



# Article Estimation of Non-Optically Active Water Quality Parameters in Zhejiang Province Based on Machine Learning

Lingfang Gao<sup>1</sup>, Yulin Shangguan<sup>1</sup>, Zhong Sun<sup>2</sup>, Qiaohui Shen<sup>2</sup> and Zhou Shi<sup>1,3,\*</sup>

- <sup>1</sup> Institute of Agricultural Remote Sensing and Information Technology Application, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China; gaolingfang@zju.edu.cn (L.G.); yulinsg@zju.edu.cn (Y.S.)
- <sup>2</sup> Zhejiang Ecological and Environmental Monitoring Center, Hangzhou 310012, China; sunspan1985@163.com (Z.S.); centralway@163.com (Q.S.)
- <sup>3</sup> Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture, Hangzhou 310058, China
- Correspondence: shizhou@zju.edu.cn

Abstract: Water parameter estimation based on remote sensing is one of the common water quality evaluation methods. However, it is difficult to describe the relationship between the reflectance and the concentration of non-optically active substances due to their weak optical characteristics, and machine learning has become a viable solution for this problem. Therefore, based on machine learning methods, this study estimated four non-optically active water quality parameters including the permanganate index (COD<sub>Mn</sub>), dissolved oxygen (DO), total nitrogen (TN), and total phosphorus (TP). Specifically, four machine learning models including Support Vector Machine Regression (SVR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and K-Nearest Neighbor (KNN) were constructed for each parameter and their performances were assessed. The results showed that the optimal models of COD<sub>Mn</sub>, DO, TN, and TP were RF ( $R^2 = 0.52$ ), SVR ( $R^2 = 0.36$ ), XGBoost ( $R^2 = 0.45$ ), and RF ( $R^2 = 0.39$ ), respectively. The seasonal 10 m water quality over the Zhejiang Province was measured using these optimal models based on Sentinel-2 images, and the spatiotemporal distribution was analyzed. The results indicated that the annual mean values of COD<sub>Mn</sub>, DO, TN, and TP in 2022 were 2.3 mg/L, 6.6 mg/L, 1.85 mg/L, and 0.063 mg/L, respectively, and the water quality in the western Zhejiang region was better than that in the northeastern Zhejiang region. The seasonal variations in water quality and possible causes were further discussed with some regions as examples. It was found that DO would decrease and COD<sub>Mn</sub> would increase in summer due to the higher temperature and other factors. The results of this study helped understand the water quality in Zhejiang Province and can also be applied to the integrated management of the water environment. The models constructed in this study can also provide references for related research.

**Keywords:** machine learning; non-optically active water quality parameters; Sentinel-2; spatiotemporal analysis

# 1. Introduction

As the source of life, water plays an important role in human production and life. However, due to the imbalance between economic development and environmental protection, current society is faced with many serious water problems, such as water pollution, water ecology fragility, and water resource shortage [1,2]. Thus, it is necessary to measure the physical, chemical, and biological properties of water, which can provide a scientific basis for pollution source control and water environment management [3].

Traditional water quality monitoring is usually based on sampling analysis. Water temperature, color, transparency, and other physical properties are mainly measured in situ, and most of the chemical and biological properties such as the concentration of water components need to be analyzed in the laboratory. Such kind of method is inefficient and lacks details in time and space [4]. Although the development of water quality monitoring



Citation: Gao, L.; Shangguan, Y.; Sun, Z.; Shen, Q.; Shi, Z. Estimation of Non-Optically Active Water Quality Parameters in Zhejiang Province Based on Machine Learning. *Remote Sens.* 2024, *16*, 514. https://doi.org/ 10.3390/rs16030514

Academic Editors: Chunqiao Song, Liping Zhu and Xingong Li

Received: 31 October 2023 Revised: 11 January 2024 Accepted: 25 January 2024 Published: 29 January 2024



**Copyright:** © 2024 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) license (https:// creativecommons.org/licenses/by/ 4.0/). stations can provide almost continuous measurement data, the point-based data still fail to fully describe the spatial characteristics [5]. With the characteristics of a large observation range and convenient data acquisition, monitoring based on remote sensing techniques can make up for such limitations.

The remote sensing technique was first utilized in water quality monitoring in the 1970s [6] and has been increasingly applied in water quality estimation with the continuous progress of satellite sensor development and other technologies. For example, both Mohammadpour and Pirasteh [7] and Cao et al. [8] used MODIS (Moderate Resolution Imaging Spectroradiometer) images to analyze the spatial and temporal variability of suspended particulate matter (SPM) in the Persian Gulf and lakes across China, respectively. Li et al. [9] used Landsat 8 data to estimate the colored dissolved organic matter (CDOM) over Saginaw Bay. Based on Sentinel-2 images, Zhang et al. [10] also investigated the water quality change in Chinese rivers. Yang et al. [11] utilized the fusion images generated with Landsat, Sentinel-2, and GaoFen-2 data to estimate chlorophyll a (Chl-a).

However, most previous studies mainly focus on the estimation of optically active components such as the Chl-a, SPM, and CDOM due to their direct strong interaction with electromagnetic radiation [5,12–15]. Since the interaction of non-optically active components such as nitrogen, phosphorus, and oxygen with light is not obvious, and the relationship between the two has not been fully studied, it is difficult to estimate the non-optically active water quality via remote sensing.

Currently, the estimation of non-optically active water quality by remote sensing can be classified into two categories: the empirical method and the machine learning method [16]. The empirical methods usually use standard linear regression or logarithmic or exponential transformations to simulate the relationship between water quality parameters and spectral reflectance or other intermediate variables. Taking the estimation of total phosphorus (TP) as an example, Wu et al. [17] constructed a linear regression equation for natural logarithmic TP concentration, and Gao et al. [18] constructed a combination of linear regression models adapted to the different regions. This kind of method is easy to establish and has achieved high accuracy in several studies [19-24]. It may be because most of these studies focused on small areas and using only a small amount of data for model building, so the water type was relatively homogeneous. However, more studies show low correlations among the non-optically active parameters and the reflectance or other optically active components, which makes empirical methods often perform poorly because of the difficulty in finding the key regressive variables [5]. The machine learning methods are an ideal solution for they could fully capture the complex nonlinear relationship [25] and have been widely applied in water quality estimation [26,27]. For example, a back-propagation (BP) neural network was successfully leveraged to estimate chemical oxygen demand (COD), the permanganate index ( $COD_{Mn}$ ), the total nitrogen (TN), and TP [28–30]. In addition, decision trees, support vector machine regression (SVR), random forest (RF), and other methods are also widely used [31–33]. Some studies compared the performance of empirical methods and machine learning methods on the same data, and the results often showed that machine learning had higher accuracy [5,28]. Although machine learning is less interpretable and not as intuitive and easy to understand as empirical methods, most people working in water environment management are more focused on the results than the principles of the model. Therefore, the use of machine learning has great potential in non-optically active water quality parameter estimation.

However, there are still certain limitations of previous studies that utilized machine learning methods to estimate non-optically active water quality parameters. Firstly, in many previous studies, the training sets were mostly from field sampling by researchers, which made the data often limited to relatively small areas (e.g., a single lake) and small in number. Although a lot of effort had been made and the data numbered in the hundreds, it was still not enough to build a robust machine learning model. Since machine learning models are typically data-driven, the richness of the training dataset would significantly affect the performance of the model. Therefore, the models built on small regions or small

amounts of data usually cannot be accurately extrapolated and show poor performance at larger regional scales, such as the provincial scales. Many water quality monitoring stations are built to realize the long-term and high-frequency monitoring of some important water bodies. The rich field measurements provided by these stations can be used as the dataset for the construction of machine learning models. In addition, previous studies used medium-resolution satellite data (e.g., Landsat series with 30 m resolution) for water quality estimation studies, but such a spatial resolution still failed not meet the needs of fine-grained water quality monitoring, which requires a spatial resolution of 10 m or several meters. Meanwhile, most studies only used one or two machine learning models for estimating water quality, and the performance of different machine learning models in non-optically active water quality parameter estimation remains unclear. Therefore, the establishment and comparison of multiple machine learning models can avoid the problem of poor estimation results due to the inapplicability of one model.

The main objectives of this study were as follows. (1) Based on the measurements from water quality monitoring stations and Sentinel-2 images, machine learning models were constructed to estimate four non-optically active water parameters of  $COD_{Mn}$ , dissolved oxygen (DO), TN, and TP, respectively. (2) The performances of the different models and the optimal models were compared to generate the seasonal 10 m water quality over the Zhejiang Province. (3) The spatial distribution of water quality over Zhejiang Province was analyzed, and the temporal characteristics were further discussed by taking some regions as examples. This study helps us to understand the overall water quality of Zhejiang Province and provides a reference for the integrated management of the regional water environment from the perspective of the whole province.

## 2. Materials

## 2.1. Study Area

Zhejiang Province, located on the southeast coast of China and the southern wing of the Yangtze River Delta, has a land area of 105,500 km<sup>2</sup>, of which the area of rivers and lakes accounts for about 5.05% (Figure 1a). It is a subtropical monsoon climate with four distinct seasons and moderate annual temperatures, with an annual average temperature of 15~18 °C. The lowest and highest temperatures occur in January and July, respectively. In the same period of rain and heat, the average annual rainfall is 980~2000 mm, of which May and June are concentrated rainfall periods [34].

West Lake, located in the north of Zhejiang Province (Figure 1b), is a typical urban landscape lake with a catchment area of 21.22 km<sup>2</sup>, an annual runoff of 14 million m<sup>3</sup>, and a water storage capacity of nearly 14 million m<sup>3</sup>. As a semi-closed still water lake, its water flow is poor, and the sediment brought by the flowing stream is constantly deposited, making the water silting situation constantly aggravated. To improve the water environment of the West Lake, the government carried out a large number of water conservancy projects, the most notable of which was the diversion of Qiantang River water into the West Lake.

Qiantang River is the largest river in Zhejiang Province. Xin'an River is the north source of the Qiantang River, and the Lan River is the south source of the Qiantang River, and the river after the intersection of the two is called the Fuchun River. This confluence (Figure 1c) is an important transportation hub in history and had an important impact on people's production and life. In addition, as the confluence of two sources, knowing the water quality of this area is also of great significance to the water quality monitoring of the Qiantang River basin.

Changtan Reservoir is located in the southeast of Zhejiang Province (Figure 1d), surrounded by green mountains, is about 3 km wide from east to west, about 12 km long from north to south, has a lake area of 36 km<sup>2</sup>, a rainwater collection area of 441.3 km<sup>2</sup>, and a total storage capacity of 732 million m<sup>3</sup>. As one of the most important water sources in Zhejiang Province, it provides safe water for about three million people. In addition to



supplying water to the city, the reservoir also serves a variety of functions such as flood control, irrigation, and power generation.

**Figure 1.** Study area and distribution of water quality monitoring stations. (a) Zhejiang province; (b) West Lake; (c) Confluence of Xin'an River, Fuchun River, and Lan River; (d) Changtan reservoir.

### 2.2. Satellite Data

Except for some large rivers and lakes, inland water bodies are mainly small- and medium-sized rivers, which require satellite images with relatively high spatial resolution for observation. Therefore, the Sentinel-2 series data, which have 13 spectral bands and a spatial resolution of 10 m, were chosen in this study and have been widely used in water environment monitoring [35,36]. The data are collected by two satellites, namely the Sentinel-2A and Sentinel-2B, and the revisit time is 10 days for a single satellite and is 5 days when both satellites are in operation at the same time. In this study, the Sentinel-2 Level-2A images for Zhejiang Province from 1 January 2022 to 15 May 2023 were obtained through the Google Earth Engine (GEE). Cloud filtering was applied and images with cloud coverage greater than 20% were excluded, resulting in a total number of 1523 valid images in 194 days (Figure 2).

#### 2.3. Monitoring Station Data

Instead of the field sampling performed by the researchers, the measurements from 115 water quality monitoring stations (Figure 1a) in Zhejiang Province were utilized. These monitoring stations are mostly located near representative rivers or lakes with important ecological and socioeconomic values, and they integrate a collection of pumps, pips, and many sensors to collect the samples from the target water body and measure them. The water quality parameters are provided every four hours, including the water temperature, electrical conductivity, turbidity, pH value, dissolved oxygen (DO), permanganate index ( $COD_{Mn}$ ), ammonia nitrogen ( $NH_3N$ ), total phosphorus (TP) and total nitrogen (TN). In this study, the  $COD_{Mn}$ , DO, TN, and TP were selected as targets and observation values.

Since the concentration of substances in water rarely changes significantly under natural circumstances, and also to better match the remote sensing images, the daily average value was used in this study, ignoring the variation of water quality in one day. Partial daily average data for several monitoring stations are listed in Table S1.



**Figure 2.** Distribution of remote sensing images with less than 20% cloud cover from 1 January 2022 to 15 May 2023.

The average values of  $COD_{Mn}$ , DO, TN, and TP during January 2022 to May 2023 were 3.0 mg/L (range: 0.2–11.6 mg/L), 8.9 mg/L (range: 1.9–21.0 mg/L), 2.69 mg/L (range: 0.08–10.27 mg/L), and 0.085 mg/L (range: 0.003–0.526 mg/L), respectively (Figure 3), indicating a relativity good water quality for Zhejiang Province. Given the reasonable distribution of the stations, these data can accurately reflect the overall water quality of Zhejiang Province, and thus the constructed models were supposed as being suitable for estimating the water quality components over Zhejiang Province.



**Figure 3.** Density distributions and density curves of the daily average values about four water quality parameters from January 2022 to May 2023. (a) CODMn; (b) DO; (c) TN; (d) TP.

The overall framework of estimating the non-optically active water quality parameters over Zhejiang Province is shown in Figure 4, which consists of the following three steps: (1) data preparation and preprocessing: prepare remote sensing and measuring data and conduct necessary preprocessing; (2) machine learning model construction: identify the optical band combination and construct four machine learning models for each water quality parameter; and (3) water quality estimation and analysis: estimating the water quality parameters using the optimal models and analysis the spatiotemporal dynamics.



Figure 4. Study workflow.

## 3.1. Data Preprocessing

For satellite images, after obtaining the required date, the QA band of Sentinel-2 was used as a mask for cloud masking. The reflectance of the corresponding position of the monitoring station was extracted from the satellite images. Then, the extracted reflectivity data and the water quality data were matched. After removing the invalid data that failed to match or contained zero values, 2013 records of data containing both water quality and reflectance information were retained and used to build the models. In this study, the data were randomly divided into the training set, validating set, and testing set according to a 7:1:2 ratio, in which the training set and validating set were used to build machine learning models and the test set for model accuracy assessment.

## 3.2. Optimal Band Identification

Only the reflectance of band 1 to band 8 data were used due to the strong absorption and relatively low reflectance of the water body in the near-infrared and longer bands. In addition, to effectively enhance the underlying information and improve the performance of the model, the band ratios were also calculated, which have been widely used as the inputs of models in previous research [37–39].

Therefore, a total number of 36 features were obtained and their potentials for estimating each water quality parameter were assessed through Pearson correlation analysis (Equation (1)). Specifically, the Pearson correlation coefficient with each water quality parameter was calculated for each feature and those with significant correlations (p < 0.01) were selected as the input of the model.

$$\mathbf{r} = \frac{\sum_{i=1}^{n} \left( X_i - \overline{X} \right) \left( Y_i - \overline{Y} \right)}{\sqrt{\sum_{i=1}^{n} \left( X_i - \overline{X} \right)^2} \sqrt{\sum_{i=1}^{n} \left( Y_i - \overline{Y} \right)^2}}$$
(1)

where  $X_i$  and  $Y_i$  represent the two input variables and  $\overline{X}$  and  $\overline{Y}$  represent their respective sample means. The range of r is -1 to 1, and the closer the |r| is to 1, the stronger the linear correlation between the two.

#### 3.3. Machine Learning Model

The relationship between non-optically active parameters and reflectance is complicated, and traditional empirical linear/non-linear regression fails to fully capture such relationships. Machine learning methods can represent the complex nonlinear relationship among data and have been widely used in the estimation of non-optically active water quality parameters. Therefore, in this study, four machine learning models that have been previously utilized for water quality estimation including the Support Vector Machine Regression (SVR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and K-Nearest Neighbor (KNN) were leveraged to estimate the non-optically active water quality parameters [11,40–42].

SVR, which maps nonlinear data to high-dimensional space via a kernel function to construct a decision function for linear regression, has unique advantages in solving small-sample, nonlinear, and high-dimensional pattern recognition problems [43], and is often used for small-sample water quality parameter estimation.

RF is an integrated model, which is essentially an improvement of the decision tree model. It is represented by the repetitive random sampling of samples from the original training sample set to generate a new training sample set, and then a classification tree is generated according to the new sample set. Finally, multiple decision trees are combined to form a random forest. Its final result is obtained by combining several weak classifiers in the form of taking the mean [44]. The error of the results depends on the classification ability of each tree and the correlation between them, which makes the results of the overall model have high accuracy and generalization performance [45].

Similar to RF, XGBoost is also an integrated learning method, which is a kind of synthesis method that combines basis function and weight to form a good data-fitting effect. Its essence is a gradient tree-based method that iteratively trains a series of weak learners (usually decision trees), each iteration attempting to correct the error of the previous iteration, and eventually combines these weak learners into a strong learner [46,47]. Unlike traditional gradient boosting decision trees (GBDT), XGBoost adds regularization terms to the loss function and uses second-order Taylor expansion of the loss function as a fitting of the loss function, so XGBoost is more efficient when dealing with large data sets and complex models while preventing overfitting and improving generalization.

KNN is an instance-based and parameter-less learner. Instead of establishing the relationship between variables through the processing or optimization of the data, it calculates the distance between the new sample and the training dataset, then selects a few of the most similar samples and uses their average as the predicted value of the new sample [48]. It can deal well with the situation of strong interdependence among multiple features and complex relationships among features.

The models were built using the scikit-learn package in Python, and the hyperparameters of each model were fine-tuned using Bayesian optimization.

#### 3.4. Accuracy Assessment

In this study, the accuracy of each machine learning model was assessed using the testing set and four evaluation metrics including the determination coefficient ( $R^2$ ), the

Mean Absolute Error (MAE), the Mean Square Error (MSE), and the Root Mean Square Error (RMSE). The formulas of each evaluation of the criterion are expressed as follows:

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (M_{i} - E_{i})^{2}}{\sum_{i=1}^{n} (M_{i} - \overline{M})^{2}}$$
(2)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |M_i - E_i|$$
(3)

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (M_i - E_i)^2$$
(4)

RMSE = 
$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} (M_i - E_i)^2}$$
 (5)

where  $M_i$  refers to the measured values,  $E_i$  refers to the estimated value by models, and n represents the total number of samples.  $R^2$  reflects the degree to which the independent variable explains the change of the dependent variable, and the closer it is to 1, the better the model fits. MSE indicates the deviation between the predicted value and the measured value, and the closer it is to zero, the better the predicted value of the model agrees with the actual measured value.

## 4. Results

#### 4.1. Optimal Band Selection

The results of the correlation analysis (Figure 5) showed that the four water quality parameters had a relatively strong correlation with B1 to B5, br24, br25, br34, and br35. The results of  $COD_{Mn}$ , TN, and TP were quite similar, all having a positive correlation with B1 to B5, and a negative correlation with br24, br25, br34, and br35. Among these three parameters, the correlation of  $COD_{Mn}$  was relatively higher, and the correlation of the TN was slightly weaker. The results of DO exhibited opposite results, with B1 to B5 showing a negative correlation and br24, br25, br34, and br35 showing a positive correlation. This might be associated with the different roles of  $COD_{Mn}$ , TN, TP, and DO components in the water ecosystem. When evaluating water quality, lower concentrations of  $COD_{Mn}$ , TN, and TP and higher the concentrations of DO are positively correlated with the water quality.

Concerning the correlation values, all four parameters presented relatively inferior relationships with reflectance and their ratios with most of the absolute values of correlation coefficients under 0.4. However, these four parameters were significantly related to most bands and band ratios in terms of significance level, suggesting a statistically significant correlation between the water quality parameters and reflectance. In this study, bands and band ratios that were significantly correlated with each water quality parameter (p < 0.01) were selected as input variables of the model. The number of input variables for the COD<sub>Mn</sub>, DO, TN, and TP models was 25, 21, 31, and 26, respectively.

#### 4.2. Evaluation of Machine Learning Models

The evaluation results of four machine learning methods for estimating each water quality parameter are shown in Table 1. Overall, all four machine learning methods presented the best performance for the  $COD_{Mn}$  parameter, followed by the TN parameter, while their performances in estimating the TP and DO were relatively poor. The accuracies of diverse machine learning approaches varied among different water quality parameters. Specifically, for  $COD_{Mn}$ , the RF and XGBoost were the best two models with R<sup>2</sup> values of 0.52 and 0.51, respectively. The SVR and KNN methods exhibited relatively inferior performances (R<sup>2</sup> = 0.45 and 0.46, respectively). Similarly, they also had poor behaviors for the TP parameter, especially for the SVR model with an R<sup>2</sup> value of 0.1. As for the TN parameter, the XGBoost showed the best performance with an R<sup>2</sup> value of 0.45, and the

RF had slightly lower accuracy, but they were both significantly higher than the other two models (e.g., the SVR and KNN). Regarding the DO, however, the SVR outperformed the other three models with  $R^2$  values of 0.36. The RF and XGBoost were both slightly inferior to the KNN.



**Figure 5.** Heat map of Pearson's correlation analysis. Bi represents the reflectance of band i of Sentinel-2 and brij represents the reflectance ratio of band i to band j of Sentinel-2.

Parameter	Model	<b>R</b> <sup>2</sup>	MAE	MSE	RMSE
COD <sub>Mn</sub>	SVR	0.45	0.807	1.18	1.09
	RF	0.52	0.757	1.03	1.02
	XGBoost	0.51	0.765	1.06	1.03
	KNN	0.46	0.796	1.16	1.08
DO	SVR	0.36	1.558	3.96	1.99
	RF	0.34	1.558	4.11	2.03
	XGBoost	0.34	1.566	4.12	2.03
	KNN	0.35	1.547	4.03	2.00
	SVR	0.31	0.890	1.36	1.164
TN	RF	0.42	0.844	1.14	1.068
	XGBoost	0.45	0.816	1.09	1.045
	KNN	0.33	0.900	1.32	1.151
TP	SVR	0.10	0.047	0.0033	0.057
	RF	0.39	0.035	0.0022	0.047
	XGBoost	0.37	0.036	0.0023	0.048
	KNN	0.35	0.036	0.0023	0.048

Table 1. Performance of machine learning models for COD<sub>Mn</sub>, DO, TN, and TP.

Except for the DO, the performances of the KNN and the SVR in other parameters were relatively poor. This might be because the KNN and the SVR were constructed based on the distance between variables [43,48]. Although they have performed well

in some previous studies, their training sets were almost two or three hundred or even dozens [32,38,49]. When the amount of the training set or the number of input variables increases, the complexity of the model significantly increases, and the training efficiency and estimation accuracy decrease. In this study, the  $COD_{Mn}$ , the TN, and the TP had more input variables than the DO, even though the models were built based on the same number of training sets. For these non-optically active water quality parameters with many input variables, the KNN and the SVR had more difficulty achieving accurate estimation.

According to the performance of the models described above, the RF model was chosen for estimating  $COD_{Mn}$  and TP, and the XGBoost and SVR models were selected for estimating TN and DO, respectively.

## 4.3. Annual Mean Water Quality Maps in Zhejiang Province

Utilizing the optimal models and Sentinel-2 images, the  $COD_{Mn}$ , DO, TN, and TP concentrations for each season over the Zhejiang Province in 2022 were derived. From the seasonal results, the annual mean distribution and statistics of water quality were further obtained are and shown in Figure 6 and Table 2. The annual mean values of  $COD_{Mn}$ , DO, TN, and TP over Zhejiang in 2022 were 2.3 mg/L, 6.6 mg/L, 1.85 mg/L, and 0.063 mg/L, respectively. They indicated that the overall water quality of Zhejiang Province in 2022 was relatively good, except for the slightly higher concentration of the TN.



**Figure 6.** The distribution of mean concentrations of  $COD_{Mn}$  (**a**), DO (**b**), TN (**c**), and TP (**d**) in Zhejiang Province in 2022, respectively; (**e1–e4**) are the details of the Thousand-island Lake, and (**f1–f4**) are of the main reach of Qiantang River.

Parameter -	Statistical Characteristics				Standards Limits			
	Minimum	Mean	Maximum		Level 2	Level 3	Level 4	
COD <sub>Mn</sub> (mg/L)	0.2	2.3	6.0	$\leq$	4	6	10	
DO(mg/L)	0.6	6.6	13.7	$\geq$	6	5	3	
TN (mg/L)	0.00	1.85	5.74	$\leq$	0.5	1.0	1.5	
TP (mg/L)	0.001	0.063	0.187	$\leq$	0.1	0.2	0.3	

**Table 2.** Statistical characteristics of average water quality in Zhejiang Province in 2022 and limits of "Surface water Environmental Quality Standards", and the limits of TN are only for lakes and reservoirs.

There were obvious differences in water quality among different regions. Compared with the western region of Zhejiang represented by the Thousand-island Lake (Figure 6(e1–e4)), the concentrations of  $COD_{Mn}$ , TN, and TP in the densely populated northern eastern Zhejiang region were significantly higher, while the concentration of DO was significantly lower. This reflected that the water quality in the western Zhejiang region was better than that in the northeastern Zhejiang region.

The change in water quality was especially obvious in the Qiantang River (Figure 6(f1–f4)). Along the way from the west to the east of the river, the concentrations of  $COD_{Mn}$ , TN, and TP gradually increased, while the concentrations of DO significantly decreased, which means the overall situation of water quality became worse. A similar change was also observed in the Oujiang River region located in the southeastern province, where the water quality gradually deteriorated toward the river (Figure S1). The results of the study of Zhang et al. [10] also showed a similar situation. The reason for this phenomenon may be that pollutants continuously enter the river on both sides of the river with the flow of the river. At the same time, the upstream flow rate is fast, while the downstream flow rate is slow, and thus the migration rate of substances in the water is reduced, resulting in more pollutants suspended or deposited in the river.

However, this trend of water quality change was opposite to the abovementioned situation in the estuary area of the Qiantang River, where the concentrations of  $COD_{Mn}$ , TN, and TP were lower and the concentration of DO was higher than those in the middle and lower reaches of this river (Figure 7b–e). This may be attributed to the fact that the water quality in the estuary of the Qiantang River was greatly affected by the tidal action, and the water quality was improved by the entry of seawater.



**Figure 7.** The estuary of the Qiantang River and the water quality in this area (annual mean in 2022). (a) The remote sensing image; (b) the distributions of  $COD_{Mn}$ ; (c) the distribution of DO; (d) the distribution of TN; and (e) the distribution of TP.

# 5. Discussion

# 5.1. Seasonal Differences of Water Quality

Obvious seasonal fluctuations were observed in the water quality parameters in Zhejiang Province due to the subtropical monsoon climate, and these differences were analyzed taking three typical areas as cases: West Lake, the Changtan Reservoir, and the confluence of Xin'an River, Fuchun River, and LAN River (Figure 1).

In West Lake, the water quality was relatively uniform in winter (January to March), and summer (July to September), while the spatial heterogeneity was quite large in spring (April to June) and autumn (October to December), especially for  $COD_{Mn}$  and TN (Figure 8). Compared with winter, the concentration of  $COD_{Mn}$  apparently increased in summer, while the concentrations of DO and TN decreased. The average concentrations of  $COD_{Mn}$  in winter and summer were 2.3 mg/L and 2.8 mg/L, of DO were 9.3 mg/L and 7.4 mg/L, and of TN were 2.66 mg/L and 2.16 mg/L, respectively. In addition, even though the change in TP concentration was relatively weak (the seasonal average concentrations were 0.069 mg/L, 0.075 mg/L, 0.076 mg/L, and 0.069 mg/L, respectively), a significant reduction in the western part of the lake from spring to autumn can still be observed. Compared with the obvious seasonal characteristics of the water quality in West Lake, the water quality in most areas of Changtan Reservoir changed little with the seasons and was stable at a good level all year round (Figure S2). However, there were some areas at the edge of the reservoir where the concentrations of  $COD_{Mn}$ , TN, and TP were higher in summer and autumn than in winter and spring.



Figure 8. Seasonal variation of COD<sub>Mn</sub> (a1–a4), DO (b1–b4), TN (c1–c4), and TP (d1–d4) concentrations in West Lake during 2022.

Previous studies also showed a consistent tendency that the DO was lower in summer than in winter. For example, a shallow lake in southern America called University Lake was investigated by Xu et al. [50], as well as Xianvu Lake [51], Dianshan Lake [52], and Tianmu Lake [53], which are all located in the southeast of China and have similar climates to the study area. The study of Dianshan Lake also showed that the  $COD_{Mn}$  and TN were higher in summer and autumn, while the TP showed an opposite trend. The research by Qian et al. showed that in the Three Gorges Reservoir Area, the TN was highest in winter and lowest in summer, while the TP increased in spring and summer, decreased first in autumn and winter, and then increased [54]. It indicated that for some lakes, the trend of  $COD_{Mn}$  was consistent; specifically,  $COD_{Mn}$  was usually high in summer and low in winter. However, the trends of the TN and TP were greatly different in different water bodies.

For rivers, it was more fluid than lakes and reservoirs, and therefore, their water quality distributions were more spatiotemporally heterogeneous. Taking the intersection of Xin'an River, Fuchun River, and Lan River as an example (Figure S3), it could be observed that there was an almost triangular area at the intersection of Xin'an River and Lan River, where the concentrations of  $COD_{Mn}$ , TN, and TP were lower than those in the surrounding areas, but the concentration of DO was higher than the adjacent regions, especially in winter. Compared with the Xin'an River section, the water quality became worse after the river entered the Fuchun River section, with the increase of river flow, and the  $COD_{Mn}$  exhibited a significantly increasing trend. However, it can also be found that, in this region, the DO was relatively low and the  $COD_{Mn}$  was high in summer.

#### 5.2. Influencing Factors of Seasonal Variation in Water Quality

Due to the comprehensive effects of the water formation process, surrounding environment, human activities, and other factors, there are great differences in the composition of different water bodies, and the causes of water quality change are also very complex [55]. The concentrations of the DO in many lakes are lower in summer than in winter, possibly because it is dominated by climate. Studies have shown that there is a clear positive correlation between water temperature and air temperature [56]; more specifically, as the water temperature increases, oxygen generally has more difficulty dissolving in the water [57]. Since the concentration of DO in surface water usually tends to saturation under natural conditions [58], the change of DO concentration in water is significantly negatively correlated with temperature. Taking West Lake as an example, its temperature was significantly higher in summer than in winter (Figure 9a); hence, the concentration of DO in summer was lower than in winter, and the decrease of DO further led to the increase of reducing substances in the water, and, subsequently, the concentration of COD<sub>Mn</sub> rose.



**Figure 9.** Seasonal changes of temperature (**a**) and rainfall (**b**) in prefecture-level cities where West Lake is located from 2000 to 2021.

Different from DO, the sources of nitrogen and phosphorus in surface water include exogenous and endogenous sources, and the variation of their concentration is related to the lake itself and the surrounding environment. When there is no external input, there is a dynamic equilibrium of adsorption and release of nitrogen and phosphorus between sediments and water bodies [59]. Studies have shown that temperature changes can affect

the release rate of nitrogen and phosphorus in sediments [60,61]. Within a certain range, the rising temperature would increase the solubility of various insoluble compounds in sediments, as well as the decomposition and mineralization of organic matter in sediments by organisms, thus increasing the concentration of nitrogen and phosphorus in water bodies [62]. However, the release mechanisms of the two elements at the sediment–water interface are not the same [63]. The phosphorus is usually controlled by the redox process of iron [64,65], while the nitrogen depends on the degree of decomposition of nitrogen compounds in the sediment [66]. At the same time, the temperature change will also affect the activity of aquatic organisms, thus affecting the rate of nitrogen and phosphorus consumption by organisms. In addition, rainfall in different periods, such as flood seasons and drought seasons, has different impacts on water quality [67]. Extreme rainfall events usually wash the surface around the lake and bring surface sediments into the lake, which will worsen water quality in a short time [68,69]. However, long-term rainfall increases the water level, which plays a role in diluting the components in the water body [70,71].

As for West Lake, its water quality was also related to human activities [72], such as surrounding tourism and commercial activities which may cause pollution [73,74]. With the diversion of the Qiantang River and the strengthening of water management, the trend of deterioration of West Lake water quality and the occurrence of large-scale algal blooms have been controlled in recent years [75]. Thanks to the strict controls on input pollutants, the changes in the TN and the TP were mainly caused by a combination of endogenous pollutants and climate. However, the seasonal variations of the TN and the TP were not the same, which was most likely due to their varying degrees of response to the same changes. In summer, the external consumption of nitrogen was greater than the release of internal sources, resulting in the reduction of TN (Figure 8(c3)), while the rate of increase and consumption of phosphorus was equal, resulting in a similar concentration of TP in a year.

## 5.3. Strengths and Limitations

Although non-optically active water quality parameters were statistically significantly correlated with reflectance and reflectance ratios, the Pearson correlation coefficient of less than 0.5 indicated that there was no significant linear relationship between them (Figure 5). In this study, for each parameter, some empirical models were selected for estimation, and their parameters were recalibrated to make them more suitable for the current study area. Their performance further indicated that such a relationship was difficult to describe with ordinary empirical formulas (Table 3).

Therefore, the machine learning models were selected to estimate the non-optically active water quality parameters, and those models had similar performances compared to previous studies. For example, in the study of Yang et al. for urban water bodies in Shanghai, the estimation models for  $COD_{Mn}$  performed better than the models for DO and TP. Specifically, the R<sup>2</sup> values of XGBoost for  $COD_{Mn}$ , DO, and TP were 0.58, 0.53, and 0.39, respectively [38]. This may be because  $COD_{Mn}$  reflected the extent of organic and inorganic oxide pollution in the water body, and among these substances may be components that interacted with light more strongly than nitrogen, phosphorus, and oxygen. In the study of Wang et al., the R<sup>2</sup> value and RMSE of XGBoost developed based on Sentinel-3 images for TP of shallow lakes in the Yangtze-Huaihe River region were 0.53 and 0.08, respectively [76]. Because the performance of machine learning was highly dependent on the dataset, it might not be very reasonable to directly compare the performance of models built for different regions and different datasets, but it could still be considered that the machine learning models constructed in this study can be used to estimate non-optically active water quality parameters over Zhejiang Province.

There are still some limitations to machine learning models. As a data-driven model, the most obvious limitation of the machine learning model is that its performance is greatly affected by the quality of the training dataset [77]. If the training dataset is unbalanced, the accuracy of the machine learning model will decrease. In this study, the data of the  $COD_{Mn}$ ,

the TN, and the TP all showed a slight positive skew (Figure 3), and, especially, the samples with high concentrations of  $COD_{Mn}$  ( $COD_{Mn} > 6.0 \text{ mg/L}$ ) and TP (TP > 0.20 mg/L) were few. This might lead to different performances of the model over different concentration ranges, especially with apparent underestimation of the large values. Although this imbalance problem can be improved by some transformations during data preprocessing, the ultimate solution is to expand the number of corresponding samples to a preset minimum threshold of various sample numbers. In addition, there was also a lack of some water types, such as black and odorous water, so the machine learning models established in this study might fail to estimate the water quality for poor-quality water bodies [59], and the estimated results are more likely to be better than the actual situation. This means that these models still need some improvement when applied to related studies such as the identification of black and odorous water bodies.

Table 3. The performance of re-calibrated linear regression models of COD<sub>Mn</sub>, DO, TN, and TP.

Parameter	Equation	Reference	R <sup>2</sup>	MAE	MSE	RMSE
COD <sub>Mn</sub> –	$Y = ax + b; x = \ln(B8) / B2$	[19]	0.121	1.107	1.945	1.395
	$Y = ax^2 + bx + c$ $x = (B3 - B5)/(B3 + B5)$	[20]	0.218	1.022	1.732	1.316
DO _	$Y = ax_1 + bx_2 + cx_3 + d$ $x_1 = B4 \times B5; \ x_2 = B4/B5; \ x_3 = B4$	[21]	0.115	1.811	5.338	2.310
	$Y = ax + b; \ x = B2$	[22]	0.100	1.843	5.429	2.330
TN _	$\ln(Y) = ax_1 + bx_2 + cx_3 + d$ x <sub>1</sub> = B4/B2; x <sub>2</sub> = B2/B4; x <sub>3</sub> = B2	[23]	0.047	1.138	2.100	1.449
	$Y = ax^b; x = B6/B8$	[24]	0.000	1.226	2.203	1.484
TP	$\ln(Y) = ax_1 + bx_2 + cx_3 + d$ x <sub>1</sub> = B4/B2; x <sub>2</sub> = B2/B4; x <sub>3</sub> = B2	[23]	0.077	0.0404	0.0033	0.0571
	$Y = ax^b; x = (B4 + B5)/B3$	[24]	0.093	0.0434	0.0032	0.0566
	Y = ax + b $x = (B3 - B5)/(B3 + B5)$	[20]	0.083	0.0437	0.0032	0.0569

Another obvious shortcoming of the machine learning methods is that they do not explicitly explain the error propagation mechanism between input and output data [77], so it is difficult to explain how each part and process affects the final output results. When there is a high concentration of Chl-a or SPM in the water body, the spectral reflectance of the water body is quite different from that of the general water body [42], and the accuracy of the results is difficult to guarantee.

# 6. Conclusions

In this study, based on Sentinel-2 images, the non-optically active water quality parameter retrieval over Zhejiang Province was estimated via four machine learning methods and automatic measurements. The optimal bands for each parameter were identified and the performances of different machine learning models were inter-compared and fully assessed. The 10 m seasonal water quality parameters were then obtained using the optimal models and their spatiotemporal distributions were analyzed. The main conclusions were as follows:

(1) The performance of the four machine learning methods was inconsistent in the estimation of the different parameters, and the optimal models of  $COD_{Mn}$ , DO, TN, and TP were RF (R<sup>2</sup> = 0.52), SVR (R<sup>2</sup> = 0.36), XGBoost (R<sup>2</sup> = 0.45) and RF (R<sup>2</sup> = 0.39), respectively;

- (2) The average annual water quality in Zhejiang Province was good, and the annual mean values of COD<sub>Mn</sub>, DO, TN, and TP in 2022 over Zhejiang Province were 2.3 mg/L, 6.6 mg/L, 1.85 mg/L, and 0.063 mg/L, respectively;
- (3) The water quality in the western Zhejiang region was better than that in the northeastern region. As for rivers, the water quality in the upper reaches was better than that in the lower reaches of rivers;
- (4) Compared with spring and autumn, the water quality in winter and summer was more uniform. Though the seasonal variations of water quality in different areas were not the same, DO and TN generally decreased in summer, while COD<sub>Mn</sub> and TP increased, and the temperature and rainfall might be the important influencing factors.

The results of this study are helpful in obtaining the basic water quality information of Zhejiang Province and provide a reference for water management and the introduction of more comprehensive policies from the perspective of the whole province. For example, the results showed that the water quality in the northeast of Zhejiang Province was relatively poor, so the government should pay more attention to these areas. Considering that the water quality deteriorates with the inflow of rivers, the management of the input of pollutants along rivers should be strengthened. The estimated results of the TN and the TP can also be used to evaluate the risk of the algal blooms.

In future studies, black odorous, eutrophication, or high-turbidity water bodies can be added to enrich the types of water bodies in the training dataset, and the estimation model of these special water bodies can be established to achieve more comprehensive water quality monitoring and evaluation. The water quality should also be studied for a longer period to explore the annual trend of water quality change and further analyze the impact of climate change and other factors on water quality, which is conducive to maintaining the sustainable development of human society.

**Supplementary Materials:** The following supporting information can be downloaded at: https: //www.mdpi.com/article/10.3390/rs16030514/s1, Figure S1: Oujiang River and the water quality in this river (annual mean in 2022). (a) the remote sensing image; (b) the distributions of CODMn; (c) the distribution of DO; (d) the distribution of TN and (e) the distribution of TP; Figure S2: Seasonal variation of CODMn (a), DO (b), TN (c), and TP (d) concentrations in the Changtan Reservoir during 2022; Figure S3: Seasonal variation of CODMn (a), DO (b), TN (c), and TP (d) concentrations at river confluence during 2022; Table S1: The daily averages of water quality parameters from 50 water quality monitoring stations (part of all stations) on 3 January 2022. T is the water temperature. DO is the dissolved oxygen. CODMn is the permanganate index. NH3N is the ammonia nitrogen. TP is the total phosphorus. TN is the total nitrogen.

Author Contributions: Conceptualization, L.G. and Y.S.; data curation, Q.S.; formal analysis, L.G.; funding acquisition, Z.S. (Zhou Shi); investigation, Z.S. (Zhong Sun); methodology, L.G.; project administration, Z.S. (Zhou Shi); resources, Z.S. (Zhong Sun) and Q.S.; software, L.G.; supervision, Z.S. (Zhou Shi); validation, L.G. and Y.S.; visualization, L.G.; writing—original draft, L.G.; writing—review and editing, L.G., Y.S. and Z.S. (Zhou Shi). All authors have read and agreed to the published version of the manuscript.

Funding: This study has been supported by the Key R&D Program of Zhejiang (2022C03078).

Data Availability Statement: Data is contained within the article or Supplementary Materials.

**Acknowledgments:** The authors are very grateful to all the anonymous reviewers for their constructive comments.

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

## References

- 1. Tang, W.; Pei, Y.; Zheng, H.; Zhao, Y.; Shu, L.; Zhang, H. Twenty years of China's water pollution control: Experiences and challenges. *Chemosphere* **2022**, 295, 133875. [CrossRef]
- 2. Xue, J.; Wang, Q.; Zhang, M. A review of non-point source water pollution modeling for the urban–rural transitional areas of China: Research status and prospect. *Sci. Total Environ.* **2022**, *826*, 154146. [CrossRef]
- 3. Chawla, I.; Karthikeyan, L.; Mishra, A.K. A review of remote sensing applications for water security: Quantity, quality, and extremes. *J. Hydrol.* **2020**, *585*, 124826. [CrossRef]
- 4. 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]
- 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]
- 6. Wrigley, R.C.; Horne, A.J. Remote sensing and lake eutrophication. *Nature* 1974, 250, 213–214. [CrossRef]
- Mohammadpour, G.; Pirasteh, S. Interference of CDOM in remote sensing of suspended particulate matter (SPM) based on MODIS in the Persian Gulf and Oman Sea. *Mar. Pollut. Bull.* 2021, 173, 113104. [CrossRef] [PubMed]
- 8. Cao, Z.; Hu, C.; Ma, R.; Duan, H.; Liu, M.; Loiselle, S.; Song, K.; Shen, M.; Liu, D.; Xue, K. MODIS observations reveal decrease in lake suspended particulate matter across China over the past two decades. *Remote Sens. Environ.* **2023**, 295, 113724. [CrossRef]
- 9. Li, J.; Yu, Q.; Tian, Y.Q.; Becker, B.L.; Siqueira, P.; Torbick, N. Spatio-temporal variations of CDOM in shallow inland waters from a semi-analytical inversion of Landsat-8. *Remote Sens. Environ.* **2018**, *218*, 189–200. [CrossRef]
- 10. Zhang, Y.; He, X.; Lian, G.; Bai, Y.; Yang, Y.; Gong, F.; Wang, D.; Zhang, Z.; Li, T.; Jin, X. Monitoring and spatial traceability of river water quality using Sentinel-2 satellite images. *Sci. Total Environ.* **2023**, *894*, 164862. [CrossRef]
- 11. Yang, H.; Du, Y.; Zhao, H.; Chen, F. Water quality Chl-a inversion based on spatio-temporal fusion and convolutional neural network. *Remote Sens.* **2022**, *14*, 1267. [CrossRef]
- IOCCG. Earth Observations in Support of Global Water Quality Monitoring; Greb, S., Dekker, A., Binding, C., Eds.; IOCCG Report Series, No. 17; International Ocean Colour Coordinating Group: Dartmouth, Canada, 2018. Available online: https://ioccg.org/ what-we-do/ioccg-publications/ioccg-reports/ (accessed on 1 June 2023).
- 13. Lee, Z.; Carder, K.L.; Arnone, R.A. Deriving inherent optical properties from water color: A multiband quasi-analytical algorithm for optically deep waters. *Appl. Opt.* **2002**, *41*, 5755–5772. [CrossRef] [PubMed]
- 14. Matthews, M.W. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. *Int. J. Remote Sens.* **2011**, *32*, 6855–6899. [CrossRef]
- 15. Liu, G.; Li, L.; Song, K.; Li, Y.; Lyu, H.; Wen, Z.; Cao, Z.; Shang, Y.; Yu, G.; Zheng, Z.; et al. An OLCI-based algorithm for semi-empirically partitioning absorption coefficient and estimating chlorophyll a concentration in various turbid case-2 waters. *Remote Sens. Environ.* **2020**, *239*, 111648. [CrossRef]
- 16. Topp, S.N.; Pavelsky, T.M.; Jensen, D.; Simard, M.; Ross, M.R. Research trends in the use of remote sensing for inland water quality science: Moving towards multidisciplinary applications. *Water* **2020**, *12*, 169. [CrossRef]
- 17. Wu, C.; Wu, J.; Qi, J.; Zhang, L.; Huang, H.; Lou, L.; Chen, Y. Empirical estimation of total phosphorus concentration in the mainstream of the Qiantang River in China using Landsat TM data. *Int. J Remote Sens.* **2010**, *31*, 2309–2324. [CrossRef]
- Gao, Y.; Gao, J.; Yin, H.; Liu, C.; Xia, T.; Wang, J.; Huang, Q. Remote sensing estimation of the total phosphorus concentration in a large lake using band combinations and regional multivariate statistical modeling techniques. *J. Environ. Manag.* 2015, 151, 33–43. [CrossRef] [PubMed]
- 19. Zhu, X.; Wen, Y.; Li, X.; Yan, F.; Zhao, S. Remote Sensing Inversion of Typical Water Quality Parameters of a Complex River Network: A Case Study of Qidong's Rivers. *Sustainability* **2023**, *15*, 6948. [CrossRef]
- Xiao, Y.; Yin, G.; Zhang, X.; Shi, Y.; Hao, F.; Fu, Y. UAV multispectral image-based urban river water quality monitoring using stacked ensemble machine learning algorithms—A case study of the Zhanghe river, China. *Remote Sens.* 2022, 14, 3272. [CrossRef]
- Padilla-Mendoza, C.; Torres-Bejarano, F.; Campo-Daza, G.; González-Márquez, L.C. Potential of Sentinel Images to Evaluate Physicochemical Parameters Concentrations in Water Bodies—Application in a Wetlands System in Northern Colombia. *Water* 2023, 15, 789. [CrossRef]
- 22. Rahul, T.S.; Brema, J.; Wessley, G.J.J. Evaluation of surface water quality of Ukkadam lake in Coimbatore using UAV and Sentinel-2 multispectral data. *Int. J. Environ. Sci. Technol.* 2023, 20, 3205–3220. [CrossRef]
- 23. Muhoyi, H.; Gumindoga, W.; Mhizha, A.; Misi, S.N.; Nondo, N. Water quality monitoring using remote sensing, Lower Manyame Sub-catchment, Zimbabwe. *Water Pract. Technol.* **2022**, *17*, 1347–1357. [CrossRef]
- 24. Wang, S.; Shen, M.; Liu, W.; Ma, Y.; Shi, H.; Zhang, J.; Liu, D. Developing remote sensing methods for monitoring water quality of alpine rivers on the Tibetan Plateau. *GIScience Remote Sens.* **2022**, *59*, 1384–1405. [CrossRef]
- Peterson, K.T.; Sagan, V.; Sidike, P.; Hasenmueller, E.A.; Sloan, J.J.; Knouft, J.H. Machine learning based ensemble prediction of water quality variables using featurelevel 1 and decision-level fusion with proximal remote sensing. *Photogramm. Eng. Remote Sens.* 2019, 85, 269–280. [CrossRef]

- Peterson, K.T.; Sagan, V.; Sidike, P.; Cox, A.L.; Martinez, M. Suspended sediment concentration estimation from landsat imagery along the lower missouri and middle Mississippi Rivers using an extreme learning machine. *Remote Sens.* 2018, 10, 1503. [CrossRef]
- Zhu, M.; Wang, J.; Yang, X.; Zhang, Y.; Zhang, L.; Ren, H.; Wu, B.; Ye, L. A review of the application of machine learning in water quality evaluation. *Eco-Environ. Health* 2022, 1, 107–116. [CrossRef] [PubMed]
- 28. Deng, C.; Zhang, L.; Cen, Y. Retrieval of chemical oxygen demand through modified capsule network based on hyperspectral data. *Appl. Sci.* **2019**, *9*, 4620. [CrossRef]
- 29. 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]
- Sun, X.; Zhang, Y.; Shi, K.; Zhang, Y.; Li, N.; Wang, W.; Huang, X.; Qin, B. Monitoring water quality using proximal remote sensing technology. *Sci. Total Environ.* 2022, *803*, 149805. [CrossRef] [PubMed]
- 31. Chatziantoniou, A.; Spondylidis, S.C.; Stavrakidis-Zachou, O.; Papandroulakis, N.; Topouzelis, K. Dissolved oxygen estimation in aquaculture sites using remote sensing and machine learning. *Remote Sens. Appl. Soc. Environ.* **2022**, *28*, 100865. [CrossRef]
- 32. Ding, L.; Qi, C.; Li, G.; Zhang, W. TP Concentration Inversion and Pollution Sources in Nanyi Lake Based on Landsat 8 Data and InVEST Model. *Sustainability* **2023**, *15*, 9678. [CrossRef]
- Tan, Z.; Ren, J.; Li, S.; Li, W.; Zhang, R.; Sun, T. Inversion of Nutrient Concentrations Using Machine Learning and Influencing Factors in Minjiang River. Water 2023, 15, 1398. [CrossRef]
- Mao, F.; Du, H.; Zhou, G.; Zheng, J.; Li, X.; Xu, Y.; Huang, Z.; Yin, S. Simulated net ecosystem productivity of subtropical forests and its response to climate change in Zhejiang Province, China. *Sci. Total Environ.* 2022, 838, 155993. [CrossRef] [PubMed]
- 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 Obs. Remote Sens.* 2021, 14, 9132–9146. [CrossRef]
- Guo, H.; Huang, J.J.; Chen, B.; Guo, X.; Singh, V.P. A machine learning-based strategy for estimating non-optically active water quality parameters using Sentinel-2 imagery. *Int. J. Remote Sens.* 2021, 42, 1841–1866. [CrossRef]
- EI-Rawy, M.; Fathi, H.; Abdalla, F. Integration of remote sensing data and in situ measurements to monitor the water quality of the Ismailia Canal, Nile Delta, Egypt. *Environ. Geochem. Health* 2020, 42, 2101–2120. [CrossRef] [PubMed]
- 38. 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]
- 39. Zhang, J.; Fu, P.; Meng, F.; Yang, X.; Xu, J.; Cui, Y. Estimation algorithm for chlorophyll-a concentrations in water from hyperspectral images based on feature derivation and ensemble learning. *Ecol. Inform.* **2022**, *71*, 101783. [CrossRef]
- 40. Du, C.; Wang, Q.; Li, Y.; Lyu, H.; Zhu, L.; Zheng, Z.; Wen, S.; Liu, G.; Guo, Y. Estimation of total phosphorus concentration using a water classification method in inland water. *Int. J. Appl. Earth Obs.* **2018**, *71*, 29–42. [CrossRef]
- Hafeez, S.; Wong, M.S.; Ho, H.C.; Nazeer, M.; Nichol, J.; Abbas, S.; Tang, D.; Lee, K.H.; Pun, L. Comparison of machine learning algorithms for retrieval of water quality indicators in case-II waters: A case study of Hong Kong. *Remote Sens.* 2019, 11, 617. [CrossRef]
- 42. Valera, M.; Walter, R.K.; Bailey, B.A.; Castillo, J.E. Machine learning based predictions of dissolved oxygen in a small coastal embayment. *J. Marine Sci. Eng.* 2020, *8*, 1007. [CrossRef]
- Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogram.* 2011, 66, 247–259. [CrossRef]
- 44. Mutanga, O.; Adam, E.; Cho, M.A. High density biomass estimation for wetland vegetation using world view-2 imagery and random forest regression algorithm. *Int. J. Appl. Earth Obs.* **2012**, *18*, 399–406. [CrossRef]
- 45. Béjaoui, B.; Ottaviani, E.; Barelli, E.; Ziadi, B.; Dhib, A.; Lavoie, M.; Gianluca, C.; Turki, S.; Solidoro, C.; Aleya, L. Machine learning predictions of trophic status indicators and plankton dynamic in coastal lagoons. *Ecol. Indic.* **2018**, *95*, 765–774. [CrossRef]
- 46. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
- 47. Ghatkar, J.G.; Singh, R.K.; Shanmugam, P. Classification of algal bloom species from remote sensing data using an extreme gradient boosted decision tree model. *Int. J. Remote Sens.* **2019**, *40*, 9412–9438. [CrossRef]
- McRoberts, R.E. Estimating forest attribute parameters for small areas using nearest neighbors techniques. *For. Ecol. Manag.* 2012, 272, 3–12. [CrossRef]
- 49. Xu, Y.; Ma, C.; Liu, Q.; Xi, B.; Qian, G.; Zhang, D. Method to predict key factors affecting lake eutrophication—A new approach based on support vector regression model. *Int. Biodeterior. Biodegrad.* **2015**, *102*, 3008–3315. [CrossRef]
- Xu, Z.; Xu, Y. Determination of trophic state changes with Diel dissolved oxygen: A case study in a shallow lake. *Water Environ. Res.* 2015, *87*, 1970–1979. [CrossRef]
- 51. Xia, W.; Zhang, M.; Zhou, M.; Wu, J.; Yao, N.; Feng, B.; Ouyang, T.; Liu, Z.; Zhang, Q. Spatio-temporal dynamics of dissolved oxygen and its influencing factors in Lake Xianny Jiangxi, China. J. Lake Sci. 2023, 35, 874–885. [CrossRef]
- 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. [CrossRef]

- 53. Zeng, C.; Haung, W.; Wang, W.; Zhu, G. Distribution and its influence factors of dissolved oxygen in Tianmuhu Lake. *Resour. Environ. Yangtze Basin* **2010**, *19*, 445–451.
- 54. Qian, T.; Huang, Q.; He, B.; Li, T.; Liu, S.; Fu, S.; Zeng, R.; Xiang, K. Seasonal variations in nitrogen and phosphorus concentration and stoichiometry of Hanfeng Lake in the Three Gorges Reservoir Area. *Environ. Sci.* **2020**, *41*, 5381–5388. [CrossRef]
- 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]
- 56. Girgibo, N.; Lü, X.; Hiltunen, E.; Peura, P.; Dai, Z. The air temperature change effect on water quality in the Kvarken Archipelago area. *Sci. Total Environ.* 2023, 874, 162599. [CrossRef]
- 57. Carstens, D.; Amer, R. Spatio-temporal analysis of urban changes and surface water quality. J. Hydrol. 2019, 569, 720–734. [CrossRef]
- 58. Hamid, A.; Bhat, S.U.; Jehangir, A. Local determinants influencing stream water quality. Appl. Water Sci. 2020, 10, 24. [CrossRef]
- 59. Li, Q.; Tian, Y.; Liu, L.; Zhang, G.; Wang, H. Research progress on release mechanisms of nitrogen and phosphorus of sediments in water bodies and their influencing factors. *Wetland Sci.* **2022**, *20*, 94–103. [CrossRef]
- 60. Jensen, H.; Andersen, F. Importance of temperature, nitrate, and pH for phosphate release from aerobic sediments of four shallow, eutrophic lakes. *Limnol. Oceanogr.* **1992**, *37*, 577–589. [CrossRef]
- Wu, Q.; Zhang, R.; Huang, S.; Zhang, H. Effects of bacteria on nitrogen and phosphorus release from river sediment. *J. Environ. Sci.* 2008, 20, 404–412. [CrossRef]
- 62. Wu, Y.; Wen, Y.; Zhou, J.; Wu, Y. Phosphorus release from lake sediments: Effects of pH, temperature and dissolved oxygen. *KSCE J. Civ. Eng.* **2014**, *18*, 323–329. [CrossRef]
- 63. Fan, C. Advances and prospect in sediment-water interface of lakes: A review. J. Lake Sci. 2019, 31, 1191–1218. [CrossRef]
- 64. Zhu, H.; Wang, D.; Cheng, P.; Fan, J.; Zhong, B. Effects of sediment physical properties on the phosphorus release in aquatic environment. *Sci. China Phys. Mech. Astron.* 2015, *58*, 1–8. [CrossRef]
- 65. Gong, M.; Jin, Z.; Wang, Y.; Lin, J.; Ding, S. Coupling between iron and phosphorus in sediments of shallow lakes in the middle and lower reaches of Yangtze River using diffusive gradients in thin films (DGT). *J. Lake Sci.* 2017, 29, 1103–1111. [CrossRef]
- 66. Valdemarsen, T.; Quintana, C.O.; Flindt, M.R.; Kristensen, E. Organic N and P in eutrophic fjord sediments–rates of mineralization and consequences for internal nutrient loading. *Biogeosciences* **2015**, *12*, 1765–1779. [CrossRef]
- 67. Fukushima, T.; Kitamura, T.; Matsushita, B. Lake water quality observed after extreme rainfall events: Implications for water quality affected by stormy runoff. *SN Appl. Sci.* **2021**, *3*, 841. [CrossRef]
- 68. Li, X.; Huang, T.; Ma, W.; Sun, X.; Zhang, H. Effects of rainfall patterns on water quality in a stratified reservoir subject to eutrophication: Implications for management. *Sci. Total Environ.* **2015**, *521*, 27–36. [CrossRef]
- 69. Jia, Z.; Chang, X.; Duan, T.; Wang, X.; Wei, T.; Li, Y. Water quality responses to rainfall and surrounding land uses in urban lakes. *J. Environ. Manag.* **2021**, *298*, 113514. [CrossRef]
- Ma, L.; Qi, X.; Zhou, S.; Niu, H.; Zhang, T. Spatiotemporal distribution of phosphorus fractions and the potential release risks in sediments in a Yangtze River connected lake: New insights into the influence of water-level fluctuation. *J. Soils Sediments* 2023, 23, 496–511. [CrossRef]
- Pang, X.; Gao, Y.; Guan, M. Linking downstream river water quality to urbanization signatures in subtropical climate. *Sci. Total Environ.* 2023, 870, 161902. [CrossRef] [PubMed]
- Ni, Z.; Wang, S.; Wu, Y.; Pu, J. Response of phosphorus fractionation in lake sediments to anthropogenic activities in China. *Sci. Total Environ.* 2020, 699, 134242. [CrossRef] [PubMed]
- 73. Zhu, W.; Huang, L.; Sun, N.; Chen, J.; Pang, S. Landsat 8-observed water quality and its coupled environmental factors for urban scenery lakes: A case study of West Lake. *Water Environ. Res.* 2020, *92*, 255–265. [CrossRef] [PubMed]
- 74. Sun, Q.; Liu, Z. Impact of tourism activities on water pollution in the West Lake Basin (Hangzhou, China). *Open Geosci.* **2020**, *12*, 1302–1308. [CrossRef]
- 75. You, A.J.; Hua, L. Optimization and Effect of Inner Water Diversion and Distribution in the West Lake of Hangzhou. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2019; Volume 264, p. 012018. [CrossRef]
- 76. Wang, X.; Jiang, Y.; Jiang, M.; Cao, Z.; Li, X.; Ma, R.; Xu, L.; Xiong, J. Estimation of total phosphorus concentration in lakes in the Yangtze-Huaihe region based on Sentinel-3/OLCI images. *Remote Sens.* **2023**, *15*, 4487. [CrossRef]
- 77. 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]

**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.