limnological review

MDPI

Article

Proximal Hyperspectral Sensing and Machine Learning for Chlorophyll-a Retrieval in Optically Complex Urban Freshwaters

Tiago A. Figueiredo [1], Bernardo T. A. Souza [1], Daniel H. C. Salim [1], Caio C. S. Mello [1], Gabriel Pereira [2] and Camila C. Amorim [1,*]

[1] Department of Sanitary and Environmental Engineering, Universidade Federal de Minas Gerais, Belo Horizonte 31270901, MG, Brazil; tiagofigueiredo.eng.amb@gmail.com (T.A.F.); bernardoo.tadeu@gmail.com (B.T.A.S.); danielhcsalim@ufmg.br (D.H.C.S.); caiocsm@ufmg.br (C.C.S.M.)

[2] Department of Geosciences, Universidade Federal de São João del-Rei, São João del-Rei 36301360, MG, Brazil; pereira@ufsj.edu.br

* Correspondence: camila@desa.ufmg.br

Abstract

Urban freshwater ecosystems affected by eutrophication and recurrent algal blooms require monitoring approaches capable of representing optical complexity and spatial heterogeneity. This study evaluated an integrated workflow combining proximal in situ hyperspectral sensing, radiometric calibration, spectral filtering, predictor-band selection, data transformation, and machine-learning regression to estimate chlorophyll-a (chl-a) in a tropical eutrophic urban reservoir. Monthly field campaigns were conducted from September 2022 to February 2023, with simultaneous chl-a measurements and hyperspectral image acquisition. After preprocessing, noise removal, and exclusion of anomalous spectra, 82 matched hyperspectral–chl-a observations were retained for model development. Predictor bands were selected using Pearson correlation and F-test analysis, identifying five relevant wavelengths: 530, 535, 682, 687, and 732 nm. Multiple Linear Regression, Random Forest Regressor, Support Vector Regressor, and XGBoost Regressor were tested under different data transformations. The Support Vector Regressor with logarithmic transformation achieved the best performance, with R^2 = 0.86 and RMSE = 6.89 μg L^{-1}. The selected wavelengths correspond to spectral regions associated with green reflectance, red chl-a absorption, and red-edge/NIR responses in productive waters. The results indicate that proximal hyperspectral sensing combined with machine learning can support chl-a estimation in optically complex urban reservoirs and provide complementary information for eutrophication monitoring and bloom-management strategies.

Keywords: proximal hyperspectral sensing; chlorophyll-a retrieval; machine learning regression; predictor-band selection; spectral preprocessing; Support Vector Regression; optically complex inland waters; eutrophic reservoirs

Check for updates

Academic Editor: Gabriel-Ionut Plavan

Received: 18 May 2026
Revised: 24 June 2026
Accepted: 29 June 2026
Published: 2 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Accurate chlorophyll-a (chl-a) monitoring in urban reservoirs remains a major challenge because these systems frequently combine eutrophication, recurrent algal blooms, high suspended-matter concentrations, colored dissolved organic matter, wastewater inputs, and strong spatial heterogeneity [1]. In such optically complex inland waters, the spectral signal recorded by sensors is not controlled by phytoplankton alone, but by the interaction among multiple optically active constituents and surface effects [2]. This complexity limits the direct transfer of algorithms developed for clearer waters or larger aquatic

systems and reinforces the need for locally adapted approaches that combine high-spectral-resolution measurements with robust empirical modelling [3].

Chlorophyll-a (chl-a) is one of the most widely used indicators of phytoplankton biomass and trophic state in aquatic ecosystems. As a photosynthetic pigment present in most phytoplanktonic organisms, chl-a plays a central role in light absorption within the water column and provides an indirect measure of algal abundance [1,4,5]. Its optical behavior is characterized by absorption features in the blue and red regions of the visible spectrum and reflectance peaks in the green and red-edge regions [3,6,7]. Therefore, chl-a is especially relevant for remote sensing applications because variations in its concentration modify the spectral response of the water surface.

However, estimating chl-a in inland waters remains challenging. Unlike clear oceanic waters, inland and transitional aquatic environments are often classified as optically complex waters, or Case-2 waters, because their spectral response is not controlled by phytoplankton alone [3,8]. In these systems, suspended sediments, colored dissolved organic matter, detrital material, cyanobacteria, and variable bottom or surface effects can interact with the radiation field and affect the spectral signal recorded by sensors [9,10]. As a result, algorithms developed for open waters or based on a limited number of spectral bands may perform poorly when applied to small, turbid, eutrophic, or urban reservoirs.

Traditional water quality monitoring remains essential, but it is often costly, labor-intensive, spatially sparse, and temporally discontinuous. Sampling campaigns usually represent only a limited number of points and dates, which may not capture the spatial heterogeneity and short-term dynamics of algal blooms, especially in reservoirs affected by wastewater inflows, hydrodynamic compartmentalization, and rapid biological responses. Remote sensing has therefore become an important complementary tool for water quality monitoring, enabling broader spatial coverage, repeated observations, and integration with statistical and computational models [3,11].

Hyperspectral remote sensing has particular advantages for aquatic applications because it captures narrow and contiguous spectral bands, allowing for detailed characterization of subtle absorption and reflectance features associated with optically active water constituents. Compared with multispectral sensors, hyperspectral systems can better represent the spectral complexity of inland waters and improve the identification of diagnostic wavelengths for chl-a, suspended solids, turbidity, colored dissolved organic matter, and other parameters [1,12,13]. Proximal or in situ hyperspectral sensing is especially useful for small water bodies, where satellite pixels may be affected by mixed land–water signals, atmospheric interference, adjacency effects, or insufficient spatial resolution.

The increasing availability of hyperspectral data has also expanded the use of machine learning algorithms for water quality retrieval. Machine learning models can capture nonlinear relationships between spectral reflectance and water quality parameters, which is particularly important in optically complex environments where the spectral signal results from multiple interacting constituents [7,14,15]. Algorithms such as Support Vector Regression, Random Forest, XGBoost, and neural networks have shown strong potential for estimating chl-a and other water quality variables from hyperspectral and multispectral datasets [5,12,13,16]. Nevertheless, model performance is highly dependent on the optical properties of the water body, the range of observed concentrations, the spectral preprocessing strategy, the selection of predictor bands, and the size and quality of the calibration dataset.

Despite these advances, important limitations remain, for example, many chl-a retrieval studies rely on satellite or UAV imagery with a limited number of spectral bands, which may not fully capture narrow absorption and red-edge features in productive inland waters [3,9,10]. In addition, machine-learning models are often evaluated mainly by predic-

tive accuracy, while less attention is given to spectral preprocessing, anomalous-spectrum removal, predictor-band interpretability, overfitting control, and the relationship between selected wavelengths and known bio-optical mechanisms [16]. Another limitation is that models calibrated for one water body may not be directly transferable to other reservoirs because optical properties vary according to trophic state, suspended sediments, cyanobacterial composition, hydrodynamics, and local sources of pollution [9]. Therefore, studies conducted in small eutrophic urban reservoirs are important not only for local monitoring, but also for testing how proximal hyperspectral data and machine-learning workflows can be structured under realistic field constraints and limited sample sizes.

In this context, eutrophic urban reservoirs provide a valuable area for evaluating hyperspectral and machine learning approaches. These systems combine high ecological relevance with strong optical complexity, frequent algal blooms, anthropogenic nutrient inputs, and marked spatial variability. Therefore, developing robust empirical models in such environments contributes not only to local management, but also to the broader methodological challenge of improving water quality monitoring in small and optically complex freshwater systems.

Thus, the present study evaluates an integrated proximal hyperspectral and machine-learning workflow for estimating chl-a concentration in a eutrophic tropical urban reservoir. The specific objectives were: (i) to process in situ hyperspectral images acquired under field conditions using radiometric calibration, spectral filtering, and anomalous-spectrum removal; (ii) to identify physically interpretable predictor bands associated with chl-a variability; (iii) to compare linear and nonlinear regression models under different data transformations; and (iv) to assess the practical potential and limitations of this approach for supporting water-quality monitoring in optically complex urban freshwater systems.

2. Materials and Methods

2.1. Study Area and Environmental Context

The study was conducted in the Pampulha Reservoir (Figure 1), commonly known as Pampulha Lake, located in the metropolitan region of Belo Horizonte, Minas Gerais, southeastern Brazil. The reservoir is embedded in a highly urbanized watershed of approximately 98 km^2, distributed between the municipalities of Contagem (55%) and Belo Horizonte (45%), with the entire water body located within Belo Horizonte. The reservoir has a surface area of approximately 2 km^2, an annual mean hydraulic retention time of 89 days, and depths ranging from 0.7 to 14 m, characterizing it as a small and shallow tropical urban reservoir [17].

The watershed is subdivided into eight sub-basins, among which the Ressaca and Sarandi streams are particularly relevant because they drain approximately 60% of the basin area and concentrate nearly 90% of the local population, exceeding 500,000 inhabitants [18]. The regional climate is classified as Cwa according to the Köppen climate classification, with a wet and warm season from October to March and a dry and cooler season from April to September. Mean annual precipitation is approximately 1524.5 mm [19].

The reservoir represents an optically complex freshwater system affected by intense urbanization, wastewater inputs, sedimentation, solid-waste discharge, and recurrent eutrophication, as documented by previous limnological and water-quality studies in the Pampulha watershed [2,20–22]. These pressures have altered the physical, chemical, and biological characteristics of the reservoir and have favored persistent algal blooms, often dominated by cyanobacteria adapted to hypereutrophic conditions. In the present study, this optical and spatial complexity was also evidenced by the field chl-a dataset, which ranged from 4.98 to 68.13 μg L^{-1} and showed marked monthly and spatial variability among sampling campaigns. The standard deviation reached 19.56 μg L^{-1} in September

2022 and 15.15 µg L^{-1} in January 2023, indicating strong heterogeneity in phytoplankton biomass across the reservoir. This environmental setting makes Pampulha Reservoir a suitable test site for evaluating proximal hyperspectral sensing and machine-learning methods under conditions of high optical complexity and spatial variability.

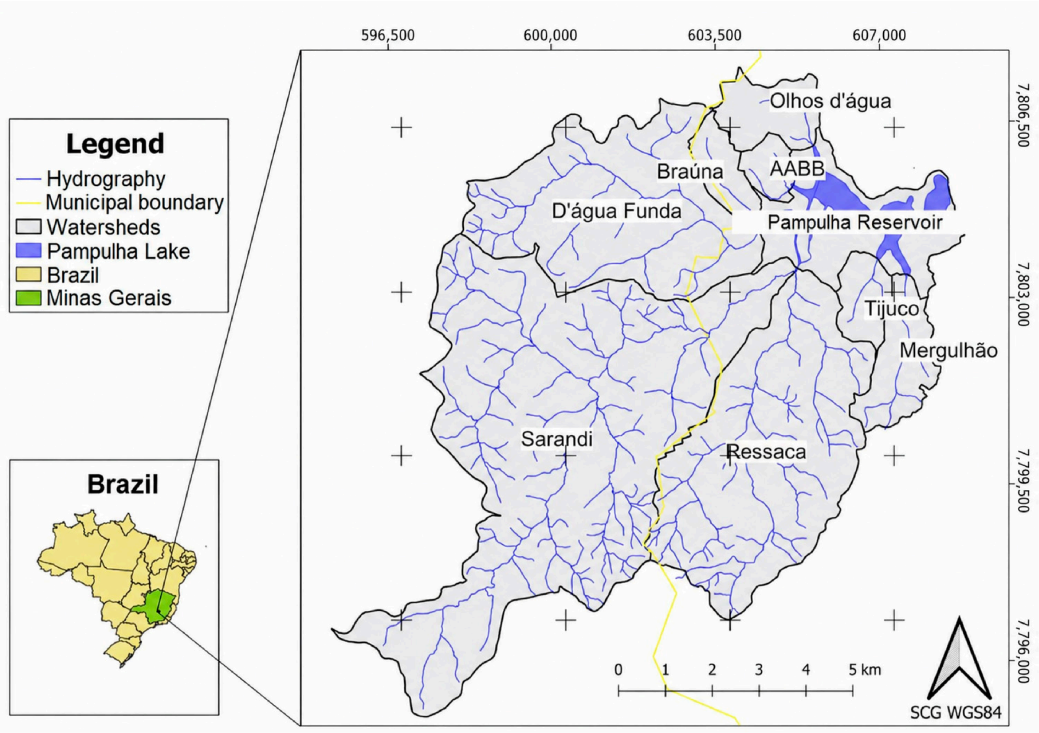


Figure 1. Location of the Pampulha Reservoir watershed and hydrographic network in Belo Horizonte, Minas Gerais, Brazil.

To provide a broader satellite-based context for the field campaigns, Sentinel-2 Level-2A surface reflectance imagery from the Harmonized Sentinel-2 MSI collection was used to derive the Normalized Difference Chlorophyll Index (NDCI) for the reservoir between 2017 and 2025 (Figure 2). Sentinel-2 MSI provides a red band centered at approximately 665 nm and a red-edge band centered at approximately 705 nm, and the Google Earth Engine harmonized surface reflectance collection provides atmospherically corrected imagery suitable for multitemporal analysis [23,24]. The index was calculated according to Equation (1):

$$\mathrm{NDCI} = (\mathrm{R705} - \mathrm{R665})/(\mathrm{R705} + \mathrm{R665}) \quad (1)$$

The NDCI formulation was originally proposed for remote estimation of chl-a in turbid and productive waters [25]. Sentinel-2-derived NDCI was used as a complementary satellite-based indicator of chlorophyll-related optical variability in the reservoir. In an exploratory step, NDCI values were converted to chl-a estimates using a literature-based relationship to provide spatial context for the reservoir-scale distribution of bloom-related conditions. However, these satellite-derived estimates were not locally calibrated or independently validated against the in situ dataset used for proximal hyperspectral modelling. Therefore, they were interpreted only as contextual spatial evidence and not as a substitute for the validated proximal hyperspectral machine-learning retrieval.

2.2. Field Campaigns and Chlorophyll-a Measurements

Field campaigns were carried out between September 2022 and February 2023, covering the main seasonal conditions observed in the region (Figure 3). Valid paired chl-a and hyperspectral observations used in the final modelling dataset were obtained for September,

October, and November 2022, and January and February 2023. December 2022 was not included in the final dataset because no valid matched hyperspectral–chl-a observations were retained after quality-control procedures. Chlorophyll-a (chl-a) concentration was measured in situ using an EXO2 multiparameter sonde (YSI) at 19 sampling points distributed across the reservoir. The spatial distribution of the sampling points was defined to represent the main limnological compartments of the water body, including fluvial zones near the main tributaries, the transition zone associated with the waste-retention curtain, and lacustrine conditions near the dam, thus designed to capture gradients associated with inflow areas, primary productivity, hydrodynamic compartmentalization, and the influence of urban drainage. These compartments are relevant because optically complex reservoirs may exhibit strong short-distance variability in phytoplankton biomass, suspended solids, colored dissolved organic matter, and surface conditions, which directly affect the relationship between spectral reflectance and chl-a concentration.

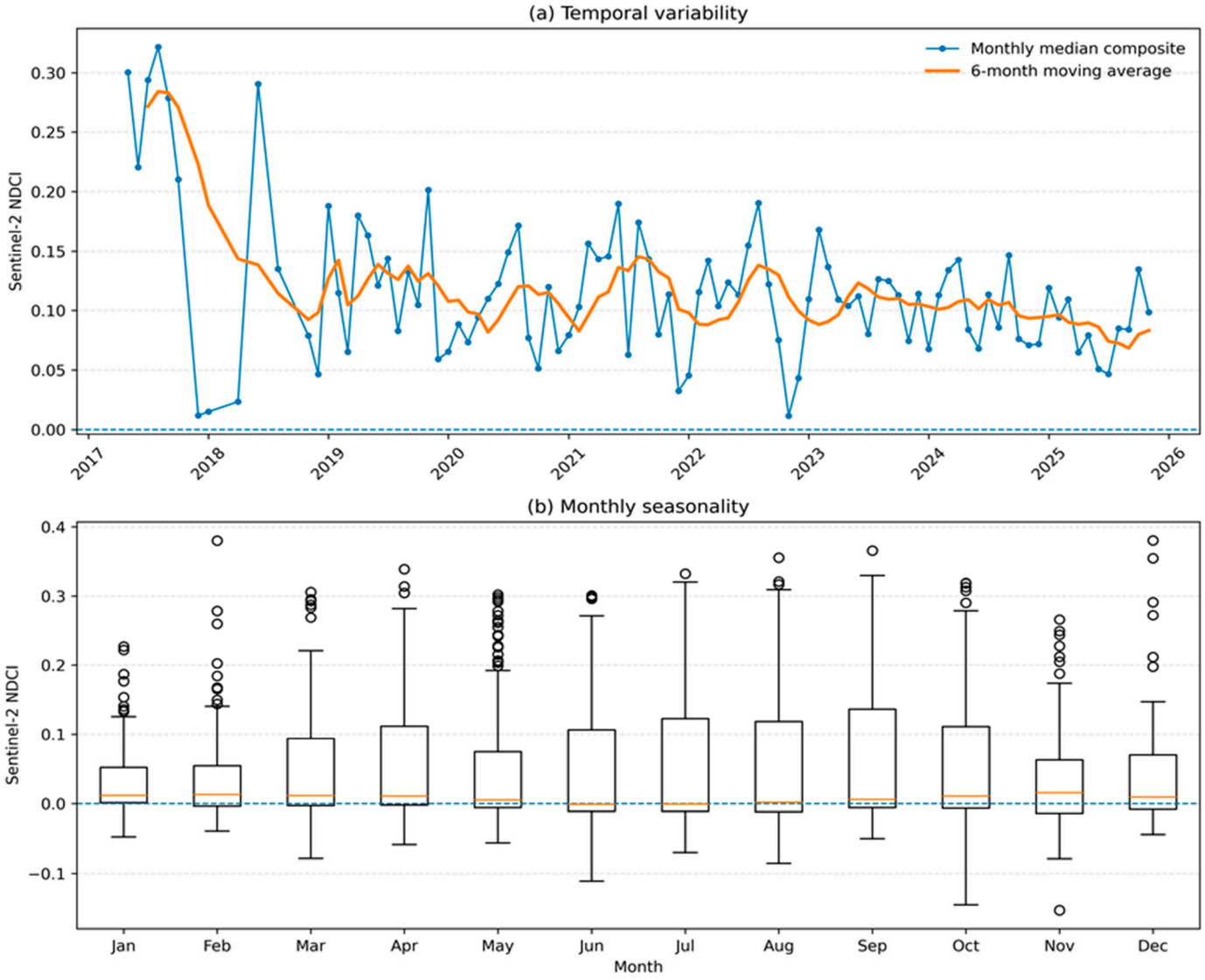


Figure 2. Sentinel-2-derived Normalized Difference Chlorophyll Index (NDCI) used as a satellite-based contextual proxy of algal-bloom variability in the Pampulha Reservoir. (**a**) Shows the monthly NDCI time series and its six-month moving average between 2017 and 2025. (**b**) Summarizes the monthly distribution of image-level NDCI values.

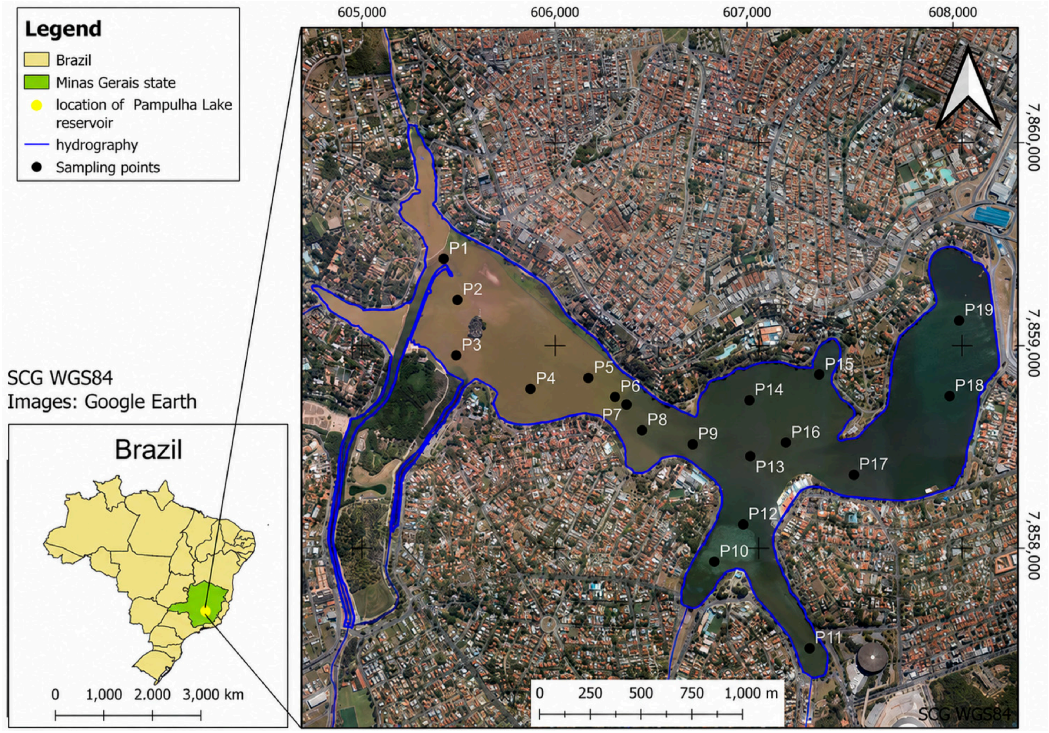


Figure 3. Distribution of in situ chl-a sampling points and hyperspectral image acquisition sites in the Pampulha Reservoir.

2.3. In Situ Hyperspectral Image Acquisition

Hyperspectral images were acquired using an OCI™-F Ultra-Compact Hyperspectral Imager (BaySpec Inc., San Jose, CA, USA), a compact CMOS-based sensor designed for high-resolution spectral imaging. The system records radiance information across narrow spectral bands, enabling detailed characterization of water-surface reflectance under proximal field conditions [26].

To minimize sun glint and specular reflection effects, the sensor was mounted vertically relative to the water column at approximately 1 m above the target, following acquisition strategies commonly adopted for proximal aquatic hyperspectral measurements [13,27]. Field acquisitions were performed during the morning, when solar illumination geometry reduces the probability that direct solar reflection enters the sensor field of view [28,29].

Chl-a measurements and hyperspectral image acquisition were performed simultaneously at each sampling point. This synchronization was adopted to reduce uncertainties related to water movement, wind-induced surface variability, and short-term changes in the spatial distribution of phytoplankton biomass. Such temporal matching is essential for empirical modelling because the predictive relationship depends on the direct correspondence between the measured water quality variable and the spectral signal recorded from the water surface [30].

Because the hyperspectral imagery was acquired from a proximal position directly above each sampling site, each retained spectrum was interpreted as a local water-surface observation associated with the corresponding in situ chl-a measurement. Thus, the modelling dataset was constructed using paired observations with close temporal and spatial correspondence between the spectral signal and the measured chl-a concentration. However, these proximal observations represent a much finer spatial support than Sentinel-2 pixels and should not be interpreted as spatially equivalent to satellite observations. For this reason, the proximal hyperspectral data were used for model calibration and validation, whereas Sentinel-2-derived products were used only as complementary reservoir-scale spatial context.

2.4. Radiometric Calibration and Spectral Preprocessing

The native CubeCreator 1.0 software provided by BaySpec Inc. was initially considered for processing the hyperspectral images. This software uses reference measurements to calibrate and correct reflectance values and can generate hyperspectral cubes containing complete spectral information for each pixel [31]. However, to avoid the loss of field information, independent computational routines were developed in R for radiometric calibration and spectral extraction. The procedure was based on white and dark reference images, using a white reflectance standard of 95%, to correct variations in ambient illumination and convert raw digital numbers into relative reflectance values. The radiometric calibration followed a normalization procedure commonly used in hyperspectral image processing [32,33]:

$$R(x,y) = [(RND(x,y) - D(x,y))/(W(x,y) - D(x,y))] \times 100 \quad (2)$$

where R(x,y) is the calibrated relative reflectance value at pixel position (x,y), RND(x,y) is the original raw image value, D(x,y) is the dark reference, and W(x,y) is the white reference. The resulting reflectance values were expressed as relative reflectance from 0% to 100%.

The original hyperspectral images had a spatial resolution of 1024 × 1280 pixels. In the raw image structure, each image column corresponded to a spectral band, resulting in 240 spectral samples. To reduce the signal-to-noise ratio and harmonize the spectral information for modelling, the spectral dimensionality was reduced to wavelengths between 413 and 955 nm, covering the visible and near-infrared regions. The final spectral dataset consisted of 109 bands with a spectral resolution of 5.1 nm. This reduction was obtained by averaging the reflectance over an approximately 11 nm window around the central wavelength of each band, following the approach described by [34].

After radiometric calibration and spectral reduction, noisy spectral regions were identified at the lower and upper wavelength limits. Such edge effects are common in hyperspectral images and may introduce uncertainty into water quality modelling [12]. Therefore, wavelengths between 400 and 470 nm and those above 850 nm were excluded from the modelling dataset because they showed higher noise incidence. After radiometric calibration and spectral reduction, additional quality-control criteria were applied to remove spectra that were not representative of water-surface reflectance. Spectra were excluded when they contained non-finite values, negative reflectance values after calibration, abnormally low responses associated with inadequate exposure, or excessive reflectance amplitudes inconsistent with the distribution of valid water spectra. Spectra affected by sharp isolated peaks or irregular shapes incompatible with the mean retained spectral behavior were also removed, as these patterns were interpreted as effects of sun glint, specular reflection, or unstable acquisition geometry. After the removal of noisy wavelength intervals and anomalous spectra, 82 valid hyperspectral–chl-a pairs were retained for predictor-band selection and machine-learning modelling [29].

2.5. Predictor-Band Selection and Machine Learning Modelling

The empirical modelling workflow was implemented in Python 3.10 and was designed to estimate chl-a concentration from the preprocessed hyperspectral reflectance data. Four regression approaches were evaluated: Multiple Linear Regression (MLR), Support Vector Regressor (SVR), Random Forest Regressor (RFR), and XGBoost Regressor (XGB). These models were selected to compare a traditional linear baseline with machine learning algorithms capable of modelling nonlinear relationships between spectral reflectance and water quality parameters.

The predictor-band selection procedure was designed to reduce the high redundancy and multicollinearity commonly observed among adjacent hyperspectral bands. First,

Pearson correlation analysis was applied to quantify the empirical association between each wavelength and measured chl-a concentration. Then, an F-test-based feature selection approach was used to identify wavelengths with higher explanatory power for chl-a variability. The final predictor set was defined by combining statistical relevance with spectral interpretability, retaining wavelengths that showed stronger association with chl-a while representing key bio-optical regions for eutrophic inland waters.

This procedure identified five predictor bands: 530, 535, 682, 687, and 732 nm. These wavelengths were subsequently used as input variables for all regression models to ensure comparability among algorithms. The selected bands correspond to spectral regions associated with green reflectance, red chl-a absorption/red-edge transition, and red-edge/NIR responses, which are particularly relevant in productive and optically complex inland waters (Table 1).

Table 1. Statistical and bio-optical interpretation of the selected hyperspectral predictor bands used for chl-a modelling.

Wavelength	Spectral Region	Bio-Optical Interpretation
530 nm	Green reflectance	Located in the green region, where chl-a absorption is comparatively lower and reflectance may increase due to phytoplankton-related scattering.
535 nm	Green reflectance	Represents the green reflectance domain commonly associated with algal biomass and productive waters.
682 nm	Red absorption/red-edge transition	Close to the red chl-a absorption feature and to the transition toward red-edge reflectance, making it sensitive to phytoplankton biomass.
687 nm	Red absorption/red-edge transition	Represents a narrow spectral interval associated with chl-a-related absorption and red-edge transition in productive waters.
732 nm	Red-edge/near-infrared	Associated with red-edge/NIR scattering responses, which may be influenced by algal biomass, cyanobacteria, turbidity, and suspended particles in optically complex waters.

Hyperparameter optimization was performed within the calibration subset before final validation. The modelling routine was implemented in Python and included Multiple Linear Regression, Random Forest Regressor, Support Vector Regressor, and XGBoost Regressor. The paired hyperspectral–chl-a dataset was divided into calibration and validation subsets. During calibration, literature-based hyperparameter ranges were tested for each machine-learning algorithm (Table 2), and the configurations producing lower RMSE values were selected for model construction. The calibration subset was then evaluated through repeated K-fold validation, using five folds and 100 repetitions, resulting in 500 performance estimates. The optimized models were subsequently applied to the independent validation subset, and model performance was assessed using R^2 and RMSE.

This validation-oriented procedure was adopted to reduce the risk of selecting models based only on training performance. This point is particularly relevant for small hyperspectral water-quality datasets, in which highly flexible algorithms may overfit the calibration data. Therefore, the final model comparison prioritized validation RMSE and validation R^2. This criterion supported the selection of the SVR model with logarithmic transformation, which achieved the best validation performance and showed better generalization capacity than the tree-based models.

The final model database was divided into calibration and validation subsets. From the 82 valid paired observations, 62 samples were assigned to calibration and 20 samples were used for validation. Within the model-selection procedure, a stratified repeated K-fold cross-validation strategy was applied. The dataset was divided into four groups, with 75%

of the data used for calibration and 25% for validation in each iteration. This procedure was repeated 2500 times across four iterations, corresponding to 10,000 Monte Carlo resampling runs. The best hyperparameter sets were selected based on the calibration R^2 and validation RMSE distributions, prioritizing model performance and internal generalization capacity within the available field dataset.

Table 2. Literature-based hyperparameter search space used for machine-learning model optimization.

Model	Hyperparameter	Search Range/Values Tested
Multiple Linear Regression	Not applicable	No hyperparameter tuning was applied
Support Vector Regressor	kernel	linear; polynomial; radial basis function (RBF)
Support Vector Regressor	C	0.1; 1; 10; 100
Support Vector Regressor	epsilon	0.01; 0.05; 0.1; 0.2; 0.5
Support Vector Regressor	gamma	scale; auto; 0.001; 0.01; 0.1; 1
Random Forest Regressor	n_estimators	100; 300; 500; 1000
Random Forest Regressor	max_depth	None; 5; 10; 20; 30
Random Forest Regressor	min_samples_split	2; 5; 10
Random Forest Regressor	min_samples_leaf	1; 2; 4
Random Forest Regressor	max_features	sqrt; log2; None
XGBoost Regressor	n_estimators	100; 300; 500; 1000
XGBoost Regressor	max_depth	2; 3; 4; 5; 6
XGBoost Regressor	learning_rate	0.01; 0.05; 0.1; 0.2
XGBoost Regressor	subsample	0.6; 0.8; 1.0
XGBoost Regressor	colsample_bytree	0.6; 0.8; 1.0
XGBoost Regressor	reg_lambda	1; 5; 10

2.6. Model Performance Assessment

Model performance was evaluated by comparing measured and predicted chl-a concentrations across the training, testing, and independent validation stages. The main performance metrics were the coefficient of determination (R^2) and the root mean square error (RMSE). R^2 was used to assess the proportion of variance in chl-a concentration explained by each model, whereas RMSE quantified the average prediction error in the same unit as the measured variable ($\mu g\ L^{-1}$).

The comparison among MLR, SVR, RFR, and XGB allowed evaluation of whether nonlinear machine-learning models improved chl-a retrieval relative to a linear regression baseline. Although training and testing metrics were used to assess model fitting and internal consistency during calibration, the independent validation metrics were used as the primary criterion for final model comparison. This decision was adopted because validation performance provides a more realistic assessment of the capacity of each model to predict unseen observations, which is particularly important in small hyperspectral water-quality datasets.

Scatterplots of measured versus estimated chl-a were generated for the best-performing models to assess agreement with the 1:1 line and to identify potential overestimation or underestimation patterns. The final model selection prioritized high validation R^2, low validation RMSE, and consistent generalization behavior between calibration and validation stages.

3. Results and Discussion

3.1. Satellite-Based Context and Hyperspectral Data Preprocessing

The Sentinel-2-derived NDCI series provides a satellite-based optical context for interpreting the field-based hyperspectral experiment (Figure 4a). The monthly NDCI dataset contained 96 valid monthly observations between May 2017 and November 2025, while

the image-level dataset included 2013 valid observations. The NDCI time series indicated persistent bloom-related optical variability in the reservoir, with monthly mean and median values of 0.116 and 0.111, respectively. These results support the interpretation that Pampulha Reservoir frequently exhibits spectral conditions associated with algal biomass variability, reinforcing the relevance of higher-spectral-resolution in situ measurements and machine-learning-based chl-a retrieval.

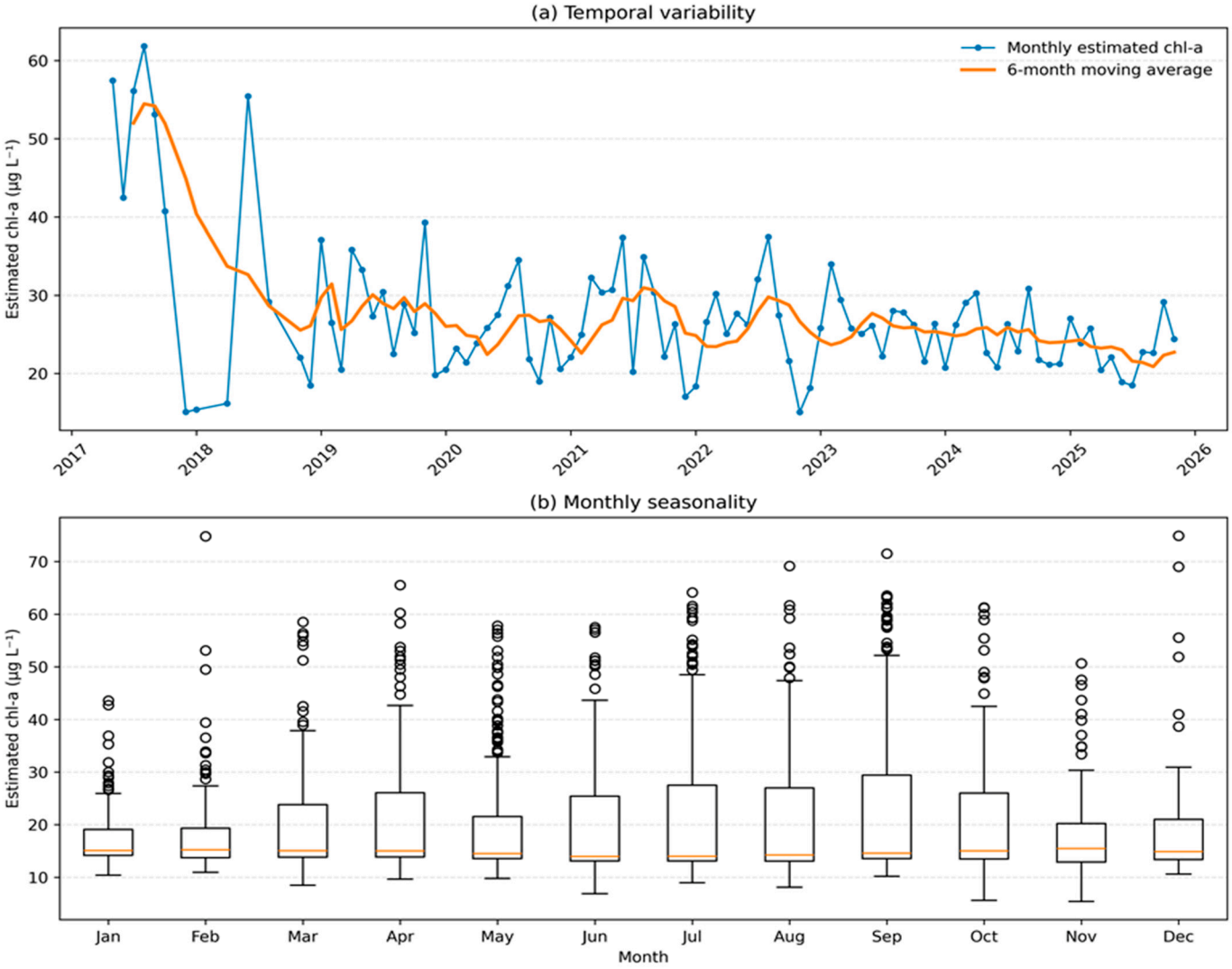


Figure 4. (**a**) Temporal variability of monthly estimated chl-a and its six-month moving average between 2017 and 2025. (**b**) Monthly seasonality of image-level estimated chl-a values.

To complement the NDCI analysis, a machine learning-based chl-a retrieval was produced to an exploratory Sentinel-2-derived analysis (Figure 4b). This product was not used to calibrate or validate the proximal hyperspectral model; instead, it was used to provide a broader temporal context for bloom-related conditions. The resulting monthly estimates ranged from approximately 15.1 to 61.9 µg L^{-1}, with a mean of 27.4 µg L^{-1} and a median of 26.0 µg L^{-1}. The seasonal distribution showed recurrent variability throughout the year, indicating that elevated bloom-related spectral responses may occur across multiple months rather than being restricted to a single fixed period.

The independent processing routine developed for the hyperspectral images allowed the extraction of spectral curves and the construction of matched reflectance–chl-a pairs. The initial spectral output contained 1280 bands per image. To reduce noise and improve model stability, the spectral dimensionality was reduced to 109 bands between approx-

imately 413 and 955 nm, with a spectral resolution of 5.1 nm. This step preserved the relevant visible and near-infrared regions for chl-a retrieval while reducing redundant and noisy information.

The computational routine generated 107 complete pairs of hyperspectral reflectance and chl-a concentration (Figure 5). After the spectral quality-control procedure described in Section 2.4, 82 valid hyperspectral–chl-a pairs were retained for model development. The retained spectra exhibited a coherent optical response for eutrophic waters, with higher reflectance in the green region, a marked response around the red absorption/red-edge transition, and increased reflectance toward the red-edge/NIR region.

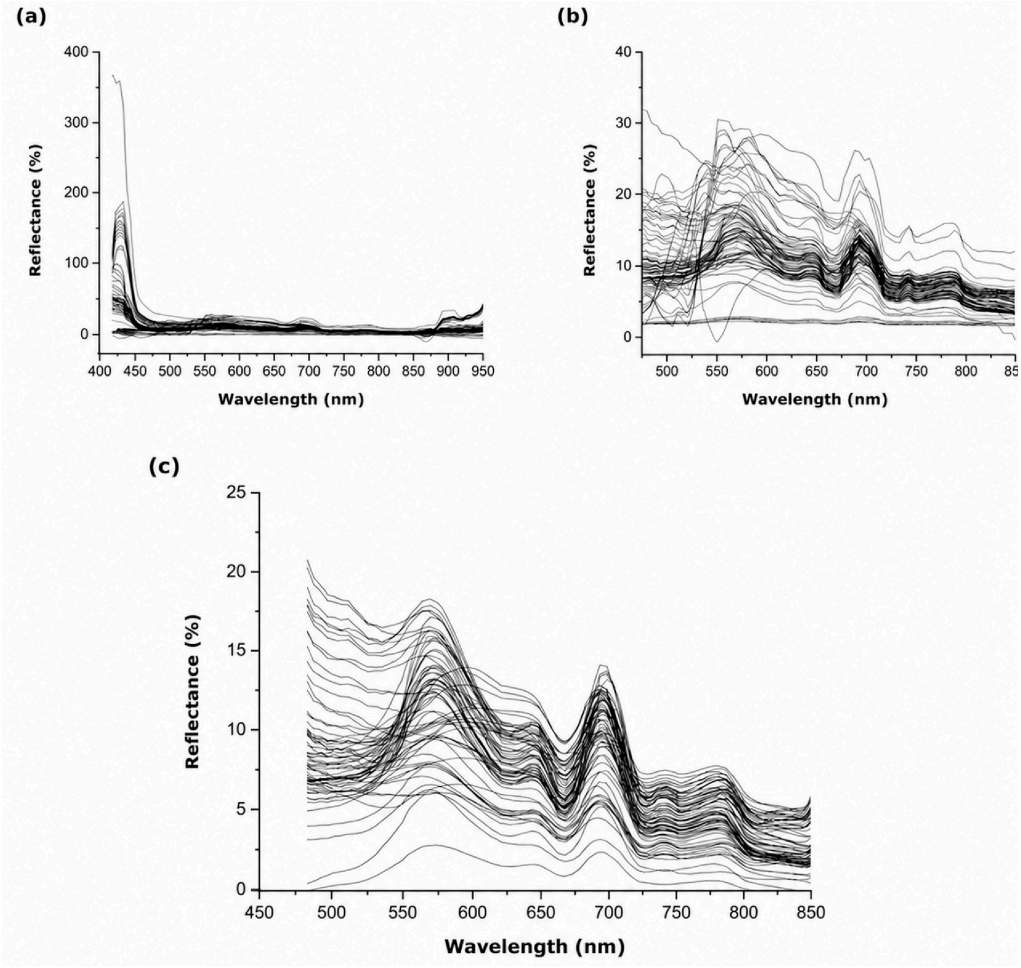


Figure 5. Hyperspectral reflectance spectra after successive processing steps: (**a**) original spectral curves affected by noise at the spectral extremes; (**b**) spectra after removal of noisy wavelength ranges; and (**c**) final spectra retained after exclusion of anomalous observations.

The retained spectra exhibited a coherent optical response for eutrophic waters. The final spectral curves show higher reflectance in the green region, a marked absorption feature in the red region, and a reflectance increase around the red-edge/NIR transition. This pattern is consistent with the spectral behavior expected for waters with high phytoplankton biomass and reinforces the suitability of the remaining spectra for empirical chl-a modeling.

3.2. Predictor-Band Selection and Machine-Learning Model Performance

The predictor-band selection procedure, based on Pearson correlation and F-test analysis, identified five hyperspectral bands with the strongest association with chl-a concentration: 530, 535, 682, 687, and 732 nm (Figure 6), representing spectral regions linked

to green reflectance, red absorption, and red-edge/near-infrared responses in eutrophic and optically complex waters. These bands were used as input variables for all regression models to ensure comparability among algorithms.

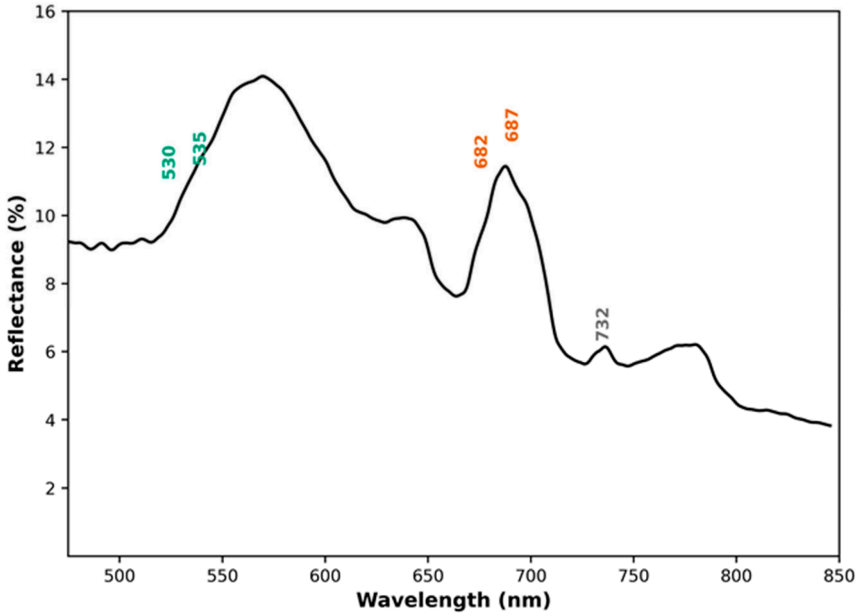


Figure 6. Mean hyperspectral reflectance spectrum of the 82 matched observations used for modeling, with the five selected predictor bands highlighted.

The selected wavelengths are physically meaningful for chl-a retrieval and are consistent with the statistical predictor-band selection procedure based on Pearson correlation and F-test analysis. The bands at 530 and 535 nm are located in the green reflectance region, where pigment absorption is comparatively lower and scattering by phytoplankton cells contributes to increased water-leaving reflectance. This spectral region has been repeatedly associated with chl-a detection in eutrophic and optically complex inland waters [5,11,13,35].

The bands at 682 and 687 nm are close to the red absorption domain of chl-a and to the transition between strong absorption and red-edge reflectance. This interval is particularly sensitive to phytoplankton biomass in productive waters and is widely used in spectral indices and empirical models for chl-a estimation [1,3,27,35,36]. In the present study, the selection of these red-region bands confirms that the hyperspectral signal captured diagnostic absorption-related information associated with chl-a variability.

The selection of the 732 nm band is also relevant. Although this wavelength is often associated with turbidity and suspended particles, it can contribute to chl-a estimation in turbid and eutrophic inland waters because algal biomass itself acts as a particulate component affecting scattering and near-infrared reflectance. In cyanobacteria-dominated systems, the relationship between chl-a, turbidity, and particle-driven optical responses may strengthen the predictive value of red-edge and NIR bands [3,36,37]. Therefore, the selected band set reflects not only pigment absorption, but also the broader optical complexity of urban eutrophic waters.

The chl-a dataset used for model development showed relevant spatial and temporal variability among the field campaigns. Concentrations ranged from 4.98 to 68.13 μg L^{-1}, with monthly medians between 20.02 and 55.79 μg L^{-1} (Table 3). The highest median was observed in October 2022, although this campaign had a smaller number of valid samples due to acquisition problems. In contrast, September 2022 and January 2023 exhibited

larger dispersion, indicating stronger spatial heterogeneity in chl-a concentration across the reservoir.

Table 3. Descriptive statistics of chl-a concentrations (μg L^{-1}) measured simultaneously with hyperspectral image acquisition.

Sampling Month	Median	Minimum	Maximum	Standard Deviation	Variance
September 2022	20.02	10.87	62.35	19.56	382.65
October 2022	55.79	45.97	68.13	6.07	36.89
November 2022	31.81	15.42	54.37	10.51	110.56
January 2023	37.80	4.98	46.23	15.15	229.79
February 2023	34.92	20.53	61.69	14.54	211.44

Although the final modeling dataset did not include the extremely high chl-a values reported in previous monitoring studies for the same reservoir, it still covered a concentration gradient sufficient to evaluate the relationship between spectral reflectance and algal biomass. Such variability is important for machine-learning calibration because models require a representative range of target values to learn the spectral response associated with chl-a changes [35].

Thus, the model-development routine tested different combinations of regression algorithms and data transformations (Table 4). Overall, the results indicate that nonlinear machine-learning models outperformed Multiple Linear Regression, confirming that the spectral response of chl-a in optically complex waters cannot be adequately represented by simple linear relationships alone.

The comparative ranking of the best-performing configuration of each model is summarized in Figure 7. This synthesis highlights the superior performance of SVR, which combined the highest validation R^2 with the lowest validation RMSE among the tested approaches. Multiple Linear Regression showed the weakest performance, with validation R^2 values between 0.17 and 0.19 and validation RMSE values above 16 μg L^{-1}. This result indicates that, for this dataset, a purely linear multivariate formulation was not sufficient to represent the interaction between hyperspectral reflectance and chl-a concentration. This limitation is expected in Case-2 waters, where reflectance is affected by phytoplankton, suspended particles, colored dissolved organic matter, and surface effects.

Random Forest Regressor and XGBoost Regressor achieved intermediate validation performance. The best Random Forest result was obtained with the exponential transformation, reaching validation R^2 = 0.68 and RMSE = 10.29 μg L^{-1}. XGBoost showed validation R^2 values between 0.57 and 0.64, with the best RMSE of 10.82 μg L^{-1} under logarithmic transformation. However, XGBoost produced R^2 = 1.00 during training across all transformations, while validation performance remained moderate. This pattern indicates severe overfitting, probably associated with the combination of a small calibration dataset, a limited number of validation samples, repeated use of highly correlated spectral predictors, and the high flexibility of gradient-boosting decision trees.

It is important to emphasize that, for model comparison and performance assessment, the validation metrics were used as the main criterion because this stage evaluates model predictions on unseen data and therefore provides a more realistic assessment of predictive capacity. In this sense, training metrics alone do not represent the actual predictive performance of the developed model, especially when highly flexible algorithms are applied to small datasets. Although hyperparameter tuning and repeated resampling were used to improve internal generalization, the results suggest that XGBoost fitted idiosyncratic patterns of the calibration subset rather than stable spectral–chl-a relationships. Therefore, XGBoost

was not selected as the preferred model, despite its perfect training performance. This finding reinforces the importance of prioritizing validation metrics over training metrics when comparing model performance, especially in small hyperspectral water-quality datasets.

Table 4. Performance of the tested models and data transformations for chl-a estimation from in situ hyperspectral reflectance.

Model	Transformation	R^2 Training	R^2 Validation	Validation RMSE (µg L^{-1})
Multiple Linear Regression	Linear	0.16	0.19	16.26
Multiple Linear Regression	Polynomial	0.17	0.19	16.35
Multiple Linear Regression	Exponential	0.08	0.17	16.49
Multiple Linear Regression	Logarithmic	0.01	0.18	16.42
Multiple Linear Regression	Min–max	0.16	0.19	16.29
Random Forest Regressor	Linear	0.91	0.62	11.19
Random Forest Regressor	Polynomial	0.90	0.65	10.65
Random Forest Regressor	Exponential	0.89	0.68	10.29
Random Forest Regressor	Logarithmic	0.91	0.66	10.57
Random Forest Regressor	Min–max	0.91	0.66	10.64
Support Vector Regressor	Linear	0.82	0.78	8.53
Support Vector Regressor	Polynomial	0.78	0.73	9.33
Support Vector Regressor	Exponential	−0.57	−0.41	21.55
Support Vector Regressor	Logarithmic	0.76	0.86	6.89
Support Vector Regressor	Min–max	0.77	0.82	7.65
XGBoost Regressor	Linear	1.00	0.62	11.13
XGBoost Regressor	Polynomial	1.00	0.61	11.32
XGBoost Regressor	Exponential	1.00	0.57	11.94
XGBoost Regressor	Logarithmic	1.00	0.64	10.82
XGBoost Regressor	Min–max	1.00	0.62	11.13

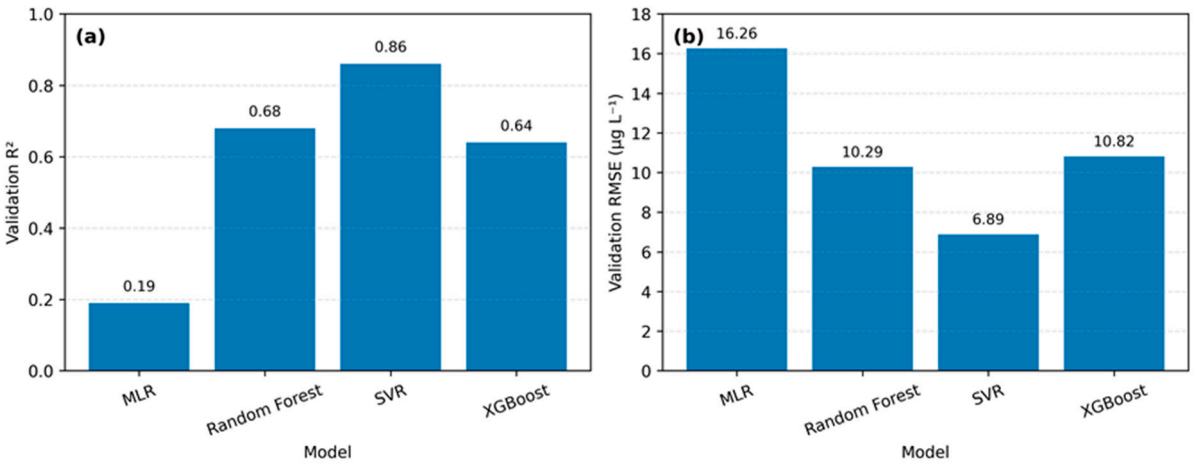


Figure 7. Comparative performance of the best configuration of each empirical and machine-learning model used for chl-a estimation: (**a**) shows validation *R*2, and (**b**) shows validation RMSE. The SVR model with logarithmic transformation achieved the highest predictive performance.

The Support Vector Regressor provided the best overall performance. The logarithmic transformation produced the highest validation accuracy, with R^2 = 0.86 and RMSE = 6.89 µg L^{-1}. The min–max transformation also performed well, with R^2 = 0.82 and RMSE = 7.65 µg L^{-1}. In contrast, the exponential transformation resulted in negative validation R^2, indicating that this transformation was unsuitable for the spectral–chl-a relationship in the present dataset. The superior performance of SVR suggests that this model was better able to capture nonlinear patterns while maintaining generalization capacity under limited sample conditions (Figure 8).

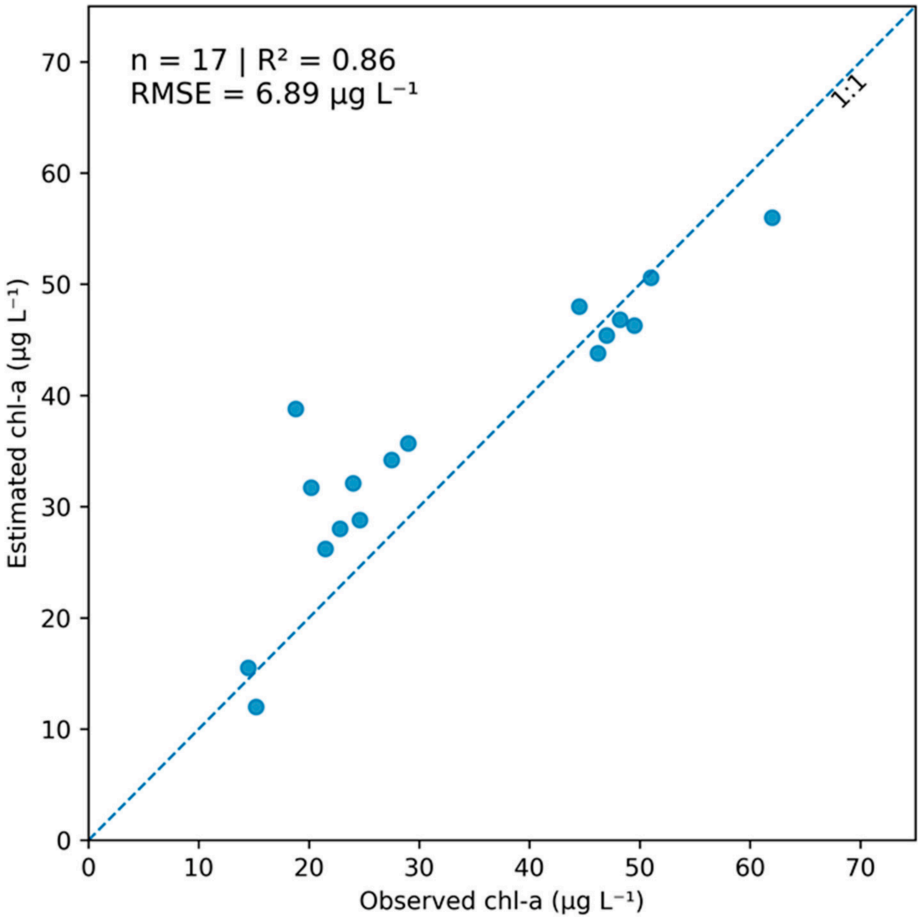


Figure 8. Relationship between observed and SVR-estimated chl-a concentrations for the validation dataset. The dashed line indicates the 1:1 relationship.

The observed-versus-estimated plot confirms the general agreement between measured and predicted chl-a values. Most validation samples were distributed close to the 1:1 line, although two samples showed larger deviations. These deviations suggest that the model may still be sensitive to specific optical conditions or localized water-quality gradients. Nevertheless, the overall dispersion pattern supports the quantitative metrics and indicates that the SVR model captured most of the chl-a variability represented in the validation set.

The better performance of the logarithmic SVR configuration can be interpreted from both statistical and bio-optical perspectives. From a statistical perspective, the logarithmic transformation likely reduced the influence of higher chl-a concentrations and compressed the variability of the response variable. This effect is relevant because chl-a concentrations in eutrophic reservoirs are often unevenly distributed, with localized high values associated with bloom patches and short-distance spatial heterogeneity. By reducing the leverage of high-concentration observations, the logarithmic transformation may have produced

a more stable relationship between the selected hyperspectral bands and measured chl-a, improving the capacity of the model to generalize to the validation dataset.

From a bio-optical perspective, the relationship between reflectance and chl-a in optically complex inland waters is inherently nonlinear. The spectral signal is not controlled only by phytoplankton pigments, but also by the combined effects of pigment absorption, phytoplankton scattering, cyanobacteria, suspended particles, colored dissolved organic matter, surface reflection, and local acquisition conditions. These interactions can modify the intensity and shape of reflectance features in the green, red, and red-edge/NIR regions, especially around the wavelengths selected in this study. Therefore, a logarithmic transformation is consistent with the expected nonlinear and heteroscedastic behavior of spectral–chl-a relationships in eutrophic Case-2 waters.

The SVR algorithm was particularly suitable under these conditions because it can model nonlinear relationships while controlling prediction errors through a regularized optimization structure. This is important for relatively small hyperspectral datasets, in which highly flexible tree-based models may fit the calibration subset too closely. In this study, the logarithmic SVR configuration provided the best balance between model flexibility and generalization capacity, whereas XGBoost showed evidence of overfitting, with perfect training performance but lower validation accuracy. Thus, the superior performance of SVR with logarithmic transformation likely resulted from the combination of variance stabilization in the chl-a response and the ability of SVR to represent nonlinear spectral relationships without excessive model flexibility.

The results are consistent with previous studies showing the potential of SVR and other machine-learning algorithms for water-quality retrieval from hyperspectral data. Peterson et al. [38] also identified SVR as a high-performing model for chl-a estimation in lakes, while Sagan et al. [3] and Keller et al. [12] demonstrated the usefulness of machine learning for estimating chl-a and other optically active constituents from proximal hyperspectral measurements [39]. The performance obtained here is therefore compatible with the broader literature, even though the dataset was collected in a small, shallow, urban, and optically complex reservoir.

Compared with tree-based models, the better performance of SVR may be related to the size of the calibration dataset. Random Forest and XGBoost can perform well with complex nonlinear data, but they may require larger and more diverse training datasets to fully exploit their flexibility. In the present study, XGBoost clearly fitted the training data too closely, while SVR achieved a better balance between model complexity and predictive generalization. This finding is relevant for environmental monitoring programs, where high-quality paired spectral and laboratory or probe-based observations are often limited.

From a spectral perspective, the model selected bands that are consistent with known chl-a optical features and with the behavior of productive inland waters [40]. The combination of green, red, and red-edge/NIR bands indicates that the model relied on multiple components of the spectral response: reduced absorption and enhanced scattering in the green region, pigment absorption in the red region, and particle-related reflectance near the red edge. This combination is particularly important for eutrophic urban waters, where chl-a retrieval is influenced not only by phytoplankton pigments but also by cyanobacterial blooms, turbidity, and suspended matter.

The study also reinforces the operational value of proximal hyperspectral sensing. In small urban reservoirs, satellite-based monitoring may be constrained by mixed pixels, atmospheric correction uncertainties, adjacency effects, cloud cover, and revisit time. In situ hyperspectral acquisition can complement conventional monitoring by increasing spectral detail and providing spatially distributed information under controlled field conditions [41,42]. When coupled with machine-learning models, this approach can support

more frequent assessments of algal biomass and improve the detection of water-quality changes that may be missed by sparse point-based sampling.

Despite these promising results, some limitations must be acknowledged. First, the final dataset contained 82 matched observations, which is adequate for exploratory empirical modeling but still limited for broad generalization. Second, the model was calibrated and validated within a single reservoir, and its transferability to other optically complex systems remains to be tested. Third, although spectral quality control was essential, future applications should formalize objective thresholds for excluding anomalous spectra and should evaluate how these decisions affect model uncertainty. Finally, independent external validation using new campaigns, additional seasons, or other reservoirs would strengthen the operational applicability of the proposed approach.

Overall, the findings demonstrate that proximal hyperspectral sensing combined with machine learning is a robust and promising strategy for chl-a estimation in optically complex urban freshwaters. The approach is especially valuable for monitoring eutrophic systems where algal blooms, cyanobacteria, suspended particles, and high spatial heterogeneity challenge conventional remote-sensing algorithms. The results support the development of intelligent, high-frequency water-quality monitoring frameworks that integrate field spectroscopy, automated preprocessing, and data-driven modeling.

3.3. Spatial Interpretation, Operational Relevance, and Limitations

To provide a broader spatial context for the proximal hyperspectral and machine-learning results, annual Sentinel-2-derived chlorophyll-a estimates were mapped for the Pampulha Reservoir from 2019 to 2025 (Figure 9). The maps revealed a spatial heterogeneity within the reservoir throughout the analyzed period, indicating that chlorophyll-related conditions were not uniformly distributed across the water body. Higher estimated chlorophyll-a values were generally concentrated in portions of the central and western/northwestern arms, whereas comparatively lower values were more frequently observed in parts of the eastern sector and in some narrow marginal areas.

The interannual comparison also showed important temporal variability. The years 2019 and 2021 presented more extensive areas with higher estimated chlorophyll-a values, suggesting stronger chlorophyll-related signals across larger portions of the reservoir. In contrast, 2022, 2024, and 2025 showed a greater dominance of moderate to lower estimated values, although localized zones of higher chlorophyll-a were still observed. Therefore, the spatial patterns shown in Figure 9 support the relevance of integrating in situ/proximal hyperspectral measurements with remote sensing products to improve the interpretation of chlorophyll-a dynamics in urban eutrophic reservoirs.

The spatial patterns observed in the Sentinel-2-derived maps are consistent with the environmental gradients previously described for the Pampulha Reservoir. The upstream sector, located near the inflow of the Ressaca and Sarandi streams, is strongly affected by urban runoff, external nutrient inputs, suspended-solid loading, and persistent eutrophic conditions [2,18,43]. These inputs help explain why chlorophyll-related signals tend to be spatially heterogeneous and more variable in the western and northwestern sectors of the reservoir, where pollutant loads from the watershed and shallow-water conditions favor phytoplankton development.

The interpretation of these spatial patterns is also consistent with previous studies reporting recurrent cyanobacterial blooms and strong phytoplankton responses to environmental drivers in the Pampulha Reservoir [21,22]. In this sense, the spatial variability shown in Figure 9 should not be interpreted only as a satellite-derived pattern, but as an expression of the reservoir's known limnological compartmentalization and chronic eutrophic condition. The persistence of bloom-prone areas reinforces the importance of

monitoring chlorophyll-related variability at spatial scales finer than those provided by sparse conventional sampling.

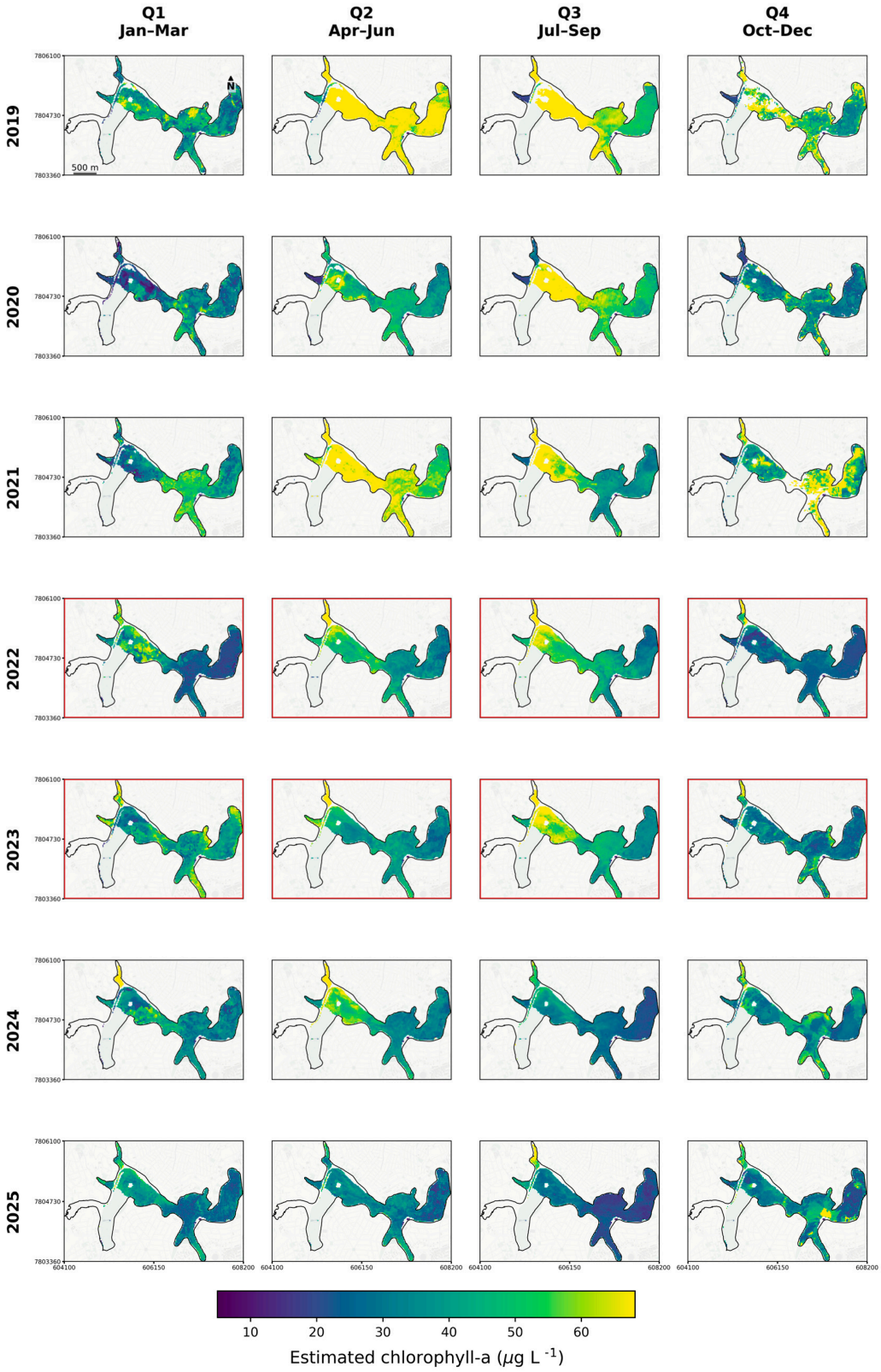


Figure 9. Annual spatial variability of Sentinel-2-derived chlorophyll-a estimates in the Pampulha Reservoir from 2019 to 2025. Each panel represents the annual mean spatial pattern calculated from quarterly Sentinel-2 composites. Red outlines highlight the years associated with the field-campaign period used in the proximal hyperspectral sensing and machine-learning experiment.

The waste-retention curtain may also contribute to the spatial organization of water-quality conditions. Previous local analyses indicated that the curtain separates the reservoir into two distinct environments, with poorer water quality upstream and comparatively lower pollution loads downstream [44]. Therefore, the spatial discontinuities observed in chlorophyll-related conditions may reflect not only gradual longitudinal gradients, but also physical and hydrodynamic compartmentalization within the reservoir.

Cluster analysis previously applied to the primary water-quality dataset identified four groups in the reservoir, including a sector located immediately downstream of the waste-retention curtain that was not directly represented by the official monitoring network [44]. This finding suggests that the spatial complexity of the Pampulha Reservoir may require additional monitoring points to better capture local transitions in water quality, especially near the curtain and in sectors influenced by lateral tributary inputs.

These findings reinforce the operational relevance of integrating satellite-based spatial context, proximal hyperspectral measurements, and machine-learning models for water-quality monitoring in small urban eutrophic reservoirs. While Sentinel-2-derived products provide a broader spatial perspective, proximal hyperspectral sensing can complement conventional monitoring by providing high-spectral-resolution observations under controlled field conditions [41,42]. Therefore, the combined use of remote sensing and proximal hyperspectral modelling may support more detailed monitoring of algal biomass and improve the detection of spatially localized water-quality changes in optically complex urban reservoirs.

4. Conclusions

This study demonstrated the potential of proximal in situ hyperspectral sensing combined with machine learning for estimating chlorophyll-a (chl-a) in an optically complex tropical urban reservoir. By integrating radiometric calibration, spectral filtering, anomalous-spectrum removal, predictor-band selection, data transformation, and regression modelling, the proposed workflow provided a structured methodological framework for chl-a retrieval under realistic field conditions.

Among the tested models, the Support Vector Regressor with logarithmic transformation achieved the best predictive performance, with $R^2 = 0.86$ and RMSE = 6.89 μg L^{-1}. This result indicates that nonlinear learning approaches can better represent the relationship between chl-a concentration and hyperspectral reflectance in optically complex inland waters than multiple linear regression. The comparatively weaker performance of Random Forest and XGBoost, particularly the overfitting observed for XGBoost, highlights the need for careful model selection and validation when working with small but information-rich hyperspectral datasets.

The selected predictor bands at 530, 535, 682, 687, and 732 nm are consistent with key spectral regions associated with phytoplankton biomass, including green reflectance, red chl-a absorption, and red-edge/NIR responses. Their selection confirms that narrow hyperspectral bands can capture bio-optical features relevant to chl-a retrieval even under high turbidity, cyanobacterial dominance, and mixed optical interference from suspended matter and dissolved organic substances.

From a practical management perspective, the proposed approach can complement conventional water-quality monitoring by providing spatially distributed and spectrally detailed information on algal biomass in eutrophic urban reservoirs. Such information may support the identification of bloom-prone areas, guide targeted field sampling, improve the interpretation of satellite-based products, and assist reservoir managers in monitoring eutrophication dynamics and evaluating mitigation actions.

https://doi.org/10.3390/limnolrev26030032

Nevertheless, some limitations should be acknowledged. The final dataset contained 82 matched hyperspectral–chl-a observations, which is suitable for exploratory empirical modelling but still limited for broad generalization. The model was calibrated and validated within a single reservoir, and its transferability to other optically complex systems remains to be tested. In addition, the exclusion of noisy and anomalous spectra was necessary to ensure spectral quality, but future studies should formalize objective thresholds for quality control and evaluate their influence on model uncertainty.

Future research should expand the temporal and spatial sampling design, include additional hydrological periods and trophic conditions, test the model in other reservoirs, and evaluate complementary dimensionality-reduction and spectral-transformation approaches, such as derivative spectroscopy and principal component analysis. External validation with independent campaigns will be essential to assess operational robustness and transferability. Overall, the results support the use of proximal hyperspectral sensing and machine learning as complementary tools for high-frequency, spatially explicit monitoring of eutrophication and algal-bloom dynamics in small, optically complex inland waters.

Author Contributions: Conceptualization, T.A.F., G.P. and C.C.A.; methodology, T.A.F., G.P. and C.C.A.; software, T.A.F. and G.P.; validation, T.A.F., B.T.A.S., D.H.C.S., C.C.S.M., G.P. and C.C.A.; formal analysis, T.A.F. and G.P.; investigation, T.A.F., B.T.A.S., D.H.C.S. and C.C.S.M.; resources, C.C.A.; data curation, T.A.F., B.T.A.S., D.H.C.S. and C.C.S.M.; writing—original draft preparation, T.A.F. and G.P.; writing—review and editing, T.A.F., B.T.A.S., D.H.C.S., C.C.S.M., G.P. and C.C.A.; visualization, T.A.F. and G.P.; supervision, C.C.A.; project administration, C.C.A.; funding acquisition, C.C.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Companhia Energética de Minas Gerais (CEMIG), Project GT-0607, titled "Intelligent Water Quality Monitoring through the Development of a Photo-optical Algorithm"; by the ANP/PETROBRAS (AquaSmart Project PT-200.20.00248); by the National Council for Scientific and Technological Development (CNPq), grant numbers 306373/2023-8 and 312467/2025-7; and by the Research Support Foundation of the State of Minas Gerais (FAPEMIG), grant number APQ-01056-21.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

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

References

1. Chen, L.; Liu, L.; Liu, S.; Shi, Z.; Shi, C. The Application of Remote Sensing Technology in Inland Water Quality Monitoring and Water Environment Science: Recent Progress and Perspectives. *Remote Sens.* **2025**, *17*, 667. [CrossRef]
2. Silva, T.F.G.; Vinçon-Leite, B.; Lemaire, B.J.; Petrucci, G.; Giani, A.; Figueredo, C.C.; de Oliveira Nascimento, N. Impact of urban stormwater runoff on cyanobacteria dynamics in a tropical urban lake. *Water* **2019**, *11*, 946. [CrossRef]
3. 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]
4. Esteves, F.A. *Fundamentos de Limnologia*; Interciência: Rio de Janeiro, Brazil, 2011; p. 790.
5. Xia, X.; Lei, S.; Lu, H.; Xu, Z.; Li, X.; Chen, X.; Hong, N.; Xu, J.; Shi, K.; Huang, J. Remote Sensing of Particle Absorption Coefficient of Pigments Using a Two-Stage Framework Integrating Optical Classification and Machine Learning. *Remote Sens.* **2025**, *17*, 1756. [CrossRef]
6. Patel, A.K.; Albarico, F.P.; Perumal, P.K.; Vadrale, A.P.; Nian, C.; Chau, H.T.B.; Anwar, C.; Wani, H.M.; Pal, A.; Saini, R.; et al. Algae as an emerging source of bioactive pigments. *Bioresour. Technol.* **2022**, *351*, 126910. [CrossRef] [PubMed]
7. Batina, A.; Krtalic, A. Integrating Remote Sensing Methods for Monitoring Lake Water Quality: A Comprehensive Review. *Hydrology* **2024**, *11*, 92. [CrossRef]
8. Cook, K.V.; Beyer, J.E.; Xiao, X.; Hambright, K.D. Ground-based remote sensing provides alternative to satellites for monitoring cyanobacteria in small lakes. *Water Res.* **2023**, *242*, 120076. [CrossRef] [PubMed]

9. He, J.; Chen, Y.; Wu, J.; Stow, D.A.; Christakos, G. Space-time chlorophyll-a retrieval in optically complex waters that accounts for remote sensing and modeling uncertainties and improves remote estimation accuracy. *Water Res.* **2020**, *171*, 115403. [PubMed]
10. Gai, Y.; Yu, D.; Zhou, Y.; Yang, L.; Chen, C.; Chen, J. An improved model for chlorophyll-a concentration retrieval in coastal waters based on UAV-Borne Hyperspectral imagery: A case study in Qingdao, China. *Water* **2020**, *12*, 2769.
11. Lu, X.; Situ, C.; Ma, W.; Dai, X.; He, C.; Li, L.; Zhou, L. Comparative analysis of four semi-analytical models for estimating chlorophyll-a concentration in case-2 waters using field hyperspectral reflectance. *Int. J. Remote Sens.* **2020**, *41*, 584–594.
12. Keller, S.; Maier, P.M.; Riese, F.M.; Norra, S.; Holbach, A.; Börsigh, N.; Wilhelms, A.; Moldaenke, C.; Zaake, A.; Hinz, S. Hyperspectral data and machine learning for estimating CDOM, chlorophyll a, diatoms, green algae and turbidity. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1881. [CrossRef] [PubMed]
13. Cui, Y.; Fei, M.; Fu, P.; Yang, X.; Zhang, Y.; Liu, P. Application of hyperspectral analysis of chlorophyll-a concentration inversion in Nansi Lake. *Ecol. Inform.* **2021**, *64*, 101360.
14. 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]
15. Cao, Q.; Yu, G.; Sun, S.; Dou, Y.; Li, H.; Qiao, Z. Monitoring water quality of the Haihe River based on ground-based hyperspectral remote sensing. *Water* **2022**, *14*, 22.
16. Lu, L.; Gong, Z.; Liang, Y.; Liang, S. Retrieval of chlorophyll-a concentrations of Class II water bodies of inland lakes and reservoirs based on ZY1-02D satellite hyperspectral data. *Remote Sens.* **2022**, *14*, 1842. [CrossRef]
17. Plec, D.F.; Silva, T.G.F.; Viçon-Leite, B.; Nascimento, N. Thermal functioning of a tropical reservoir assessed through three-dimensional modelling and high-frequency monitoring. *Rev. Bras. Recur. Hidr.* **2021**, *26*, e15.
18. Furtado, A.P.F.V.; Monte-Mor, R.C.A.; Couto, E.A. Evaluation of reduction of external load of total phosphorus and total suspended solids for rehabilitation of urban lakes. *J. Environ. Manag.* **2021**, *296*, 113339. [CrossRef]
19. CBHSF. *Resumo Executivo do Plano de Recursos Hídricos da Bacia Hidrográfica do Rio São Francisco 2016–2025*; Comitê da Bacia Hidrográfica do Rio São Francisco: Belo Horizonte, Brazil, 2016; p. 300.
20. Campos, M.O. Fatores Que Influenciam a Distribuição Espacial do Fitoplâncton na Lagoa da Pampulha—BH, MG. Master's Thesis, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, 2010; p. 75.
21. Batista, A.M.M.; Figueiredo, C.C.; Giani, A. Variability in a permanent cyanobacterial bloom: Species-specific responses to environmental drivers. *FEMS Microbiol. Ecol.* **2018**, *94*, fiy197. [CrossRef]
22. Barçante, B.; Nascimento, N.O.; Silva, T.F.G.; Leandro, A.R.; Giani, A. Cyanobacteria dynamics and phytoplankton species richness as a measure of waterbody recovery: Response to phosphorus removal treatment in a tropical eutrophic reservoir. *Ecol. Indic.* **2020**, *117*, 106702. [CrossRef]
23. European Space Agency. *Sentinel-2 User Handbook; ESA Standard Document*; European Space Agency: Paris, France, 2015. Available online: https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook (accessed on 12 May 2026).
24. Google Earth Engine. Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A Surface Reflectance. Available online: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED (accessed on 12 May 2026).
25. Mishra, S.; Mishra, D.R. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sens. Environ.* **2012**, *117*, 394–406. [CrossRef]
26. Riba, J. Application of image sensors to detect and locate electrical discharges: A review. *Sensors* **2022**, *22*, 5886. [CrossRef] [PubMed]
27. Wang, Z.; Kawamura, K.; Sakuno, Y.; Fan, X.; Gong, Z.; Lim, J. Retrieval of chlorophyll-a and total suspended solids using iterative stepwise elimination partial least squares (ISE-PLS) regression based on field hyperspectral measurements in irrigation ponds in Higashihiroshima, Japan. *Remote Sens.* **2017**, *9*, 264. [CrossRef]
28. Kay, S.; Hedley, J.D.; Lavender, S. Sun glint correction of high and low spatial resolution images of aquatic scenes: A review of methods for visible and near-infrared wavelengths. *Remote Sens.* **2009**, *1*, 697–730. [CrossRef]
29. Chong, W.S.; Zaki, N.H.M.; Hossain, M.S.; Muslin, A.M.; Pour, A.B. Introducing Theil-Sen estimator for sun glint correction of UAV data for coral mapping. *Geocarto Int.* **2021**, *37*, 4527–4556. [CrossRef]
30. Su, T.C. A study of a matching pixel by pixel (MPP) algorithm to establish an empirical model of water quality mapping, based on unmanned aerial vehicle (UAV) images. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *58*, 213–224. [CrossRef]
31. Rossiter, T.; Furey, T.; McCarthy, T.; Stengel, D.B. UAV-mounted hyperspectral mapping of intertidal macroalgae. *Estuar. Coast. Shelf Sci.* **2020**, *242*, 106789. [CrossRef]
32. Noviyanto, A.; Abdulla, W.H. Segmentation and calibration of hyperspectral imaging for honey analysis. *Comput. Electron. Agric.* **2019**, *159*, 129–139. [CrossRef]
33. Shaikh, M.S.; Jaferzadeh, K.; Thornberg, B.; Casselgren, J. Calibration of a hyperspectral imaging system using a low-cost reference. *Sensors* **2021**, *21*, 3738. [PubMed]

34. Khezrabad, M.A.; Zoej, M.J.V.; Safdarinezhad, A.A. A new approach for geometric correction of UAV-based pushbroom images through the processing of simultaneously acquired frame images. *Measurement* **2021**, *199*, 111431.
35. Cheng, C.; Wei, Y.; Lu, G.; Xu, N. Remote sensing estimation of chlorophyll-a concentration in Taihu Lake considering spatial and temporal variations. *Environ. Monit. Assess.* **2019**, *191*, 84. [CrossRef] [PubMed]
36. Yang, H.; Kong, J.; Hu, H.; Du, Y.; Gao, M.; Chen, F. A review of remote sensing for water quality retrieval: Progress and challenges. *Remote Sens.* **2022**, *14*, 1770. [CrossRef]
37. Murugan, P.; Sivakumar, R.; Pandiyan, R.; Annadurai, M. Comparison of in-situ hyperspectral and Landsat ETM+ data for chlorophyll-a mapping in Case-II water (Krishnarajapuram Lake, Bangalore). *J. Indian Soc. Remote Sens.* **2016**, *44*, 949–957. [CrossRef]
38. 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 feature-level and decision-level fusion with proximal remote sensing. *Photogramm. Eng. Remote Sens.* **2019**, *85*, 269–280.
39. Jeong, B.; Lee, S.; Heo, J.; Lee, J.; Lee, M.-J. Deep Learning-Based Retrieval of Chlorophyll-a in Lakes Using Sentinel-1 and Sentinel-2 Satellite Imagery. *Water* **2025**, *17*, 1718. [CrossRef]
40. Pan, W.; Lin, C.; Zhong, L.; Ye, Z. Cross-Water-Body Validation of Chlorophyll-a Retrieval Using Synergistic UAV Hyperspectral and Satellite Multispectral Data in Eutrophic Inland Waters. *Water* **2026**, *18*, 159. [CrossRef]
41. Goodrich, S.; Schaeffer, B.; Meyers, K.; Salls, W.; King, T.; Seegers, B.; Reif, M. Sentinel-2 for chlorophyll-a water quality monitoring: A review of validation evidence and application potential. *Int. J. Remote Sens.* **2026**, *47*, 3820–3845. [CrossRef] [PubMed]
42. Ngwenya, N.; Bangira, T.; Sibanda, M.; Gurmessa, S.K.; Mabhaudhi, T. UAV-based remote sensing of chlorophyll-a concentrations in inland water bodies: A systematic review. *Geocarto Int.* **2025**, *40*, 2452246. [CrossRef]
43. Bezerra, M.P.; Viana, E.A.P.; Brandão, L.P.M.; McGinnis, D.F.; Bezerra-Neto, J.F.; Barbosa, F.A.R. Water quality evaluation and dissolved organic matter characterization of a tropical hypereutrophic reservoir and its streams treated with Phoslock® and microbial bioremediation Enzilimp®. *Environ. Sci. Pollut. Res.* **2022**, *29*, 1375–1390. [CrossRef] [PubMed]
44. Figueiredo, T.A. Determinação de Clorofila-A Utilizando Sensor Hiperespectral e Análise da Qualidade da Água da Lagoa da Pampulha-MG. Master's Thesis, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, 2023.

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.