanninkhof, R.; Cai, W.-J.; Barbero, L.; Pierrot, D. A machine learning approach to estimate surface ocean pCO₂ from satellite measurements. *Remote Sens. Environ.* **2019**, *228*, 203–226. [[CrossRef](#)]
43. Du, Z.; Qi, J.; Wu, S.; Zhang, F.; Liu, R. A spatially weighted neural network based water quality assessment method for large-scale coastal areas. *Environ. Sci. Technol.* **2021**, *55*, 2553–2563. [[CrossRef](#)] [[PubMed](#)]
44. Li, T.; Shen, H.; Yuan, Q.; Zhang, L. Geographically and temporally weighted neural networks for satellite-based mapping of ground-level PM_{2.5}. *ISPRS-J. Photogramm. Remote Sens.* **2020**, *167*, 178–188. [[CrossRef](#)]
45. Wu, S.; Qi, J.; Yan, Z.; Lyu, F.; Lin, T.; Wang, Y.; Du, Z. Spatiotemporal assessments of nutrients and water quality in coastal areas using remote sensing and a spatiotemporal deep learning model. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102897. [[CrossRef](#)]
46. Álvarez, X.; Valero, E.; Santos, R.M.B.; Varandas, S.G.P.; Sanches Fernandes, L.F.; Pacheco, F.A.L. Anthropogenic nutrients and eutrophication in multiple land use watersheds: Best management practices and policies for the protection of water resources. *Land Use Policy* **2017**, *69*, 1–11. [[CrossRef](#)]
47. Xu, L.; Yang, D.; Benthuyssen, J.A.; Yin, B. Key Dynamical Factors Driving the Kuroshio Subsurface Water to Reach the Zhejiang Coastal Area. *J. Geophys. Res. Ocean* **2018**, *123*, 9061–9081. [[CrossRef](#)]
48. Cui, T.W.; Zhang, J.; Wang, K.; Wei, J.W.; Mu, B.; Ma, Y.; Zhu, J.H.; Liu, R.J.; Chen, X.Y. Remote sensing of chlorophyll a concentration in turbid coastal waters based on a global optical water classification system. *ISPRS-J. Photogramm. Remote Sens.* **2020**, *163*, 187–201. [[CrossRef](#)]
49. Xie, M.; Feng, Z.; Wang, S.; Wang, S.; Chang, X. Spatiotemporal patterns and nonlinear drivers of carbon emission intensity under urban shrinkage and growth: A machine learning approach with interpretability. *Sustain. Cities Soc.* **2025**, *130*, 106593. [[CrossRef](#)]
50. Xiong, X.; Guo, X.; Zeng, P.; Zou, R.; Wang, X. A Short-Term Wind Power Forecast Method via XGBoost Hyper-Parameters Optimization. *Front. Energy Res.* **2022**, *10*, 905155. [[CrossRef](#)]
51. Zhao, Y.; Zhang, Z.; Hu, B.; Liu, J.; Wang, X.; Zou, L.; Yu, T. Interpretable data-driven chemometric approach for predicting non-optically active water quality parameters using ultraviolet-visible-near infrared absorption spectroscopy and physical-chemical measurements. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2025**, *331*, 125768. [[CrossRef](#)] [[PubMed](#)]
52. Wu, Y.; Knudby, A.; Pahlevan, N.; Lapen, D.; Zeng, C. Sensor-generic adjacency-effect correction for remote sensing of coastal and inland waters. *Remote Sens. Environ.* **2024**, *315*, 114433. [[CrossRef](#)]

53. Alvera-Azcárate, A.; Barth, A.; Rixen, M.; Beckers, J.M. Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: Application to the Adriatic Sea surface temperature. *Ocean Model.* **2005**, *9*, 325–346. [\[CrossRef\]](#)
54. Beckers, J.M.; Rixen, M. EOF Calculations and Data Filling from Incomplete Oceanographic Datasets*. *J. Atmos. Ocean Technol.* **2003**, *20*, 1839–1856. [\[CrossRef\]](#)
55. Nie, J.; Feng, H.; Witherell, B.B.; Alebus, M.; Mahajan, M.D.; Zhang, W.; Yu, L. Causes, Assessment, and Treatment of Nutrient (N and P) Pollution in Rivers, Estuaries, and Coastal Waters. *Curr. Pollut. Rep.* **2018**, *4*, 154–161. [\[CrossRef\]](#)
56. Zhang, P.; Chen, Y.; Peng, C.; Dai, P.; Lai, J.; Zhao, L.; Zhang, J. Spatiotemporal variation, composition of DIN and its contribution to eutrophication in coastal waters adjacent to Hainan Island, China. *Reg. Stud. Mar. Sci.* **2020**, *37*, 101332. [\[CrossRef\]](#)
57. Liu, H.; Li, Q.; Bai, Y.; Yang, C.; Wang, J.; Zhou, Q.; Hu, S.; Shi, T.; Liao, X.; Wu, G. Improving satellite retrieval of oceanic particulate organic carbon concentrations using machine learning methods. *Remote Sens. Environ.* **2021**, *256*, 112316. [\[CrossRef\]](#)
58. Zhang, S.; Zhang, H. Anthropogenic impact on long-term riverine COD(Mn), BOD, and nutrient flux variation in the Pearl River Delta. *Sci. Total Environ.* **2023**, *859*, 160197. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Liang, W.; Wang, Y.; Liu, S.; Wang, M.; Zhao, L.; Liu, C.; Zhang, X.; Wu, N.; Wang, L.; Zhu, D.; et al. Nutrient dynamics and coupling with biological processes and physical conditions in the Bohai Sea. *Front. Mar. Sci.* **2022**, *9*, 1025502. [\[CrossRef\]](#)
60. Wang, D.; Cui, Q.; Gong, F.; Wang, L.; He, X.; Bai, Y. Satellite Retrieval of Surface Water Nutrients in the Coastal Regions of the East China Sea. *Remote Sens.* **2018**, *10*, 1896. [\[CrossRef\]](#)
61. Fofonoff, N.P. Physical properties of seawater: A new salinity scale and equation of state for seawater. *J. Geophys. Res. Ocean.* **2012**, *90*, 3332–3342. [\[CrossRef\]](#)
62. Humphreys, A.F.; Halfar, J.; Ingle, J.C.; Manzello, D.; Raymond, C.E.; Westphal, H.; Riegl, B. Effect of seawater temperature, pH, and nutrients on the distribution and character of low abundance shallow water benthic foraminifera in the Galápagos. *PLoS ONE* **2018**, *13*, e0202746. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Long, Y.; Guo, X.; Zhu, X.-H.; Li, Z. Nutrient streams in the North Pacific. *Prog. Oceanogr.* **2022**, *202*, 102756. [\[CrossRef\]](#)
64. Hu, W.; Miao, D.; Zhang, C.; He, Z.; Gu, W.; Shao, D. Retrieval of water quality parameters based on IOA-ML models and their response to short-term hydrometeorological factors. *J. Hydrol. Reg. Stud.* **2025**, *57*, 102118. [\[CrossRef\]](#)
65. Grekousis, G. Geographical-XGBoost: A new ensemble model for spatially local regression based on gradient-boosted trees. *J. Geogr. Syst.* **2025**, *27*, 169–195. [\[CrossRef\]](#)
66. Yang, X.; Yang, L.; Huang, D.; Chen, L.; Yang, Y.; Luo, Y.; Liu, Y.; Na, J.; Ding, H. Revealing the Impact of Urban Morphology Evolution on the Urban Heat Island Effect in the Main Urban Area of Guangzhou: Insights from Local Climate Zones. *Remote Sens.* **2025**, *17*, 2959. [\[CrossRef\]](#)
67. Qin, Z.; Peng, Q.; Jin, C.; Xu, J.; Xing, S.; Zhu, P.; Yang, G. Geographically weighted random forest fusing multi-source environmental covariates for spatial prediction of soil heavy metals. *Environ. Pollut.* **2025**, *385*, 127135. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Yang, Y.; Wang, L.; Zhai, X.; Zheng, X.; Zhao, G.; Yang, Q.; Du, Y.; Ling, F. A novel Improved Geographically Weighted Random Forest (IGWRF) model for low-resolution soil moisture data downscaling in Africa. *Agric. Water Manag.* **2026**, *323*, 110034. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.