

Article

A Machine Learning-Based Assessment of Proxies and Drivers of Harmful Algal Blooms in the Western Lake Erie Basin Using Satellite Remote Sensing

Neha Joshi ¹, Armeen Ghorkhanian ^{1,2}, Jongmin Park ^{1,3,4}, Kaiguang Zhao ³ and Sami Khanal ^{1,*} 

¹ Department of Food, Agricultural and Biological Engineering, The Ohio State University, Columbus, OH 43210, USA; joshi.304@osu.edu (N.J.); ghorkhanian.5@osu.edu (A.G.)

² Department of Computer Science and Engineering, The Ohio State University, Columbus, OH 43210, USA

³ School of Environment and Natural Resources, The Ohio State University, Columbus, OH 43210, USA; zhao.1423@osu.edu

⁴ Department of Environmental Engineering, Korea National University of Transportation, Chungju 27469, Republic of Korea; jmpark1@ut.ac.kr

* Correspondence: khanal.3@osu.edu

Abstract

The western region of Lake Erie has been experiencing severe water-quality issues, mainly through the infestation of algal blooms, highlighting the urgent need for action. Understanding the drivers and the intricacies associated with algal bloom phenomena is important to develop effective water-quality remediation strategies. In this study, the influences of multiple bloom drivers were explored, together with Harmonized Landsat Sentinel-2 (HLS) images, using the datasets collected in Western Lake Erie from 2013 to 2022. Bloom drivers included a group of physicochemical and meteorological variables, and Chlorophyll-a (Chl-a) served as a proxy for algal blooms. Various combinations of these datasets were used as predictor variables for three machine learning models, including Support Vector Regression (SVR), Extreme Gradient Boosting (XGB), and Random Forest (RF). Each model is complemented with the SHapley Additive exPlanations (SHAP) model to understand the role of predictor variables in Chl-a estimation. A combination of physicochemical variables and optical spectral bands yielded the highest model performance (R^2 up to 0.76, RMSE as low as 8.04 $\mu\text{g/L}$). The models using only meteorological data and spectral bands performed poorly ($R^2 < 0.40$), indicating the limited standalone predictive power of meteorological variables. While satellite-only models achieved moderate performance (R^2 up to 0.48), they could still be useful for preliminary monitoring where field data are unavailable. Furthermore, all 20 variables did not substantially improve model performance over models with only spectral and physicochemical inputs. While SVR achieved the highest R^2 in individual runs, XGB provided the most stable and consistently strong performance across input configurations, which could be an important consideration for operational use. These findings are highly relevant for harmful algal bloom (HAB) monitoring, where Chl-a serves as a critical proxy. By clarifying the contribution of diverse variables to Chl-a prediction and identifying robust modeling approaches, this study provides actionable insights to support data-driven management decisions aimed at mitigating HAB impacts in freshwater systems.

Keywords: harmful algal blooms; Harmonized Landsat Sentinel-2; water quality; machine learning



Academic Editor: Xiaodong Li

Received: 22 March 2025

Revised: 14 June 2025

Accepted: 18 June 2025

Published: 24 June 2025

Citation: Joshi, N.; Ghorkhanian, A.; Park, J.; Zhao, K.; Khanal, S.

A Machine Learning-Based Assessment of Proxies and Drivers of Harmful Algal Blooms in the Western Lake Erie Basin Using Satellite Remote Sensing. *Remote Sens.* **2025**, *17*, 2164. <https://doi.org/10.3390/rs17132164>

Copyright: © 2025 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/>).

1. Introduction

Harmful algal blooms (HABs) have been a serious threat to water quality for multiple decades now, with significant implications on the aquatic ecosystem [1,2], public health [3,4], and the economy [5,6] worldwide. To develop effective mitigation measures to resolve HAB-driven problems, it is important to understand the factors to which the occurrence and spread of these blooms are particularly sensitive. Either alone or in combination, physicochemical factors [7], such as Total Phosphorus (TP), Total Nitrogen (TN), turbidity, and Microcystin, among many others, and climatic factors such as wind speed, air pressure, humidity, sunshine, and air temperature [8] have been found to modulate the bloom phenomenon.

The surge in algal growth is a complex challenge, particularly in areas like the Lake Erie basin, where various stressors from anthropogenic activities such as agricultural intensification and industrialization [9] add nutrients into the water, resulting in an elevated level of eutrophication. With over 72% of the Maumee River Basin being used for agriculture, the widespread use of chemical fertilizers significantly contributes to non-point-source nutrient pollution in the Western Lake Erie Basin, fueling most blooms [7,10]. Additionally, meteorological conditions such as precipitation and wind can influence the transport of nutrients from agricultural fields and industries to water resources [8], resulting in the fluctuation of bloom magnitude.

Identifying bloom-inducing drivers by collecting water-quality samples and various meteorological and biophysical parameters is a costly endeavor [11,12]. To address this, space-borne remote sensing has been introduced as a complementary approach, which leverages the sensitivity of bloom proxies, such as Chlorophyll-a (Chl-a), to certain spectral characteristics captured by satellite sensors. Chl-a has been found to be optically sensitive to the blue (450–510 nm), green (530–590 nm), red (640–670 nm), and near-infrared (850–880 nm) regions of the spectrum [13,14]. Satellite images offer low-cost, continuous global spatiotemporal coverage, and real-time monitoring capabilities, surpassing traditional data collection methods. With recent free access to historical satellite datasets, it has become easier to conduct time series analyses of any events of interest, including algal blooms [15–17].

To date, various data analytical approaches have been put in practice to understand bloom occurrences. The most common ones include (i) traditional statistical models such as linear and multiple regression [18,19], (ii) bio-optical models that utilize the optical properties of water and bloom constituents [14], and (iii) process-based ecological models that consider the interactions of physical, chemical, and environmental variables that simulate bloom phenomena [20,21]. While each of these approaches has its strengths and limitations, machine learning (ML) algorithms stand out as a flexible approach in that they are capable of capturing the complex interaction among a wide range of variables involved in the process, and their predictive capabilities provide them an edge over their counterpart approaches [22,23].

For HAB detection, monitoring, and forecasting, several studies have been conducted based on ML models [24–26]. Huang et al. [27] used the RF model to analyze the relationship between environmental factors (nutrients, latitude and longitude, precipitation, altitude, and temperature) and Chl-a concentration in different lakes in China for three years (2000, 2005, and 2010), while Chegoonian et al. [28] utilized Support Vector Regression (SVR) to detect Chl-a from Sentinel-2 images in the shallow and eutrophic inland water body of Buffalo Pound Lake, Saskatchewan, Canada. Some other studies have used deep learning methods, mainly long short-term memory (LSTM), for assessing the temporal dependencies in bloom dynamics [29–31]. Despite these advances, the application of ML for HAB assessment remains relatively new, and its full potential has yet to be fully realized [25].

One major challenge in developing reliable ML models lies in preventing overfitting, especially when working with limited or imbalanced data. This can be addressed through careful model tuning and rigorous validation. Another common criticism of ML models is their lack of interpretability compared to more transparent, mechanistic approaches. Recently, explainable machine learning (eXML) methods have gained traction as a way to bridge this gap, providing clearer insights into how models function and why specific predictions are made.

Despite growing interest, few studies have harnessed eXML techniques to unravel the complex interplay between physicochemical water-quality parameters, meteorological factors, and spectral bands and indices influencing Chl-a prediction. This study addresses this gap by applying three ML algorithms, including SVR, RF, and XGB, to evaluate the potential of various predictors in assessing Chl-a. Using harmonized Landsat 30 m (L30) and Sentinel-2 30 m (L30) images, which have a spatial resolution of 30 m and frequent temporal coverage (every 1–3 days), this study captures nuanced spatial and temporal dynamics. Additionally, we apply explainability methods that reveal both overall model behavior and the reasoning behind individual predictions. Using SHapley Additive exPlanations (SHAP), we can rank how different spectral bands, meteorological factors, and physicochemical variables contribute to Chl-a estimates, showing their relative importance across the models, as well as within specific predictions.

2. Materials and Methods

2.1. Study Area

This study is focused on the Western Lake Erie Basin (Figure 1), which is vulnerable to nutrient and sediment loading from both agricultural and urban runoff [32]. Lake Erie, the fourth-largest North American Great Lake, borders Canada to the north and the U.S. states of New York, Ohio, Pennsylvania, and Michigan along its eastern, southern, and western shores. The lake is naturally divided into eastern, central, and western basins based on differences in bathymetry, geology, thermal structure, and trophic characteristics [33]. Compared to the other two, the western basin experiences the most frequent and intense algal blooms, largely driven by elevated nutrient and sediment inputs from urbanization, industrialization, and agricultural runoff [34].

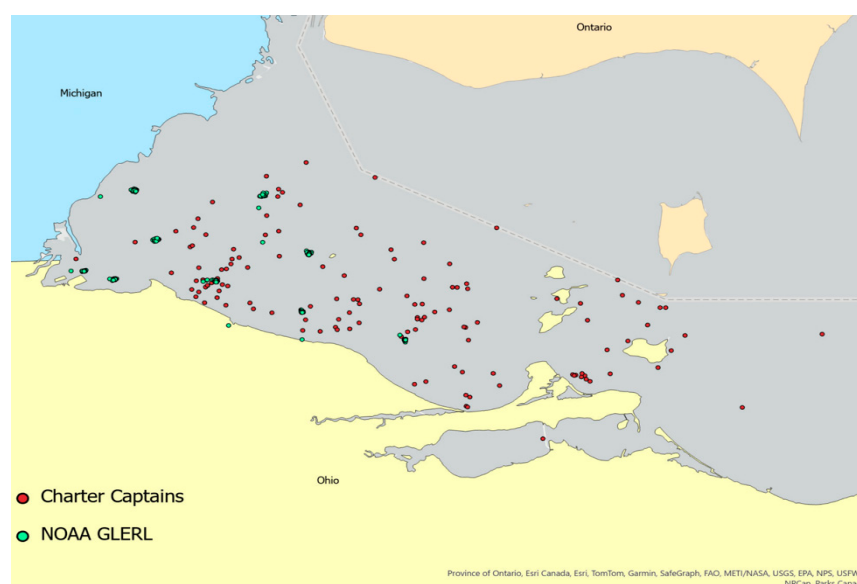


Figure 1. Study region and the distribution of ground-truth water-quality sampling locations within the Western Lake Erie Basin.

2.2. Datasets

2.2.1. Physicochemical Water-Quality Dataset

This study used eight physicochemical variables collected from either ground-based stations or via manual sampling between 2013 and 2022: Chl-a ($\mu\text{g/L}$), Total Phosphorus (TP; $\mu\text{g P/L}$), Total Nitrogen (TN; $\mu\text{g N/L}$), Secchi depth (m), Microcystin (MC; $\mu\text{g/L}$), Colored Dissolved Organic Matter (CDOM; $\mu\text{g/L}$), water temperature ($^{\circ}\text{C}$), and Total Suspended Solids (TSSs; g/L). There were a total of 2138 ground-based water-quality samples, mainly Chl-a, that were made available by the Stone Lab Algal and Water Quality Laboratory [35]. These data were collected at a frequency ranging from every other week during bloom events to once a month. Chl-a, being a dominant pigment in harmful algal blooms, serves as a robust indicator of HAB occurrence and, hence, was used as a predictor variable, while other variables were used as response variables. The temporal patterns of Chl-a are discussed in the later part of the Materials and Methods section.

2.2.2. Meteorological Dataset

Six meteorological variables used in this study include solar radiation ($\text{kWh/m}^2/\text{day}$), wind speed (m/s), wind direction (degrees), air temperature ($^{\circ}\text{C}$), humidity (g/kg), and precipitation (mm). All these variables, except precipitation, were accessed through the NASA POWER web interface. The datasets are available at daily, monthly, and annual temporal scales, with solar radiation from 1984 and the rest from 1981. The spatial resolution of the solar data is a $1^{\circ} \times 1^{\circ}$ latitude/longitude grid, while that of the meteorological data is a $0.5^{\circ} \times 0.625^{\circ}$ latitude/longitude grid [36].

2.2.3. Harmonized Landsat Sentinel (HLS) Dataset

This study used HLS images, which include reflectance data from the Operational Land Imager (OLI) sensors on the Landsat-8 satellites and the Multi-Spectral Instrument (MSI) sensors on the Copernicus Sentinel-2A and Sentinel-2B satellites. The harmonized products, L30 and S30, are radiometrically harmonized, with a common pixel resolution of 30 m and a common projection system based on the Sentinel-2 Military Grid Reference System (MGRS) [37]. These images were collected from 2013 to 2022, with L30 datasets available every 16 days from 2013 onwards, while S30 images were available every 5 days from 2015, corresponding with the launch of Sentinel-2A in 2015. The source images were downloaded using a bash script released by NASA EarthData. Among the available datasets, the images were selected from three tiles, including 17TKG, 17TLF, and 17TLG, that covered the entire study area.

Due to differences in the number of spectral bands available via two OLI and MSI sensors (11 versus 13) (NASA USGS), the spectral bands common to these sensors, which include blue, green, red, near-infrared (NIR), short-wave infrared 1 (SWIR 1), and short-wave infrared 2 (SWIR 2), were selected. [34]

2.2.4. Spectral Band Indices

Eight commonly used spectral band indices were derived to detect the Chl-a concentration in the lake (Table 1).

Table 1. Remotely sensed indices used in assessing Chl-a in water.

Indices	Formulae	Reference
Normalized Difference Vegetation Index (NDVI)	$\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$	[38]
Green Normalized Difference Vegetation Index (GNDVI)	$\frac{\text{NIR} - \text{Green}}{\text{NIR} + \text{Green}}$	[39]
Blue Normalized Difference Vegetation Index (BNDVI)	$\frac{\text{NIR} - \text{Blue}}{\text{NIR} + \text{Blue}}$	[40]
Normalized Difference Turbidity Index (NDTI)	$\frac{\text{Red} - \text{Green}}{\text{Red} + \text{Green}}$	[41]
Green Chlorophyll Index (GCI)	$\left(\frac{\text{NIR}}{\text{Green}} \right) - 1$	[42]
Surface Algal Bloom Index (SABI)	$\frac{(\text{NIR} - \text{Red})}{(\text{Blue} + \text{Green})}$	[43]
Adjusted Floating Algal Index (AFAI)	$\text{NIR} - \text{Red} + (\text{SWIR} - \text{Red}) \times 0.5$	[44]
Enhanced Vegetation Index (EVI)	$2.5 \times ((\text{NIR} - \text{Red}) / (\text{NIR} + 6 \times \text{Red} - 7.5 \times \text{Blue} + 1))$	[45]

Spectral band indices simplify the model and enhance the informational content of the satellite imagery [46]. By incorporating different spectral bands, these indices capture the unique characteristics of algal blooms with varying effectiveness [47]. For instance, NDVI is the most common vegetation index [48,49]. NDVI values range from -1 to 1 , with positive values indicating the presence of vegetation, which, in this study, is an indication of the presence of Chl-a, while negative values indicate their absence or lower presence. Similarly, GNDVI is used to monitor moisture and nitrogen concentration on the surface [50]. BNDVI works better if cyanobacteria dominate, as cyanobacterial blooms have a high absorption peak in the blue spectrum [40]. NDTI can detect turbidity in water [51]; clean water exhibits relatively high reflectance in the green region and low reflectance in the red region [52]. As turbidity increases with an increase in the concentration of suspended particles, the reflectance in the red region increases [53]. EVI was developed to resist atmospheric aerosols, enhancing the accuracy and precision of the spectral reflectance. It is sensitive in high-biomass regions, unlike NDVI, which is saturated [54].

2.3. Data Quality Control

Several missing values in both ground- and satellite-based variables presented data gap challenges for our analysis. To address this, several data quality control measures were used. For instance, for satellite imagery, noise-affected pixels, including cloud, cloud shadow, and snow, were removed using the HLS Fmask band [37]. For details, please refer to the Supplementary Materials (Table S4). This quality control step eliminated 62 ground-truth samples. To fill the resulting data gaps, the remaining ground samples were matched with the closest-in-time satellite pixels acquired within a ± 3 days of the field sampling, following the approach used in Maeda et al. [55]. This resulted in a total of 535 in situ satellite sample pairs for further analysis. The temporal distribution of these matched dates can be inferred from Figure S1.

Similarly, the relationship (Equation (1)) between water temperature and air temperature [56] was used to fill gaps in water temperature (T_{water}) data, using daily MERRA-2 Air Temperature at 2 m height (T_{air}) from the NASA Power Project:

$$T_{\text{water}} = 5.0 + 0.75 * T_{\text{air}} \quad (1)$$

where temperature was measured in $^{\circ}\text{C}$.

TSS data were available for only 155 observations, which is about 29% of the total dataset with Chl-a observations. Over 70% of the dataset would be lost if TSS is included. Imputing values via regression-based, multiple imputation, or Bayesian methods to fill in gaps under such extreme sparsity would introduce high uncertainty and unstable inferences.

To evaluate whether excluding TSS would affect model performance, we trained two RF models on this 155-sample subset: one including TSS and one without. Although the small sample size constrained the overall R^2 , both models produced virtually identical chlorophyll-a predictions (with TSS: $R^2 = 0.4795$, RMSE = 8.783; without TSS: $R^2 = 0.4785$, RMSE = 8.791). These results demonstrate that excluding TSS does not meaningfully degrade prediction accuracy, justifying its removal from subsequent analyses.

2.4. Exploratory Data Analysis

2.4.1. Annual and Seasonal Patterns in Chl-a Concentration

Chlorophyll-a concentrations in the Western Lake Erie Basin varied markedly over the studied period of 2013–2022 (Figure 2). In 2013, the median Chl-a was around 22 $\mu\text{g/L}$, with a few values exceeding 30 $\mu\text{g/L}$. The most intense bloom years were 2014 and 2015, where medians rose to about 45 $\mu\text{g/L}$ and 35–40 $\mu\text{g/L}$, respectively, with peak observations surpassing 100 $\mu\text{g/L}$ (2015 reaching ~115 $\mu\text{g/L}$). From 2016 to 2018, medians dropped to roughly 8–12 $\mu\text{g/L}$, and maxima rarely exceeded 30 $\mu\text{g/L}$, indicating declines in bloom severity. Since 2019, the medians have stabilized around 10–15 $\mu\text{g/L}$, while upper values typically range from 20 to 40 $\mu\text{g/L}$, showing that annual HAB events remain present but are less extreme than mid-decade peaks.

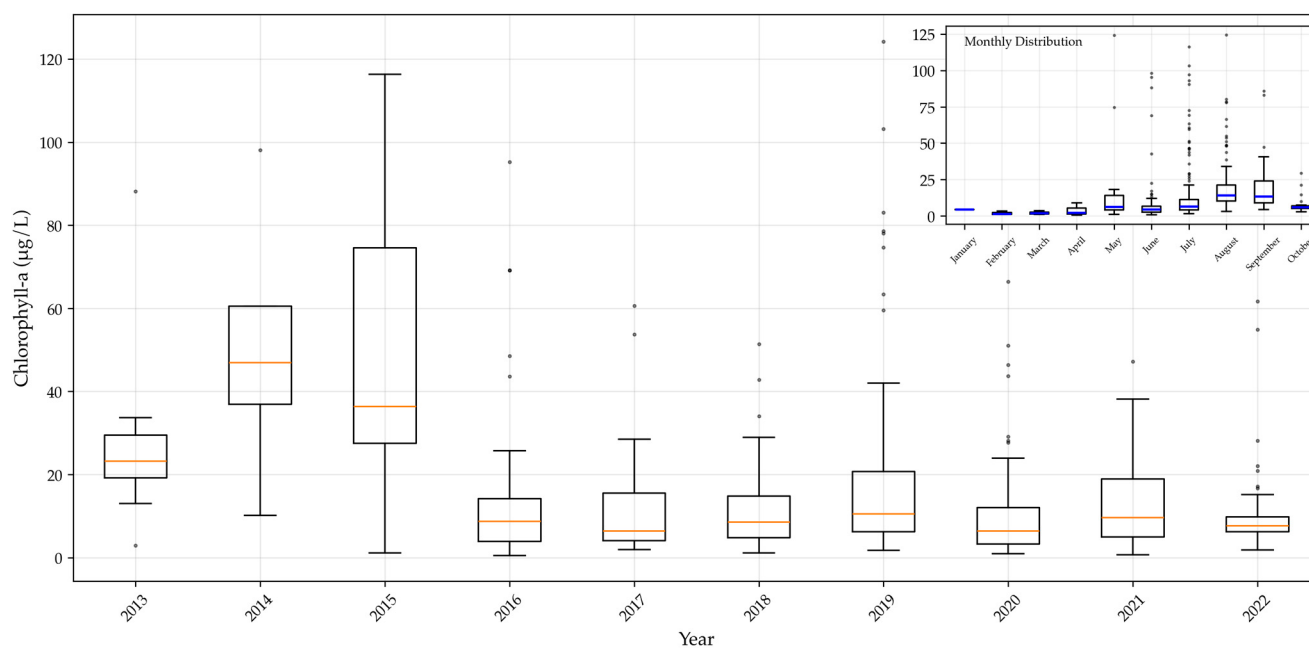


Figure 2. Boxplots of yearly chlorophyll-a concentrations from January to October, showing the median, interquartile range (25th–75th percentiles), whiskers, and outliers. The monthly distribution for all years is provided in the top right.

As far as the monthly pattern for when data from all years were combined, Chl-a stayed very low in January–March, rising modestly in April. Bloom onset begins in May, intensifies in June, and peaks in July–August. By September, median levels fall with occasional high outliers, with a steady decline starting in October.

Other key variables, such as TP, ranged from 1.16 $\mu\text{g P/L}$ to 177.91 $\mu\text{g P/L}$, with a mean of 31.58 $\mu\text{g P/L}$, and MC ranged from 0 $\mu\text{g/L}$ to 21.53 $\mu\text{g/L}$, with a mean of 0.98 $\mu\text{g/L}$. Basic statistics reporting the minimum, maximum, mean, median, and standard deviation of other variables considered in this study are reported in the Supplementary Materials (Table S3).

2.4.2. Relationship of Chl-a with Physicochemical and Meteorological Variables

Correlation matrices suggested that TP, MC, and TN had the highest correlation with Chl-a (Figure 3). Meteorological parameters showed low correlations with Chl-a and among themselves. Several bands (Bs) and band indices (BIs) exhibited strong associations. To avoid potential multicollinearity, only selected variables were included based on their significance in water-quality monitoring, inferred from the prior literature, and the correlation coefficients (r) between variables. For instance, SWIR1 and SWIR2 had $r = 0.97$, so only SWIR2 was considered. Similarly, NDVI was chosen over BNDVI and GNDVI ($r \geq 0.90$). Likewise, SABI was selected over GCI, and EVI was selected over AFAI.

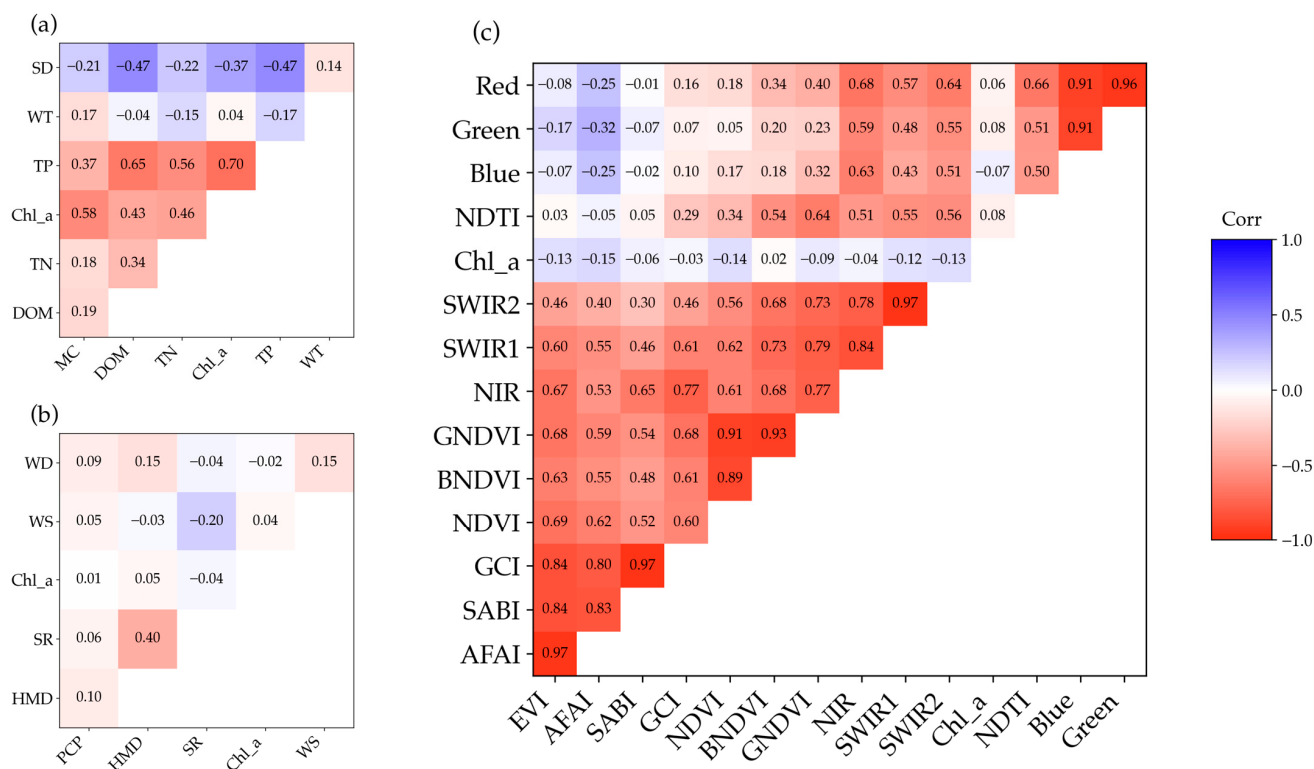
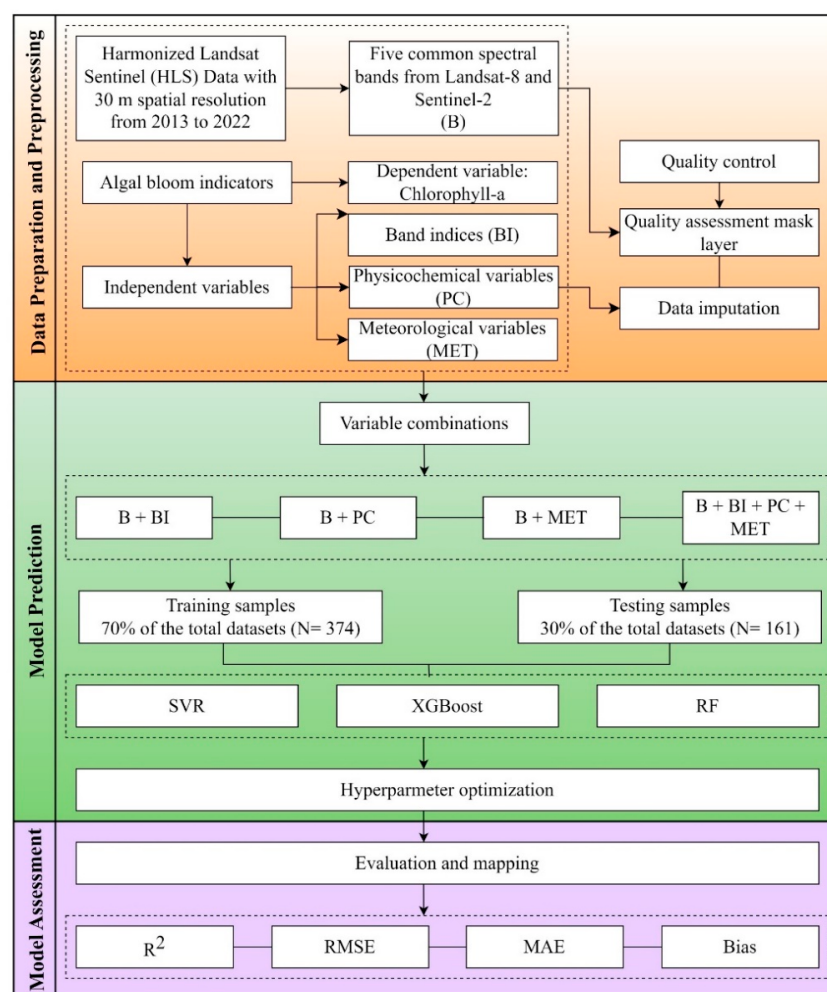


Figure 3. Correlation matrices illustrating the relationship between Chlorophyll-a and (a) physicochemical parameters, (b) spectral bands (B) and vegetation indices (also referred to as BI), and (c) meteorological variables.

A total of twenty predictor variables, including five raw band reflectance and four band indices derived from HLS images, six physicochemical variables, and five meteorological variables, were considered for developing ML models (Table 2). Specifically, the ML models were grouped into four categories based on various combinations of input variables—(i) band indices (B + BI), (ii) physicochemical water-quality variables (B + PC), (iii) meteorological variables (B + MET), and (iv) a combination of all categories (B + BI + PC + MET). All models had raw band reflectance variables in common. In addition to the environmental drivers affecting Chl-a, the decision to use a combination of raw bands and band indices was motivated by their ease of access and availability [57]. The working methodology is shown in Figure 4.

Table 2. List of input variables for the machine learning models.

Category	Parameter	Abbr.	Unit	Data Source
Physicochemical water-quality factors	Total Phosphorus	TP	$\mu\text{g P/L}$	(Ohio Sea Grant, 2022)
	Total Nitrogen	TN	$\mu\text{g NP/L}$	
	Secchi Depth	SD	M	
	Microcystin	MC	$\mu\text{g/L}$	
	Dissolved Organic Matter	DOM	$\mu\text{g/L}$	
	Water Temperature	WT	$^{\circ}\text{C}$	
Meteorological factors	Solar Radiation or Sunlight	SR	$\text{kWh/m}^2/\text{day}$	(NASA POWER)
	Wind Speed	WS	m/s	
	Wind Direction	WD	Degrees	
	Humidity	HMD	g/kg	
	Precipitation	PCP	Mm	
Spectral bands	Red	Red	Nm	(NASA EarthData)
	Green	Green	Nm	
	Blue	Blue	Nm	
	NIR	NIR	Nm	
	SWIR2	SWIR2	Nm	
Band indices	NDVI	NDVI	-	(NASA EarthData)
	NDTI	NDTI	-	
	SABI	SABI	-	
	EVI	EVI	-	

**Figure 4.** Working methodology illustrating different stages of this study, involving data preparation and pre-processing, model prediction, and model assessment.

2.5. Machine Learning Algorithms

Three widely used ML algorithms—SVR, RF, and XGB—were used for the analysis. SVR maps input data into a higher-dimensional feature space using kernels such as linear, polynomial, and radial kernels [58]. The radial kernel was chosen in this study to capture non-linear relationships between the variables [15,59]. RF is an ensemble learning method that aggregates results from multiple decision trees [60], reducing overfitting by averaging their outputs. XGB is a decision tree-based boosting algorithm that combines weak prediction models iteratively, focusing on models with higher residuals to improve accuracy [61]. It uses L1 (Lasso) to eliminate less important features and L2 (Ridge) to reduce the impact of individual features, making the model robust to noise [62].

These models were selected for their unique approaches to learning patterns in the dataset. The comparison of tree-based (XGB and RF) and non-tree-based (SVR) methods provides insight into their different approaches to estimation. Tree-based models divide the feature space into smaller non-overlapping regions defined by specific rules, while non-tree-based models find optimal hyperplanes that fit the regression problem [24]. This comparison assesses trade-offs like interpretability and data handling.

2.6. Model Optimization and Evaluation

Prior to model fitting, all predictor variables were normalized using min–max scaling. The datasets were then randomly split into 70% training and 30% testing sets, with the 30% independent holdout set reflecting a real-world scenario. To optimize the model performance, the hyperparameters of each of the models were tuned by testing various ranges of values in the hyperparameter space (Table 3). Hyperparameter optimization was implemented using the hyperopt 0.2.7 library in Python 3.13.3.

Table 3. List of hyperparameters used for the models.

SVR	XGB	RF
Cost, Gamma, Epsilon, Degree, Kernel	n_estimators, max_depth, learning_rate, gamma, reg_alpha (L1), reg_lambda (L2), min_child_weight, subsample, colsample_bytree	n_estimators, max_depth, min_samples_leaf, min_samples_split

Model performance was evaluated using test datasets, comparing models with statistical metrics, including R^2 , RMSE, bias, and MAE. R^2 indicates the proportion of variance explained by the model, while RMSE and MAE measure error magnitudes and their direction. RMSE is sensitive to outliers due to squaring errors, whereas MAE is more resistant by taking absolute values. Bias indicates the mean deviation of predicted values from true values, with consistent negative or positive bias suggesting the overfitting or underfitting of a model. Overall, combining these metrics provides a comprehensive understanding of model accuracy and errors.

To support the robustness of our observations and reduce the risk of reporting results biased by a single train-test split, we repeated the train–test split ten times with different random seeds, ran each of the models ten times, and reported the mean performance metrics.

2.7. Model Interpretation

The impact of the input variables on the model’s performance was assessed using SHAP (SHapley Additive exPlanations), which is based on game theory [63]. Traditionally, the importance of variables in the model was interpreted using the Feature Importance

approach [7,22], which is limited to tree-based methods like RF and XGB and only provides a ranking of variables without indicating the direction or local magnitude of their impact. SHAP (SHapley Additive exPlanations) overcomes these limitations by calculating the relative importance of the variables and portraying their relationship with the predicted target variable, offering both global and local interpretations of the model. For the SVR model, we used SHAP's KernelExplainer rather than the TreeExplainer, which is used for tree-based models, such as RF and XGB. Regardless, KernelExplainer provides the same local and global attributes consistent with our tree-based models.

The Shapely value of a feature is calculated as the average contribution of that feature across all possible combinations. The SHAP plots showcase a cluster of dots representing the magnitude and direction of variables' global contribution to the prediction outcomes. Each dot represents a single sample within the model, with colors indicating low (blue) to high (red) values. The absolute mean SHAP value, displayed next to the variable's name on plots, represents the global importance of each variable. Based on these values, the input variables are ranked in order of importance. The interpretability of SHAP makes it versatile across various ML algorithms [64–66].

For model interpretation using SHAP, we selected the most stable and best-performing model, i.e., XGB, which was evaluated with two predictor sets (B + PC and B + BI + PC + Met). After each model run (e.g., 10 model runs using XGB for the B + PC predictor set), we calculated the SHAP values to quantify every variable's contribution to the model's output. By averaging these values across all runs, we identified the variables that consistently drove performance, revealing both their global influence on predictions and their local impact for individual samples.

3. Results

3.1. Variability in Model Performance Across Input Groups

The distribution of model performance metrics (R^2 , RMSE, MAE, and bias) across ten random seeds for the SVR, XGB, and RF models, using different input variable groups (B + BI, B + BC, B + MET, and All) for Chl-a predictions with test datasets, is summarized in Figure 5. Tables S6 and S7 in the Supplementary Materials report model performance metrics across random seeds by model and variable combinations.

Across all metrics, models incorporating physicochemical variables (B+ PC and All) outperformed those relying solely on spectral bands (B + BI) or meteorological variables (B + MET). These models achieved notably higher R^2 values, with median values consistently exceeding 0.6. In contrast, models based on B + BI or B + MET inputs yielded a significantly lower R^2 , underscoring the critical role of physicochemical variables in improving predictive accuracy and indicating the limited predictive utility of meteorological variables alone.

Among the models, XGB and RF demonstrated more stable performance across random seeds, as reflected by narrower interquartile ranges, while SVR showed greater variability, particularly in bias and R^2 , for the B + PC and ALL groups, indicating its higher sensitivity to data splits. Bias distributions further revealed that SVR tended to systematically underpredict when using B + BI and B + MET inputs, while XGB and RF maintained more balanced predictions across variable groups.

Given the observations of model performance across seeds, we selected seed 300 for a detailed comparison and visualization of observed vs. predicted Chl-a (discussed in the next section), as it provided balanced and representative performance across all model and input groups.

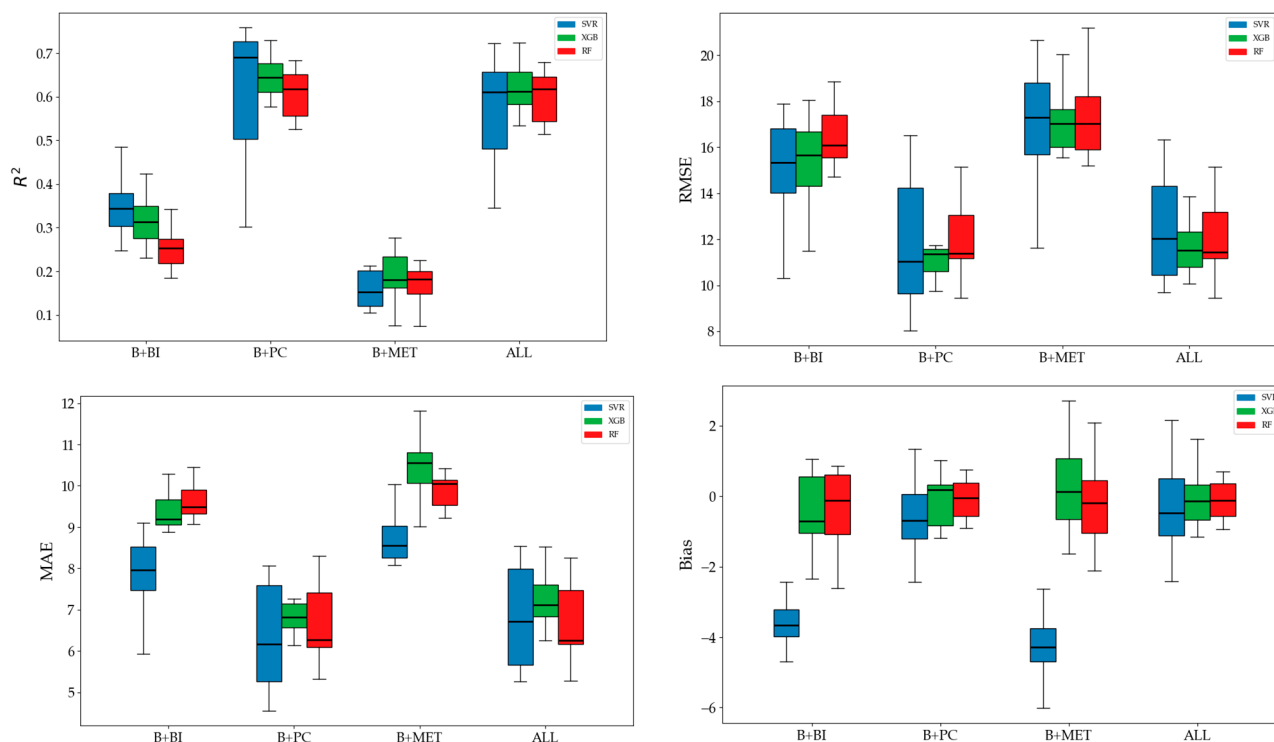


Figure 5. Distribution of R^2 , RMSE, MAE, and bias values across 10 random seeds for SVR, XGB, and RF using four variable groups. The black line error bars represent the maximum and minimum performance values for each of the three models, and the line in the middle of each boxplot shows the median values.

3.1.1. Model Performance Using Spectral Bands and Band-Derived Indices

Figure 6 shows the scatter plots of observed versus predicted Chl-a for SVR, XGB, and RF models when trained on spectral bands and band index (B + BI) variables in the test dataset selected using seed 300. Consistent with the trends observed across multiple random seeds (Tables S6 and S7), XGB achieved the highest R^2 (0.42) and lowest RMSE ($13.88 \mu\text{g L}^{-1}$) and MAE ($8.88 \mu\text{g L}^{-1}$), indicating better overall predictive performance compared to SVR and RF. SVR had a slightly lower R^2 (0.38) and higher error metrics, and it exhibited a notable bias ($-3.66 \mu\text{g L}^{-1}$), indicating systematic overprediction. RF yielded the lowest R^2 (0.27), with the largest RMSE and MAE among the three models. Predictions showed greater scatter around the 1:1 line, especially at higher Chl-a, indicating more frequent overestimation and underestimation compared to SVR and XGB. In contrast, predictions from XGB are tightly grouped around the 1:1 line at low to moderate concentrations, with occasionally underpredicted peak values. Overall, while XGB remained the top-performing model for this input group, the limited performance across all models using B + BI alone also reinforces the value of incorporating additional predictors, such as physicochemical variables, to improve Chl-prediction accuracy.

3.1.2. Model Performance Using Bands and Physicochemical Parameters

Figure 7 shows the observed vs. predicted Chl-a comparison for the same three models, but with both spectral bands and physiochemical parameters (B + PC) as predictors. Incorporating physicochemical parameters markedly improved all models, as reflected by the higher R^2 and lower RMSE and MAE. SVR achieved the highest R^2 and the lowest MAE, followed by RF and XGB with comparable performances ($R^2 \sim 0.65\text{--}0.66$). The bias is minimal across all models, indicating balanced predictions and far less dispersion from the 1:1 line compared to Figure 6. The consistent improvement across all models highlights

that spectral data alone is insufficient for optimal Chl-a prediction and that integrating key physicochemical drivers substantially improves model performance and generalizability.

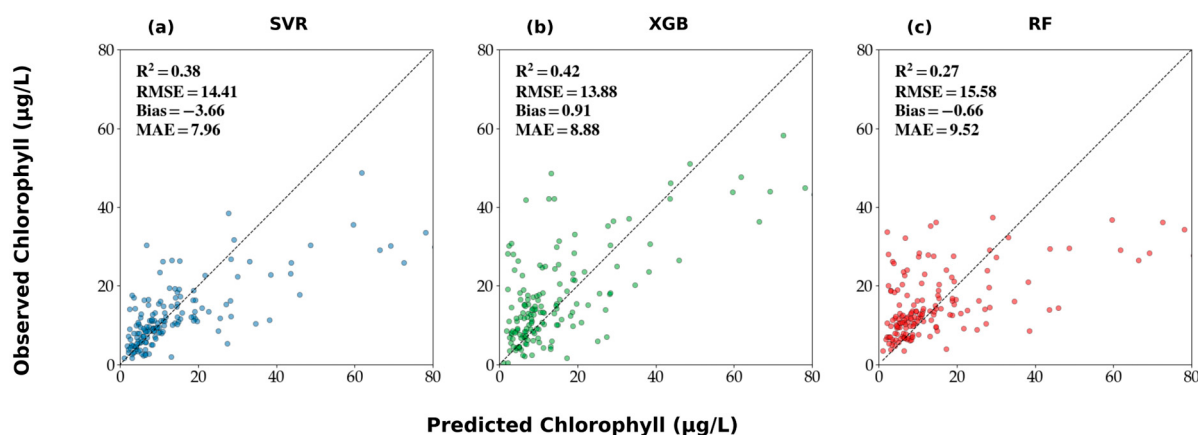


Figure 6. Scatter plots showing the performance of (a) SVR, (b) XGB, and (c) RF models when using raw band reflectance and multiple band indices. Key performance statistics are reported in the inset.

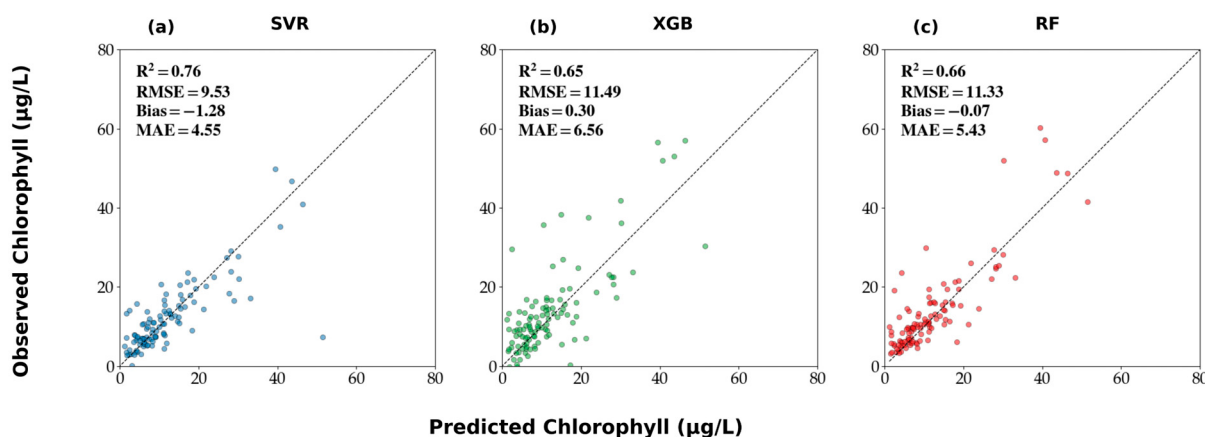


Figure 7. Scatter plots showing the performance of (a) SVR, (b) XGB, and (c) RF models when using raw spectral bands and physicochemical variables.

3.1.3. Model Performance Using Bands and Meteorological Parameters

Compared with models that included B + BI (Figure 6) or B + PC (Figure 7), models using band reflectance and meteorological variables for Chl-a prediction performed substantially worse, with lower explained variance and higher error. With this configuration, the explained variance of the models dropped by roughly half or more, with none of the models achieving an R^2 above 0.20 (Figure 8). This highlighted that, while meteorological variables contribute some predictive power, they cannot substitute for in-water measurements and spectral indices when estimating Chl-a.

3.1.4. Model Performance Using All Input Variables

The performance of three models when using spectral bands, band indices, and physicochemical and meteorological parameters (i.e., ALL input group) (Figure 9) was superior to the models using only spectral bands and band indices (Figure 6), followed by the models using spectral bands and meteorological variables (Figure 8). These models provided slightly more balanced performance among the three algorithms. SVR again achieved the highest R^2 (0.70), with a low RMSE ($10.63 \mu\text{g L}^{-1}$) and MAE ($5.35 \mu\text{g L}^{-1}$), followed closely by RF and XGB (both $R^2 \sim 0.65$ – 0.66). Bias remains small across models, indicating well-centered predictions. Compared to the B + PC combination, the inclusion of additional

variables (meteorological factors) did not substantially improve model performance and, in some cases, slightly increased RMSE, particularly for SVR. These results, consistent with the trends observed in the multi-seed stability analysis (Tables S6 and S7), suggest that the addition of all variables provides marginal gains beyond the strong predictive power already achieved with B + PC.

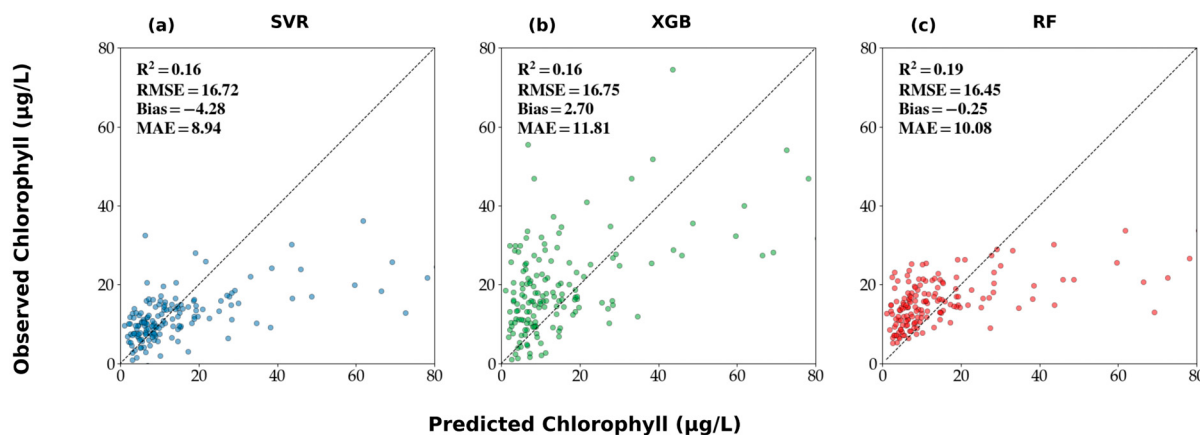


Figure 8. Scatter plot showing the performance of (a) SVR, (b) XGB, and (c) RF models when using raw bands from HLS along with meteorological variables.

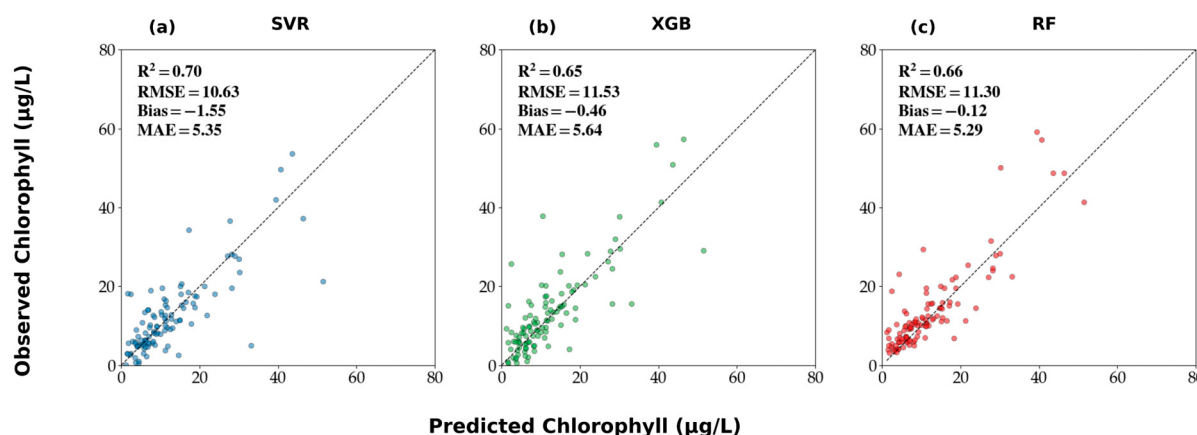


Figure 9. Scatter plot showing the performance of raw bands from HLS along with all predictor variables for the models: (a) SVR, (b) XGB, and (c) RF.

3.2. Global and Local Explanations of Models with Varying Input Variables

The SHAP technique was applied to all models across the different input variable combinations and random seeds described in Section 3.1. For each model and variable combination, SHAP values were averaged across 10 seeds to produce summary plots that provide a clear visualization of both the magnitude and direction of each predictor's influence on model outputs. Although SHAP plots were generated for three models and four input combinations, the discussion below focuses on model configurations for two variable combinations (B + PC and ALL) that achieved a higher R^2 and lower RMSE in order to highlight the predictors driving the best-performing models. The SHAP plots for the remaining variable combinations (B + BI and B + MET) are provided in the Supplementary Materials (Figures S2 and S3).

3.2.1. Global Explanation of Models

Consistent with the model's performance trends discussed previously, the inclusion of physicochemical variables, particularly TP and MC, drives much of the predictive accuracy

across all models (Figure 10). These two variables consistently rank as the top predictors, with large mean SHAP values across models and input groups, underscoring their critical role in accurate and robust Chl-a prediction across ML models. Additional physicochemical variables such as DOM, TN, SD, and the blue band also contributed meaningfully to model predictions, though with somewhat greater variability in their importance rankings across models. The addition of meteorological and vegetation variables in the ALL-input group only provides modest incremental value; variables such as sunlight and wind speed appear in the SHAP plots but do not compete with the core physicochemical predictors in influence. These patterns align with the performance results from the previous figures (Figures 7 and 9), where models using B + PC or ALL inputs substantially outperformed those using spectral bands alone.

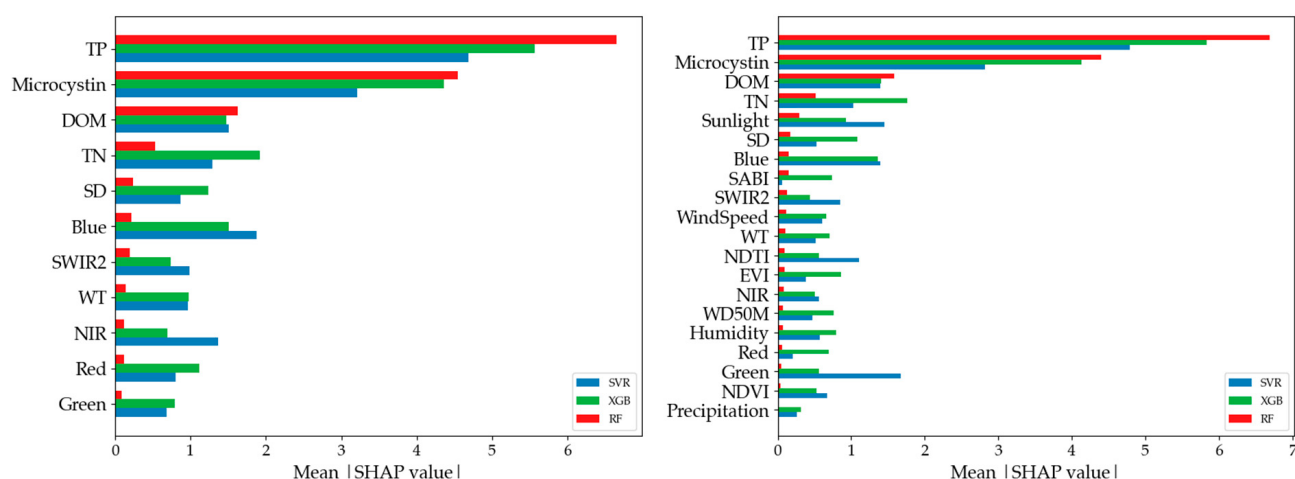


Figure 10. SHAP summary plots showing the relative influence of predictor variables for Chl-a prediction across SVR, XGB, and RF models, using the B + PC (left) and ALL input (right) group.

3.2.2. Local Explanation of Models

Figure 11 presents SHAP dependence plots illustrating how individual feature values impact Chl-a predictions across SVR, XGB, and RF models for both the B + PC input group (a–c) and the ALL-input group (d–f). These plots reveal not only the magnitude of variable influence but also the direction (positive or negative) and variation in local effects across samples. Across both input groups and all models, TP exhibits a strong, consistent positive relationship with Chl-a predictions, with higher TP values leading to markedly higher predicted Chl-a levels. Similarly, Microcystin shows a generally positive association across models, reflecting its close linkage with algal biomass and bloom conditions.

Other key variables exhibit more nuanced or variable relationships. Specifically, DOM shows a mostly positive effect, particularly in RF and SVR, though with some scatter at low DOM levels. Total Nitrogen (TN) also tends to have a positive contribution, though with greater variation across models, stronger in XGB, weaker in RF. Several spectral bands exhibit more complex patterns. Blue reflectance generally contributes negatively to Chl-a predictions. Higher blue reflectance (often associated with lower algal concentrations) tends to reduce predicted Chl-a. NIR and SWIR bands show mixed or non-linear effects depending on the model, with NIR often having a negative contribution at higher values (reflecting water and vegetation interference).

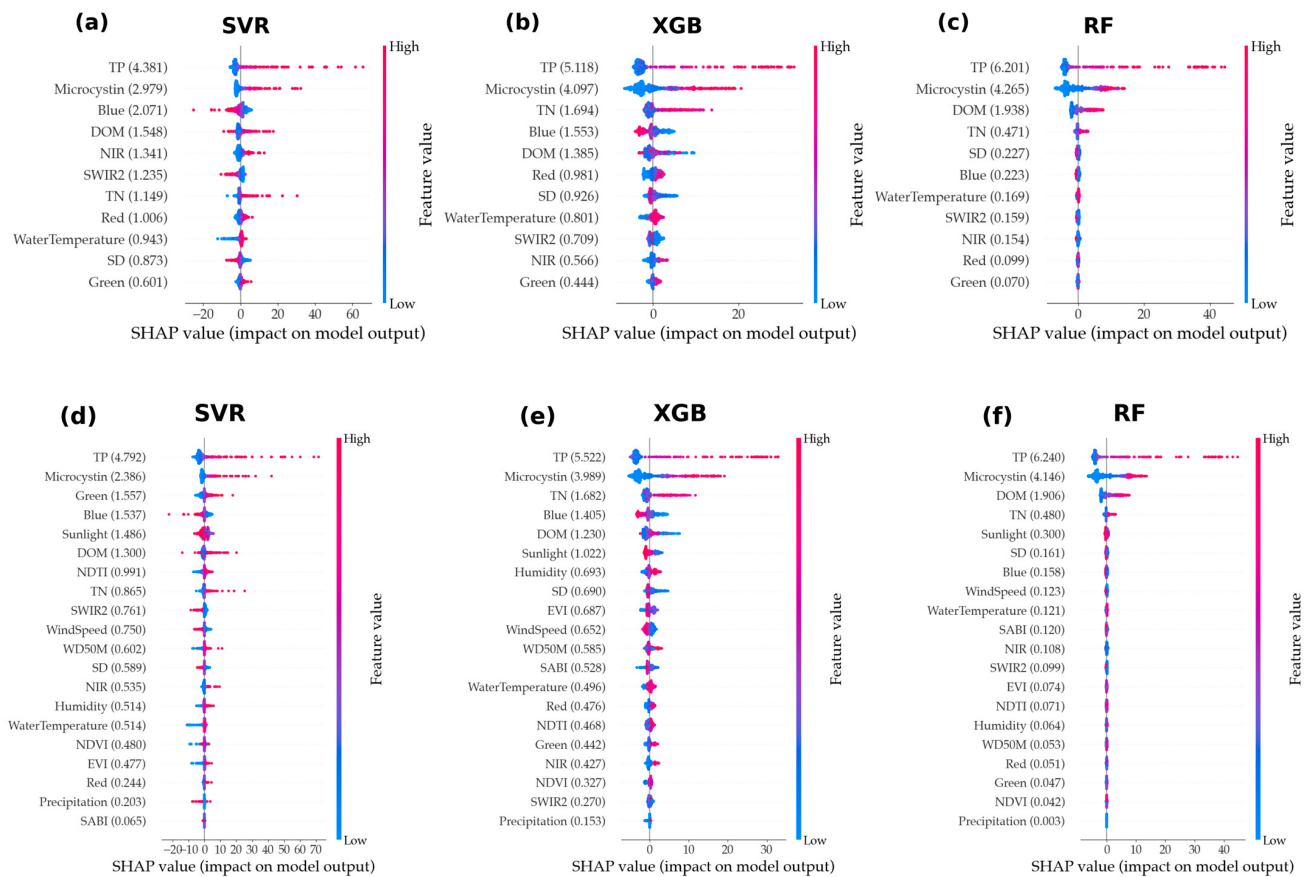


Figure 11. SHAP plots illustrating the local effects of key predictors on Chl-a predictions across SVR, XGB, and RF models using B + PC (a–c) and ALL (d–f) input groups.

In the ALL-input group, additional variables such as sunlight show a clear positive effect on chlorophyll predictions across models, consistent with the role of light availability in promoting algal growth. Conversely, meteorological variables such as wind speed and precipitation display weak or inconsistent effects, often centered around zero, suggesting minimal contributions to Chl-a variability within our modeling framework.

3.3. Explaining Model Behavior: Effects of Key Individual Predictors and Interactions

To further examine the interactions between key physicochemical variables and spectral bands identified as consistently important across models and to elucidate their combined influence on Chl-a prediction, we generated SHAP interaction plots using the XGB model with the B + PC input configuration (Figure 12). In these interaction plots, we assessed the interaction of Microcystin with variables, including TP, TN, DOM, and the blue band, given that it is a key indicator of algal bloom toxicity.

Based on SHAP interaction plots, it was observed that TP and MC have a synergetic effect, where an elevated TP further amplified the positive influence of MC on Chl-a (Figure 12a–c). The mean SHAP interaction of TP and MC is 0.264, meaning that beyond what TP and MC each contribute individually, their combined effect adds or subtracts, on average, an extra 0.264 $\mu\text{g/L}$ to the Chl-a prediction. Similarly, MC and TN interactions indicated that higher TN levels strengthened the positive contribution of MC (interaction of 0.55), though with more variability across the TN gradient, indicating that TN moderates the MC effect at certain ranges (Figure 12d–f). In contrast, DOM exhibited a more complex relationship. At low MC levels, increasing DOM often led to negative SHAP contributions, while at high MC levels, the positive effect of MC on Chl-a remained dominant regardless of DOM variations (Figure 12g–i). The interaction between MC and blue band reflectance

was relatively weak, with higher blue reflectance generally contributing negatively to Chl-a predictions, likely reflecting clearer water conditions with lower phytoplankton biomass, while MC continued to exert a strong positive effect (Figure 12j–l). Collectively, these interactions highlight the intertwined roles of nutrient availability, water optical properties, and bloom toxicity in shaping Chl-a dynamics within the modeling framework.

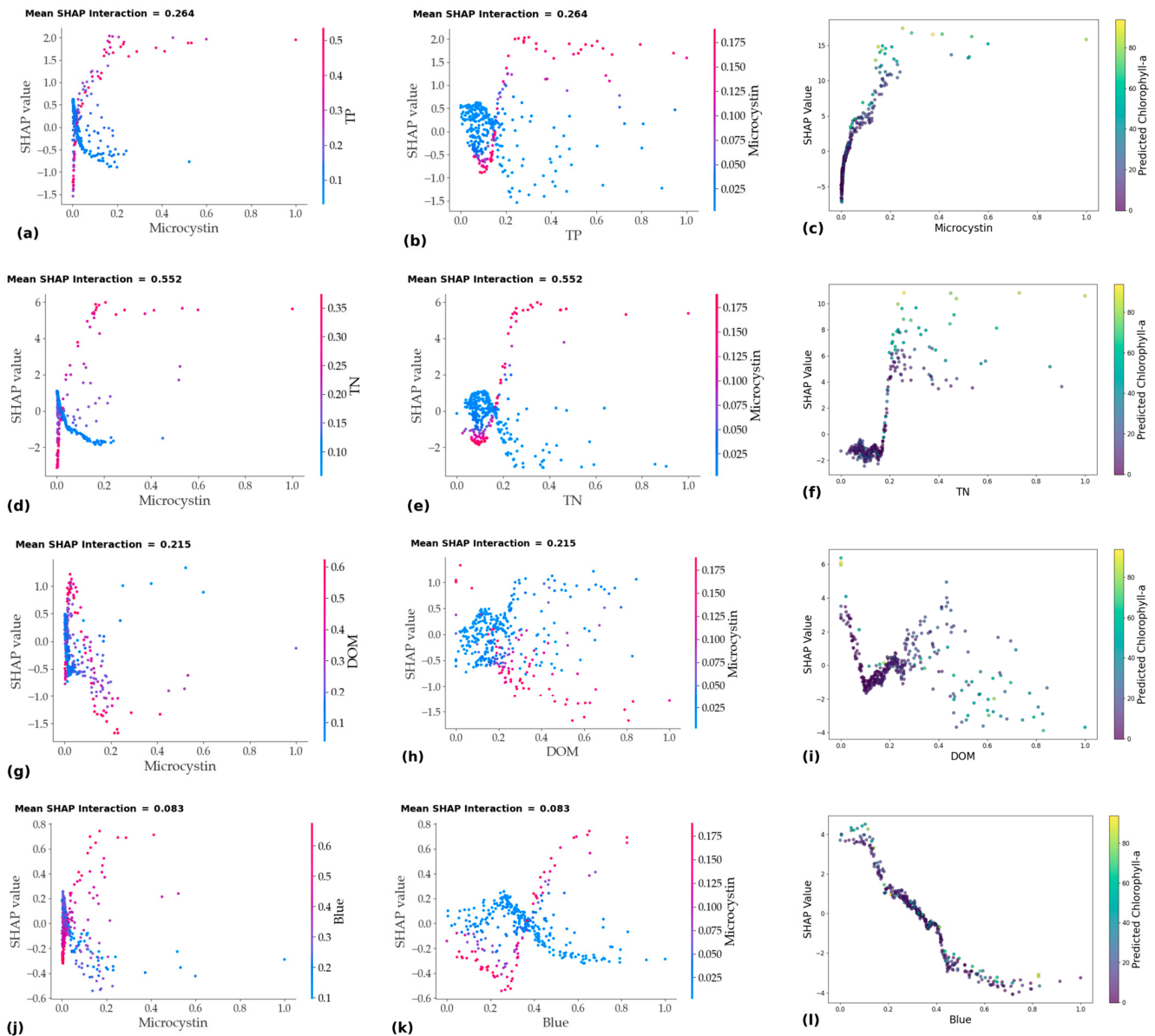


Figure 12. SHAP interaction dependence plots for the XGB model using the B + PC variable combination, illustrating pairwise interactions between TP and Microcystin (MC) (a,b), MC and predicted Chl-a (c), TN and MC (d,e), TN and predicted Chl-a (f), DOM and MC (g,h), DOM and predicted Chl-a (i), blue and MC (j,k), and Blue and predicted Chl-a (l). The primary Y-axis displays the SHAP value of the variable shown on the X-axis, while the secondary Y-axis indicates the corresponding values of the interacting variable.

4. Discussion

4.1. Comparative Evaluation of Satellite-Based Chlorophyll-a Estimation Approaches in Western Lake Erie

Several studies have used satellite-based approaches to estimate chlorophyll-a concentrations in Western Lake Erie, reporting a range of R^2 values depending on the sensor's

type and resolution, algorithm complexity, and modeling approach. Most existing work has focused on MERIS satellite data and semi-analytical or empirical models. For instance, Ali et al. [67] applied semi-analytical models to MERIS imagery and reported R^2 values between 0.61 and 0.65. Zolfagjari and Duguay [68] developed linear mixed models using MERIS-derived band ratios, achieving R^2 values ranging from 0.49 to 0.56. Pirasteh et al. [17] applied the Sentinel-2 MSI Maximum Chlorophyll Index (MCI, exploiting Sentinel's red-edge bands) and Sentinel-3 OLCI with empirical band ratios, achieving an R^2 of 0.92 for summer 2016–2017 data; their ground-truth Chl-a data ranged from 0.13 $\mu\text{g/L}$ to 88 $\mu\text{g/L}$, with a mean of 16.5 $\mu\text{g/L}$. In comparison, our prior study [34] employed RF models using selected spectral bands from Sentinel-3 OLCI to estimate Chl-a between April and October from 2016 to 2021, resulting in an R^2 of 0.55 based on ground-truth data ranging from 0.53 to 124.55, with a mean of 14.84 $\mu\text{g/L}$.

In this context, our current study leverages the HLS product, which offers a notable advantage for water-quality monitoring. By integrating the high spatial resolution of Landsat and Sentinel-2 (10–30 m) with an enhanced revisit frequency of 2–3 days, HLS provides significantly improved spatiotemporal coverage compared to individual sensors such as Sentinel-2 and Sentinel-3 or legacy sensors like MERIS. While Sentinel-3 and MERIS provide spectrally rich data optimized for aquatic applications, their coarse resolution (300–1200 m) limits effectiveness in capturing nearshore variability and small-scale bloom dynamics. In contrast, HLS enables consistent, cloud-gap-filled observations that are critical for tracking rapid water-quality changes in systems like Western Lake Erie.

Furthermore, our models built on the HLS dataset and physicochemical water-quality variables achieved an R^2 of up to 0.68 with the RF model and up to 0.73 with the SVR and XGB models. These results are on par with previously reported MERIS-data-based studies and align with the lower bound of performance reported for Sentinel-2 MCI methods. Although Sentinel-2 MCI approaches outperformed our models in absolute R^2 (0.90–0.92) [17], they benefited from the targeted use of red-edge bands and indices specifically designed for Case 2 waters, as well as higher spatial resolutions and narrower temporal windows. In contrast, our approach integrates multi-sensor data with physicochemical variables over a broader seasonal scale, offering practical accuracy ($R^2 = 0.68$ –0.73) suitable for operational water-quality screening and trend detection.

In summary, although differences in sensor resolution, algorithm complexity, and seasonal sampling windows contribute to variability in the reported R^2 , making direct comparisons across studies challenging, our results remain comparable to, and in some cases exceed, earlier benchmarks for chlorophyll-a estimation in Western Lake Erie.

4.2. Model Performances Across Various Variable Combinations

An analysis of R^2 variability across random seeds provides further insights into the stability and robustness of the model's performance. Among the three models, XGB demonstrated the most consistent performance, with relatively small fluctuations in R^2 across seeds for the ALL and B + PC input groups (ranges of ~0.57–0.72 and ~0.58–0.73, respectively). This suggests that XGB is less sensitive to data splits and consistently captures key patterns driving chlorophyll-a variations.

In contrast, SVR exhibited greater variability across seeds, particularly for the B + PC input group, where R^2 ranged from 0.30 to 0.76, a substantial jump reflecting sensitivity to different training/test splits. While SVR often achieved the highest R^2 in individual seeds, this variability indicates that its performance is less stable and potentially more influenced by the sample's composition.

RF showed intermediate stability for the ALL and B + PC combinations (ranges of ~0.51–0.68 and ~0.52–0.68, respectively), though it did not reach the top R^2 values as

frequently as SVR or XGB. RF performance was more stable than SVR but less robust than XGB. For the weaker input combinations (B + BI and B + MET), all models exhibited relatively low and unstable R^2 values across seeds, underscoring the limited predictive value of these input groups.

Similar comparisons were made by Tian et al. [15], where four models—XGB, SVR, RF, and ANN—were used to predict water-quality variables, and XGB was found to be the best performer. The study suggested that XGB has a superior ability to handle model complexity and, along with its use of regularization and a gradient-boosting approach, helps prevent overfitting, setting it apart from the other models. Similarly, Hafeez et al. [49] reported that among four ML models, Artificial Neural Network (ANN), SVR, RF, and Cubist, used to estimate chlorophyll-a concentrations from satellite-derived surface reflectance, the RF model yielded the poorest performance, likely due in part to limitations imposed by a relatively small training sample. Fan et al. [69] reported similar observations, where XGB outperformed SVR in predicting solar radiation, showing both stability and better computational efficiency.

In summary, while SVR achieved the best R^2 in more individual seeds, XGB provided the most stable and consistently high performance across seeds and input configurations, an important consideration for practical deployment where model robustness to data variability is critical.

4.3. Model Interpretability for Chl-a Prediction Using SHAP Analyses

The contribution of input variables to Chl-a prediction varied across models, reflecting both expected relationships and the complexity of aquatic systems. For instance, for the models that were solely based on imagery-based input variables (i.e., B + BI), SHAP analyses (Figure S2) confirmed the consistent predictive importance of the blue band, which showed an inverse relationship with Chl-a across all models, capturing the high absorption characteristic of blue light and hence low reflectance in Chl-a-rich water [19,70]. SWIR2 also appeared as an informative predictor in SVR and XGB models. SWIR2 absorption is high in turbid water, with close to a hundred percent in very turbid water [70,71]. The inverse pattern observed between these bands and mean absolute SHAP values suggests that high SWIR2 levels might indicate that Chl-a is a major contributor to increased turbidity.

In contrast, traditional band indices (NDVI and EVI) did not improve model performance as anticipated, likely due to the sensitivity of environmental factors (e.g., sun glint, solar geometry) [72], which were not fully addressed in the preprocessing of datasets; only aerosol and cloud cover were masked in this study. This suggests that raw spectral bands may be more reliable than composite indices for Chl-a prediction in this context.

Meteorological variables, despite their known influence on algal blooms [9,73], were found to be the weakest predictors of Chl-a, likely due to their coarse resolution (~110 km/pixel) relative to the finer spatial detail of spectral and physicochemical data. Almost all meteorological parameters, except for sunlight, ranked low in the SHAPLEY analyses. Although sunlight is known to influence algal growth and Chl-a intensity [8,46], its relationship with Chl-a was unclear in our models (Figure 11), with both high and low sunlight values associated with high Chl-a. This suggests that sunlight's influence on Chl-a may be closely tied to other environmental and biological factors like nutrient availability, seasons, and the presence of other organisms like zooplankton, which feed on algae [74], in turn influencing Chl-a. For instance, zooplankton grazing on algae increases over the summer, reducing algal biomass, which results in lower Chl-a concentrations.

High wind speeds, conversely, were found to have an inverse relationship with Chl-a in SVR and XGB models (Figure S3), aligning with prior studies [34] that show that low wind speeds promote water stratification and nutrient release, fueling algal growth [75].

Low turbulence within a water column restricts oxygen from the surface layers from reaching the bottom effectively, causing oxygen depletion in deeper layers. This hypoxia triggers the anaerobic decomposition of organic matter, releasing nutrients like phosphorus and nitrogen, which promotes eutrophication, fueling algal growth and causing an increase in Chl-a concentration.

Among the 20 input variables tested in the ALL-variable models, only a subset of variables, including TP, MC, TN (physicochemical), SR (meteorological), and blue (spectral), showed consistent impacts on Chl-a predictions, with TP and MC having the strongest influence (Figure 11), which is consistent with the findings of prior studies [13,76,77]. The role of other variables was difficult to interpret, suggesting the complexity of aquatic ecosystems, where Chl-a and/or HAB can be influenced by interactions among seasonal meteorological, physical, and biological factors that are often challenging to capture via ML models. Higher values of TP corresponded to higher Chl-a, which seems to make sense given that TP is one of the primary nutrients that favor HAB growth [13,78,79]. On the other hand, an increase or decrease in Chl-a was proportional to the corresponding MC concentration: in this case, in both directions of the SHAP plot. As most phytoplankton species utilize Chl-a for photosynthesis, incorporating MC in this study adds clarity in distinguishing HABs from other phytoplankton or identifying the non-toxic algae with no MC in them [80–82].

The variable importance patterns observed in this study point to several takeaways. First, incorporating high-resolution physicochemical and spectral data appears to be critical for improving Chl-a predictions. Second, raw spectral bands, particularly in the blue and SWIR regions, may offer more stable predictive power than traditional vegetation indices under varying environmental conditions. Third, meteorological data at finer spatial scales may be needed to better capture their effects on Chl-a dynamics.

Overall, the observed discrepancies in the ranking of variables for Chl-a prediction across models with various combinations of input variables do not indicate the superiority of one model over another. Instead, they reflected the distinct approaches each model uses. Nevertheless, it would be useful to focus on these variables to enhance the model's robustness for Chl-a predictions in the future.

4.4. Model Selection Rationale and Future Directions

We acknowledge that the three ML models used in this study, including RF, SVR, and XGB, are well-established and widely used in remote sensing and environmental modeling. Our rationale for selecting these models was twofold: (1) they offer interpretability and robustness when applied to heterogeneous data sources such as HLS imagery combined with physicochemical covariates, and (2) they allow for direct benchmarking against prior studies, many of which used similar modeling approaches.

While emerging deep learning models (e.g., convolutional neural networks or transformer-based architectures) offer potential improvements, their performance gains are often contingent on larger datasets, which were not available in this study. Moreover, such models may introduce additional complexity and reduce interpretability, factors that might limit their adoption in operational water-quality monitoring frameworks. Nonetheless, we recognize the importance of exploring more advanced techniques and plan to investigate hybrid and deep learning models as higher-resolution, temporally dense datasets become more available in future work.

5. Conclusions

This study provided insights into the potential of integrating spectral, physicochemical, and meteorological data with three ML models—SVR, XGB, and RF—to predict Chl-a

concentrations in Western Lake Erie. Besides assessing the models' performances, we used SHAP techniques to show that various input variables contribute to Chl-a predictions through their relative importance across models. The key findings of the study are listed below:

- When combining physicochemical (i.e., water chemistry) data with spectral satellite information, the models achieved an R^2 of up to 0.76 and an RMSE down to 8.04 $\mu\text{g/L}$, underscoring the value of combining high-spectral and physicochemical inputs for optimal model performance.
- Models that rely solely on meteorological inputs with spectral bands perform considerably worse ($R^2 < 0.40$ across all three algorithms). This suggests that meteorological variables by themselves have limited power to predict Chl-a in our study area.
- Using only satellite-derived variables (no ground chemistry and no meteorology) resulted in moderate performance, with R^2 up to 0.48. While these values are lower than those of models with BI + PC variables, satellite-only approaches could still be practical for preliminary monitoring, especially where field sampling is difficult or expensive.
- When we used all 20 variables (physicochemical + meteorological + spectral), we did not necessarily see an increase in model performance compared to models with B + PC variables.
- While SVR achieved the highest R^2 in more individual runs, XGB demonstrated the most stable and consistently strong performance across various input configurations, a key advantage for practical applications where model robustness to data variability is essential.
- The differences observed in variable rankings across models emphasize that algorithm selection influences not only predictive performance but also the interoperability and stability of the resulting models, which can be a key consideration for operational water-quality monitoring.

These findings are significant in HAB monitoring, as Chl-a concentration is one of the key proxies of blooms. By understanding the role and strength of a multitude of variables in Chl-a prediction through the proper modeling approaches carried out in this study, stakeholders will be able to make an informed decision in addressing HAB-related challenges.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs17132164/s1>, Figure S1: Temporal variation in chlorophyll-a concentration from June 2013 to October 2022; Figure S2: SHAP plots showing the influence of spectral bands and derived indices on Chl-a prediction; Figure S3: SHAP plots showing the influence of spectral bands and meteorological variables on Chl-a prediction; Table S1: Minimum, maximum, mean, median and standard deviation by month for Chl-a concentration; Table S2: Minimum, maximum, mean, median, and standard deviation by year for Chl-a concentration; Table S3: Minimum, maximum, mean, median, and standard deviation for all variables over the study period; Table S4: Quality assessment to evaluate a pixel by indicating different conditions associated with different bit combinations; Table S5: List of hyperparameters and their values used for the models; Table S6: Distribution of R^2 across three models/variable combination and ten random seeds; Table S7: Distribution of RMSE across three models/variable combination and ten random seeds.

Author Contributions: Conceptualization, S.K. and K.Z.; methodology, N.J., A.G., S.K., K.Z., and J.P.; software, N.J.; validation, N.J., A.G., S.K., K.Z., and J.P.; formal analysis, N.J. and A.G.; investigation, N.J.; resources, S.K.; data curation, N.J. and A.G.; writing, N.J., A.G., and S.K.; writing—review and editing, S.K., A.G., K.Z., and J.P.; visualization, N.J. and A.G.; supervision, S.K.; project administration, S.K.; funding acquisition, S.K. and K.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by the Harmful Algal Bloom Research Initiative grant from the Ohio Department of Higher Education and the OSU Graduate School Fellowship programs, including College Allocated and Environmental Fellowships.

Data Availability Statement: Physicochemical water-quality data (Total Phosphorus, Total Nitrogen, Secchi depth, Microcystin, Dissolved Organic Matter, and water temperature) can be obtained from the Ohio Sea Grant's Lake Erie monitoring program at <https://ohioseagrant.osu.edu/research/live/water> (accessed on 19 May 2025). Meteorological data (solar radiation, wind speed, wind direction, humidity, and precipitation) can be obtained from the NASA POWER database at <https://power.larc.nasa.gov> (accessed on 19 May 2025). Spectral bands (Red, Green, Blue, NIR, SWIR2) can be downloaded from the NASA EarthData portal at <https://hls.gsfc.nasa.gov/documents/> (accessed on 19 May 2025). The final dataset can be made available upon request.

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

References

- Hao, Y.; Tang, D.; Yu, L.; Xing, Q. Nutrient and Chlorophyll a Anomaly in Red-Tide Periods of 2003–2008 in Sishili Bay, China. *Chin. J. Oceanol. Limnol.* **2011**, *29*, 664–673. [CrossRef]
- Tewari, M.; Kishtawal, C.M.; Moriarty, V.W.; Ray, P.; Singh, T.; Zhang, L.; Treinish, L.; Tewari, K. Improved Seasonal Prediction of Harmful Algal Blooms in Lake Erie Using Large-Scale Climate Indices. *Commun. Earth Environ.* **2022**, *3*, 195. [CrossRef]
- Carmichael, W.W.; Boyer, G.L. Health Impacts from Cyanobacteria Harmful Algae Blooms: Implications for the North American Great Lakes. *Harmful Algae* **2016**, *54*, 194–212. [CrossRef] [PubMed]
- Backer, L.C.; Manassaram-Baptiste, D.; Leprell, R.; Bolton, B. Cyanobacteria and Algae Blooms: Review of Health and Environmental Data from the Harmful Algal Bloom-Related Illness Surveillance System (HABISS) 2007–2011. *Toxins* **2007**, *7*, 1048–1064. [CrossRef] [PubMed]
- Larkin, S.L.; Adams, C.M. Economic Consequences of Harmful Algal Blooms: Literature Summary. *EDIS* **2013**, *2013*, 1–10. [CrossRef]
- Hoagland, P.; Anderson, D.M.; Kaoru, Y.; White, A.W. The Economic Effects of Harmful Algal Blooms in the United States: Estimates, Assessment Issues, and Information Needs. *Estuaries* **2002**, *25*, 819–837. [CrossRef]
- Ai, H.; Zhang, K.; Sun, J.; Zhang, H. Short-Term Lake Erie Algal Bloom Prediction by Classification and Regression Models. *Water Res.* **2023**, *232*, 119710. [CrossRef]
- Mu, M.; Li, Y.; Bi, S.; Lyu, H.; Xu, J.; Lei, S.; Miao, S.; Zeng, S.; Zheng, Z.; Du, C. Prediction of Algal Bloom Occurrence Based on the Naive Bayesian Model Considering Satellite Image Pixel Differences. *Ecol. Indic.* **2021**, *124*, 107416. [CrossRef]
- Michalak, A.M.; Anderson, E.J.; Beletsky, D.; Boland, S.; Bosch, N.S.; Bridgeman, T.B.; Chaffin, J.D.; Cho, K.; Confesor, R.; Daloglu, I.; et al. Record-Setting Algal Bloom in Lake Erie Caused by Agricultural and Meteorological Trends Consistent with Expected Future Conditions. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 6448–6452. [CrossRef]
- Mohamed, M.N.; Wellen, C.; Parsons, C.T.; Taylor, W.D.; Arhonditsis, G.; Chomicki, K.M.; Boyd, D.; Weidman, P.; Mundle, S.O.C.; Cappellen, P.V.; et al. Understanding and Managing the Re-Eutrophication of Lake Erie: Knowledge Gaps and Research Priorities. *Freshw. Sci.* **2019**, *38*, 675–691. [CrossRef]
- Papenfus, M.; Schaeffer, B.; Pollard, A.I.; Loftin, K. Exploring the Potential Value of Satellite Remote Sensing to Monitor Chlorophyll-a for US Lakes and Reservoirs. *Environ. Monit. Assess.* **2020**, *192*, 808. [CrossRef] [PubMed]
- Mishra, S.; Stumpf, R.P.; Schaeffer, B.; Werdell, P.J.; Loftin, K.A.; Meredith, A. Evaluation of a Satellite-Based Cyanobacteria Bloom Detection Algorithm Using Field-Measured Microcystin Data. *Sci. Total Environ.* **2021**, *774*, 145462. [CrossRef]
- Gholizadeh, M.H.; Melesse, A.M.; Reddi, L. A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors* **2016**, *16*, 1298. [CrossRef] [PubMed]
- Shanmugam, P. A New Bio-Optical Algorithm for the Remote Sensing of Algal Blooms in Complex Ocean Waters. *J. Geophys. Res. Ocean.* **2011**, *116*, 4016. [CrossRef]

15. 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]
16. 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]
17. Pirasteh, S.; Mollaei, S.; Narges Fathollahi, S.; Li, J. Estimation of Phytoplankton Chlorophyll-a Concentrations in the Western Basin of Lake Erie Using Sentinel-2 and Sentinel-3 Data. *Can. J. Remote Sens.* **2020**, *46*, 585–602. [CrossRef]
18. Hafeez, S.; Wong, M.S.; Abbas, S.; Asim, M. Evaluating Landsat-8 and Sentinel-2 Data Consistency for High Spatiotemporal Inland and Coastal Water Quality Monitoring. *Remote Sens.* **2022**, *14*, 3155. [CrossRef]
19. Salem, S.I.; Higa, H.; Kim, H.; Kobayashi, H.; Oki, K.; Oki, T. Assessment of Chlorophyll-a Algorithms Considering Different Trophic Statuses and Optimal Bands. *Sensors* **2017**, *17*, 1746. [CrossRef]
20. Verhamme, E.M.; Redder, T.M.; Schlea, D.A.; Grush, J.; Bratton, J.F.; DePinto, J.V. Development of the Western Lake Erie Ecosystem Model (WLEEM): Application to Connect Phosphorus Loads to Cyanobacteria Biomass. *J. Great Lakes Res.* **2016**, *42*, 1193–1205. [CrossRef]
21. Walsh, J.J.; Penta, B.; Dieterle, D.A.; Bissett, W.P. Predictive Ecological Modeling of Harmful Algal Blooms. *Hum. Ecol. Risk Assess. Int. J.* **2001**, *7*, 1369–1383. [CrossRef]
22. Yu, P.; Gao, R.; Zhang, D.; Liu, Z.P. Predicting Coastal Algal Blooms with Environmental Factors by Machine Learning Methods. *Ecol. Indic.* **2021**, *123*, 107334. [CrossRef]
23. Wen, J.; Yang, J.; Li, Y.; Gao, L. Harmful Algal Bloom Warning Based on Machine Learning in Maritime Site Monitoring. *Knowl.-Based Syst.* **2022**, *245*, 108569. [CrossRef]
24. Izadi, M.; Sultan, M.; Kadiri, R.E.; Ghannadi, A.; Abdelmohsen, K. A Remote Sensing and Machine Learning-Based Approach to Forecast the Onset of Harmful Algal Bloom. *Remote Sens.* **2021**, *13*, 3863. [CrossRef]
25. Khan, R.M.; Salehi, B.; Mahdianpari, M.; Mohammadimanesh, F.; Mountrakis, G.; Quackenbush, L.J. A Meta-Analysis on Harmful Algal Bloom (Hab) Detection and Monitoring: A Remote Sensing Perspective. *Remote Sens.* **2021**, *13*, 4347. [CrossRef]
26. Park, J.; Khanal, S.; Zhao, K.; Byun, K. Remote Sensing of Chlorophyll-a and Water Quality over Inland Lakes: How to Alleviate Geo-Location Error and Temporal Discrepancy in Model Training. *Remote Sens.* **2024**, *16*, 2761. [CrossRef]
27. Huang, H.; Wang, W.; Lv, J.; Liu, Q.; Liu, X.; Xie, S.; Wang, F.; Feng, J. Relationship between Chlorophyll a and Environmental Factors in Lakes Based on the Random Forest Algorithm. *Water* **2022**, *14*, 3128. [CrossRef]
28. Chegoonian, A.M.; Zolfaghari, K.; Baulch, H.M.; Duguay, C.R. Support Vector Regression for Chlorophyll-a Estimation Using Sentinel-2 Images in Small Waterbodies. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Brussels, Belgium, 11–16 July 2021; pp. 7449–7452. [CrossRef]
29. Korff, B. Von Assessing and Forecasting Chlorophyll Abundances in Minnesota Lakes Using Remote Sensing and Statistical Approaches. Master's Thesis, Minnesota State University, Mankato, MN, USA, 2016.
30. Dagtekin Bsc, O. Deep Learning for the Early Detection of Harmful Algal Blooms and Improving Water Quality Monitoring. Ph.D. Thesis, University of Hull, Hull, UK, 2022.
31. Ly, Q.V.; Nguyen, X.C.; Lê, N.C.; Truong, T.D.; Hoang, T.H.T.; Park, T.J.; Maqbool, T.; Pyo, J.C.; Cho, K.H.; Lee, K.S.; et al. Application of Machine Learning for Eutrophication Analysis and Algal Bloom Prediction in an Urban River: A 10-Year Study of the Han River, South Korea. *Sci. Total Environ.* **2021**, *797*, 149040. [CrossRef]
32. Chaffin, J.; Bratton, J.F.; Verhamme, E.M.; Bair, H.B.; Beecher, A.A.; Binding, C.E.; Birbeck, J.A.; Bridgeman, T.B.; Chang, X.; Crossman, J.; et al. The Lake Erie HABs Grab: A Binational Collaboration to Characterize the Western Basin Cyanobacterial Harmful Algal Blooms at an Unprecedented High-Resolution Spatial Scale. *Harmful Algae* **2021**, *108*, 102080. [CrossRef]
33. Bartish, T. A Review of Exchange Processes Among the Three Basins of Lake Erie. *J. Great Lakes Res.* **1987**, *13*, 607–618. [CrossRef]
34. Joshi, N.; Park, J.; Zhao, K.; Londo, A.; Khanal, S. Monitoring Harmful Algal Blooms and Water Quality Using Sentinel-3 OLCI Satellite Imagery with Machine Learning. *Remote Sens.* **2024**, *16*, 2444. [CrossRef]
35. Stone Lab Algal and Water Quality Laboratory. Available online: <https://ohioseagrant.osu.edu/research/live/water> (accessed on 19 June 2025).
36. Stackhouse, P. Methodology. Available online: <https://power.larc.nasa.gov/docs/methodology/> (accessed on 19 June 2025).
37. Claverie, M.; Ju, J.; Masek, J.G.; Dungan, J.L.; Vermote, E.F.; Roger, J.C.; Skakun, S.V.; Justice, C. The Harmonized Landsat and Sentinel-2 Surface Reflectance Data Set. *Remote Sens. Environ.* **2018**, *219*, 145–161. [CrossRef]
38. Kahru, M.; Leppanen, J.M.; Rud, O. Cyanobacterial Blooms Cause Heating of the Sea Surface. *Mar. Ecol. Prog. Ser.* **1993**, *101*, 1–8. [CrossRef]
39. Gitelson, A.A.; Merzlyak, M.N. Remote Estimation of Chlorophyll Content in Higher Plant Leaves. *Int. J. Remote Sens.* **1997**, *18*, 2691–2697. [CrossRef]
40. Van der Merwe, D.; Price, K.P. Harmful Algal Bloom Characterization at Ultra-High Spatial and Temporal Resolution Using Small Unmanned Aircraft Systems. *Toxins* **2015**, *7*, 1065–1078. [CrossRef] [PubMed]

41. Lacaux, J.P.; Tourre, Y.M.; Vignolles, C.; Ndione, J.A.; Lafaye, M. Classification of Ponds from High-Spatial Resolution Remote Sensing: Application to Rift Valley Fever Epidemics in Senegal. *Remote Sens. Environ.* **2007**, *106*, 66–74. [\[CrossRef\]](#)
42. Gitelson, A.A.; Viña, A.; Ciganda, V.; Rundquist, D.C.; Arkebauer, T.J. Remote Estimation of Canopy Chlorophyll Content in Crops. *Geophys. Res. Lett.* **2005**, *32*, 1–4. [\[CrossRef\]](#)
43. Alawadi, F. Detection of Surface Algal Blooms Using the Newly Developed Algorithm Surface Algal Bloom Index (SABI). In *Remote Sensing of the Ocean, Sea Ice, and Large Water Regions 2010*; SPIE: Bellingham, WA, USA, 2010; Volume 7825, p. 782506. [\[CrossRef\]](#)
44. Fang, C.; Song, K.S.; Shang, Y.X.; Ma, J.H.; Wen, Z.D.; Du, J. Remote Sensing of Harmful Algal Blooms Variability for Lake Hulun Using Adjusted FAI (AFAI) Algorithm. *J. Environ. Inform.* **2019**, *34*, 108–122. [\[CrossRef\]](#)
45. Huete, A.R.; Liu, H.Q.; Batchily, K.; Van Leeuwen, W. A Comparison of Vegetation Indices over a Global Set of TM Images for EOS-MODIS. *Remote Sens. Environ.* **1997**, *59*, 440–451. [\[CrossRef\]](#)
46. Zhang, Y.; Ma, R.; Duan, H.; Loiselle, S.A.; Xu, J.; Ma, M. A Novel Algorithm to Estimate Algal Bloom Coverage to Subpixel Resolution in Lake Taihu. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 3060–3068. [\[CrossRef\]](#)
47. Cao, M.; Qing, S.; Jin, E.; Hao, Y.; Zhao, W. A Spectral Index for the Detection of Algal Blooms Using Sentinel-2 Multispectral Instrument (MSI) Imagery: A Case Study of Hulun Lake, China. *Int. J. Remote Sens.* **2021**, *42*, 4514–4535. [\[CrossRef\]](#)
48. Alharbi, B. Remote Sensing Techniques for Monitoring Algal Blooms in the Area between Jeddah and Rabigh on the Red Sea Coast. *Remote Sens. Appl. Soc. Environ.* **2023**, *30*, 100935. [\[CrossRef\]](#)
49. Kislik, C.; Dronova, I.; Kelly, M. UAVs in Support of Algal Bloom Research: A Review of Current Applications and Future Opportunities. *Drones* **2018**, *2*, 35. [\[CrossRef\]](#)
50. Elhag, M.; Gitas, I.; Othman, A.; Bahrawi, J.; Psilovikos, A.; Al-Amri, N. Time Series Analysis of Remotely Sensed Water Quality Parameters in Arid Environments, Saudi Arabia. *Environ. Dev. Sustain.* **2021**, *23*, 1392–1410. [\[CrossRef\]](#)
51. 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. [\[CrossRef\]](#)
52. Malinowski, R.; Groom, G.; Schwanghart, W.; Heckrath, G. Detection and Delineation of Localized Flooding from WorldView-2 Multispectral Data. *Remote Sens.* **2015**, *7*, 14853–14875. [\[CrossRef\]](#)
53. Yip, H.D.; Johansson, J.; Hudson, J.J. A 29-Year Assessment of the Water Clarity and Chlorophyll-a Concentration of a Large Reservoir: Investigating Spatial and Temporal Changes Using Landsat Imagery. *J. Great Lakes Res.* **2015**, *41*, 34–44. [\[CrossRef\]](#)
54. Fensholt, R.; Sandholt, I.; Stisen, S. Evaluating MODIS, MERIS, and VEGETATION Vegetation Indices Using in Situ Measurements in a Semiarid Environment. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1774–1786. [\[CrossRef\]](#)
55. Maeda, E.E.; Lisboa, F.; Kaikkonen, L.; Kallio, K.; Koponen, S.; Brotas, V.; Kuikka, S. Temporal Patterns of Phytoplankton Phenology across High Latitude Lakes Unveiled by Long-Term Time Series of Satellite Data. *Remote Sens. Environ.* **2019**, *221*, 609–620. [\[CrossRef\]](#)
56. Stefan, H.G.; Preud'homme, E.B. Stream Temperature Estimation From Air Temperature. *JAWRA J. Am. Water Resour. Assoc.* **1993**, *29*, 27–45. [\[CrossRef\]](#)
57. Alvarez-Vanhard, E.; Corpetti, T.; Houet, T. UAV & Satellite Synergies for Optical Remote Sensing Applications: A Literature Review. *Sci. Remote Sens.* **2021**, *3*, 100019. [\[CrossRef\]](#)
58. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learning* **1995**, *20*, 273–297. [\[CrossRef\]](#)
59. Pamula, A.S.P.; Gholizadeh, H.; Krzmarzick, M.J.; Mausbach, W.E.; Lampert, D.J. A Remote Sensing Tool for near Real-Time Monitoring of Harmful Algal Blooms and Turbidity in Reservoirs. *JAWRA J. Am. Water Resour. Assoc.* **2023**, *59*, 929–949. [\[CrossRef\]](#)
60. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
61. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [\[CrossRef\]](#)
62. Melkumova, L.E.; Shatskikh, S.Y. Comparing Ridge and LASSO Estimators for Data Analysis. *Procedia Eng.* **2017**, *201*, 746–755. [\[CrossRef\]](#)
63. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the Advances in neural information processing systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
64. Jeong, B.; Chapeta, M.R.; Kim, M.; Kim, J.; Shin, J.; Cha, Y. Machine Learning-Based Prediction of Harmful Algal Blooms in Water Supply Reservoirs. *Water Qual. Res. J.* **2022**, *57*, 304. [\[CrossRef\]](#)
65. Kim, Y.; Kim, T.H.; Shin, J.; Lee, D.S.; Park, Y.S.; Kim, Y.; Cha, Y.K. Validity Evaluation of a Machine-Learning Model for Chlorophyll a Retrieval Using Sentinel-2 from Inland and Coastal Waters. *Ecol. Indic.* **2022**, *137*, 108737. [\[CrossRef\]](#)
66. Park, J.; Lee, W.H.; Kim, K.T.; Park, C.Y.; Lee, S.; Heo, T.Y. Interpretation of Ensemble Learning to Predict Water Quality Using Explainable Artificial Intelligence. *Sci. Total Environ.* **2022**, *832*, 155070. [\[CrossRef\]](#)
67. Ali, K.; Witter, D.; Ortiz, J. Application of Empirical and Semi-Analytical Algorithms to MERIS Data for Estimating Chlorophyll a in Case 2 Waters of Lake Erie. *Env. Earth Sci.* **2014**, *71*, 4209–4220. [\[CrossRef\]](#)

68. Zolfaghari, K.; Duguay, C. Estimation of Water Quality Parameters in Lake Erie from MERIS Using Linear Mixed Effect Models. *Remote Sens.* **2016**, *8*, 473. [CrossRef]
69. Fan, J.; Wang, X.; Wu, L.; Zhou, H.; Zhang, F.; Yu, X.; Lu, X.; Xiang, Y. Comparison of Support Vector Machine and Extreme Gradient Boosting for Predicting Daily Global Solar Radiation Using Temperature and Precipitation in Humid Subtropical Climates: A Case Study in China. *Energy Convers. Manag.* **2018**, *164*, 102–111. [CrossRef]
70. Shi, W.; Wang, M. An Assessment of the Black Ocean Pixel Assumption for MODIS SWIR Bands. *Remote Sens. Environ.* **2009**, *113*, 1587–1597. [CrossRef]
71. Vanhellemont, Q.; Ruddick, K. Advantages of High Quality SWIR Bands for Ocean Colour Processing: Examples from Landsat-8. *Remote Sens. Environ.* **2015**, *161*, 89–106. [CrossRef]
72. Hu, C. A Novel Ocean Color Index to Detect Floating Algae in the Global Oceans. *Remote Sens. Environ.* **2009**, *113*, 2118–2129. [CrossRef]
73. Guimarães, D.B.M.M.; Lima Neto, I.E. Chlorophyll-a Prediction in Tropical Reservoirs as a Function of Hydroclimatic Variability and Water Quality. *Environ. Sci. Pollut. Res.* **2023**, *30*, 91028–91045. [CrossRef]
74. Adams, H.; Ye, J.; Persaud, B.; Slowinski, S.; Kheyrollah Pour, H.; Van Cappellen, P. Chlorophyll-a Growth Rates and Related Environmental Variables in Global Temperate and Cold-Temperate Lakes. *Earth Syst. Sci. Data* **2021**, *14*, 5139–5156. [CrossRef]
75. Deng, J.; Paerl, H.W.; Qin, B.; Zhang, Y.; Zhu, G.; Jeppesen, E.; Cai, Y.; Xu, H. Climatically-Modulated Decline in Wind Speed May Strongly Affect Eutrophication in Shallow Lakes. *Sci. Total Environ.* **2018**, *645*, 1361–1370. [CrossRef] [PubMed]
76. Qin, B.; Yang, G.; Ma, J.; Wu, T.; Li, W.; Liu, L.; Deng, J.; Zhou, J. Spatiotemporal Changes of Cyanobacterial Bloom in Large Shallow Eutrophic Lake Taihu, China. *Front. Microbiol.* **2018**, *9*, 451. [CrossRef]
77. Lee, G.F.; Jones-Lee, A.; Rast, W.; Macero, A. El Secchi Depth as a Water Quality Parameter. Available online: https://www.gfredlee.com/Nutrients/Secchi_Depth.pdf (accessed on 19 June 2025).
78. Deng, J.; Chen, F.; Hu, W.; Lu, X.; Xu, B.; Hamilton, D.P. Variations in the Distribution of Chl-a and Simulation Using a Multiple Regression Model. *Int. J. Environ. Res. Public Health* **2019**, *16*, 4553. [CrossRef]
79. Stow, C.A.; Cha, Y. Are Chlorophyll a -Total Phosphorus Correlations Useful for Inference and Prediction? *Environ. Sci. Technol.* **2013**, *47*, 3768–3773. [CrossRef]
80. Hollister, J.W.; Kreakie, B.J.; Wilson, A.E.; Marion, J.W. Associations between Chlorophyll a and Various Microcystin Health Advisory Concentrations. *F1000Research* **2016**, *5*, 151. [CrossRef] [PubMed]
81. Cunha, D.G.F.; Dodds, W.K.; Arthur, S.; Loiselle, S.A. Factors Related to Water Quality and Thresholds for Microcystin Concentrations in Subtropical Brazilian Reservoirs. *Inland Water* **2018**, *8*, 368–380. [CrossRef]
82. Francy, D.S.; Brady, A.M.G.; Stelzer, E.A.; Cicale, J.R.; Hackney, C.; Dalby, H.D.; Struffolino, P.; Dwyer, D.F. Predicting Microcystin Concentration Action-Level Exceedances Resulting from Cyanobacterial Blooms in Selected Lake Sites in Ohio. *Environ. Monit. Assess* **2020**, *192*, 513. [CrossRef] [PubMed]

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.