remote sensing

MDPI

Article

Retrieval of Water Quality Parameters in Dianshan Lake Based on Sentinel-2 MSI Imagery and Machine Learning: Algorithm Evaluation and Spatiotemporal Change Research

Lei Dong [1,2], Cailan Gong [1,*], Hongyan Huai [3], Enuo Wu [4], Zhihua Lu [3], Yong Hu [1], Lan Li [1] and Zhe Yang [1,2]

[1] Key Laboratory of Infrared System Detection and Imaging Technologies, Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China; fcdl@mail.ustc.edu.cn (L.D.); lilan@mail.sitp.ac.cn (L.L.)
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Shanghai Environment Monitoring Center, Shanghai 200235, China; huaihy@sheemc.cn (H.H.); luzh@saes.sh.cn (Z.L.)
[4] Shanghai Academy of Environmental Sciences, Shanghai 200233, China; wuan@sheemc.cn
* Correspondence: gcl@mail.sitp.ac.cn

Abstract: According to current research, machine learning algorithms have been proven to be effective in detecting both optical and non-optical parameters of water quality. The use of satellite remote sensing is a valuable method for monitoring long-term changes in the quality of lake water. In this study, Sentinel-2 MSI images and in situ data from the Dianshan Lake area from 2017 to 2023 were used. Four machine learning methods were tested, and optimal detection models were determined for each water quality parameter. It was ultimately determined that these models could be applied to long-term images to analyze the spatiotemporal variations and distribution patterns of water quality in Dianshan Lake. Based on the research findings, integrated learning algorithms, especially CatBoost, have achieved good results in the retrieval of all water quality parameters. Spatiotemporal analysis reveals that the overall distribution of water quality parameters is uneven, with significant spatial variations. Permanganate index (COD_{Mn}), Total Nitrogen (TN), and Total Phosphorus (TP) show relatively small interannual differences, generally exhibiting a decreasing trend in concentrations. In contrast, chlorophyll-a (Chl-a), dissolved oxygen (DO), and Secchi Disk Depth (SDD) exhibit significant interannual and inter-year differences. Chl-a reached its peak in 2020, followed by a decrease, while DO and SDD showed the opposite trend. Further analysis indicated that the distribution of water quality parameters is significantly influenced by climatic factors and human activities such as agricultural expansion. Overall, there has been an improvement in the water quality of Dianshan Lake. The study demonstrates the feasibility of accurately monitoring water quality even without measured spectral data, using machine learning methods and satellite reflectance data. The research results presented in this paper can provide new insights into water quality monitoring and water resource management in Dianshan Lake.

Keywords: machine learning; water quality parameters; spatiotemporal distribution; Dianshan Lake; Sentinel-2

Citation: Dong, L.; Gong, C.; Huai, H.; Wu, E.; Lu, Z.; Hu, Y.; Li, L.; Yang, Z. Retrieval of Water Quality Parameters in Dianshan Lake Based on Sentinel-2 MSI Imagery and Machine Learning: Algorithm Evaluation and Spatiotemporal Change Research. *Remote Sens.* **2023**, *15*, 5001. https://doi.org/10.3390/rs15205001

Academic Editors: Miro Govedarica, Flor Alvarez-Taboada and Gordana Jakovljević

Received: 5 September 2023
Revised: 12 October 2023
Accepted: 13 October 2023
Published: 18 October 2023

1. Introduction

The effective provision of water resources is closely intertwined with the progress of cities, ecological equilibrium, and economic prosperity [1,2]. Inland water bodies such as lakes are vital in maintaining ecological balance, supporting industrial production, and ensuring human well-being [3,4]. However, in recent years, the compounded impacts of human activities and climate change have posed severe threats to the ecological equilibrium of water bodies, resulting in intensified global freshwater eutrophication and deterioration of water quality [2,5]. Against this backdrop, the effective assessment of lake water quality

is paramount in maintaining ecosystem stability. This evaluation relies on key nutrient indicators, namely, Chl-a, TP, TN, SDD, and COD_{Mn} [6–8]. Chl-a, a primary pigment in phytoplankton, functions as a biomarker for phytoplankton biomass, thereby significantly influencing the overall health of the ecosystem [9]. SDD, quantified using the Secchi disk transparency method, provides insights into the nutrient status of the lake and assumes a critical role in monitoring water quality [3,10,11]. Elevated levels of TN and TP serve as indicators of potential eutrophication concerns [12,13]. The measurement of DO, which is closely correlated with Chl-a, plays a pivotal role in evaluating water quality and its impact on aquatic life [14,15]. The proper management and interpretation of these key indicators are imperative for ensuring sustainable water resource management and safeguarding the delicate balance of lake ecosystems [16].

Traditional water quality monitoring involves manual in situ sampling and lab analysis, providing accurate data but with limited spatial coverage and efficiency. Unlike time-consuming conventional techniques, satellites offer high-frequency, wide-ranging, and long-term water quality data, thus overcoming limitations [4,17–21]. Specialized watercolor satellites have been developed for aquatic environments and are widely used [22,23]. However, lakes smaller than 100 sq. km constitute 63% of the total lake area [24]. Due to watercolor satellites' relatively low spatial resolution, smaller lakes may not be fully monitored. In contrast, Landsat and Sentinel satellite data have higher spatial resolution and are more suitable for monitoring small inland water bodies [18,25,26]. Some studies have effectively employed Sentinel-2 and Landsat imagery for coastal and inland lake water quality monitoring [27–30].

Methods for evaluating water quality parameters using satellite remote sensing data can be categorized into two types: empirical modeling and bio-optical modeling [14]. In recent years, bio-optical modeling has made some progress; however, it is severely constrained by data limitations and challenges in atmospheric correction accuracy [18], because atmospheric correction is a factor that must be considered in aquatic remote sensing [31–35]. A subset of researchers has initiated exploration into direct modeling methods utilizing satellite reflectance data. Their goal is to mitigate errors and uncertainties arising from atmospheric correction to the greatest extent possible. In recent years, with the development of the field of artificial intelligence, the application of machine learning algorithms in water quality assessment has been increasing gradually [14]. Machine learning models can uncover underlying complex nonlinear relationships, thus providing a general and optimized approach for water quality parameter detection [36–38]. Its application in water quality modeling and detection shows a continuous growth trend [39–42]. Common machine learning methods used for water quality assessment include Support Vector Machine Regression (SVR) and Random Forest Regression (RF). In recent years, XGBoost Regression (XGBoost) and CatBoost Regression (CatBoost) have also gained increasing popularity.

Current research utilizing machine learning combined with satellite data for the retrieval of water quality parameters has been successfully applied in multiple cases [14,18,24,25,29,42]. However, there are significant differences in the water quality parameters used, and the spatial and radiometric resolution of sensors in different regions, leading to variations in retrieval algorithms [43]. Dianshan Lake, which receives water from Taihu Lake and is influenced by agricultural activities and residential wastewater discharge in the surrounding areas, has experienced several water pollution incidents over the past two decades. Water quality monitoring has been a focal point for government water authorities and the research community [44]. Presently, there is limited research on the spatiotemporal characteristics of water quality evolution and driving factors in Dianshan Lake using remote sensing algorithms, making it challenging to provide targeted recommendations for environmental protection, management, and control measures. Therefore, the central objective of this study is to directly utilize satellite reflectance data to develop and validate models for retrieving water quality parameters. The specific objectives are as follows: (1) Utilize four machine learning methods (RF, XGBoost, CatBoost, and SVR) to establish optimal retrieval models for various water quality parameters (Chl-a,

COD_{Mn}, DO, SDD, TN, TP). (2) Employ Sentinel-2 satellite remote sensing imagery from 2017 to 2023 to retrieve various water quality parameters for spatiotemporal change analysis. The study aims to provide a scientific basis for lake management and environmental protection efforts.

2. Materials and Methods

2.1. Study Area

Dianshan Lake (31°04′–31°12′N, 120°54′–121°01′E) is situated on the border of Qingpu District in Shanghai and Kunshan City in Jiangsu Province, China. Its location in China is shown in Figure 1a. With an area of approximately 62 square kilometers and an average depth of 2.5 m, the lake plays a pivotal role in various social and ecological functions. It serves as the receiving end of water from the Wujiang area of Taihu Lake and functions as the headwaters of the Huangpu River.

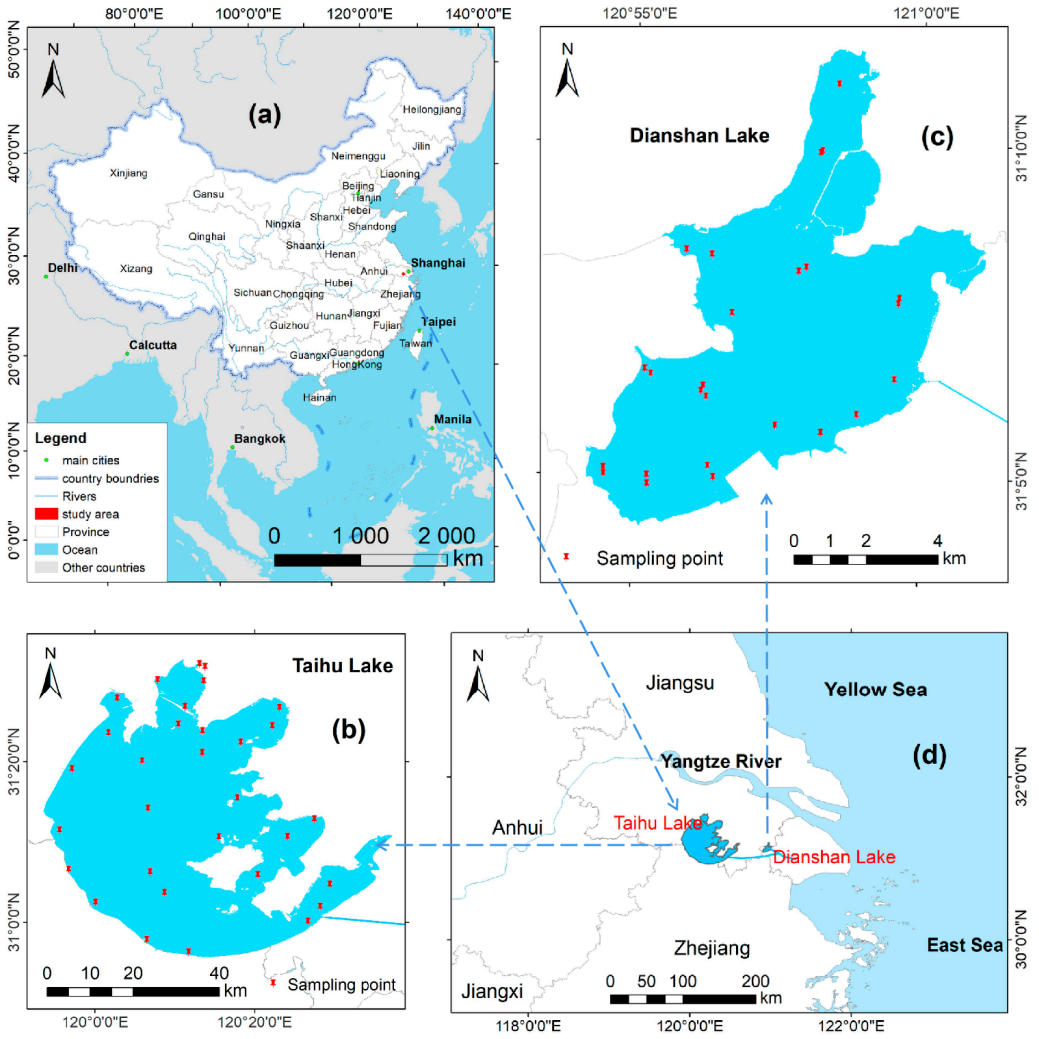

Figure 1. (**a**) Location schematic diagram of the study area, (**b**) Distribution of sampling points in Taihu Lake, (**c**) Distribution of sampling points in Dianshan Lake, (**d**) Schematic diagram of the relative positions of Taihu Lake and Dianshan Lake.

2.2. Dataset

This study employed three types of datasets: (1) Sentinel-2 MSI satellite imagery data spanning the period from 2017 to 2023, utilized to retrieve water quality parameters; (2) Concentration data of Chl-a, COD_{Mn}, DO, SDD, TN, and TP acquired through sampling in Dianshan Lake. These data were employed for the development and evaluation of ma-

chine learning methods; (3) Measured Chl-a, COD_{Mn}, DO, SDD, TN, and TP concentration data from Taihu Lake were utilized to further validate the model's applicability.

2.2.1. Satellite Data

Both Sentinel-2 MSI and Landsat offer high-resolution remote sensing image data for Earth observation and environmental monitoring. Considering Sentinel-2 MSI's distinct advantages over Landsat, which include shorter revisit periods, a greater number of spectral bands, higher spatial resolution, and an open data policy, this study harnessed the capabilities of Sentinel-2 MSI. Specifically, we utilized a dataset of 100 Sentinel-2 MSI images acquired from the Copernicus Open Access Hub (https://scihub.copernicus.eu/, accessed on 15 June 2023) spanning the period from 2017 to April 2023. The selection of these downloaded images adhered to strict criteria, ensuring cloud-free conditions above the lake and minimal sun glint on the lake surface. The distribution of data according to the quantity of time is shown in Figure 2.

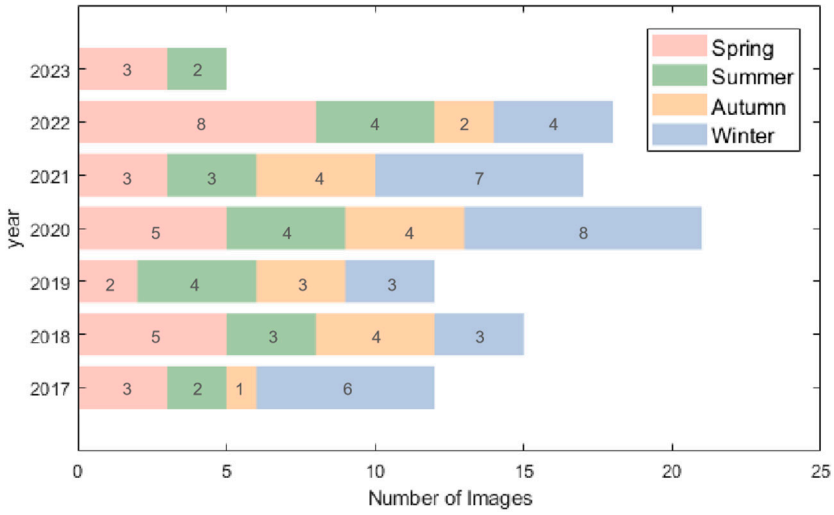

Figure 2. Temporal and Quantitative Distribution of Sentinel-2 MSI Images Used in This Study.

The radiation received by sensors at the top of the atmosphere (TOA) can be primarily attributed to Rayleigh scattering and aerosol scattering [45]. Atmospheric correction is a process aimed at mitigating the impacts of Rayleigh scattering, Mie scattering, atmospheric absorption, and aerosol influence on remote sensing images. Some researchers have proposed that using uncorrected TOA images can yield superior results compared to images that have undergone atmospheric correction [46]. In this study, we employed the SNAP software for Rayleigh correction of the images, resulting in dimensionless Rayleigh-corrected reflectance. Following this, the image resolution was resampled to 20 m, and the Normalized Difference Water Index (NDWI) [47] was utilized to delineate water regions. Before performing water quality modeling, and to mitigate uncertainties stemming from aerosols and other factors, an enhanced MD09 method [48,49] was implemented for aerosol correction. This method involves a straightforward Rayleigh reflectance correction technique that entails subtracting the minimum value from the shortwave infrared band (Band 11 in MSI images) within the visible and near-infrared bands. The resulting value is then divided by π.

2.2.2. Field Data

From 2017 to 2022, a monthly routine water sampling campaign was conducted in Dianshan Lake to collect data on water quality parameters. The study specifically selected data points falling within a $\pm$5-day range of the satellite overpass time as the focal dataset, resulting in a total of 398 datasets. The statistical description of the data is shown in Table 1. The precise locations of the sampling sites within Dianshan Lake are illustrated in Figure 1c.

Table 1. Statistical description of measured water quality parameters in Dianshan Lake.

Water Quality Parameter	Range	Mean ± Std	Median	CV	N
Chl-a (mg/m^3)	1.34–51	15.04 ± 10.35	12.80	0.69	398
COD_{Mn} (mg/L)	2.10–7.00	3.96 ± 0.80	3.80	0.20	398
DO (mg/L)	3.90–13.84	8.73 ± 1.97	8.60	0.23	398
TN (mg/L)	0.33–5.23	2.04 ± 1.00	1.87	0.49	398
TP (mg/L)	0.03–0.26	0.10 ± 0.05	0.090	0.45	398
SDD (m)	0.1–1.1	0.42 ± 0.4	0.17	0.41	398

The water samples collected during in situ experiments were transported to the laboratory for analysis of water quality parameters. The laboratory analysis methods adhered to the water quality parameter determination procedures outlined in the Chinese National Standard GB3838-2002. Table A1 presents a compilation of the names of different water quality parameters, alongside their corresponding determination methods.

To assess the transferability of the optimal model to different geographical regions, additional data were collected from 2018 to 2022 at 32 monitoring stations situated around Taihu Lake. Due to the high level of eutrophication in Taihu Lake, surface blooms of cyanobacteria are frequent. To ensure water body consistency as much as possible, we utilized a visual interpretation method to identify sampling points unaffected by cyanobacterial blooms in satellite true-color images as supplementary data. There were a total of 161 validation points in Taihu Lake. The statistical description of the data is shown in Table 2 and the sampling site locations in the Taihu Lake region are visually depicted in Figure 1b.

Table 2. Statistical description of measured water quality parameters in Taihu Lake.

Water Quality Parameter	Arrange	Mean ± Std	Median	CV	N
Chl-a (mg/m^3)	6.34–63.38	21.41–9.63	19.62	0.45	130
COD_{Mn} (mg/L)	3.37–5.15	4.24–0.42	4.30	0.10	130
DO (mg/L)	6.10–11.55	7.98–1.20	7.70	0.15	130
TN (mg/L)	0.24–0.53	0.36–0.07	0.34	0.19	130
TP (mg/L)	0.83–3.51	1.73–0.54	1.60	0.31	130
SDD (m)	0.066–0.329	0.112–0.029	0.111	0.26	130

2.3. Modeling

Based on the latitude and longitude coordinates of the actual measurement sites, the corresponding image reflectance for the respective dates is extracted. To ensure data consistency, a 3 × 3 pixel window surrounding each site is considered. The average reflectance within this window is then computed and utilized as the matched data.

In the investigation of the six water quality parameters, our study explored four distinct machine learning methods, namely: (1) Random Forest Regression (RF), (2) XGBoost Regression (XGB), (3) CatBoost Regression (CatBoost), and (4) SVR. The selection of these methods was grounded in their performance and characteristics across various data scenarios. Moreover, these techniques have been demonstrated as successful applications in estimating water quality parameters in several inland lakes previously [18,24,29,38,50–54].

These methods possess distinct characteristics. In the landscape of ensemble learning techniques, Random Forest Regression (RF) has garnered substantial interest due to its commendable performance and robust characteristics. By constructing multiple decision trees and aggregating their predictions, RF not only mitigates the risk of overfitting but also accommodates a diverse range of data types, including both continuous and categorical features. In contrast, XGBoost Regression (XGB) distinguishes itself through its efficient gradient boosting algorithm, which facilitates exceptional performance on large-scale datasets. XGB incorporates regularization techniques to control model complexity and

exhibits considerable proficiency in handling missing values and feature engineering. Conversely, CatBoost Regression (CatBoost) specializes in the treatment of categorical features, autonomously affecting feature transformations without necessitating additional preprocessing steps. This confers it with advantages in certain domains. Support Vector Regression (SVR) is one of the most frequently used methods in recent years. SVR excels in regression with high dimensions, noise, and nonlinearity. Its adaptable kernels and robustness with small datasets contribute to its significance in ensemble learning.

For each model, an identical set of input features was chosen to assess the ultimate outcomes. In this study, we utilized the Pearson correlation coefficient to ascertain the relationships between various water quality parameters and some widely employed spectral band combinations.

Prior researchers have demonstrated the robustness of band ratio algorithm ($R_{rs}(\lambda 1) - R_{rs}(\lambda 2)$) and band difference algorithm ($\frac{R_{rs}(\lambda 1)}{R_{rs}(\lambda 2)}$) when applied to the retrieval of water quality in optical complex inland lakes [38]. In this study, we also incorporated the Normalized Difference Band Calculation algorithm ($\frac{R_{rs}(\lambda 2)-R_{rs}(\lambda 1)}{R_{rs}(\lambda 2)+R_{rs}(\lambda 1)}$) [40] and the three-band combination form ($R_{rs}(\lambda 3) \times \left(\frac{1}{R_{rs}(\lambda 2)} - \frac{1}{R_{rs}(\lambda 1)}\right)$) [39] to assess their correlations with water quality parameters. The objective was to identify the optimal inputs for the machine learning models. In the process of constructing retrieval models for each water quality parameter, a comprehensive set of 13 input features was employed. Among these input variables, the combination of these 13 variables exhibited the most optimal performance. These encompassed the initial 9 visible and near-infrared bands from the MSI image, alongside the band combinations from each method that exhibited the highest correlation with the concentration of water quality parameters. Please refer to Table 3 for the most relevant band combinations for each water quality parameter.

Table 3. Input features for various water quality parameters (only wavelength combinations listed).

Band Combination Form	Chl-a	COD_{Mn}	DO	SDD	TN	TP
$R_{rs}(\lambda 1) - R_{rs}(\lambda 2)$	B7 [1] B9	B7 B2	B6 B7	B5 B2	B2 B3	B7 B6
$\frac{R_{rs}(\lambda 1)}{R_{rs}(\lambda 2)}$	B4 B5	B6 B7	B6 B7	B2 B5	B6 B7	B7 B6
$\frac{R_{rs}(\lambda 2)-R_{rs}(\lambda 1)}{R_{rs}(\lambda 2)+R_{rs}(\lambda 1)}$	B4 B5	B6 B7	B6 B7	B5 B2	B7 B6	B6 B7
$R_{rs}(\lambda 3) \times \left(\frac{1}{R_{rs}(\lambda 2)} - \frac{1}{R_{rs}(\lambda 1)}\right)$	B5 B4 B2	B7 B6 B5	B6 B7 B1	B3 B5 B6	B6 B7 B1	B7 B6 B2

[1] The wavelengths of Sentinel-2 MSI image bands.

It is noteworthy that the selection of hyperparameters in machine learning substantially influences the model's performance and generalization capability. This process directly impacts the model's robustness and governs its complexity to mitigate overfitting. In our study, the Python programming language was employed to conduct a grid search technique for determining the model's hyperparameters. Each training session incorporated a fivefold cross-validation strategy to comprehensively evaluate the model's performance.

In SVR, the C parameter controls the degree of regularization, the kernel parameter defines the type of kernel function, and the gamma parameter influences the range of the kernel function's impact. By judiciously adjusting these parameters, a balance between model complexity and regularization can be achieved to enhance performance. Optimizing the performance of the XGBoost model depends on the selection of several key hyperparameters. Smaller learning rates and larger gamma values contribute to improved generalization performance, while parameters like min child weight, max depth, and reg alpha help stabilize the model, preventing overfitting. Random Forest (RF) can effectively control the number and depth of trees in the forest by tuning parameters such as n estimators, max depth, min samples split, min samples leaf, and max features, thereby enhancing model performance. CatBoost can optimize model complexity and regulariza-

tion by adjusting parameters like iterations, learning rate, depth, and l2 leaf reg, resulting in improved performance.

The grid search strategies for each model are summarized in Table 4, and the optimal parameters chosen for different water quality parameters in each model are presented in Table A2.

Table 4. Hyperparameter grid search table for each model.

Model	Hyperparameters	Options
RF	n_estimators	np.arange [1] (10, 600, 10)
	max_depth	np.arange (10, 50, 5)
	min_samples_split	np.arange (1, 50, 1)
	min_samples_leaf	np.arange (1, 12, 1)
SVR	C	np.arange (1, 10, 0.01)
	kernel	['linear', 'rbf','sigmoid']
	gamma	np.arange (1, 100, 0.001)
XGBoost	learning_rate	np.arange (0.15, 0.2, 0.005)
	gamma	np.arange (0.001, 0.005, 0.001)
	min_child_weight	np.arange (5, 10, 1)
	max_depth	np.arange (2, 10, 1)
	sub_sample	[0.8, 1]
	reg_alpha	[0.001, 0.01, 0.1, 1]
CatBoost	iterations	np.arange (50, 500, 10)
	learning_rate	np.arange (0.01, 0.05, 0.01)
	depth	np.arange (2,10,1)
	l2_leaf_reg	np.arange (1,10,1)

[1] 'np.arange(10, 600, 10)' generates a sequence of numbers, starting at 10 and increasing by 10 at each step, until it is just below 600.

2.4. Accuracy Evaluation

The metrics chosen for assessing the models' performance encompassed the coefficient of determination (R^2), mean absolute percentage error (*MAPE*), root mean squared error (*RMSE*), and bias.

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

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \quad (2)$$

$$RMSE\ (y, \hat{y}) = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \quad (3)$$

$$bias = \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y}) \quad (4)$$

where N represents the sample size, y_i is the value of the i-th observed data point, $\hat{y}_i$ is the value of the i-th predicted data point, and $\overline{y}$ is the mean value of N observed data points.

3. Results and Analysis

3.1. Model Calibration and Validation

Out of the entire synchronized dataset, 80% of the data (N = 318) was randomly allocated for constructing the models, whereas the remaining 20% of the data (N = 80) was employed to assess the models' performance. It is essential to emphasize that a consistent training dataset was utilized across all experiments for training and validation.

Regarding Chl-a estimation (Figure 3), it was observed that all models tended to underestimate high-concentration values, possibly due to the limited availability of data points for such values. Nevertheless, the CatBoost, RF, and XGBoost models exhibited significantly

improved performance in accurately predicting true values across both the training and test datasets when compared to the SVR models. In particular, the CatBoost model showcased a well-distributed scatter around the 1:1 line for both the training set (*RMSE* = 3.26 mg/m^2, *MAPE* = 15.18%) and the test set (*RMSE* = 11.11 mg/m^2, *MAPE* = 28.12%). This signifies a higher level of accuracy. Consequently, the CatBoost model emerges as the optimal choice for Chl-a retrieval.

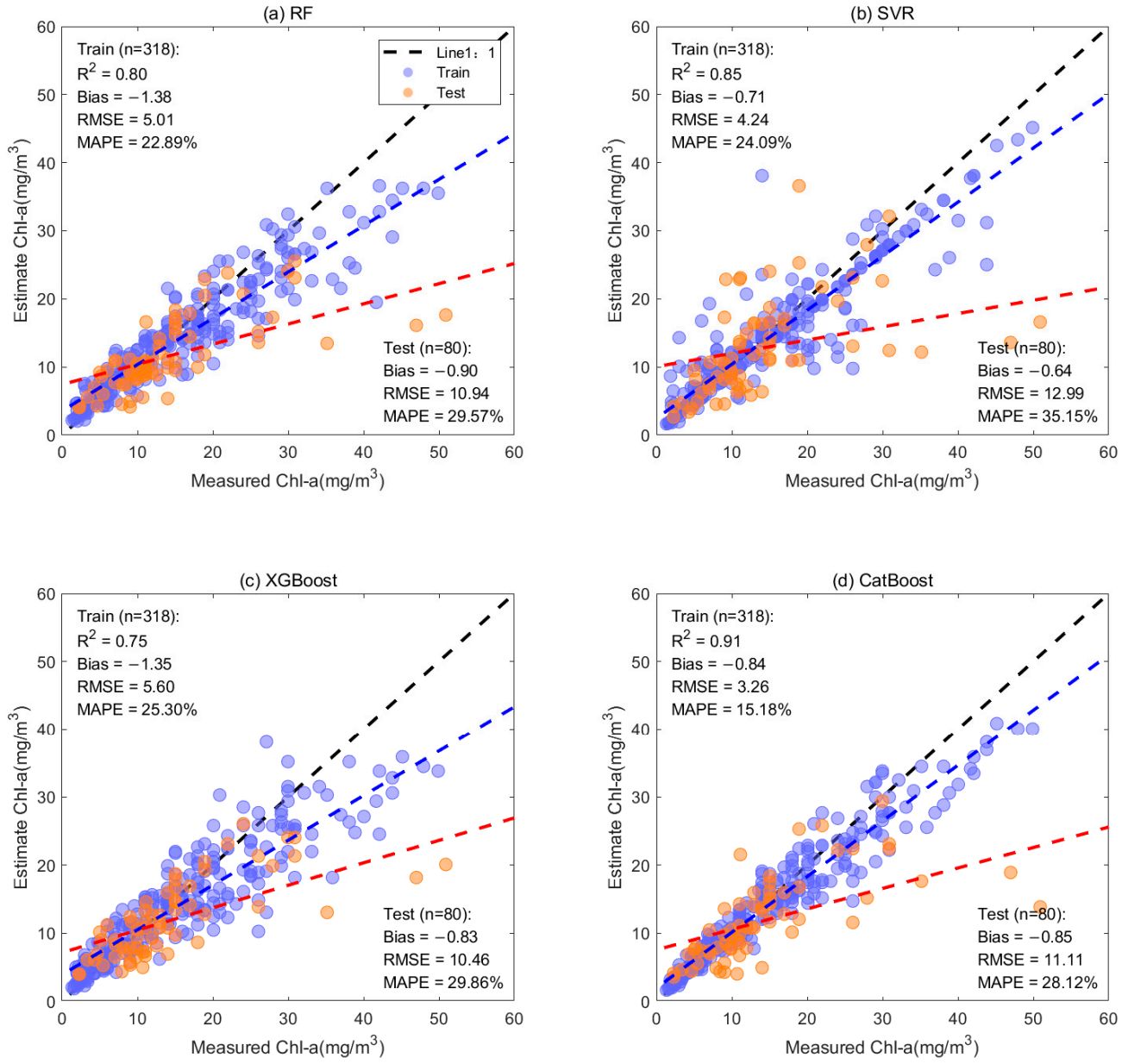

Figure 3. For the training set (n = 318) and test set (n = 80) in Dianshan Lake, scatter plots of the results of (**a**) RF, (**b**) SVR, (**c**) XGBoost, and (**d**) CatBoost for Chl-a retrieval are presented. The black, blue, and red lines represent the 1:1 line and regression lines between measured and estimated values on the training and test datasets, respectively. The blue dots and red dots represent the training set and test set, respectively.

Regarding the COD_{Mn} index (Figure 4), all models exhibited an overestimation of values with COD_{Mn} < 4.5 mg/L and an underestimation of values with COD_{Mn} > 4.5 mg/L. This phenomenon was particularly prominent in the SVR model. Although the *MAPE* values for all models remained below 15%, the performance of CatBoost stood out as notably superior to that of XGBoost, RF, and SVR. Among these models, CatBoost yielded the most

favorable results for COD_{Mn} estimation (training set: *RMSE* = 0.33 mg/L, *MAPE* = 6.85%; test set: *RMSE* = 0.55 mg/L, *MAPE* = 10.55%). Thus, the RF model emerges as a preferred choice for COD_{Mn} estimation.

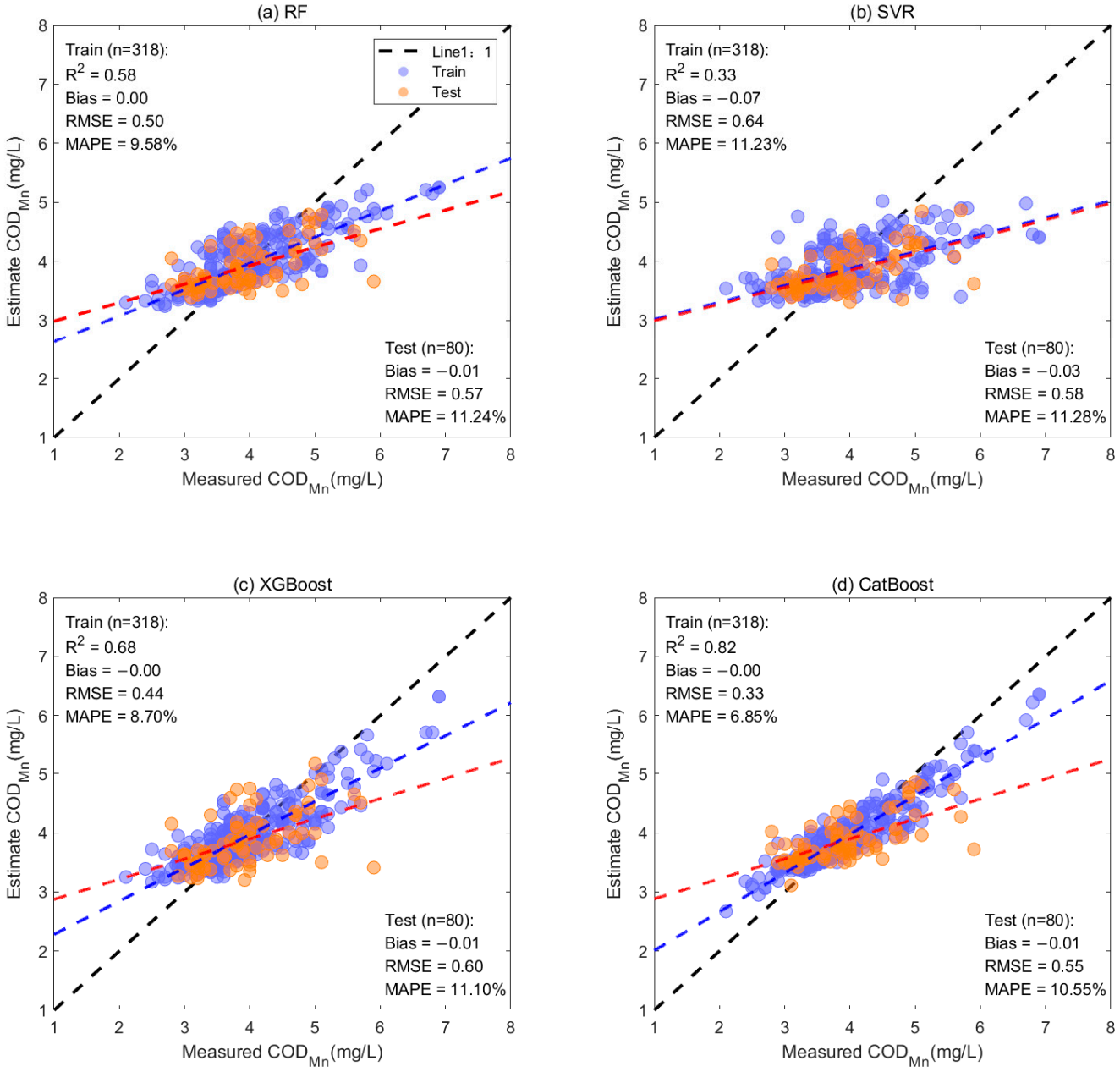

Figure 4. For the training set (n = 318) and test set (n = 80) in Dianshan Lake, scatter plots of the results of (**a**) RF, (**b**) SVR, (**c**) XGBoost, and (**d**) CatBoost for COD_{Mn} retrieval are presented. The black, blue, and red lines represent the 1:1 line and regression lines between measured and estimated values on the training and test datasets, respectively. The blue dots and red dots represent the training set and test set, respectively.

Concerning the DO index (Figure 5), all models consistently displayed a slight overestimation of low-concentration values and an underestimation of high-concentration values. It is important to highlight that all models exhibited a high degree of accuracy in estimating DO concentrations on both the training and test sets (*RMSE* < 1.5 mg/L, *MAPE* < 15%). In terms of various error metrics, it is obvious that the SVR model yields the poorest performance. Although the training set results are similar for CatBoost and XGBoost, XGBoost

performs slightly better than CatBoost on the test set. As a result, the XGBoost model (training set: *RMSE* = 1.01 mg/L, *MAPE* = 9.78%; test set: *RMSE* = 1.2 mg/L, *MAPE* = 12.11%) is considered the optimal choice for DO retrieval.

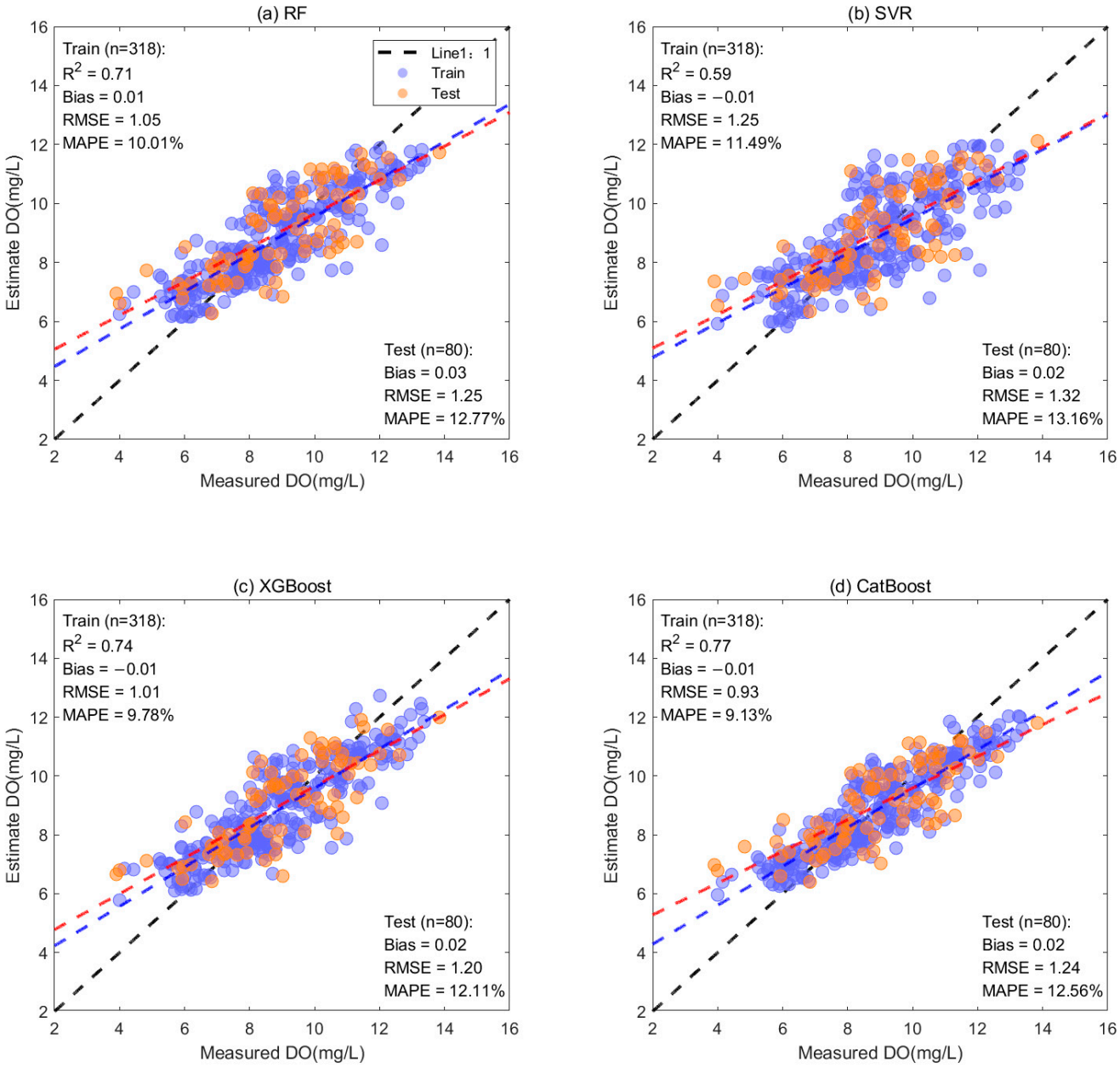

Figure 5. For the training set (n = 318) and test set (n = 80) in Dianshan Lake, scatter plots of the results of (**a**) RF, (**b**) SVR, (**c**) XGBoost, and (**d**) CatBoost models for DO retrieval are presented. The black, blue, and red lines represent the 1:1 line and regression lines between measured and estimated values on the training and test datasets, respectively. The blue dots and red dots represent the training set and test set, respectively.

In the case of the transparency index (Figure 6), the XGBoost, RF, and CatBoost models exhibited favorable results in the training set. Notably, all four models tended to overestimate SDD values when SDD < 0.4 m and underestimate values when SDD > 0.6 m. In summary, the XGBoost model showcased the best performance across both the training and test sets (training set: *RMSE* = 0.07 m, *MAPE* = 15.12%; test set: *RMSE* = 0.155 m, *MAPE* = 34.14%). Consequently, it is deemed the optimal choice for SDD retrieval.

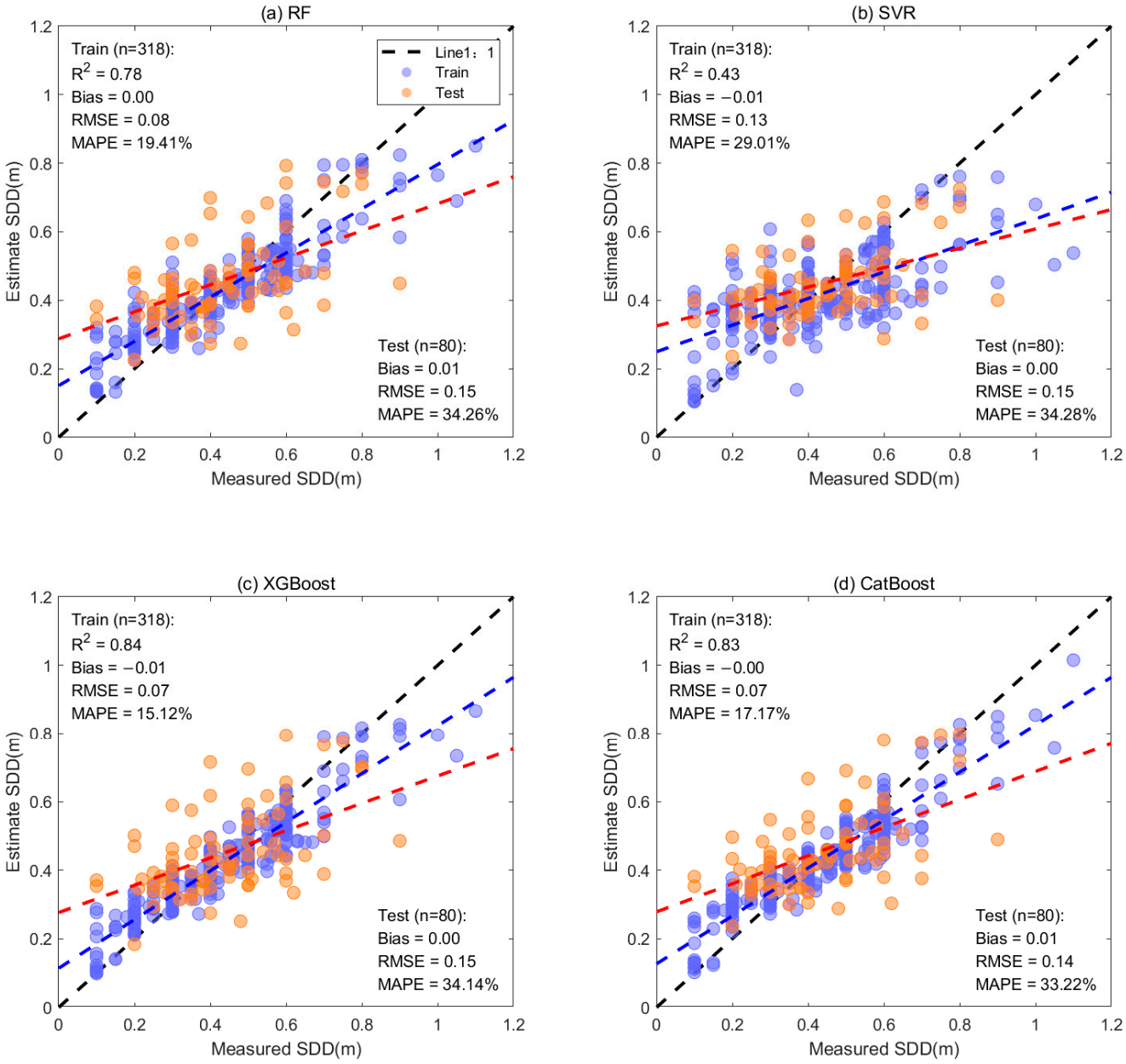

Figure 6. For the training set (n = 318) and test set (n = 80) in Dianshan Lake, scatter plots of the results of (**a**) RF, (**b**) SVR, (**c**) XGBoost, and (**d**) CatBoost models for SDD retrieval are presented. The black, blue, and red lines represent the 1:1 line and regression lines between measured and estimated values on the training and test datasets, respectively. The blue dots and red dots represent the training set and test set, respectively.

In terms of the TN index (Figure 7), the results retrieved by the four models exhibit a notable similarity. Concerning the training dataset, both the Random Forest (RF) and XGBoost models show superior performance. Their *MAPE* is below 20%. Analyzing the bias, RF outperforms all other models. Specifically, for the training set, RF demonstrates an *RMSE* of 0.45 mg/L, a *MAPE* of 19.45%, and a bias of 0. For the test set, the metrics are an *RMSE* of 0.54 mg/L, a *MAPE* of 21.83%, and a bias of −0.01. These outcomes underscore RF's heightened accuracy and stability in predictions, compared to the alternative models.

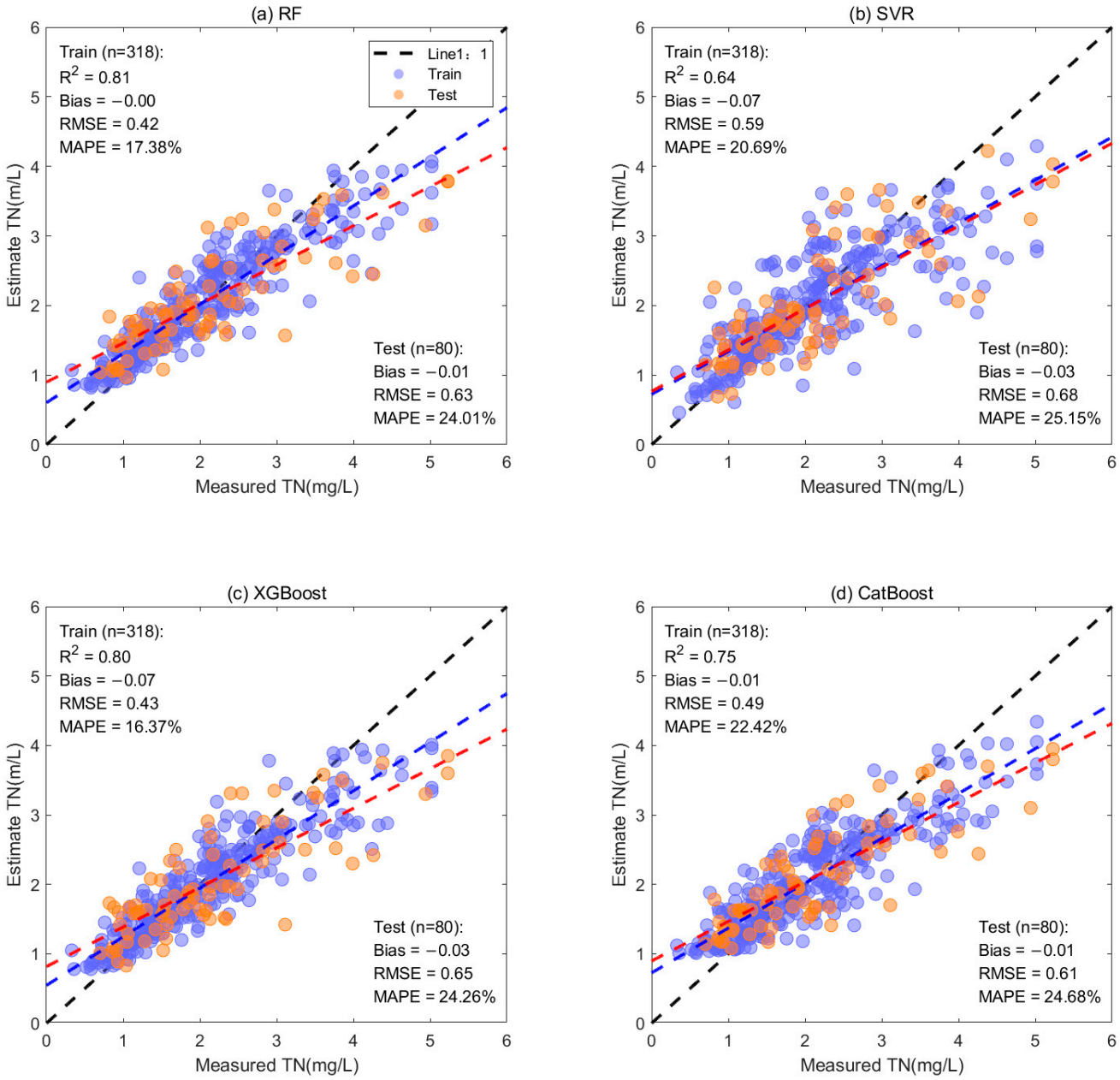

Figure 7. For the training set (n = 318) and test set (n = 80) in Dianshan Lake, scatter plots of the results of (**a**) RF, (**b**) SVR, (**c**) XGBoost, and (**d**) CatBoost models for TN retrieval are presented. The black, blue, and red lines represent the 1:1 line and regression lines between measured and estimated values on the training and test datasets, respectively. The blue dots and red dots represent the training set and test set, respectively.

Regarding the TP index (Figure 8), CatBoost notably outperformed the other models, exhibiting the best outcomes (training set: *RMSE* = 0.02 mg/L, *MAPE* = 19.2%; test set: *RMSE* = 0.036 mg/L, *MAPE* = 29.34%). Notably, the R2 values for both the training and test sets surpassed 0.75, and the *MAPE* values remained below 30%. Conversely, SVR showcased less favorable results, yielding *MAPE* values exceeding 65% across the training and test sets.

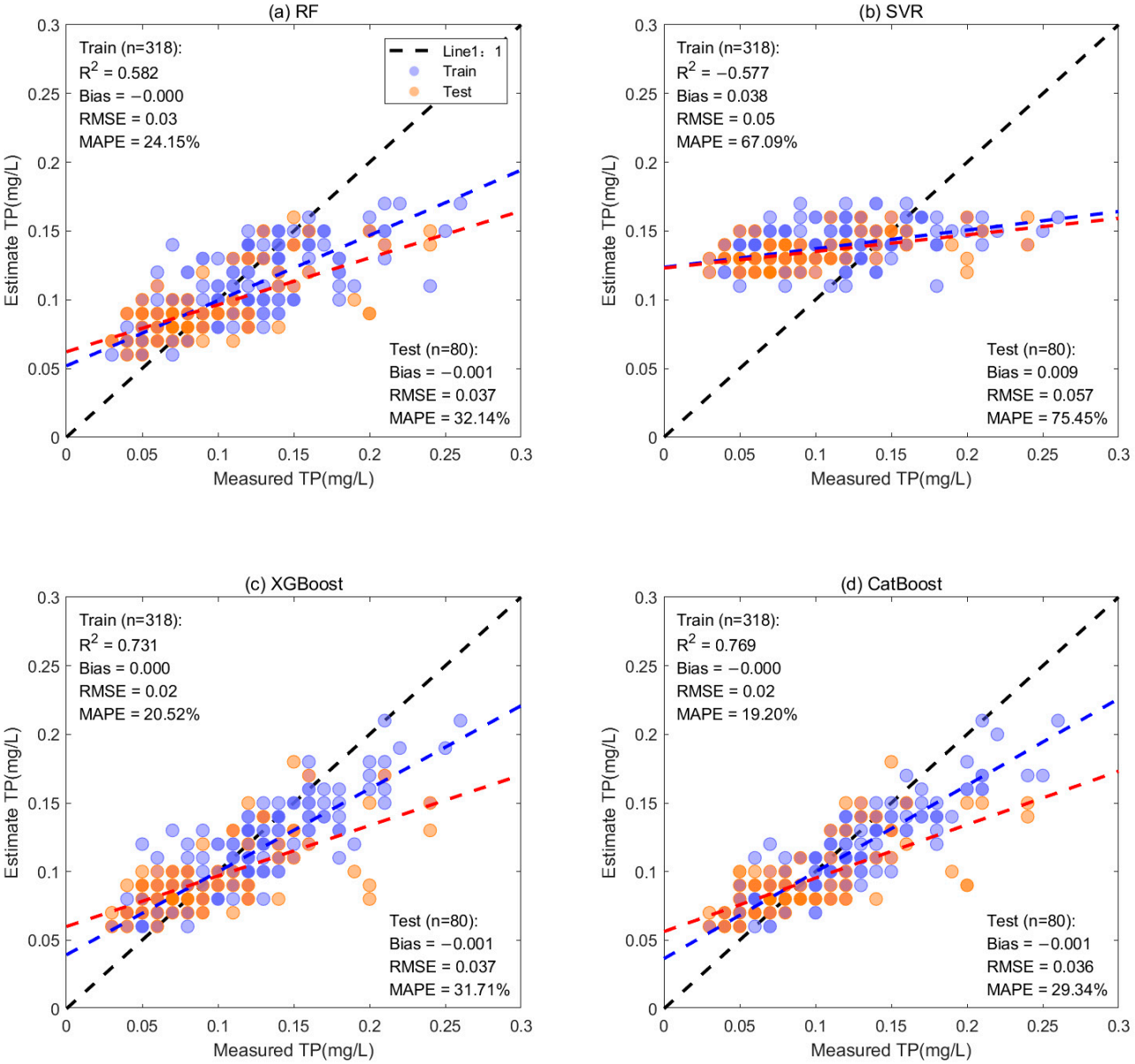

Figure 8. For the training set (n = 318) and test set (n = 80) in Dianshan Lake, scatter plots of the results of (**a**) RF, (**b**) SVR, (**c**) XGBoost, and (**d**) CatBoost models for TP retrieval are presented. The black, blue, and red lines represent the 1:1 line and regression lines between measured and estimated values on the training and test datasets, respectively.

Distinct characteristics are observed among various machine learning algorithms when predicting water quality parameters. By ranking the assessment results of the six water quality parameters, it is evident that CatBoost consistently achieves the most favorable outcomes across all four instances. XGBoost ranks within the top two positions in five out of six cases, whereas SVR consistently yields relatively inferior results across all six water quality parameters. Overall, in the evaluation of retrieval results for the six water quality parameters, CatBoost performs the best, followed by XGBoost in second place, RF in third, and SVR in the last position.

3.2. Spatiotemporal Patterns of Diandao Lake Water Quality Based on Sentinel-2

3.2.1. Temporal Variation

According to Section 3.1, it can be observed that the best models for Chl-a, CODMn, DO, SDD, TN, and TP are CatBoost, CatBoost, XGBoost, XGBoost, RF, and CatBoost,

respectively. For ease of understanding and readability, we shall refer to them as BM-Chl-a, BM-CODMn, BM-DO, BM-SDD, BM-TN, and BM-TP. In this section, the best models were employed to estimate the concentrations of Chl-a, DO, CODMn, SDD, TN, and TP. Yearly average images (Figure 9) for these water quality parameters were calculated from 2017 to 2022 (data for 2023 were available only for the first four months and were excluded from this analysis). In addition, we also plotted the overall monthly average image (Figure 10) from 2017 to 2023. To gain a more intuitive understanding of the temporal changes in various water quality parameters, we have compiled their quarterly averages for each year (Figure 11).

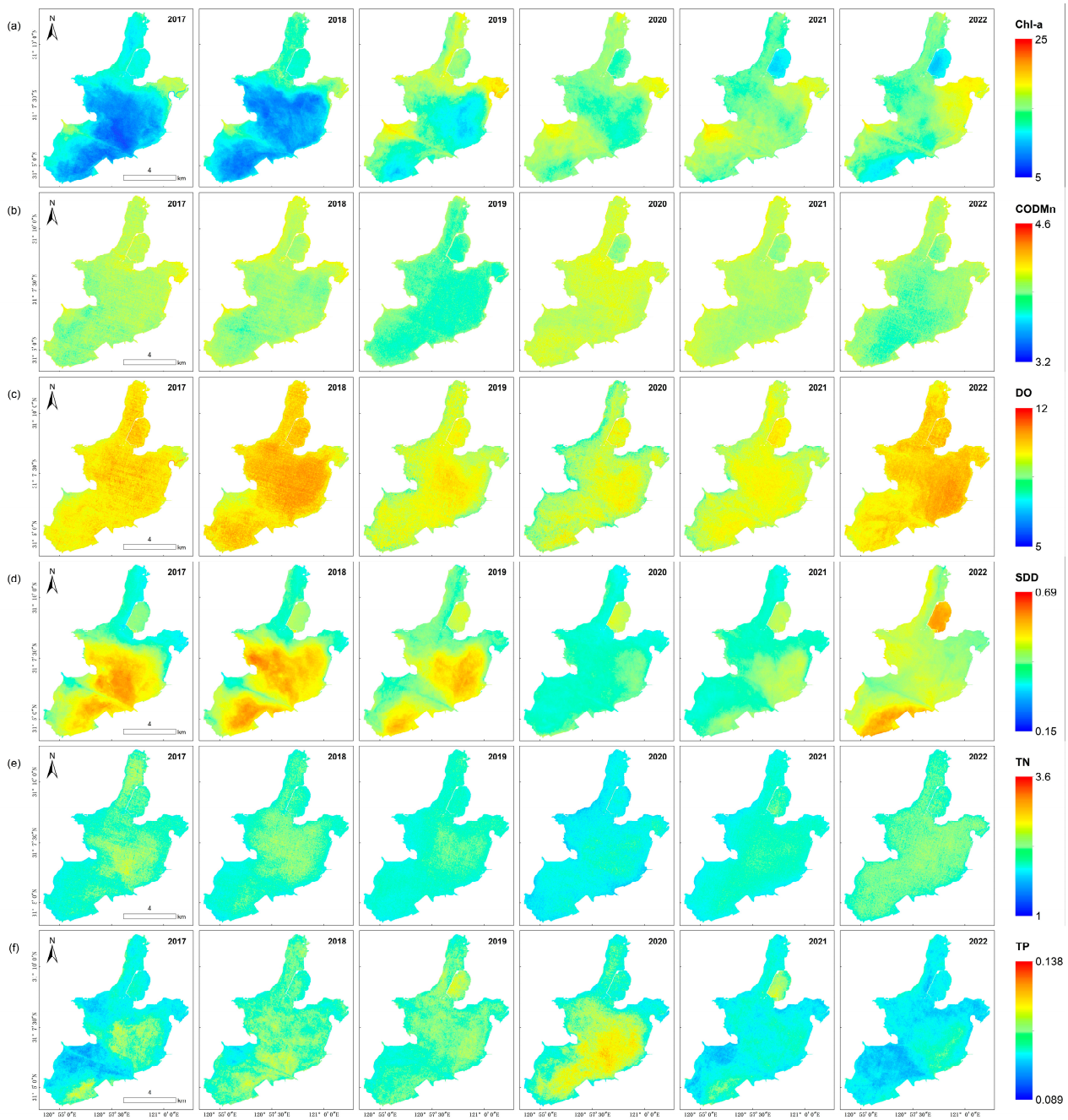

Figure 9. Images depicting the annual average concentrations of (**a**) Chl-a, (**b**) COD_{Mn}, (**c**) DO, (**d**) SDD, (**e**) TN, and (**f**) TP in Dianshan Lake, retrieved using Sentinel-2 MSI imagery, for the years 2017 to 2022.

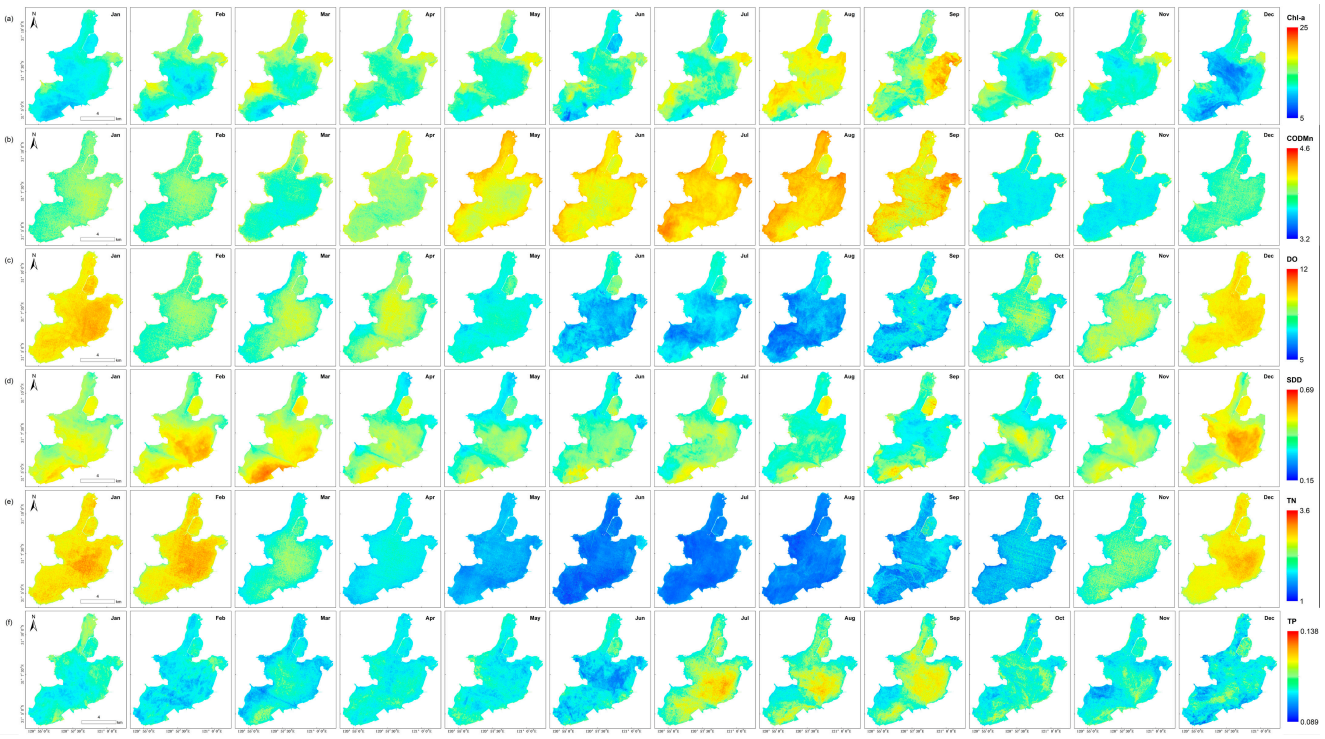

Figure 10. Images depicting the monthly average concentrations of (**a**) Chl-a, (**b**) COD_{Mn}, (**c**) DO, (**d**) SDD, (**e**) TN, and (**f**) TP in Dianshan Lake, retrieved using Sentinel-2 MSI imagery, for the years 2017 to 2023.

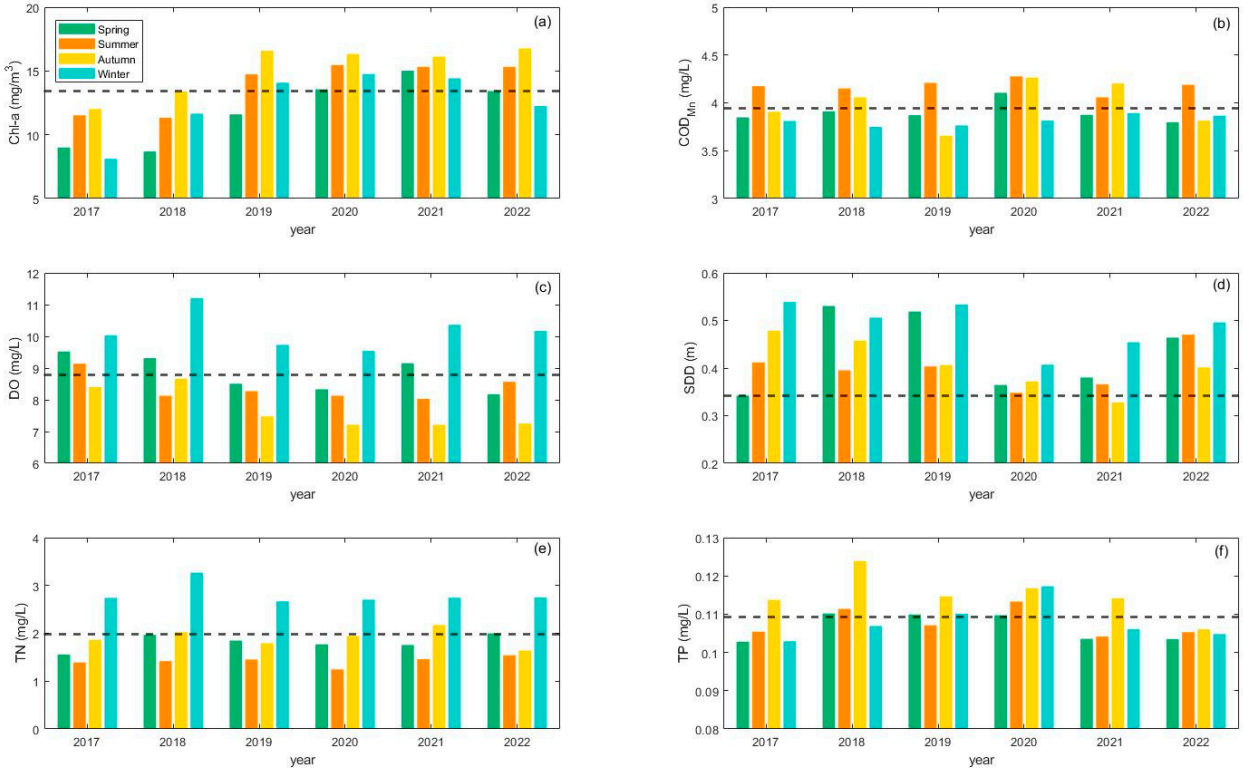

Figure 11. Bar charts illustrating the seasonally average concentration distribution of (**a**) Chl-a, (**b**) COD_{Mn}, (**c**) DO, (**d**) SDD, (**e**) TN, and (**f**) TP in Dianshan Lake from 2017 to 2023. The black dashed line represents the average value of water quality parameters calculated using six years of data from 2017 to 2022.

Upon analysis, the average Chl-a concentration over the six years was found to be 13.53 ± 2 mg/m^3. The lowest recorded value occurred in 2017 at 10.86 mg/m^3, while the highest was observed in 2020 at 15.03 mg/m^3. There was a continuous upward trend in Chl-a concentration from 2017 to 2020, with relatively minor interannual differences between 2020 and 2022. However, during the summer, autumn, and winter seasons, the concentrations showed a decreasing trend compared to 2020 (Figure 11a). The average COD_{Mn} concentration was determined to be 3.94 ± 0.4 mg/L. COD_{Mn} exhibited overall small fluctuations, with seasonal averages ranging between 3.5 and 4.5 mg/L across the years (Figure 11b). The lowest value was observed in 2019 at 3.9 mg/L, while the highest was recorded in 2020 at 4.04 mg/L. The average DO concentration amounted to 9.89 ± 0.42 mg/L. The lowest concentration was observed in 2020 at 9.37 mg/L, while the highest concentration was noted in 2018 at 10.38 mg/L. DO also showed a declining trend from 2017 to 2020, with an increase in concentration observed in the spring and winter of 2021, followed by another decrease in 2022 (Figure 11c). For SDD, the average value was 0.44 ± 0.04 m. SDD did not exhibit a clear pattern of change, but overall, it showed a trend of initially decreasing and then increasing. SDD values were higher in 2017–2019, lower in 2020 and 2021, and increased again in 2022 (Figure 11d). The average TN concentration was calculated to be 2.08 ± 0.1 mg/L. The lowest concentration occurred in 2019 at 1.91 mg/L, and the highest was observed in 2021 at 2.21 mg/L. TN displayed relatively small interannual differences, indicating stable changes over the years (Figure 11e). Lastly, the average TP concentration was measured at 0.109 ± 0.003 mg/L. The lowest value was registered in 2022 at 0.105 mg/L, whereas the highest value was recorded in 2020 at 0.111 mg/L.

The seasonal variations in water quality parameters mirror their monthly fluctuations. Chl-a, COD_{Mn}, and TN concentrations exhibit higher levels in the summer and autumn, while they demonstrate lower levels in the spring and winter. Conversely, other water quality parameters display the opposite trend (Figures 10 and 11).

3.2.2. Spatial Variation

To explore the spatial variations of various water quality parameters within Dianshan Lake, we conducted a comprehensive analysis by computing the mean values based on data collected from 100 images. The annual average values obtained for Dianshan Lake were 13.73 mg/m^3 for Chl-a, 3.94 mg/L for COD_{Mn}, 8.92 mg/L for DO, 0.44 m for SDD, 2.09 mg/L for TN, and 0.11 mg/L for TP, respectively. The corresponding standard deviations were recorded as 4.4 mg/m^3, 0.29 mg/L, 1.47 mg/L, 0.11 m, 0.74 mg/L, and 0.013 mg/L.

The mean images of each water quality parameter reveal distinct spatial patterns within Dianshan Lake. There is a clear negative correlation between Chl-a and SDD distributions, wherein areas with higher Chl-a concentrations tend to exhibit lower transparency (Figure 12a,d). Within the lake area, the northern and southwestern regions demonstrate elevated Chl-a concentrations, particularly near the entrances of Jishuigang Harbor in the northeast and the western region, where Chl-a concentrations reach their peaks. In contrast, the central open areas of the lake and the southeastern region exhibit lower Chl-a concentrations.

Similar to Chl-a, the spatial distribution of COD_{Mn} and TN also displays a correlation (Figure 12b,e). COD_{Mn} exhibits a discernible distribution pattern throughout the lake, with higher concentrations observed along the lake's edges and lower concentrations in the open areas within the lake. DO concentrations are lowest near the entrance of Jishuigang Harbor in the southwestern region, while the central open areas and northern regions of the lake demonstrate higher DO concentrations (Figure 12c).

Regarding nitrogen content, TN concentrations are lower in the lake's edge regions and higher in the open areas within the lake (Figure 12e). Similarly, the spatial distribution trend of TP resembles that of TN (Figure 12f). The southwestern and northeastern regions

of the lake exhibit lower TP concentrations, while the eastern areas and central open regions display higher TP concentrations.

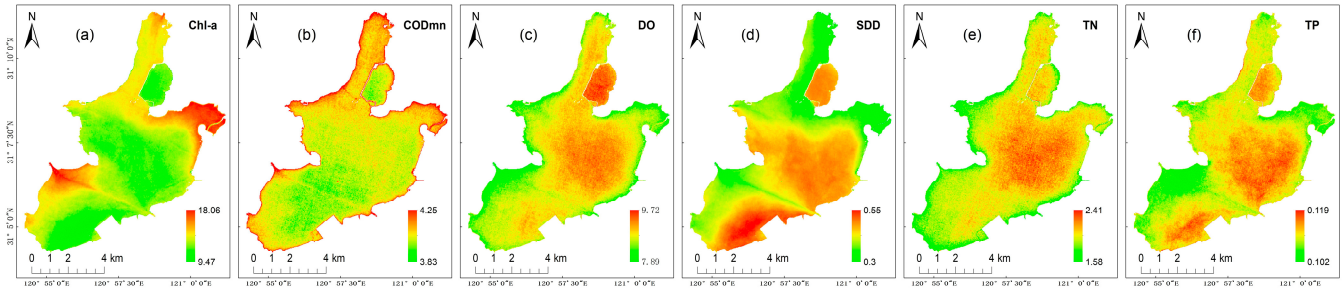

Figure 12. Average concentration maps of (**a**) Chl-a, (**b**) COD_{Mn}, (**c**) DO, (**d**) SDD, (**e**) TN, and (**f**) TP in Dianshan Lake from 2017 to 2022.

To investigate the spatial variations of Dianshan Lake more comprehensively, we calculated the coefficient of variation map for the entire lake area (Figure 13). It can be observed that regions with higher concentrations of DO, TN, and TP tend to exhibit larger variability. Similarly, SDD follows a similar pattern, with regions showing higher values appearing in red hues, indicating greater coefficients of variation. In contrast, the standard deviation of COD_{Mn} remains relatively consistent across the entirety of the lake, suggesting an overall lower variability, which is in line with its temporal variation image. The situation for Chl-a is slightly different, whereby regions with lower average concentrations across the lake display relatively unstable conditions, implying significant variability.

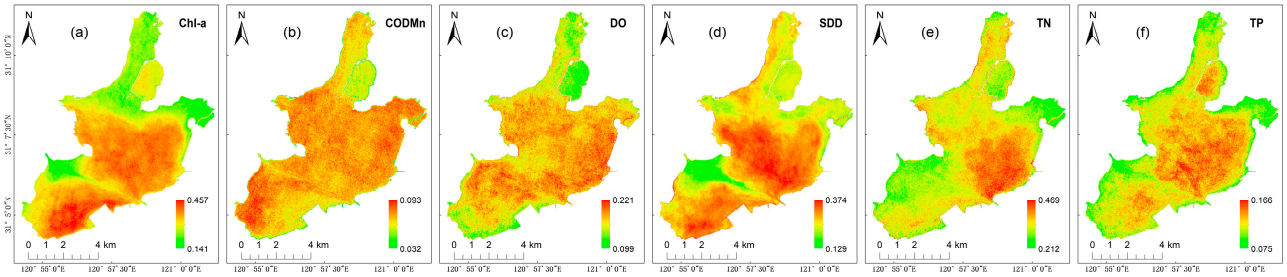

Figure 13. Coefficient of variation maps of (**a**) Chl-a, (**b**) COD_{Mn}, (**c**) DO, (**d**) SDD, (**e**) TN, and (**f**) TP concentrations in Dianshan Lake from 2017 to 2022.

4. Discussion

4.1. Applicability of the Models

In the practical application of Dianshan Lake, promising outcomes have been achieved through the construction of models utilizing both actual measurement data from Dianshan Lake and satellite reflectance data, enabling the prediction of various water quality parameters. To comprehensively assess the applicability of the established best models for various water quality parameters, further in-depth research was conducted. Considering Dianshan Lake as a representative small lake with poor-to-low nutrient levels, we extended our investigation to Taihu Lake—a larger lake characterized by higher nutrient levels. The primary objective was to validate the generality and stability of BM-Chl-a, BM-CODMn, BM-DO, BM-SDD, BM-TN, and BM-TP across diverse environmental contexts. The performance of the best model for each parameter in the Taihu Lake dataset is shown in Table 5.

Based on the outcomes, noteworthy shifts were observed in the prediction performance of Chl-a. The *RMSE* of Chl-a escalated to 19.88 mg/m^2, while the *MAPE* surged to 45%, nearly doubling the previous values. Consequently, BM-Chl-a exhibited substantial errors in predicting Chl-a, signifying a diminished predictive capacity for this parameter. The prediction errors of BM-CODMn and BM-DO also displayed some increase. The *RMSE* values (0.74 mg/L and 1.69 mg/L) and *MAPE* values (15% and 15%) both grew by

approximately half. Although the prediction outcomes remained within a certain range, in comparison to the prior test data, these two models exhibited heightened uncertainty in predicting COD_{Mn} and DO. Regarding BM-SDD, the model's predictive performance showed improvement, as reflected by diminished *RMSE* (0.07 m) and *MAPE* (16%) values, indicating enhanced accuracy in forecasting water transparency concentrations. However, in the instances of BM-TN and BM-TP, the models' performance faltered. The predictive errors for these two indicators markedly increased when contrasted with the test data from Dianshan Lake, particularly the *MAPE* values (55% and 68%), which tripled. Consequently, BM-TN and BM-TP demonstrated marked limitations, with their forecasts significantly diverging from actual conditions.

Table 5. Performance of the best model of each parameter on Taihu Lake.

Water Quality Parameter	*RMSE*	*MAPE*	Bias
Chl-a	19.88 mg/m^3	44.88%	−12.14 mg/m^3
COD_{Mn}	0.74 mg/L	14.61%	0.08 mg/L
DO	1.69 mg/L	14.88%	−1.25 mg/L
SDD	0.07 m	15.7%	−0.013 m
TN	1.5 mg/L	54.78%	10.45 mg/L
TP	0.05 mg/L	67.56%	0.034 mg/L

In light of the models' predictive outcomes compared to the authentic test data from Taihu Lake, it can be deduced that the predictive performance of parameters such as Chl-a, COD_{Mn}, DO, TN, and TP exhibited varying degrees of decline or fluctuation. Only the predictive performance for SSD maintained a relatively favorable state. In essence, the best models for Dianshan Lake displayed specific restrictions and inadequacies in predicting water quality parameters for Taihu Lake. Further enhancement and optimization are imperative through the incorporation of localized data to augment its predictive prowess.

4.2. Performance and Evaluation of Machine Learning Algorithms

4.2.1. Analysis of Error Sources Affecting Model Performance

In the realm of machine learning, the quality of the dataset has a direct bearing on the performance of the model. Additionally, the congruence between field estimations and satellite-derived estimations stands as a crucial criterion for evaluating the efficacy of the proposed Chl-a algorithm. Given the foundation of our study in employing satellite reflectance and measurement data to construct retrieval models, the quality of satellite reflectance data becomes particularly salient.

First and foremost, atmospheric correction presents itself as a principal source of error. Particularly, in comparison to expansive oceanic regions, atmospheric correction for inland water body imagery proves to be a more intricate endeavor due to intricate interactions with neighboring land pixels. Rectifying atmospheric effects over water surfaces is particularly demanding, often requiring a higher degree of precision compared to correction procedures applied over terrestrial areas. In this study, a specialized atmospheric correction method tailored for inland water bodies [48,49] was employed. Following Rayleigh correction, the "dark pixel method" was transferred to the shortwave near-infrared band [32,45,50,55] to mitigate aerosol effects. This approach ensures a maximum level of correction accuracy, even in the presence of unavoidable errors.

Furthermore, the alignment of time windows [56] and pixel window [57] sizes for on-site estimations and satellite data introduces error sources within the retrieval model. While Sentinel-2 MSI imagery follows a five-day orbital cycle, practical limitations arising from adverse weather conditions considerably restrict the number of images that effectively align with measured data. According to statistics from pertinent water management authorities in Shanghai, Dianshan Lake has an approximate water turnover cycle of seven days. Accordingly, we extended the time window to five days to secure a more substantial dataset alignment. Apart from Chl-a, the correlation between other water quality parameters and

single-band reflectance remains relatively low (Figure 14). Encouragingly, correlations involving combinations of bands display improved trends. Looking ahead, enhancing the frequency of in situ measurements to shorten the time window and acquire more aligned data warrants consideration.

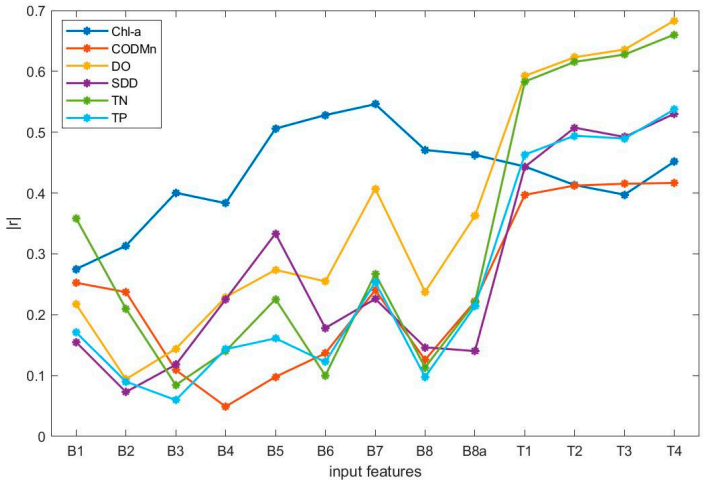

Figure 14. Machine learning input features and correlation coefficients of each water quality parameter. B1~B8a represent the corresponding bands of the Sentinel-2 MSI image, and T1~T4 represent the band combination with the greatest correlation with each water quality parameter ($R_{rs}(\lambda 1) - R_{rs}(\lambda 2)$, $\frac{R_{rs}(\lambda 1)}{R_{rs}(\lambda 2)}$, $\frac{R_{rs}(\lambda 2) - R_{rs}(\lambda 1)}{R_{rs}(\lambda 2) + R_{rs}(\lambda 1)}$, $R_{rs}(\lambda 3) \times (\frac{1}{R_{rs}(\lambda 2)} - \frac{1}{R_{rs}(\lambda 1)})$).

Moreover, to demonstrate the effectiveness of satellite reflectance pixel windows, different window sizes were employed in constructing retrieval models, including single-pixel windows, 3 × 3 pixel window averages, and 5 × 5 pixel window averages.

RMSE and *MAPE* are employed as performance metrics. As illustrated in Table 6, for both *RMSE* and *MAPE*, the 3 × 3 pixel window consistently yields the lowest error across various water quality parameter retrievals. This aligns with our predicted outcomes. The single-pixel window exhibits higher variability, potentially stemming from greater noise. The 5 × 5 pixel window, with a ground resolution of 100 m × 100 m, might excessively "smooth" the data, thus diminishing spectral features. In contrast, the 3 × 3 pixel window effectively eliminates noise while retaining significant water feature information in the region, thereby maximizing water quality uniformity.

Table 6. Accuracy display of water quality parameters using different pixel windows and different methods to build models.

Parameters	Methods	1 × 1		3 × 3		5 × 5	
		RMSE	***MAPE* (%)**	***RMSE***	***MAPE* (%)**	***RMSE***	***MAPE* (%)**
Chl-a (mg/m^3)	CatBoost	11.19	29.12	11.11	28.12	11.21	29.12
	RF	10.94	29.87	10.94	29.57	10.99	31.57
	SVR	13.99	38.15	12.99	35.15	13.99	39.15
	XGBoost	10.46	29.86	10.46	30.86	10.96	31.86
CODMn (mg/L)	CatBoost	0.58	11.29	0.57	11.24	0.59	11.87
	RF	0.60	11.33	0.58	11.28	0.62	12.19
	SVR	0.61	11.63	0.60	11.10	0.61	12.11
	XGBoost	0.57	10.60	0.55	10.55	0.57	11.32
DO (mg/L)	CatBoost	1.27	12.89	1.25	12.77	1.31	13.16
	RF	1.32	13.24	1.32	13.16	1.35	13.68
	SVR	1.24	12.29	1.20	12.11	1.24	12.29
	XGBoost	1.34	13.14	1.24	12.56	1.28	12.97

Table 6. *Cont.*

Parameters	Methods	1 × 1		3 × 3		5 × 5	
		RMSE	***MAPE* (%)**	***RMSE***	***MAPE* (%)**	***RMSE***	***MAPE* (%)**
SDD (m)	CatBoost	14.44	34.38	14.73	34.26	14.65	34.63
	RF	14.69	34.44	14.77	34.28	14.75	34.90
	SVR	14.80	34.14	14.74	34.14	14.81	34.57
	XGBoost	14.38	34.03	14.15	33.22	14.35	34.31
TN (mg/L)	CatBoost	0.72	25.83	0.63	24.01	0.68	26.14
	RF	0.82	26.08	0.68	25.15	0.65	26.33
	SVR	0.73	24.80	0.65	24.26	0.62	25.07
	XGBoost	0.69	25.81	0.61	24.68	0.61	24.76
TP (mg/L)	CatBoost	0.04	33.13	0.04	32.14	0.04	32.91
	RF	0.06	78.09	0.06	75.45	0.06	74.07
	SVR	0.04	32.47	0.04	31.71	0.04	32.02
	XGBoost	0.04	30.20	0.04	29.34	0.04	31.88

4.2.2. Evaluation of the Models

Based on the previous analysis, it is evident that CatBoost has the greatest potential for application in inland water quality assessment. It demonstrated superior performance in predicting Chl-a, COD_{Mn}, SDD, and TP. The convincing results are particularly evident in Table 6, where changes in pixel window size did not affect the accuracy trend. The main challenge of machine learning modeling is the need for extensive samples. CatBoost excels with small datasets, effectively curbing overfitting and providing valuable insights into feature importance, aiding in understanding model performance and predictions [58].

Among the models, SVR exhibited the weakest performance. It notably erred significantly in predicting TP (Figure 8). SVR's performance might excel with high-quality data [38], which could explain its unsatisfactory performance when modeling satellite reflectance data due to atmospheric correction errors. XGBoost and RF also show promise, as prior studies highlight their utility in inland lake water quality assessment [24,38].

4.3. Spatiotemporal Change Analysis

After analysis, it was found that there were minimal overall differences in the temporal variations of COD_{Mn}, TN, and TP, which may be related to the nutrient status of Dianshan Lake. The lake's overall eutrophication is not severe, with occasional occurrences of algae blooms in late summer and early autumn. Except for TN, almost all other parameters indicate better water quality during the winter and spring seasons compared to the summer and autumn seasons. TN, in particular, exhibits a pronounced seasonal variation trend over six years, with notably high values during the winter. This phenomenon may result from the combined influence of multiple factors.

Environmental and meteorological factors such as water temperature, air temperature, and precipitation can affect water quality [59]. We compiled monthly average values of environmental factors (water temperature, pH, conductivity) and meteorological factors (air temperature, precipitation, wind speed) for Dianshan Lake and examined their relationships with water quality parameters through correlation analysis (Table 7). Water temperature, air temperature, and precipitation showed strong correlations with various water quality parameters, while pH and conductivity were significantly correlated only with Chl-a and TP, and wind speed exhibited weak correlations with all parameters. Due to climate-related factors, both air and water temperatures reach their lowest points in winter (December, January, February) and peak in late summer and early autumn (August, September). Similarly, precipitation is very high in summer and very low in winter. These factors can influence the intensity of chemical reactions in the water, as well as the variation of nutrients and chemicals released from sediments, leading to changes in water chemistry and characteristics. Precipitation also drives the input of nutrients from the lake's

surroundings, which can promote algal growth and increase concentrations of Chl-a, thus improving water quality in spring and winter compared to summer and autumn.

Table 7. Pearson correlation coefficient of water quality parameters and environmental factors.

Index	Chl-a	COD_{Mn}	DO	SDD	TN	TP
Water temperature	0.38	0.71 *	−0.94 *	−0.86 *	−0.32	−0.1
PH	0.73 *	0.46	−0.35	−0.19	−0.24	−0.25
Conductivity	−0.4	−0.12	0.43	0.56	0.38	0.39
Air temperature	0.41	0.82 *	−0.97 *	−0.82 *	−0.35	−0.13
Precipitation	0.78 *	0.45	−0.72 *	−0.42	−0.67 *	−0.37
Wind speed	0.15	0.43	−0.42	−0.21	−0.49	−0.33

* Represents significant correlation, Pearson correlation coefficient > 0.6.

Agricultural nonpoint source pollution is widely recognized as one of the most important nutrient sources contributing to water quality deterioration [60]. Therefore, changes in land use, especially in the area of farmland, can also impact water quality parameters [43]. Dianshan Lake is designated as a protected area for drinking water resources, with restrictions on industrial development and a ban on livestock farming. Previous research has indicated that agriculture is the primary source of pollution leading to deteriorating water quality in this region [61]. One of the primary reasons for agricultural pollution of water quality is the widespread use of agricultural chemicals such as fertilizers and pesticides. These chemicals are washed into lakes by rainwater, leading to an increase in the concentrations of TN and TP in the water, thereby triggering eutrophication issues. As TN and TP levels rise, the content of Chl-a also increases, resulting in the overgrowth of algae in the water body. This leads to a decline in water quality, characterized by severe eutrophication, and a potential decrease in the concentration of DO, adversely affecting aquatic organisms. Additionally, due to soil erosion and wastewater discharges, there may be an increase in suspended solids in the water, leading to increased water turbidity and reduced water transparency. These changes in the range of water quality parameters can be attributed to the adverse impact of agricultural activities on the water body. Data obtained from the Statistical Yearbook website (http://www.tjnjw.com/, accessed on 4 September 2023) show that the farmland area in Qingpu District decreased from 25,466 hectares in 2017 to 20,581 hectares in 2019 and then increased to 25,400 hectares in 2021. The farmland area decreased initially and then increased, with the smallest area recorded in 2019, coinciding with the year when various parameters reached extreme values over the six years (as observed in Section 3.2.1). This suggests a strong correlation between the area of farmland around Dianshan Lake and water quality: when the farmland area decreases, water quality improves, and when the farmland area increases, water quality deteriorates.

As indicated in Section 3.2.2, water quality parameters exhibit noticeable spatial differences, which could be related to the uneven distribution of water flow, sediment, and nutrient inputs. It is noteworthy that there are significant differences between the water quality parameters at the inlet and outlet of Dianshan Lake. The inlet is typically a critical area for water quality changes, as it is directly influenced by the surrounding environment, while the outlet may be influenced by internal lake ecosystems and processes. Based on the mean values of water quality parameters (Figure 12), at the inlet of Dianshan Lake, the Chl-a concentration and COD_{Mn} concentration are higher, while DO and SDD concentrations are lower. Conversely, at the outlet, the opposite trend is observed. Jishui Port takes the shipping channel as the main water function, and the traffic discharge enters the lake area along with the entrance. The influx of water from tributaries introduces a large amount of suspended sediment and organic matter, increasing the concentrations of Chl-a and COD_{Mn}. The water near the inlet of the lake is much turbid compared to the open areas within the lake, resulting in lower transparency.

4.4. Strengths and Limitations of the Study

This study possesses significant strengths across various dimensions. Firstly, our study adopts a direct model built upon image reflectance data. This strategy mitigates the influence of atmospheric correction on research outcomes to a certain extent, effectively curtailing error propagation from modeled measured data to satellite image application. Consequently, this approach bolsters the stability of the proposed model. Secondly, our study transcends the limitations of assessing algorithm performance solely based on individual water quality parameters. Instead, we amalgamate multiple pivotal water quality parameters and gauge the efficacy of six distinct machine learning methods. In comparison to appraising methods solely on a singular water quality parameter, our all-encompassing evaluation strategy is more holistic and precise. This curtails uncertainty in evaluation findings and enhances the trustworthiness of research conclusions. Lastly, our study ventures beyond the exploration of modeling techniques. It encompasses a spatiotemporal analysis of diverse water quality parameters within Dianshan Lake. This examination of spatiotemporal distribution furnishes invaluable insights into water quality retrieval within small lakes. Moreover, our study deepens the comprehension of small lake ecosystems, delivering substantial groundwork for informed decisions regarding lake water quality management and preservation. The analytical results furnish novel viewpoints and avenues for research and practical applications in related spheres.

While this study has made commendable advancements, it is important to address a few noteworthy considerations. Firstly, regarding the distribution of the dataset, we acknowledge that there are certain limitations in terms of data samples within high- and low-concentration ranges. In our application of the model to the Taihu Lake region, we observed that predictions only for SDD (water transparency) were notably accurate. Therefore, there might be room for improvement in predicting other water quality parameters. This disparity could potentially arise from the distinct characteristics of Dianshan Lake and Taihu Lake, but it has not impeded the model's practical applicability across different regions. Secondly, when machine learning is used for prediction, the precision of measured parameter concentrations can significantly impact the model's performance, as evident in the prediction of TP and SDD. Since TP concentrations are typically quite low, often below 0.1 mg/L, and our actual measurements are controlled only up to 0.001 mg/L, this results in multiple identical TP concentration values in the measured data. This accuracy issue leads to a situation where a specific TP concentration may correspond to multiple different reflectance spectral data, increasing the difficulty for the model to distinguish similar values and, consequently, resulting in poorer model performance. The same holds true for SDD. Future work could consider improving measurement precision based on the concentration distribution of measured water quality parameters to enhance model performance. Finally, we employed a time window of five days to synchronize the measured and satellite data. It is acknowledged that changes within the water body could transpire during this period, and future research could contemplate the integration of additional observational data to expand the dataset's scope.

5. Conclusions

This study utilized satellite data and in situ measurements to determine the optimal models for Chl-a, COD_{Mn}, DO, SDD, TN, and TP from four machine learning models (RF, SVR, XGBoost, and CatBoost), which were identified as CatBoost, CatBoost, XGBoost, XGBoost, RF, and CatBoost, respectively. The applicability of these models was validated using data from Taihu Lake. These models were then applied to Sentinel-2 imagery from 2017 to 2023 to obtain the spatiotemporal distribution of water quality parameters in Dianshan Lake. Image inversion results indicated that the overall distribution of water quality parameters in the study area was uneven, with significant spatial variation, relatively minor interannual differences, and significant seasonal patterns. Further analysis revealed that the spatiotemporal variation of water quality parameters was influenced by climatic factors such as temperature and precipitation, as well as human activities including agriculture

and industry. The results of this study indicate that constructing models using multispectral satellite image reflectance and in situ water quality parameter sampling data is effective. Furthermore, in the future, model enhancement can be further achieved by improving the precision of in situ data, reducing the data time window, such as utilizing multisource satellite data, and implementing other methods. In conclusion, our study demonstrates advantages in methodology, data processing, and practical implementation. It provides valuable practical experience for the accurate monitoring of water quality parameters in small water bodies using satellite data and offers essential data support for local water resource management and environmental protection.

Author Contributions: Conceptualization, C.G. and L.D.; methodology, L.D.; validation, L.L., Z.Y. and C.G.; resources, H.H., E.W. and Z.L.; data curation, C.G.; writing—original draft preparation, L.D.; writing—review and editing, L.D.; supervision, Z.Y.; project administration, Y.H.; funding acquisition, E.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shanghai 2021 "Science and Technology Innovation Action Plan" Social Development Science and Technology Research Project (21DZ1202500), the Jiangsu Provincial Water Conservancy Science and Technology Research Project (No. 2020068), and the Science and Technology Project of the Shanghai Municipal Water Bureau (Shanghai Branch 2021-10, Shanghai Branch 2018-07).

Data Availability Statement: Not applicable.

Acknowledgments: We are also thankful to all anonymous reviewers for their constructive comments provided on the study.

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

Appendix A

Table A1. Methods for Determining Some Water Quality Parameters in Chinese National Standard GB3838-2002.

Water Quality Parameter	Determination Method
Chl-a	Nitrite Reduction Method and Continuous-Flow Analysis
COD_{Mn}	High-Temperature Oxidation Method and Continuous-Flow Analysis
DO	Electrode Method and Continuous-Flow Analysis
SDD	Potassium Permanganate Spectrophotometric Method and Continuous-Flow Analysis
TN	Ether Extraction–Spectrophotometric Method
TP	Transparency Meter Measurement

Table A2. The best hyperparameters found by grid optimization of the models.

Model	Hyperparameters	Chl-a	COD_{Mn}	DO	SDD	TN	TP
RF	n_estimators	450	500	360	390	490	300
	max_depth	40	25	10	45	20	20
	min_samples_split	5	4	5	11	7	3
	min_samples_leaf	3	5	7	2	3	7
SVR	C	4.91	2	8.67	9.82	7.52	2.94
	kernel	'rbf'	'rbf'	'rbf'	'rbf'	'rbf'	'linear'
	gamma	88.109	58.907	29.329	80.078	66.251	16.369
XGBoost	learning_rate	0.16	0.015	0.085	0.04	0.035	0.155
	gamma	0.001	0.003	0.003	0.001	0.001	0.003
	min_child_weight	9	5	8	8	9	6
	max_depth	2	2	10	6	6	8
	sub_sample	1	1	0.8	1	0.8	1
	reg_alpha	0.1	1	1	0.01	1	0.01
CatBoost	iterations	200	170	370	430	230	450
	learning_rate	0.03	0.03	0.01	0.04	0.02	0.01
	depth	6	9	8	8	6	9
	l2_leaf_reg	2	1	2	9	2	2

References

1. Ho, J.C.; Michalak, A.M.; Pahlevan, N. Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature* **2019**, *574*, 667–670. [CrossRef]
2. Yang, Z.; Gong, C.; Ji, T.; Hu, Y.; Li, L. Water Quality Retrieval from ZY1-02D Hyperspectral Imagery in Urban Water Bodies and Comparison with Sentinel-2. *Remote Sens.* **2022**, *14*, 5029. [CrossRef]
3. Pi, X.; Feng, L.; Li, W.; Zhao, D.; Kuang, X.; Li, J. Water clarity changes in 64 large alpine lakes on the Tibetan Plateau and the potential responses to lake expansion. *ISPRS-J. Photogramm. Remote Sens.* **2020**, *170*, 192–204. [CrossRef]
4. Ma, Y.; Song, K.; Wen, Z.; Liu, G.; Shang, Y.; Lyu, L.; Du, J.; Yang, Q.; Li, S.; Tao, H.; et al. Remote Sensing of Turbidity for Lakes in Northeast China Using Sentinel-2 Images with Machine Learning Algorithms. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2021**, *14*, 9132–9146. [CrossRef]
5. Cao, Z.; Ma, R.; Melack, J.M.; Duan, H.; Liu, M.; Kutser, T.; Xue, K.; Shen, M.; Qi, T.; Yuan, H. Landsat observations of chlorophyll-a variations in Lake Taihu from 1984 to 2019. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *106*, 102642. [CrossRef]
6. Chen, J.; Lyu, Y.; Zhao, Z.; Liu, H.; Zhao, H.; Li, Z. Using the multidimensional synthesis methods with non-parameter test, multiple time scales analysis to assess water quality trend and its characteristics over the past 25 years in the Fuxian Lake, China. *Sci. Total Environ.* **2019**, *655*, 242–254. [CrossRef]
7. Wang, J.; Fu, Z.; Qiao, H.; Liu, F. Assessment of eutrophication and water quality in the estuarine area of Lake Wuli, Lake Taihu, China. *Sci. Total Environ.* **2019**, *650*, 1392–1402. [CrossRef]
8. Wang, Y.; Guo, Y.; Zhao, Y.; Wang, L.; Chen, Y.; Yang, L. Spatiotemporal heterogeneities and driving factors of water quality and trophic state of a typical urban shallow lake (Taihu, China). *Environ. Sci. Pollut. Res.* **2022**, *29*, 53831–53843. [CrossRef]
9. Pahlevan, N.; Smith, B.; Alikas, K.; Anstee, J.; Barbosa, C.; Binding, C.; Bresciani, M.; Cremella, B.; Giardino, C.; Gurlin, D.; et al. Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. *Remote Sens. Environ.* **2022**, *270*, 112860. [CrossRef]
10. Shi, K.; Zhang, Y.; Zhu, G.; Qin, B.; Pan, D. Deteriorating water clarity in shallow waters: Evidence from long-term MODIS and in-situ observations. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *68*, 287–297. [CrossRef]
11. Lee, Z.; Shang, S.; Hu, C.; Du, K.; Weidemann, A.; Hou, W.; Lin, J.; Lin, G. Secchi disk depth: A new theory and mechanistic model for underwater visibility. *Remote Sens. Environ.* **2015**, *169*, 139–149. [CrossRef]
12. Sun, D.; Qiu, Z.; Li, Y.; Shi, K.; Gong, S. Detection of Total Phosphorus Concentrations of Turbid Inland Waters Using a Remote Sensing Method. *Water Air Soil Pollut.* **2014**, *225*, 1953. [CrossRef]
13. Xu, W.; Duan, L.; Wen, X.; Li, H.; Li, D.; Zhang, Y.; Zhang, H. Effects of Seasonal Variation on Water Quality Parameters and Eutrophication in Lake Yangzong. *Water* **2022**, *14*, 2732. [CrossRef]
14. 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]
15. Chen, K.; Duan, L.; Liu, Q.; Zhang, Y.; Zhang, X.; Liu, F.; Zhang, H. Spatiotemporal Changes in Water Quality Parameters and the Eutrophication in Lake Erhai of Southwest China. *Water* **2022**, *14*, 3398. [CrossRef]
16. Tian, S.; Guo, H.; Xu, W.; Zhu, X.; Wang, B.; Zeng, Q.; Mai, Y.; Huang, J.J. Remote sensing retrieval of inland water quality parameters using Sentinel-2 and multiple machine learning algorithms. *Environ. Sci. Pollut. Res.* **2023**, *30*, 18617–18630. [CrossRef] [PubMed]
17. Duan, W.; Takara, K.; He, B.; Luo, P.; Nover, D.; Yamashiki, Y. Spatial, and temporal trends in estimates of nutrient and suspended sediment loads in the Ishikari River, Japan, 1985 to 2010. *Sci. Total Environ.* **2013**, *461–462*, 499–508. [CrossRef]
18. Li, S.; Song, K.; Wang, S.; Liu, G.; Wen, Z.; Shang, Y.; Lyu, L.; Chen, F.; Xu, S.; Tao, H.; et al. Quantification of chlorophyll-a in typical lakes across China using Sentinel-2 MSI imagery with machine learning algorithm. *Sci. Total Environ.* **2021**, *778*, 146271. [CrossRef]
19. Peterson, K.T.; Sagan, V.; Sloan, J.J. Deep learning-based water quality estimation and anomaly detection using Landsat-8/Sentinel-2 virtual constellation and cloud computing. *GISci. Remote Sens.* **2020**, *57*, 510–525. [CrossRef]
20. 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]
21. 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]
22. Palmer, S.C.J.; Kutser, T.; Hunter, P.D. Remote sensing of inland waters: Challenges, progress and future directions. *Remote Sens. Environ.* **2015**, *157*, 1–8. [CrossRef]
23. Topp, S.N.; Pavelsky, T.M.; Jensen, D.; Simard, M.; Ross, M.R.V. Research Trends in the Use of Remote Sensing for Inland Water Quality Science: Moving Towards Multidisciplinary Applications. *Water.* **2020**, *12*, 169. [CrossRef]
24. Cao, Z.; Ma, R.; Duan, H.; Pahlevan, N.; Melack, J.; Shen, M.; Xue, K. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. *Remote Sens. Environ.* **2020**, *248*, 111974. [CrossRef]
25. Swain, R.; Sahoo, B. Improving river water quality monitoring using satellite data products and a genetic algorithm processing approach. *Sustain. Water Qual. Ecol.* **2017**, *9–10*, 88–114. [CrossRef]
26. Xu, S.; Li, S.; Tao, Z.; Song, K.; Wen, Z.; Li, Y.; Chen, F. Remote Sensing of Chlorophyll-a in Xinkai Lake Using Machine Learning and GF-6 WFV Images. *Remote Sens.* **2022**, *14*, 5136. [CrossRef]

27. Bramich, J.; Bolch, C.J.S.; Fischer, A. Improved red-edge chlorophyll-a detection for Sentinel 2. *Ecol. Indic.* **2021**, *120*, 106876. [CrossRef]
28. Shi, X.; Gu, L.; Jiang, T.; Jiang, M.; Butler, J.J.; Xiong, X.J.; Gu, X. Retrieval of chlorophyll-a concentration based on Sentinel-2 images in inland lakes. In Proceedings of the Earth Observing Systems XXVII, San Diego, CA, USA, 23–25 August 2022; Volume 12232.
29. Shi, X.; Gu, L.; Jiang, T.; Zheng, X.; Dong, W.; Tao, Z. Retrieval of Chlorophyll-a Concentrations Using Sentinel-2 MSI Imagery in Lake Chagan Based on Assessments with Machine Learning Models. *Remote Sens.* **2022**, *14*, 4924. [CrossRef]
30. Yang, F.; He, B.; Zhou, Y.; Li, W.; Zhang, X.; Feng, Q. Trophic status observations for Honghu Lake in China from 2000 to 2021 using Landsat Satellites. *Ecol. Indic.* **2023**, *146*, 109898. [CrossRef]
31. Gordon, H.R.; Clark, D.K.; Brown, J.W.; Brown, O.B.; Evans, R.H.; Broenkow, W.W. Phytoplankton pigment concentrations in the Middle Atlantic Bight: Comparison of ship determinations and CZCS estimates. *Appl. Opt.* **1983**, *22*, 20–36. [CrossRef]
32. Schroeder, T.; Schaale, M.; Lovell, J.; Blondeau-Patissier, D. An ensemble neural network atmospheric correction for Sentinel-3 OLCI over coastal waters providing inherent model uncertainty estimation and sensor noise propagation. *Remote Sens. Environ.* **2022**, *270*, 112848. [CrossRef]
33. Mouw, C.B.; Greb, S.; Aurin, D.; DiGiacomo, P.M.; Lee, Z.; Twardowski, M.; Binding, C.; Hu, C.; Ma, R.; Moore, T.; et al. Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. *Remote Sens. Environ.* **2015**, *160*, 15–30. [CrossRef]
34. Mobley, C.; Werdell, J.; Franz, B.A.; Ahmad, Z.; Bailey, S. *Atmospheric Correction for Satellite Ocean Color Radiometry*; NASA: Washington, DC, USA, 2016.
35. Kuhn, C.; de Matos Valerio, A.; Ward, N.; Loken, L.; Sawakuchi, H.O.; Kampel, M.; Richey, J.; Stadler, P.; Crawford, J.; Striegl, R.; et al. Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-a and turbidity. *Remote Sens. Environ.* **2019**, *224*, 104–118. [CrossRef]
36. Niroumand-Jadidi, M.; Bovolo, F.; Bresciani, M.; Gege, P.; Giardino, C. Water Quality Retrieval from Landsat-9 (OLI-2) Imagery and Comparison to Sentinel-2. *Remote Sens.* **2022**, *14*, 4596. [CrossRef]
37. He, Y.; Gong, Z.; Zheng, Y.; Zhang, Y. Inland Reservoir Water Quality Inversion and Eutrophication Evaluation Using BP Neural Network and Remote Sensing Imagery: A Case Study of Dashahe Reservoir. *Water* **2021**, *13*, 2844. [CrossRef]
38. Shen, M.; Luo, J.; Cao, Z.; Xue, K.; Qi, T.; Ma, J.; Liu, D.; Song, K.; Feng, L.; Duan, H. Random forest: An optimal chlorophyll-a algorithm for optically complex inland water suffering atmospheric correction uncertainties. *J. Hydrol.* **2022**, *615*, 128685. [CrossRef]
39. Ioannou, I.; Gilerson, A.; Gross, B.; Moshary, F.; Ahmed, S. Deriving ocean color products using neural networks. *Remote Sens. Environ.* **2013**, *134*, 78–91. [CrossRef]
40. Chang, N.; Xuan, Z.; Yang, Y.J. Exploring spatiotemporal patterns of phosphorus concentrations in a coastal bay with MODIS images and machine learning models. *Remote Sens. Environ.* **2013**, *134*, 100–110. [CrossRef]
41. Chang, N.; Vannah, B.W.; Yang, Y.J.; Elovitz, M. Integrated data fusion and mining techniques for monitoring total organic carbon concentrations in a lake. *Int. J. Remote Sens.* **2014**, *35*, 1064–1093. [CrossRef]
42. Arias-Rodriguez, L.F.; Duan, Z.; Sepúlveda, R.; Martinez-Martinez, S.I.; Disse, M. Monitoring Water Quality of Valle de Bravo Reservoir, Mexico, Using Entire Lifespan of MERIS Data and Machine Learning Approaches. *Remote Sens.* **2020**, *12*, 1586. [CrossRef]
43. Yuan, X.; Wang, S.; Fan, F.; Dong, Y.; Li, Y.; Lin, W.; Zhou, C. Spatiotemporal dynamics and anthropologically dominated drivers of chlorophyll-a, TN and TP concentrations in the Pearl River Estuary based on retrieval algorithm and random forest regression. *Environ. Res.* **2022**, *215*, 114380. [CrossRef] [PubMed]
44. Xiong, G.; Wang, G.; Wang, D.; Yang, W.; Chen, Y.; Chen, Z. Spatio-Temporal Distribution of Total Nitrogen and Phosphorus in Dianshan Lake, China: The External Loading and Self-Purification Capability. *Sustainability* **2017**, *9*, 500. [CrossRef]
45. Feng, L.; Hou, X.; Li, J.; Zheng, Y. Exploring the potential of Rayleigh-corrected reflectance in coastal and inland water applications: A simple aerosol correction method and its merits. *ISPRS-J. Photogramm. Remote Sens.* **2018**, *146*, 52–64. [CrossRef]
46. Olmanson, L.G.; Brezonik, P.L.; Bauer, M.E. Evaluation of medium to low resolution satellite imagery for regional lake water quality assessments. *Water Resour. Res.* **2011**, *47*. [CrossRef]
47. McFEETERS, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [CrossRef]
48. Werther, M.; Odermatt, D.; Simis, S.G.H.; Gurlin, D.; Jorge, D.S.F.; Loisel, H.; Hunter, P.D.; Tyler, A.N.; Spyrakos, E. Characterising retrieval uncertainty of chlorophyll-a algorithms in oligotrophic and mesotrophic lakes and reservoirs. *ISPRS-J. Photogramm. Remote Sens.* **2022**, *190*, 279–300. [CrossRef]
49. Wang, X.; Gong, C.; Ji, T.; Hu, Y.; Li, L. Inland water quality parameters retrieval based on the VIP-SPCA by hyperspectral remote sensing. *J. Appl. Remote Sens.* **2021**, *15*, 42609. [CrossRef]
50. Lo, Y.; Fu, L.; Lu, T.; Huang, H.; Kong, L.; Xu, Y.; Zhang, C. Medium-Sized Lake Water Quality Parameters Retrieval Using Multispectral UAV Image and Machine Learning Algorithms: A Case Study of the Yuandang Lake, China. *Drones* **2023**, *7*, 244. [CrossRef]
51. Guan, Q.; Feng, L.; Hou, X.; Schurgers, G.; Zheng, Y.; Tang, J. Eutrophication changes in fifty large lakes on the Yangtze Plain of China derived from MERIS and OLCI observations. *Remote Sens. Environ.* **2020**, *246*, 111890. [CrossRef]

52. He, J.; Chen, Y.; Wu, J.; Stow, D.A.; Christakos, G. Space-time chlorophyll-a retrieval in optically complex waters that accounts for remote sensing and modeling uncertainties and improves remote estimation accuracy. *Water Res.* **2020**, *171*, 115403. [CrossRef]
53. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS-J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]
54. Haghiabi, A.H.; Nasrolahi, A.H.; Parsaie, A. Water quality prediction using machine learning methods. *Water Qual. Res. J.* **2018**, *53*, 3–13. [CrossRef]
55. Shenglei, W.; Junsheng, L.; Bing, Z.; Qian, S.; Fangfang, Z.; Zhaoyi, L. A simple correction method for the MODIS surface reflectance product over typical inland waters in China. *Int. J. Remote Sens.* **2016**, *37*, 6076–6096. [CrossRef]
56. Zhang, Y.; Shi, K.; Cao, Z.; Lai, L.; Geng, J.; Yu, K.; Zhan, P.; Liu, Z. Effects of satellite temporal resolutions on the remote derivation of trends in phytoplankton blooms in inland waters. *ISPRS-J. Photogramm. Remote Sens.* **2022**, *191*, 188–202. [CrossRef]
57. Li, J.; Gao, M.; Feng, L.; Zhao, H.; Shen, Q.; Zhang, F.; Wang, S.; Zhang, B. Estimation of Chlorophyll-a Concentrations in a Highly Turbid Eutrophic Lake Using a Classification-Based MODIS Land-Band Algorithm. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2019**, *12*, 3769–3783. [CrossRef]
58. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [CrossRef]
59. Chen, Z.; An, C.; Tan, Q.; Tian, X.; Li, G.; Zhou, Y. Spatiotemporal analysis of land use pattern and stream water quality in southern Alberta, Canada. *J. Contam. Hydrol.* **2021**, *242*, 103852. [CrossRef] [PubMed]
60. Huang, J.; Zhang, Y.; Bing, H.; Peng, J.; Dong, F.; Gao, J.; Arhonditsis, G.B. Characterizing the river water quality in China: Recent progress and on-going challenges. *Water Res.* **2021**, *201*, 117309. [CrossRef] [PubMed]
61. Wang, S.; Ma, X.; Fan, Z.; Zhang, W.; Qian, X. Impact of nutrient losses from agricultural lands on nutrient stocks in Dianshan Lake in Shanghai, China. *Water Sci. Eng.* **2014**, *7*, 373–383.