# OPTIMIZED CHLOROPHYLL-A ESTIMATION AND TROPHIC STATE MAPPING IN COASTAL AQUACULTURE ZONES: A SENTINEL-2 AND IN SITU SPECTRAL APPROACH IN THAI BINH, VIETNAM

HOAN, N. T.[1,2] – PHIEN, T. D.[1,2*] – DUNG, N. V.[1,2] – QUYNH, H. T.[1] – VIET, N. Q.[3]

[1]*Institute of Earth Sciences, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Nghia Do, Hanoi, Vietnam*

[2]*Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Nghia Do, Hanoi, Vietnam*

[3]*University of Sciences, Hue University, 77 Nguyen Hue, Hue, Vietnam*

**Corresponding author*
*e-mail: phienrs@ies.vast.vn*

(Received 25th Jul 2025; accepted 28th Oct 2025)

**Abstract.** Coastal aquaculture in Thai Binh province, Vietnam, is economically vital, yet its environmental sustainability hinges on effective water quality monitoring. Chlorophyll-a (Chl-a) is a critical indicator of aquatic ecosystem health, making its accurate assessment essential for managing these dynamic environments. This study developed an optimized model for estimating Chl-a concentrations in coastal aquaculture areas, leveraging in situ Chl-a data, field spectral measurements, and Sentinel-2 MultiSpectral Instrument (MSI) imagery. Various evaluation metrics, including the coefficient of determination ($R^2$), root mean square error (RMSE), and bias, were utilized to assess the models' performance. Relationships between in situ Chl-a and various spectral indices were established, and an optimized Multiple Linear Regression (MLR) model was derived. This model was subsequently applied to Sentinel-2 imagery to generate spatial Chl-a and Trophic State Index (TSI) distribution maps. The optimized Chl-a model demonstrated high performance ($R^2 > 0.94$, low RMSE and bias), confirming Sentinel-2's significant potential as a free, high-spatial-resolution tool for Chl-a research. The resulting Chl-a and TSI maps provide critical insights into the spatial heterogeneity and trophic states of coastal aquaculture zones, facilitating the identification of areas susceptible to eutrophication. These findings enhance the understanding of aquatic environmental dynamics and support sustainable aquaculture practices.
**Keywords:** *eutrophication, water quality monitoring, TSI, optimized Chl-a model, multiple linear regression*

## Introduction

Eutrophication is recognized as the most significant ecological threat to the health of aquatic ecosystems globally, serving as a primary driver of water quality degradation (Chislock et al., 2013; Kakade et al., 2021; Khan et al., 2014; Vinh et al., 2022). This phenomenon arises from increased nutrient inputs, which foster excessive algal and macroalgal growth (Prepas and Charette, 2003). Although phytoplankton blooms are natural components of aquatic ecosystems (Carstensen et al., 2015), their excessive proliferation can severely reduce water clarity and deplete dissolved oxygen levels (Prepas and Charette, 2003). These blooms impair light penetration and nutrient access for submerged aquatic vegetation, often leading to hypoxia and the subsequent death of aquatic fauna (Mishra et al., 2022). This cascade, in turn, results in widespread degradation across the aquatic ecosystem. Furthermore, these blooms are frequently dominated by cyanobacteria, which can produce toxins (Sivonen, 1996). Therefore, to effectively address these issues, it is crucial to regularly monitor the eutrophication and algal blooms over space and time.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

Chlorophyll a (Chl-a), the main photosynthetic pigment in phytoplankton, serves as an effective parameter for monitoring the trophic status of aquatic systems. In addition, Chl-a is the main indicator of phytoplankton abundance and biomass (Moses et al., 2009; Woźniak et al., 2014; Zhang et al., 2014) and can be commonly used to assess the water clarity, water quality and eutrophication level of a water body (Salem et al., 2017). Recent trends reveal that elevated eutrophication levels in aquatic environments have been attributed to a combination of factors, including climate change (J. Wang et al., 2015) and excessive human-induced aquaculture activities (Du et al., 2024; Duan et al., 2009; Guo et al., 2017; Lao et al., 2023a, b; Qin et al., 2022). Anthropogenic activities accelerate both the rate and spatial extent of nitrogen and phosphorus loading into aquatic environments. This increased nutrient loading consequently leads to the proliferation of algae in inland water bodies, which is directly reflected in an increase in chlorophyll-a (Chl-a) concentration (Smith, 2003; Wang et al., 2018). Given these dynamics, continuous monitoring of Chl-*a* concentration and comprehensive studies on water eutrophication are indispensable for understanding and managing current ecological conditions.

Traditional water quality monitoring based on in situ point sampling and laboratory techniques have been used for decades and provide accurate, localized measurements of chlorophyll-a at specific times and locations (Jang et al., 2024; Ogashawara et al., 2021). However, conventional Chl-a assessment strategies are inherently labor-intensive, costly, time-consuming, and lack the synoptic capability to capture algal bloom conditions across broad spatial and temporal scales (Bertone et al., 2024; Duan et al., 2010; Li et al., 2024; Ogashawara et al., 2021). Consequently, there has been an increasing demand for robust and effective approaches to quantify chlorophyll-a concentration across various aquatic systems. To address this monitoring requirement, satellite-based techniques have been implemented and demonstrated as a crucial tool for aquatic environment monitoring, including chlorophyll-a (Chl-a) surveillance (Akbarnejad Nesheli et al., 2024). Remote sensing techniques offer several advantages for monitoring chlorophyll-a (Chl-a) concentration. Specifically, satellite-derived Chl-a provides substantial spatial coverage, surpassing the limitations of in situ measurements and enabling data acquisition from inaccessible aquatic environments (Ogashawara et al., 2021; Salls et al., 2024). Furthermore, remote sensing facilitates the reconstruction of historical trends through past satellite imagery and offers rapid assessment of ecological water quality status compared to laboratory-based techniques (Ha et al., 2017; Ogashawara et al., 2021).

Moderate resolution sensors, including the Medium Resolution Imaging Spectrometer (MERIS) and Sentinel-3 OLCI (300 m), along with the Moderate Resolution Imaging Spectroradiometer (MODIS) (250-1000 m), have been extensively utilized for inland Chl-a estimation. These sensors are characterized by their frequent data acquisition capabilities (1-3 days) and the inclusion of critical red and near-infrared spectral bands, making them particularly well-suited for Chl-a detection in complex waters and at elevated Chl-a concentrations (Salls et al., 2024). While effective in large inland lakes (>1000 m wide), the limited spatial resolution of these sensors restricts their utility in smaller lakes (≤1000 m wide) due to potential contamination from stray light (Hestir et al., 2015; Salls et al., 2024). To overcome these spatial limitations, efforts have focused on finer-scale sensors originally designed for terrestrial applications, such as the Landsat series (30 m) (Chen et al., 2024; Duan et al., 2007; Kuhn et al., 2019; Vinh et al., 2022). However, these finer-scale sensors often lack the specific spectral bands required to quantify the Chl-a red edge peak (Salls et al., 2024). A promising advancement in remote sensing for water quality monitoring is offered by the Sentinel-2A and 2B (S2)


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

MultiSpectral Instrument (MSI) sensors. Launched in 2015 and 2017, these sensors provide a spatial resolution of 10-60 m and a combined rapid revisit frequency of 5 days. Significantly, among their 13 spectral bands covering the blue to shortwave infrared (SWIR) spectrum, three bands are dedicated to the red-edge region. Additionally, Sentinel-2 features improved 12-bit radiometric digitization, which is particularly beneficial for applications in low-radiance aquatic environments. These attributes make S2 particularly valuable for monitoring water quality in small and shallow water bodies, such as ponds and urban lakes (Jang et al., 2024; Li et al., 2024; Ogashawara et al., 2021; Quang Vinh et al., 2024; Salls et al., 2024; Tóth et al., 2021).

Remote sensing methods for determining Chlorophyll-a (Chl-a) concentration can be broadly categorized into three main groups: physical-based, empirical, and machine learning (ML) algorithms. Physical-based algorithms leverage bio-optical models to simulate spectral radiance based on specific water constituents; however, despite their high sensitivity, they are computationally demanding. While offering promising accuracy and robustness, machine learning inherently carries the risk of overfitting during model development (Cao et al., 2020; Chen et al., 2024; Li et al., 2021; Tran and Liou, 2022, 2024). Meanwhile, empirical algorithms utilize statistical techniques, including regression, to correlate Chl-a levels with spectral reflectance data from the sensors' narrow bands or various band ratios. Despite their generally lower sensitivity when compared to bio-optical models, the straightforward development and ease of application associated with empirical methods frequently make them a more viable option for practical deployment (Akbarnejad Nesheli et al., 2024; Matthews, 2011). Given the constraint of a relatively small dataset from field measurements (41 samples), this research primarily employed empirical methods over other recognized methods for determining Chlorophyll-a (Chl-a) concentration, prioritizing practicality over the computational demands of physical-based algorithms or the overfitting risks of machine learning. This decision was driven by a practical balance between model performance, data constraints, and complexity.

Focusing on empirical methods for estimating Chl-a from remote images, the specific techniques are band ratios/combinations, multi-band indices, and Multiple Linear Regression (MLR) Johansen et al. (2022) observed that a significant proportion of empirically based remote sensing algorithms could be distilled into a limited number of general formulas. Methods including the Normalized Difference Chlorophyll Index (NDCI), two-band algorithms, three-band algorithms, Chlorophyll Index (CI), and Maximum Chlorophyll Index (MCI) are particularly notable. For decades, these algorithms have been instrumental in deriving Chl-a concentrations from a wide array of satellite imagers (Johansen et al., 2024). Simultaneously, multiple linear regression (MLR) offers distinct advantages due to its inherent simplicity and interpretability (Tran and Liou, 2024). This model characteristic facilitates an intuitive understanding of inter-variable relationships, thereby providing clear insight into how independent variables exert influence on the dependent variable (Jang et al., 2024; Kim et al., 2020; Prieto et al., 2017). Furthermore, the incorporation of diverse band ratio combinations has been shown in previous investigations to significantly enhance the performance capability of Chl-a estimation models in complex aquatic environments (Kim et al., 2016; Matus-Hernández et al., 2018).

The Coastal Region of Thai Binh Province, Vietnam, which is the focus of this study, encompasses numerous small ponds primarily used for aquaculture. As aquaculture is one of the most prominent and rapidly developing economic sectors and a significant aquaculture hub in Thai Binh province, investigating eutrophication is critical and


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

providing early warnings is helpful in preventing fish mortality due to oxygen depletion. Consequently, the Sentinel-2 satellite data is well-suited for this purpose, offering appropriate spatial, temporal, and spectral resolutions. The primary objective of this research to develop an optimized model for Chl-a estimation in the Coastal Region of Thai Binh Province, Vietnam. To achieve this, the research focused on establishing correlations between in situ Chl-a measurements and a variety of spectral band ratios and multi-band indices. Subsequently, multiple linear regression (MLR) was employed to effectively combine these parameters and derive an optimal Chl-a estimation model. This optimized model was then applied to Sentinel-2 MSI imagery, captured concurrently with the in situ survey, to generate spatial Chl-a and Trophic State Index (TSI) distribution maps for the study region.

## Data and methods

### *Study area*

Comprising the Tien Hai and Thai Thuy districts, the study region is located in the eastern part of Thai Binh province, Vietnam, characterized by a long coastline bordering the East Sea (*Fig. 1*). This region holds a strategic geographical position, offering significant potential for marine economic development, while also providing favorable conditions for agriculture and, notably, aquaculture. With a combined total area of approximately 499 km², this research region is home to around 470,000 people (https://danso.info/dan-so-thai-binh/). The relatively high population density reflects the concentration of economic and social activities here. The study region is characterized by a tropical monsoon climate, heavily influenced by its coastal proximity. The average annual temperature is about 23.4∘C, creating warm conditions conducive to various types of production. Annual rainfall is quite high, ranging from 1600 to 1700 mm, primarily concentrated in the summer months (Tran et al., 2017). An average of 1600 to 1700 sunshine hours per year provides ample energy for crops and other economic activities (Tran et al., 2017). Winters here are typically colder and drier, lasting from November of the previous year to April of the following year. Notably, the study area also boasts significant mangrove forest coverage, playing a crucial role in coastal dike protection, erosion control, and climate regulation for the entire region.

As a prominent and rapidly developing economic sector in the research area, aquaculture renders it a key aquaculture hub in Thai Binh province. The total area dedicated to aquaculture is extensive, covering thousands of hectares, including diverse types such as freshwater, brackish water, and clam farming. The region particularly focuses on promoting high-tech shrimp farming, implementing advanced models like circular pond systems to boost productivity while minimizing disease risks. Proactive management in the production and supply of local seedstock is also a crucial factor, ensuring the quality of fry and reducing production costs. Thanks to its favorable natural conditions and investment in development, the aquaculture sector has significantly contributed to the region's overall agricultural output, providing stable incomes and improving the livelihoods of local residents. Given our research area's prevalence of small ponds used for fish and shrimp farming, eutrophication studies are crucial. These studies facilitate early warning systems, thereby preventing fish mortality due to oxygen depletion. Consequently, Sentinel-2 satellite data is well-suited for this research, offering appropriate spatial, temporal, and spectral resolutions.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

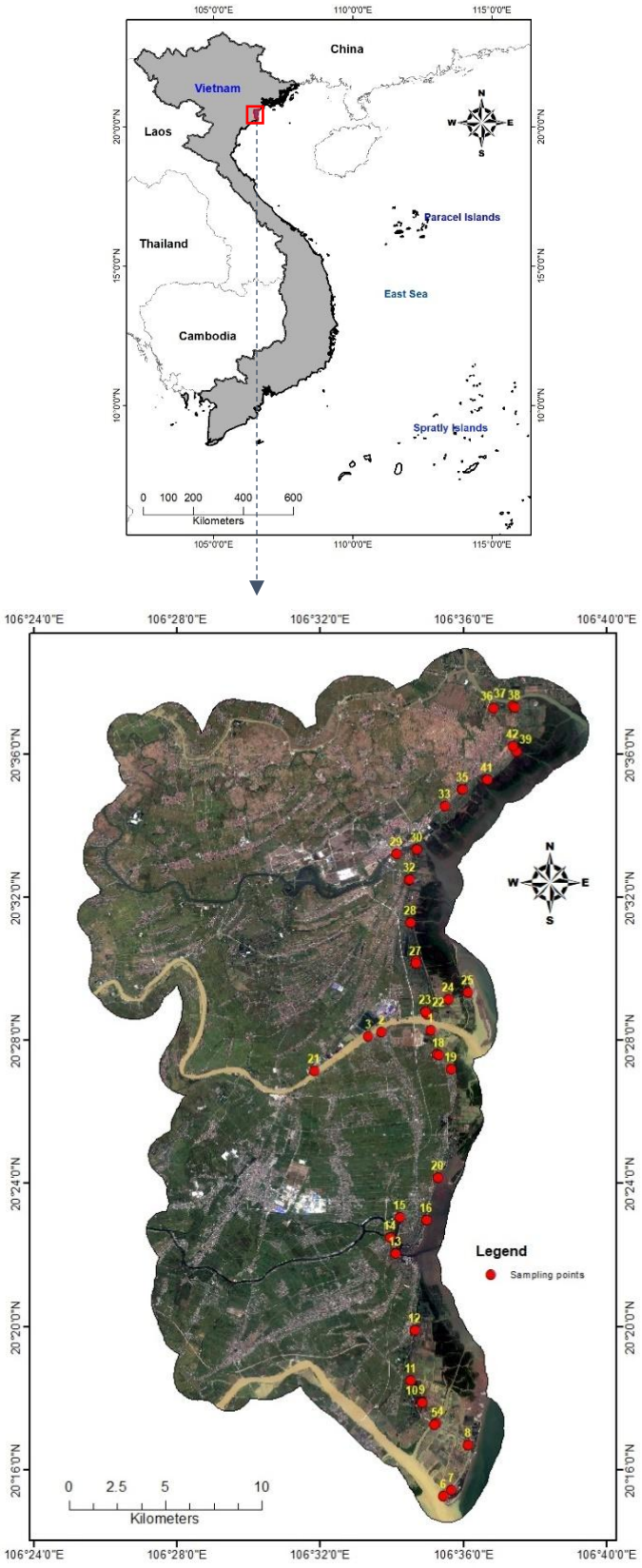


***Figure 1.*** *Location of the study area and field survey sampling points*

APPLIED ECOLOGY AND ENVIRONMENTAL RESEARCH 23(6):11541-11569.
http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

### *Field spectral measurements and water sampling data*

Forty-two in situ samples were obtained from ponds and lakes within the study area over a three-day period (Sep 30 – Oct 2, 2024) (*Fig. 1*). However, an anomaly was identified following the evaluation of sample TB22's spectral reflectance curve. Consequently, to ensure the highest level of accuracy, data from this sample (TB22) will be excluded from all subsequent calculations.

#### *In situ hyperspectral data from the spectroradiometer*

At each measurement point, field spectra were collected with a handheld spectroradiometer (PSR-2500), operating within a spectral range from 350 nm to 2500 nm, at a 1 nm spectral resolution (Manufacturer: Spectral Evolution, USA).

Calibration and Spectral Range: Prior to each measurement, white light balancing was performed to ensure accuracy and to serve as a radiometric calibration step for the spectroradiometer. Besides, at each field site, the water reflectance spectrum was measured three times and averaged to reduce variability. Although the PSR-2500 spectroradiometer is capable of measuring spectra from 350 nm to 2500 nm, at a 1 nm spectral resolution, only the 400–900 nm range was used in this study to align with Sentinel-2's spectral bands, effectively filtering out unusable data.

Spectral Resampling: The recorded spectral data were then transferred to a computer and processed using specialized software to analyze the reflectance spectra of different water bodies within the study area. A crucial step was the spectral resampling (simulation) of Sentinel-2 reflectances from field spectra. Specifically, spectroradiometer-measured reflectances (in situ hyperspectral data) were spectrally averaged based on the central waveband locations and wavelength ranges of the Sentinel-2 bands (e.g., Band 2, 3, 4, 5, 6, 7, 8, and 8a) to simulate the surface reflectances ($R_{rs}$) that the Sentinel-2 satellite sensor would record. This process transforms the high-resolution field data into Synthetic Sentinel-2 surface reflectances that are directly comparable to the satellite imagery.

Exclusion Criteria for Spectral Points: rigorous exclusion criteria were applied to maintain data quality. Field-measured spectral points were visually and statistically screened for quality assurance. Any data point exhibiting extreme spectral abnormalities (e.g., sudden spikes or drops inconsistent with known water properties) or displaying substantial discrepancies compared to the majority of observations was excluded from the dataset. This step ensured that only representative and high-quality in situ data were utilized.

#### *Laboratory analyses for determination of chlorophyll-a concentrations*

At each sampling site, water samples were collected in clean bottles for laboratory analysis. In this study, chlorophyll-a (Chl-a) was the primary chemical parameter selected for measurement and evaluation. Chlorophyll-a (Chl-a) in water samples was calculated using the standard methodology described by Eaton et al. (1995). An initial step involved the immediate filtration of water samples post-collection, utilizing a filtration unit. The volume filtered ranged from 0.1 L to 2 L, depending on the algal concentration. Prior to completing the filtration process, 2 mL of 1% $MgCO_3$ solution was added to preserve the sample. The filter was then transferred to a tissue grinder, where 2–3 mL of 90% acetone solution was added. This solution was prepared by mixing acetone with saturated $MgCO_3$ solution (prepared by dissolving 1 g of fine $MgCO_3$ powder in 100 mL of distilled water) at a 9:1 ratio. The sample was ground for 1 min at 500 rpm. The homogenized sample was then transferred into a centrifuge tube, rinsed with acetone solution, and the volume


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

was adjusted to 10 mL with acetone. The sample was stored in the dark at 4°C for at least 2 h. Subsequently, centrifugation was performed for 20 min at 3000 rpm, after which only the clear supernatant was retained and any remaining residues were discarded. Absorbance measurements were then conducted at wavelengths of 664, 665, and 750 nm, both before and after the acidification process. For acidification, 3 mL of the sample was transferred to a cuvette, and absorbance was measured. Next, 0.1 mL of 0.1 M HCl was introduced, gently mixed, and after 90 s, the absorbance was measured again. Finally, the chlorophyll-a concentration was calculated using the following formula:

$$Chlorophyll-a = \frac{26.7 \times (664b - 665a) \times V_1}{V_2 \times L} mg/m^3 \quad \text{(Eq.1)}$$

where: 664b: Absorbance measured at 664 nm (corrected by subtracting the absorbance at 750 nm) before acidification; 665a: Absorbance measured at 665 nm (corrected by subtracting the absorbance at 750 nm) after acidification; $V_1$**:** Volume of the extract; $V_2$**:** Volume of the filtered sample; and L**:** Path length of the cuvette.

### *Satellite data and image processing*

To ensure the highest possible accuracy in estimating Chlorophyll a, field sampling should be conducted precisely at the time the satellite passes over and captures imagery of the study area. However, in practice, due to limitations in personnel and equipment, the timing of field measurements often does not coincide with the satellite overpass. Furthermore, even when sampling is conducted at the same time as the satellite overpass, the resulting imagery may be unusable due to cloud cover or haze. In our study, field sampling was carried out over a three-day period (September 30, October 1, and October 2, 2024), during which a total of 41 samples were collected. However, upon reviewing satellite imagery acquired around the time of field survey (sampling), we found that the available images were significantly affected by cloud cover (*Fig. 2*). Fortunately, the Sentinel-2 image acquired on October 6, 2024, was of sufficiently high quality. Therefore, this image was selected for further analysis of mapping the distribution of chlorophyll-a concentration over the study area. The Sentinel-2 imagery acquired on October 6, 2024, was employed in this study. This decision was based on the underlying assumption regarding the overall concentration and spatial pattern of chlorophyll-a remained relatively stable over a period of several days, despite the image not being acquired at the exact time of field sampling. While acknowledging that this assumption may not hold true across the entire region under investigation because of dynamic water circulation, employing this imagery is considered acceptable for the specific objectives of this research.

More specifically, in this work a single scene of Sentinel-2 imagery (Level-1C) was acquired over the study area at approximately 10:20 local time on 06 October 2024, during a period when lake water composition in the region was relatively stable. The image was cloud-free and of high quality. In the present study, Sentinel-2 Level-1C (L1C) imagery was obtained from Google Earth Engine. This platform provides top-of-atmosphere (TOA) reflectance data that has undergone orthorectification, georeferencing, and radiometric calibration. The data is projected in UTM with the WGS84 datum, and is available at spatial resolutions of 10, 20, and 60 m (*Table 1*). The Sentinel-2 imagery underwent resampling to a 10 m spatial resolution via the nearest neighbor method to preserve radiometric integrity.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

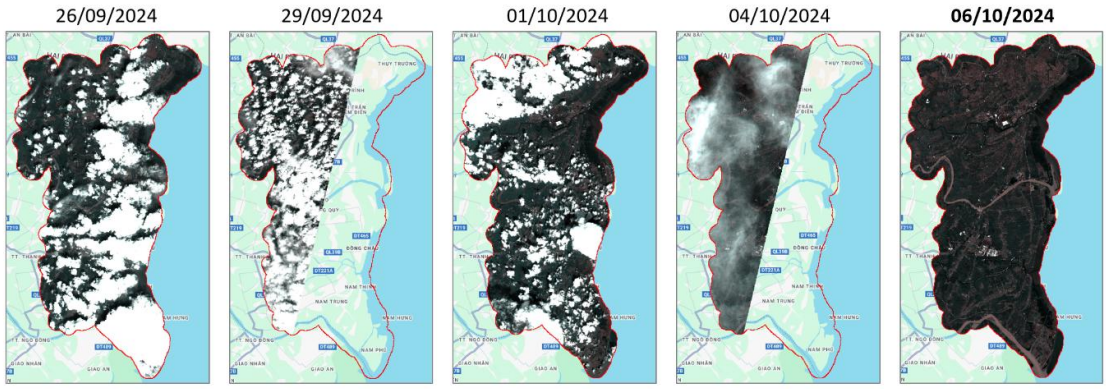


***Figure 2.*** *Satellite imagery acquired around the time of field survey*

***Table 1.*** *Description of Sentinel-2 multispectral instrument (MSI)*

| Band | Wavelength range (nm) | Central wavelength (nm) | Spatial resolution (m) | Description |
|---|---|---|---|---|
| B1 | 433–453 | 443 | 60 | Aerosol |
| B2 | 458–523 | 490 | 10 | Blue |
| B3 | 543–578 | 560 | 10 | Green |
| B4 | 650–680 | 665 | 10 | Red |
| B5 | 698–713 | 705 | 20 | Red Edge 1 |
| B6 | 734–748 | 740 | 20 | Red Edge 2 |
| B7 | 773–793 | 783 | 20 | Red Edge 3 |
| B8 | 785–900 | 842 | 10 | NIR |
| B8a | 855–875 | 865 | 20 | Red Edge 4 |
| B9 | 930–950 | 940 | 60 | Water vapor |
| B10 | 1365–1385 | 1375 | 60 | Cirrus |
| B11 | 1565–1655 | 1610 | 20 | SWIR1 |
| B12 | 2100–2280 | 2190 | 20 | SWIR2 |

Sentinel-2 data can be processed using a variety of atmospheric correction algorithms, notably Sen2cor and ACOLITE. Nonetheless, the Sen2cor atmospheric correction process was not designed for water bodies (Toming et al., 2016) and has been demonstrated to be inappropriate for estimating reflectances within the Near-Infrared (NIR) region. While ACOLITE, an atmospheric correction method specifically developed for water bodies (Vanhellemont and Ruddick, 2016), offered improved NIR reflectance estimations, its performance in the visible region was not consistently stable or accurate (Martins et al., 2017). Therefore, the Empirical Line Method (ELM) was chosen due to its established precision and successful application in previous studies, particularly for water bodies (Ha et al., 2017; Smith and Milton, 1999). The ELM assumes that the atmospheric conditions (aerosols, water vapor, etc.) are spatially constant across the entire study area. This allows a single set of linear regression coefficients, derived from the in situ sites, to be applied uniformly to the entire image. The ELM functions by removing atmospheric effects through the establishment of a statistical relationship between satellite data and ground measurements. Specifically, a linear regression is created for each individual Sentinel-2 band. This regression is established using two key


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

inputs: the Top-Of-Atmosphere (TOA) reflectance (taken directly from the satellite image pixel) and the co-located field-measured reflectance (the actual spectral data recorded on the ground). In this research, these linear functions were developed using the in situ reflectance points, and their corresponding Sentinel-2 TOA pixel values (acquired on October 6, 2024). Finally, the resulting linear regression function is applied to convert the TOA reflectance of the entire Sentinel-2 image into accurate surface reflectance, effectively removing the influence of the atmosphere. The primary constraints of the ELM stem from its direct reliance on the quality of the reference measurements. Specifically, the accuracy of the ELM is entirely dependent on the quality and representativeness of the in situ field-measured reflectance data. A key practical limitation arising from this dependency is the necessary exclusion of spectral points that show significant discrepancies from the majority of observations. This rigorous selection process is crucial, as failure to identify and manually remove poor quality or outlier field data could severely compromise the final atmospheric correction results.

### *Algorithms for estimation of chlorophyll a*

This investigation sought to construct an optimized model for quantifying chlorophyll-a (Chl-a) concentrations within the coastal region of Thai Binh Province, Vietnam. The methodology involved establishing a robust correlation between in situ Chl-a measurements and an array of remote sensing parameters, specifically various spectral band ratios and multi-band indices (*Table 2*). Following this, multiple linear regression (MLR) analysis was employed to synergistically integrate these independent variables, thereby yielding a refined Chl-a estimation model. This meticulously optimized model was subsequently applied to concurrent Sentinel-2 MSI satellite imagery to produce a comprehensive spatial distribution map of Chl-a across the designated study region.

#### *Pearson correlation analysis*

The statistical method is widely recognized as the preferred approach for establishing a correlation between various band combinations and in situ Chlorophyll-a concentration (Ha et al., 2017; Vinh et al., 2022). In the present study, the Pearson correlation analysis was conducted between in situ Chlorophyll measurements and the possible bands/band combinations (including single-band model, two-band ratios, and other band combinations/indices) derived from the synthetic Sentinel-2 surface reflectances. A great number of indices and band combinations have been developed for the determination of Chl-a. However, by reviewing previous studies, we selected those indices and band combinations that are frequently employed and compatible with the spectral channels of Sentinel 2 (*Table 2*). This analysis aimed to thoroughly examine the relationship and identify the suitable band ratio or combination for accurate Chlorophyll estimation.

#### *Multivariate regression analyses*

Based on the results of the Pearson correlation analysis, we selected a subset of variables with the highest correlation coefficient (r) from each category (single-band model, two-band ratio, and band combinations/indices). These selected variables were then used as input variables in the subsequent Multiple Linear Regression (MLR) to develop different models in estimating Chlorophyll-a and find an optimal model, which subsequently applied to a Sentinel-2 imagery to create a spatial distribution map of Chl-a across the study region.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

***Table 2.*** *Different common spectral indices for estimating Chl-a concentrations based on Sentinel 2-bands*

| Band combination | | Algorithm name | Equation | Reference |
|---|---|---|---|---|
| Two-band ratio | Green-Blue Ratio | Green-Blue | $R_{rs}(560)/R_{rs}(490)$ | Ha et al. (2017) |
| | Green-Red Ratio | Green-Red | $R_{rs}(560)/R_{rs}(665)$ | Ha et al. (2017) |
| | Blue-Green Ratio | Blue-Green | $R_{rs}(490)/R_{rs}(560)$ | Moses et al. (2009) |
| | VNIR1-Red Ratio | VNIR1-Red | $R_{rs}(705)/R_{rs}(665)$ | Duan et al. (2012); Gilerson et al. (2010); Gurlin et al. (2011) |
| | VNIR2-Red Ratio | VNIR2-Red | $R_{rs}(740)/R_{rs}(665)$ | Gitelson et al. (2008) |
| | VNIR2-VNIR1 Ratio | VNIR2-VNIR1 | $R_{rs}(740)/R_{rs}(705)$ | Li et al. (2021) |
| | NIRn-Red Ratio | NIRn-Red | $R_{rs}(865)/R_{rs}(665)$ | Duan et al. (2007); Tóth et al. (2021) |
| | VNIR1-Green Ratio | VNIR1-Green | $R_{rs}(705)/R_{rs}(560)$ | Cairo et al. (2019) |
| | VNIR2-Green Ratio | VNIR2-Green | $R_{rs}(740)/R_{rs}(560)$ | Cairo et al. (2019) |
| Band combinations/indices | Normalized difference chlorophyll index | NDCI | $(R_{rs}(705) - R_{rs}(665))/(R_{rs}(705) + R_{rs}(665))$ | Mishra and Mishra (2012) |
| | Florescence Line Height | FLH | $R_{rs}(705) - R_{rs}(665)$ | Zhao et al. (2022) |
| | Advanced Remote Sensing Sediment Parameter | Advanced_RSSP | $(R_{rs}(705) + R_{rs}(665))/(R_{rs}(705)/R_{rs}(665))$ | Pan et al. (2025) |
| | Water chlorophyll-a index | Advanced_WCI | $(R_{rs}(665) - R_{rs}(705))/(2 * R_{rs}(705))$ | Pan et al. (2025) |
| | Chlorophyll Index | ChlI | $R_{rs}(560) - (R_{rs}(443) + (560 - 443)/(665 - 443) * (R_{rs}(665) - R_{rs}(443)))$ | Hu et al. (2012); Smith et al. (2018) |
| | Maximum Chlorophyll Index | MCI | $(R_{rs}(705) - R_{rs}(665)) - ((705 - 665)/(740 - 665) * (R_{rs}(740) - R_{rs}(665)))$ | Gower et al. (2005) |
| | Cyanobacteria Index | CI | $-((R_{rs}(665) - R_{rs}(560)] - ((665 - 560)/(705 - 560) * (R_{rs}(705) - R_{rs}(560))))$ | Akbarnejad Nesheli et al. (2024); Wynne et al. (2008) |
| | Maximum Peak Height | MPH | $(R_{rs}(705) - R_{rs}(665)) - ((705 - 665)/(865 - 665) * (R_{rs}(865) - R_{rs}(665)))$ | Akbarnejad Nesheli et al. (2024); Matthews et al. (2012) |
| | Surface Algal Bloom Index | SABI | $(R_{rs}(865) - R_{rs}(665))/(R_{rs}(490) + R_{rs}(560))$ | Akbarnejad Nesheli et al. (2024); Alawadi (2010) |

Multiple Linear Regression (MLR) stands as a widely employed statistical methodology for examining the intricate connections between a single dependent variable and a collection of two or more independent variables. This particular analytical approach fundamentally operates under the premise that the relationships among the various predictor variables are both linear and additive (Jang et al., 2024). Essentially, enables the construction of a mathematical framework or an approximating equation capable of characterizing a real-world phenomenon. Notwithstanding their fundamental straightforwardness, MLR models have garnered widespread adoption and consistently


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

delivered favorable outcomes across diverse research endeavors (Jang et al., 2024; Prieto et al., 2017; Quang Vinh et al., 2024; Sousa et al., 2007). The association linking the dependent variable (Y) and the predictor variables ($X_1$, $X_2$, …, $X_n$) is generally expressed by the following equation (Jang et al., 2024; Prieto et al., 2017):

$$Y = a_0 + a_1 X_1 + a_2 X_2 + ... + a_k X_k \quad \text{(Eq.2)}$$

where $a_0$ denotes the constant term, while $a_1$, $a_2$,. . ., $a_k$ represent the regression coefficients. These coefficients, derived via the least-squares method, quantify the magnitude and direction of each explanatory variable's impact on the target variable, Y (Chlorophyll-a concentration). The independent variables ($X_k$) encompass various factors, including simulated Sentinel-2-based reflectance, two-band ratios, and other band combinations or indices.

In the present study, a total of 41 in situ samples (Chlorophyll-a and independent variables from simulated Sentinel-2 reflectance measurements) were randomly divided into two parts: the training dataset (65%) (N = 27) for model calibration and the test dataset (35%) (N = 14) for model validation. The calibration dataset exhibited a more extensive chlorophyll range than its validation counterpart, which ensured comprehensive coverage and enhanced the applicability of the chlorophyll estimation model.

*Performance assessment*

Statistical analyses were conducted using various evaluation metrics such as coefficient of determination ($R^2$), root mean square error (RMSE), and bias (systematic error) to assess the performance of developed models. The optimal statistical model for chlorophyll estimation was subsequently selected based on these metrics. $R^2$ serves as a key indicator of a model's goodness of fit, representing the ratio of variance in the target variable that can be predicted from the explanatory variables. Its value, ranging from 0 to 1, reflects the proximity of the regression line to the actual data points; a higher $R^2$ value suggests a more accurate model fit. Besides, the RMSE serves as an indicator of the typical magnitude of the prediction deviations, with reduced RMSE values signifying enhanced model performance. When the primary objective of a model is prediction, RMSE is considered a crucial criterion for assessing its accuracy. Bias quantifies the systematic discrepancy observed between the estimated parameter and its actual measurement. A zero bias indicates an unbiased estimator. Generally, values approaching zero signify superior model performance. $R^2$, RMSE and Bias were calculated by the following formula (Akbarnejad Nesheli et al., 2024):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \quad \text{(Eq.3)}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad \text{(Eq.4)}$$

$$Bias = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i) \quad \text{(Eq.5)}$$


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

where $y_i$ is the $i^{th}$ measured chlorophyll-a value, $\hat{y}_i$ is the $i^{th}$ predicted chlorophyll-a value, $\bar{y}$ is the average value of observed values, and $n$ is the number of samples. In the present work, the training and testing process was performed employing the Linear Regression algorithm from the Scikit-learn package with Python programing language. Among the chlorophyll estimation models that were developed, the most appropriate model was chosen for application to a Sentinel-2 image acquired on October 6, 2024, with the aim of enhancing the understanding of the spatial variability of chlorophyll across the study area.

### *Estimation of trophic state index (TSI)*

Various mathematical methodologies are employed for the assessment of lake eutrophication. Among these, the Trophic State Index (TSI), as conceptualized by Carlson (1977), is widely recognized as the most acceptable method for evaluating lake eutrophication (Duan et al., 2007; Xing et al., 2005). The TSI derives continuous values, typically scaled from 0 to 100, for lakes, primarily utilizing parameters such as Secchi disk transparency, chlorophyll-a (Chl-a) concentration, or total phosphorus content. Conceptually, the TSI quantifies the "greenness" of a lake, reflecting the concentration of algal biomass within its water column. However, Carlson's original TSI has been criticized for neglecting the influence of factors other than phytoplankton, such as water color, dissolved matter, and suspended matter, on Secchi disk transparency (Duan et al., 2007; Zhang et al., 2003). The aforementioned limitation was subsequently resolved with the introduction of the modified Carlson's $TSI_M$ (Aizaki et al., 1981). It is noted that methods for assessing lake eutrophication types are adaptable to diverse geographical locations, environmental conditions, and human activities.

This work utilized Shu's modified model, an adaptation of the Trophic State Index Method ($TSI_M$), to compute the trophic state index and assess the eutrophication. The calculation followed by Shu's method (Duan et al., 2007) as follows:

$$TSI_M(chla) = 10 \times \left( 2.46 + \frac{\ln(chla)}{\ln(2.5)} \right) \quad \text{(Eq.6)}$$

where chla means Chl-a content (mg/m$^3$). Each segment of the 0-100 scale indicated a different trophic state: oligotrophic (0–20), lower mesotrophic (20–30), mesotrophic (30–40), upper-mesotrophic (40–50), eutrophic (50–70), hypereutrophic (70–80), and extremely hypereutrophic (80–100).

## Results and discussion

### *Chlorophyll status and in situ reflectance properties*

Chlorophyll-a, a key phytoplankton pigment, is universally present across all algal groups inhabiting inland aquatic environments. The distribution of Chlorophyll-a concentrations, derived from 41 field measurements, is presented as a violin plot in *Figure 3*. This plot illustrates the range of observed values, with a minimum concentration of 1 mg/m³ and a maximum of 197.0 mg/m³. Descriptive statistics further indicate a mean Chla concentration of 27.63 mg/m³ and a median of 12.0 mg/m³. The computed standard deviation of 39.84 mg/m³ underscores the significant spread inherent in the Chla measurement data.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

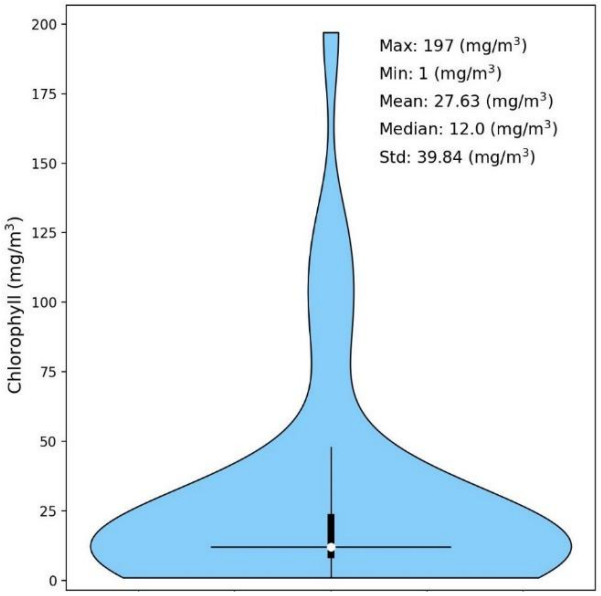


***Figure 3.*** *Violin plot of field-measured Chl-a concentrations in the coastal area of Thai Binh Province*

*Figure 4* presents the in situ reflectance spectra of water bodies located in the coastal area of Thai Binh province, corresponding to 41 samples exhibiting varying chlorophyll-a concentrations. In the blue spectral region, the acquired field-measured reflectance spectrum consistently exhibits low values at wavelengths under 520 nm. This observed decrease in reflectance is principally a consequence of the synergistic absorption by various water constituents, including algal pigments (e.g., chlorophyll-a) and colored dissolved organic matter (CDOM) (Gitelson et al., 1993). Within the spectral range of approximately 450-460 nm, an absorption feature is observed, which is attributable to phytoplankton pigments (Duan et al., 2007). A notable increase in reflectance was observed within the green spectral region, culminating in a maximum value around 570 nm. This spectral peak is likely a consequence of reduced absorption by phytoplankton pigments at this wavelength, where the reflectance signal was predominantly controlled by non-algal particles (NAP) and CDOM (Gurlin et al., 2011). Two distinct reflectance minima, evident at approximately 638 nm and 675 nm in the red and near-infrared (NIR) spectral regions, align with the primary absorption peaks of phytoplankton pigments (Gitelson and Kondratyev, 1991). The reduced reflectance around 638 nm is attributed to absorption by cianopycocyanine, while the minimum around 675 nm is associated with the strong absorption of chlorophyll-a (Duan et al., 2007; Gurlin et al., 2011). Within the 690–720 nm wavelength range, a prominent reflectance peak is observed. This peak is attributed to a localized minimum in the cumulative absorption of light by both phytoplankton pigments and water (Duan et al., 2007; Gitelson, 1992; Gurlin et al., 2011).

## *Performance of the algorithms to estimate the chlorophyll*

### *Single-band model for chlorophyll estimation*

The graph illustrates a fluctuating pattern in the correlation coefficient (r) between chlorophyll concentrations and remote-sensing reflectance (Rrs(λ)) across the observed


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

spectrum, with r values varying from -0.018 to 0.417 (N = 41) assessment (*Fig. 5*). It begins with a very weak or negative correlation in the Aerosol band (B1, -0.018), then increases in the Blue/Green range (B2, 0.085; B3, 0.178). Subsequently, there is a sharp drop in correlation in the red range (B4, -0.007). Following this dip, the correlation coefficient rises significantly into a strong positive correlation in the red-edge range, starting with Red Edge 1 band (B5, 0.306), reaching its maximum in Red Edge 2 (B6, 0.417), and remaining high through Red Edge 3 (B7, 0.406), NIR (B8, 0.393), and Red Edge 4 (B8a, 0.383). This overall pattern clearly indicates that the variable being correlated has a substantially stronger positive relationship with the data captured in the near-infrared spectrum compared to the visible light spectrum. Additionally, the findings underscore the limitations of a single-band approach in achieving accurate Chlorophyll assessment (*Fig. 5*).

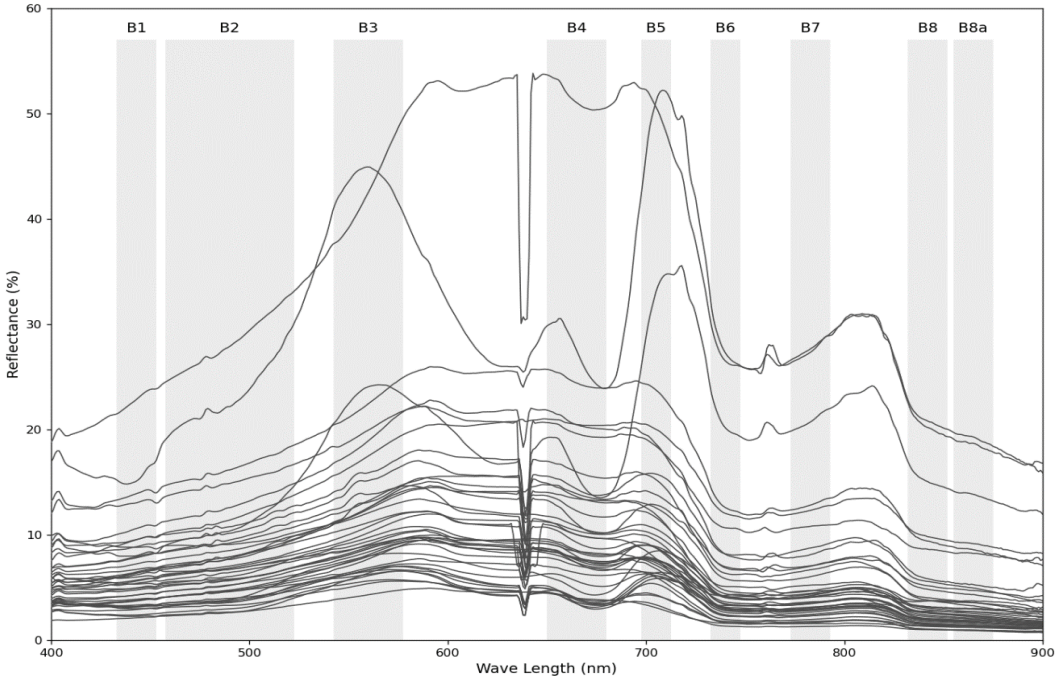


***Figure 4.*** *In situ reflectance spectra over the coastal region of Thai Binh province*

*Two-band ratio model for chlorophyll estimation*

Further analyses were performed utilizing nine different band ratios, which are widely recognized for their efficacy in estimating chlorophyll-a concentration (Chl-a), including, but not limited to, Green-Blue (Rrs(560)/Rrs(490)) (Ha et al., 2017), Green-Red (Rrs(560)/Rrs(665)) (Ha et al., 2017), Blue-Green (Rrs(490)/Rrs(560)) (Moses et al., 2009), VNIR1-Red (Rrs(705)/Rrs(665)) (Duan et al., 2012; Gilerson et al., 2010; Gitelson et al., 2008; Gurlin et al., 2011), VNIR2-Red (Rrs(740)/Rrs(665)) (Gitelson et al., 2008), VNIR2-VNIR1 (Rrs(740)/Rrs(705)) (Li et al., 2021), NIRn-Red (Rrs(865)/Rrs(665)) (Duan et al., 2007; Tóth et al., 2021), VNIR1-Green (Rrs(705)/Rrs(560)) (Cairo et al., 2019), and VNIR2-Green (Rrs(740)/Rrs(560)) (Cairo et al., 2019). *Figure 6* presents scatter plots of Chl-a against various band ratios. As illustrated in *Figure 6,* the analysis demonstrates the highest correlation between chlorophyll-a concentration (Chl-a) and the VNIR1-Red (Rrs(705)/Rrs(665)) band ratio (Gurlin et al., 2011), yielding a remarkable


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

correlation coefficient of r = 0.945 ($p < 0.001$). This specific ratio incorporates the Red Edge 1 (VNIR1, B5) and red (B4) spectral bands. Similarly, VNIR2-Red (Rrs(740)/Rrs(665)) (Gitelson et al., 2008) also exhibits a strong correlation (r = 0.923, $p < 0.001$). These findings underscore the critical role of the red and red-edge spectral regions in Chl-a estimation, aligning with the known spectral properties of chlorophyll. While other ratios also show statistically significant correlations, their strength is notably lower. The data collectively suggests that ratios incorporating red and NIR bands are an effective way for Chl-a retrieval, leveraging the unique spectral response of chlorophyll. Nevertheless, it is crucial to acknowledge that Chl-a values within the study area are subject to the influence of various interacting factors. Therefore, sole reliance on this single band ratio may not provide sufficient accuracy for the construction of a robust Chl-a estimation model. To enhance the precision of Chl-a estimation, it is imperative to account for the impact of these additional influencing factors, which can be elucidated through the comprehensive analysis of diverse band ratios.

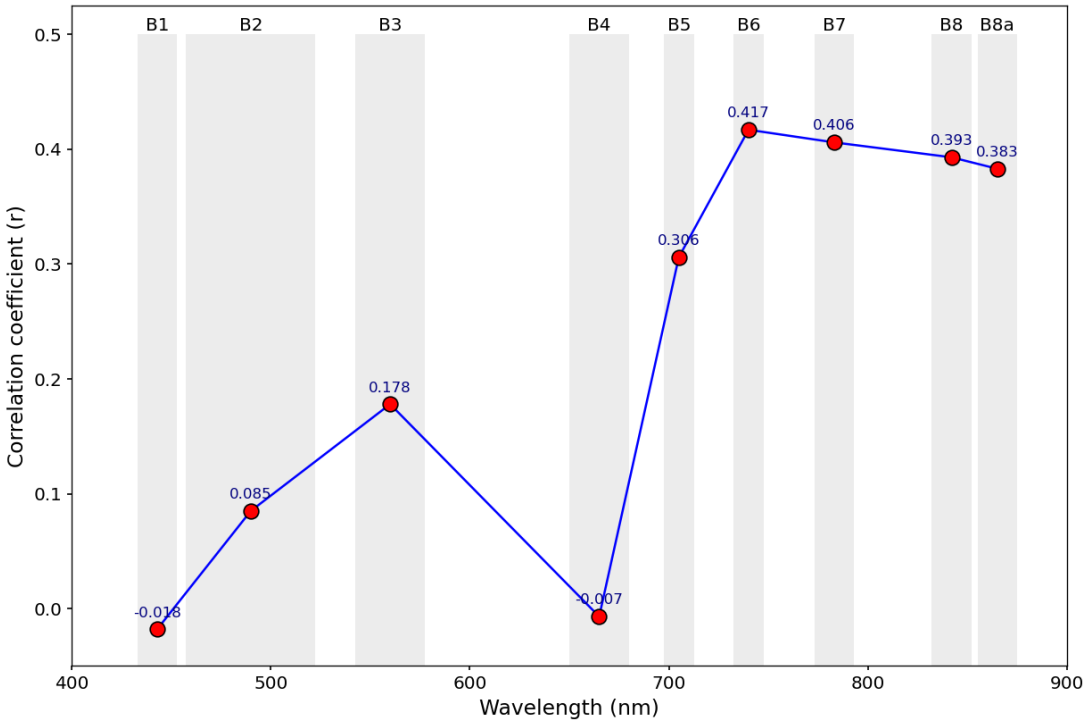


***Figure 5.*** *Correlation between chlorophyll-a and single band spectral*

*Band combinations or indices for chlorophyll estimation*

In the present research, we also explored the suitability of different band combinations or indices commonly employed for estimating Chla (Gitelson et al., 2011; Gower et al., 2005; Mishra and Mishra, 2012; Quang Vinh et al., 2024; Yang et al., 2010) (*Table 2*). *Figure 7* presents a comprehensive analysis of the correlation between field-measured Chlorophyll-a (Chla) concentrations and a diverse set of three-band remote sensing models or spectral indices. Each scatter plot, labeled (a-i), depicts this relationship for a specific model or index, accompanied by its Pearson correlation coefficient (r) and corresponding p-value (p), offering insights into the strength and statistical significance of these relationships. The scatter plots visually illustrate the data distribution and the fit


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

of each model, allowing for a direct assessment of each method's potential accuracy in estimating Chlorophyll-a concentrations.

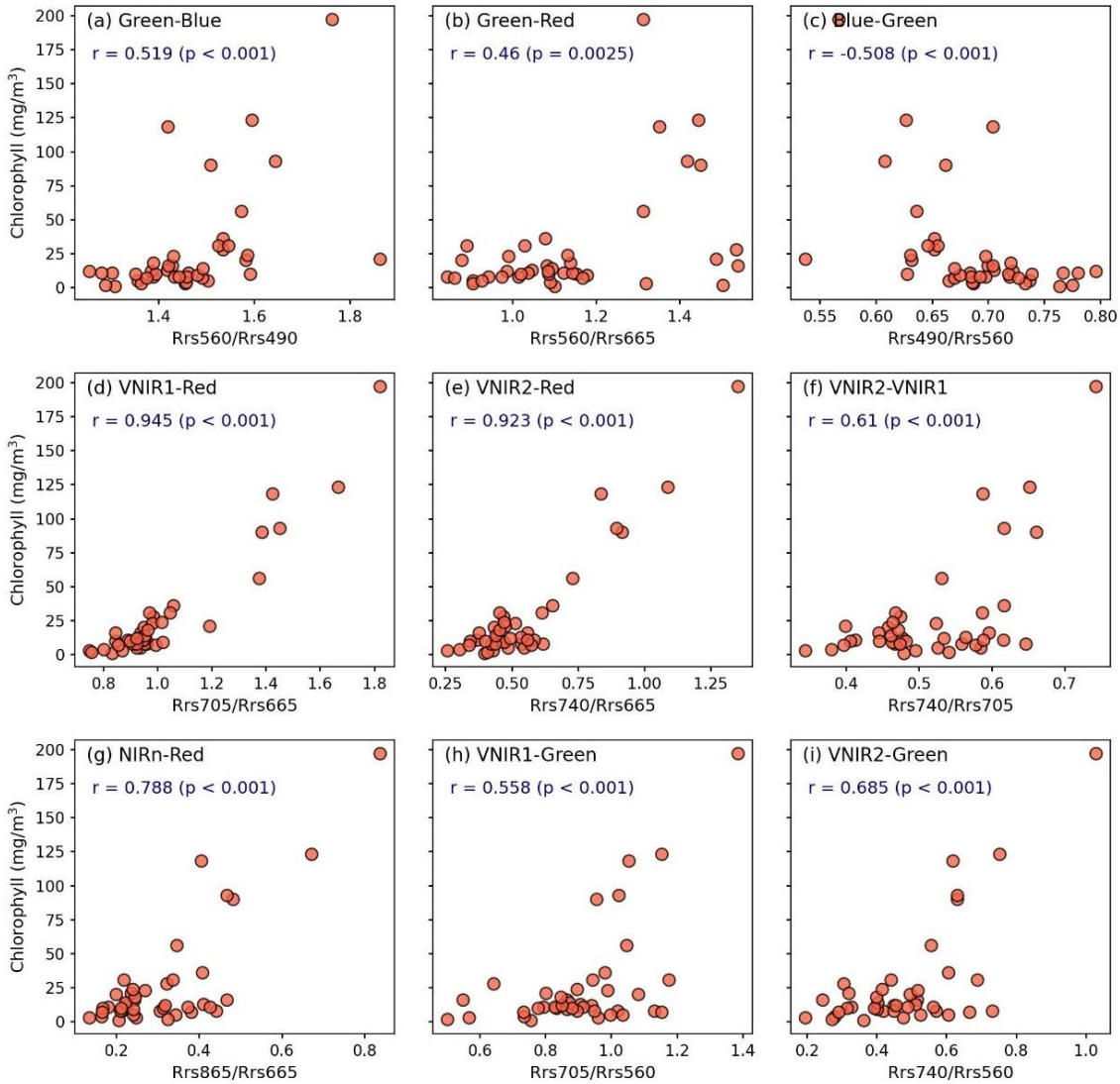


***Figure 6.*** *Correlation between chlorophyll-a and band ratios models (N = 41): (a) Green-Blue; (b) Green-Red; (c) Blue-Green; (d) VNIR1-Red; (e) VNIR2-Red; (f) VNIR2-VNIR1; (g) NIRn-Red; (h) VNIR1-Green; and (i) VNIR2-Green*

Overall, several models exhibit robust correlations with Chla concentrations. Notably, the Normalized Difference Chlorophyll Index (NDCI) (a) model yields the highest positive correlation coefficient ($r = 0.91$, $p < 0.001$), demonstrating a strong linear relationship where increasing index values correspond to a marked increase in Chla concentration. Similarly, the Florescence Line Height (FLH) (b) with $r = 0.827$, $p < 0.001$ and Cyanobacteria Index (CI) (f) with $r = 0.809$, $p < 0.001$ also demonstrate significant positive correlations, reinforcing their potential for Chla estimation. Conversely, Advanced_WCI (d) exhibits a strong negative correlation ($r = -0.867$, $p < 0.001$), implying that Chla concentration increases as this index's value decreases, showcasing an inverse yet highly significant relationship. However, not all models perform with equal


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

strength. Advanced_RSSP (c) displays a weak and statistically insignificant correlation ($r = -0.126$, $p = 0.4307$), suggesting it is not a reliable predictor for Chla concentration within this dataset. Meanwhile, models such as Maximum Chlorophyll Index (MCI) (e) ($r = 0.532$, $p < 0.001$), Maximum Peak Height (MPH) (g) ($r = 0.743$, $p < 0.001$), Surface Algal Bloom Index (SABI) (h) ($r = 0.736$, $p < 0.001$), and Chlorophyll Index (Chl-I) (i) ($r = 0.564$, $p < 0.001$) all demonstrate statistically significant correlations, though some are less robust than the leading indicators. In summary, *Figure 7* reveals a wide spectrum of correlations among the methods, ranging from strong positive and negative relationships to weak and non-significant ones. The methods NDCI, FLH, CI and Advanced_WCI consistently demonstrate robust and statistically significant correlations, suggesting strong predictive or associative power. The remaining methods exhibit weaker yet still significant correlations, with the exception of Advanced_RSSP, which does not show a statistically meaningful relationship ($r = -0.126$, $p = 0.4307$).

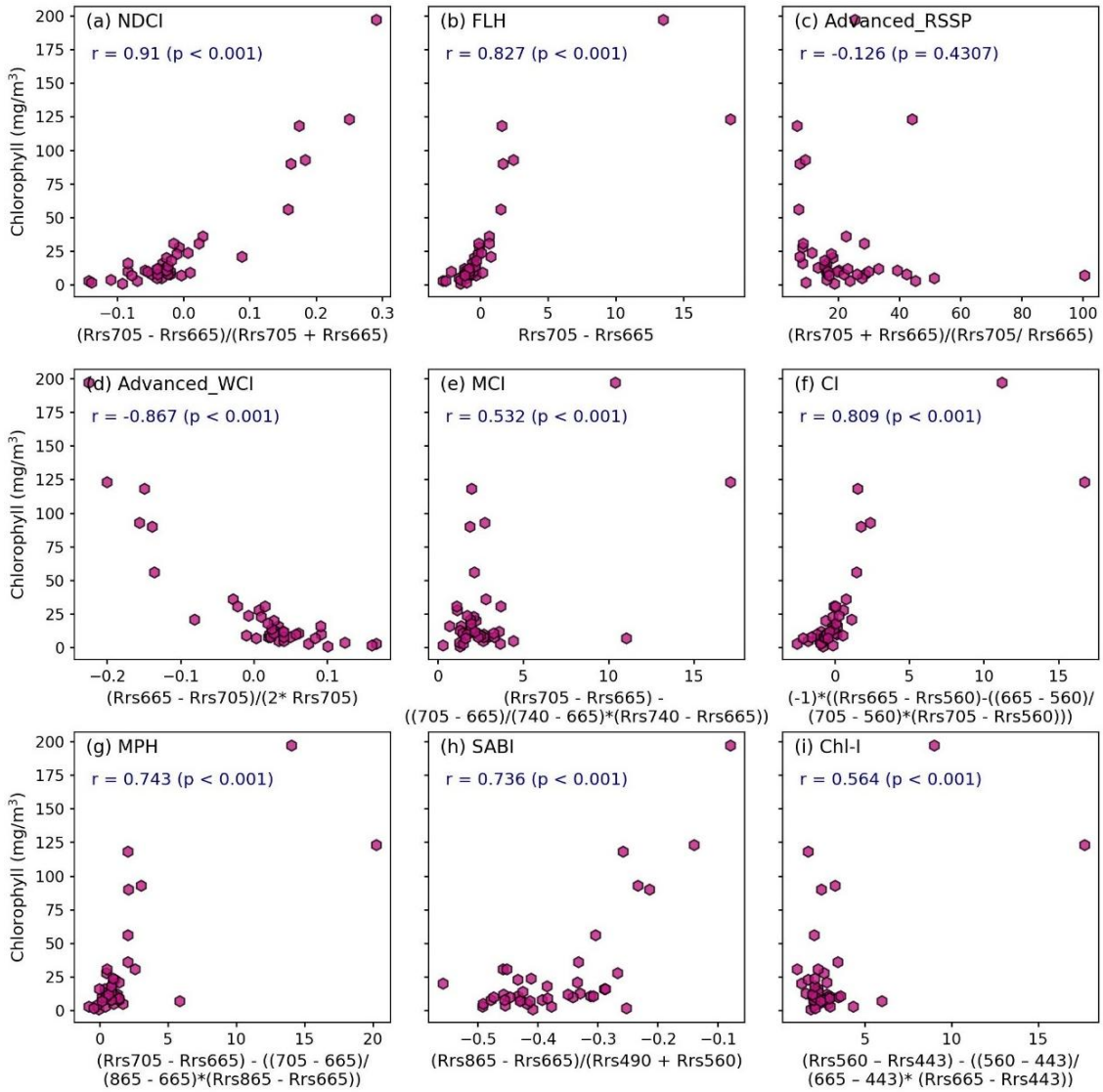


***Figure 7.*** *Correlation between Chl-a and band combinations or indices (N = 41): (a) NDCI (b) FLH; (c) Advanced_RSSP; (d) Advanced_WCI; (e) MCI; (f) CI; (g) MPH; (h) SABI; and (i) Chl-I*


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

*Multivariate regression analyses*

The interplay between Chlorophyll-a (Chl-a) concentrations and various spectral band ratios and multi-band indices was explored through the application of multivariate regression analyses. Based on the results of the Pearson correlation analysis, we selected a subset of variables with the highest correlation coefficient (r) from each category (single-band model, two-band ratio, and band combinations/indices). These selected variables were then used as input variables in the subsequent Multiple Linear Regression (MLR) to develop different models (*Table 3*) in estimating Chlorophyll-a and find an optimal model, which subsequently applied to a Sentinel-2 imagery to create a spatial distribution map of Chl-a across the study area.

*Table 3* systematically outlines the distinct sets of independent variables, also referred to as features or predictors, employed in three separate Multiple Linear Regression (MLR) models developed for the purpose of estimating Chlorophyll-a concentrations. Specifically, Model 1 is characterized by the selection of VNIR1, VNIR2, VNIR3, and NIR as its predictive variables. In contrast, Model 2 employs a distinct combination of predictors, namely VNIR1-Red, VNIR2-Red, and NIRn-Red. Finally, Model 3 is constructed utilizing VNIR1, VNIR2, NDCI, and VNIR1-Red as its input variables. The differentiation in predictor sets among these models signifies an investigative approach to assess the individual and combined contributions of various spectral bands and indices towards accurately estimating Chlorophyll-a concentrations through MLR techniques.

*Table 4* and *Figure 8* provide a comprehensive overview of the performance of three distinct MLR models (Model 1, Model 2, and Model 3) with diverse input variables in estimating Chlorophyll-a concentrations. The table presents key statistical metrics, including the $R^2$, r, RMSE, and Bias, for both the training and validation phases of each model. These metrics are crucial for assessing model fit, predictive accuracy, and generalization capability. Moreover, the observed Chl-a concentrations within both datasets encompassed a broad range, from low to high values, thereby enhancing the model's applicability across diverse conditions.

***Table 3.*** *Chlorophyll-a estimation MLR models with diverse input variables*

| Model name | Features/predictors |
|---|---|
| Model 1 | VNIR1, VNIR2, VNIR3, and NIR |
| Model 2 | VNIR1-Red, VNIR2-Red, and NIRn-Red |
| Model 3 | VNIR1, VNIR2, NDCI, and VNIR1-Red |

***Table 4.*** *Different MLR models in estimating Chl-a level (RMSE and bias are in mg/m$^3$)*

| Model | Training phase (N = 27) | | | | Testing phase (N = 14) | | | |
|---|---|---|---|---|---|---|---|---|
| | $R^2$ | r | RMSE | Bias | $R^2$ | r | RMSE | Bias |
| Model 1 | 0.792 | 0.89 | 19.385 | < 0.001 | 0.773 | 0.879 | 22.731 | 9.241 |
| Model 2 | 0.896 | 0.946 | 13.745 | < 0.001 | 0.937 | 0.968 | 10.664 | 5.508 |
| Model 3 | 0.950 | 0.975 | 9.534 | < 0.001 | 0.944 | 0.972 | 9.366 | 3.028 |

The left column of *Figure 9(a, c, e)* illustrates the performance of the models during the training phase (N = 27), with corresponding statistics presented within each subplot.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

During the training phase, visually, all three MLR models demonstrated a strong fit to the training data, with data points clustering relatively closely around the 1:1 line (dashed black line) and the regression line (solid magenta line). Model 1 (*Fig. 8a*) achieved an $R^2$ of 0.792 and an r of 0.89, with an RMSE of 19.385 mg/m$^3$. Model 2 (*Fig. 8c*) showed a noticeable improvement, yielding an $R^2$ of 0.896 and an r of 0.946, alongside a reduced RMSE of 13.745 mg/m$^3$. Notably, Model 3 (*Fig. 8e*) exhibited the most robust performance in the training phase, with the highest $R^2$ of 0.950 and r of 0.975, coupled with the lowest RMSE of 9.534 mg/m$^3$. A Bias of 0 was observed for all models in this phase, which is expected as the models are optimized on the training data. While the close proximity of data points to the 1:1 line and high $R^2$ values generally indicated that all models were well-calibrated to their respective training datasets, Model 3, with its regression equation of $y = 0.950x + 1.417$, specifically exhibited the tightest fit and highest explanatory power.

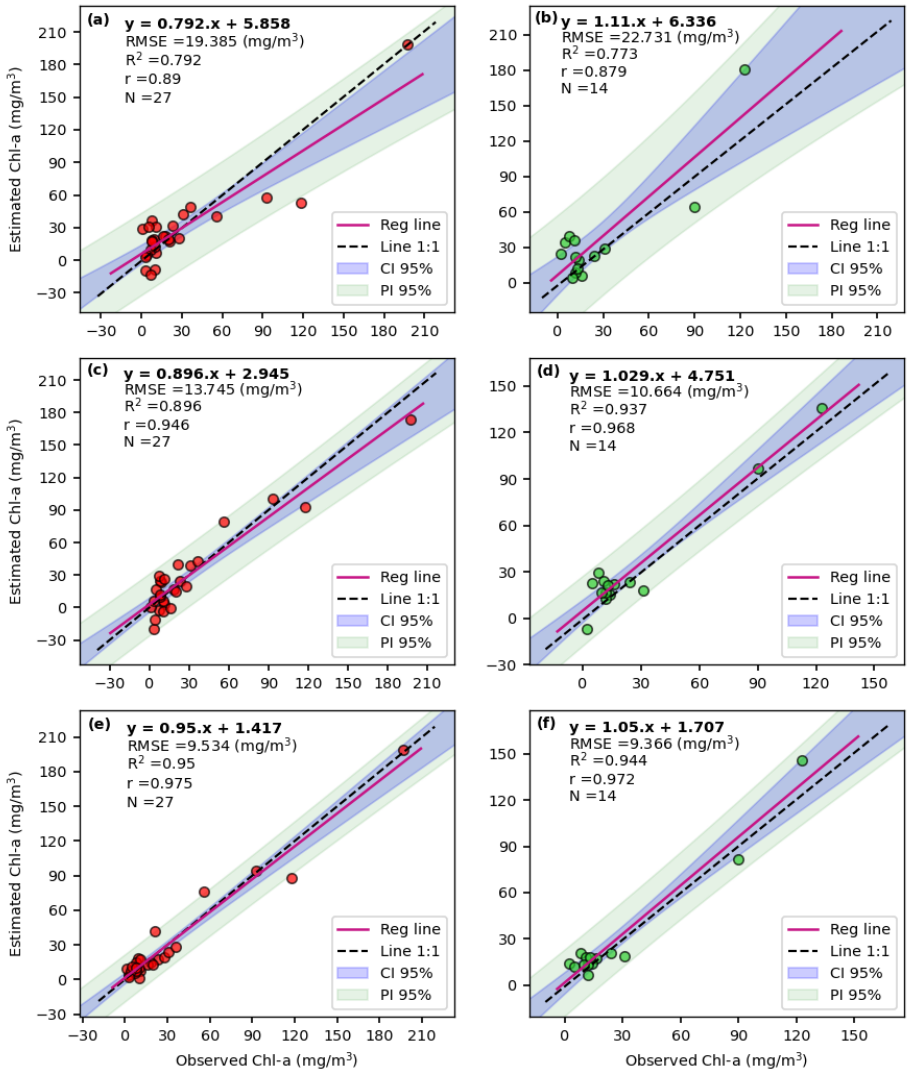


***Figure 8.*** *Relationships between observed and estimated Chl-a concentrations across different models during training and testing phases: (a, b) training and testing phases for Model 1; (c, d) training and testing phases for Model 2; and (e, f) training and testing phases for Model 3, with the 1:1 line (dashed black line) and the regression line (solid magenta line) (CI 95%: confidence Interval, PI 95%: prediction interval)*


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

The right column of *Figure 8(b, d, f)* illustrate the models' capacity to predict unseen data during the testing phase (N = 14). The testing phase is critical for assessing the models' generalization to previously unobserved data. In this phase, Model 1's performance (*Fig. 8b*) slightly declined, showing an $R^2$ of 0.773, an r of 0.879, and an increased RMSE of 22.731 mg/m$^3$, along with a Bias of 9.241 mg/m$^3$. The regression line of Model 1 (y = 1.11x + 6.336) indicates a tendency for underestimation at lower concentrations and overestimation at higher concentrations, with wider scatter. Model 2 (*Fig. 8d*), however, demonstrated exceptional generalization, with an $R^2$ of 0.937 and an r of 0.968, both slightly higher than its training performance, and a significantly lower RMSE of 10.664 mg/m$^3$, although it exhibited a Bias of 5.508 mg/m$^3$. Visually, the regression line for Model 2 (y = 1.029x + 4.751) closely aligns with the 1:1 line, and the data points show good consistency, although a slight positive bias can be observed. Crucially, Model 3 (*Fig. 8f*) consistently maintained its strong performance, achieving the highest $R^2$ of 0.944 and r of 0.972, along with the lowest RMSE of 9.366 mg/m$^3$ and the lowest Bias of 3.028 mg/m$^3$ among all models. Its regression line (y = 1.05x + 1.707) is the closest to the ideal 1:1 line among all models, and the data points exhibit the tightest cluster around this line, indicating superior accuracy and precision. The visual distribution of points in *Figure 8f* confirms the minimal bias and high predictive accuracy quantified by the statistical metrics.

As for Confidence Interval (CI 95%) and the Prediction Interval (PI 95%), Model 1's performance (*Fig. 8b*) has a dramatically wider PI 95% compared to Model 2 (*Fig. 8d*) and Model 3 (*Fig. 8f*). This signifies that a single prediction from model (b) is highly unreliable, and the expected error range is very large at 95% confidence. Model 3 (*Fig. 8f*) successfully minimizes the residual error (RMSE 9.366 mg/m$^3$), resulting in the narrowest PI 95% among the three. This indicates that the model structure in *Figure 8f* is the most effective at capturing the relationship within the small N = 14 dataset, making its individual predictions the most precise. The PI gap clearly dictates which model is viable for deployment. Only models Model 2 (*Fig. 8d*) and Model 3 (*Fig. 8f*) offer sufficiently narrow PI to be considered for reliable decision-making. The CI 95% (blue shaded area) represents the uncertainty in the estimation of the true mean regression line. Despite the similar sample size, Model 3 (*Fig. 8f*) has the highest $R^2$ and the lowest scatter, leading to a slightly narrower CI near the center of the data compared to Model 1 (*Fig. 8b*). This shows that a better-fitting model structure can slightly reduce epistemic uncertainty, even without additional data. The wide CI at the tails signals a high risk of structural error (bias) if any of these models are used for extrapolation (i.e., predicting Chl-a values far beyond 150 mg/m$^3$).

A comprehensive comparison across both training and testing phases (*Table 4* and *Fig. 8*) reveals that Model 3 consistently outperforms Model 1 and Model 2 across nearly all evaluation metrics. While Model 2 shows impressive generalization in terms of $R^2$ and r, Model 3 consistently delivers the highest predictive accuracy (lowest RMSE) and the least systematic error (lowest Bias) on unseen data. Furthermore, the visual coherence of data points around the 1:1 line in *Figure 8f* for Model 3, coupled with its minimized bias, underscores its reliability and strong practical applicability for Chlorophyll-a estimation. Therefore, based on its superior and more stable performance in both fitting and generalization capabilities, Model 3 is identified as the optimal Multiple Linear Regression model for estimating Chlorophyll-a concentrations among the evaluated options. Its robust performance across all key metrics highlights its reliability and suitability for practical applications. Model 3 is expressed as follows:


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

$$\text{Chl-a} = -3.27 * \text{VNIR1} + 5.02 * \text{VNIR2} - 741.04 * \text{NDCI} + 458.9 * \text{VNIR1-Red} - 434.48 \ (\text{mg/m}^3) \quad \text{(Eq.6)}$$

where VNIR1 and VNIR2 represent the remote sensing reflectance, Rrs(λ), derived from Sentinel-2 (S2) imagery at band 5 (705 nm) and band 6 (740 nm), respectively. NDCI refers to the Normalized Difference Chlorophyll Index, while VNIR1-Red signifies the ratio of Rrs(705) to Rrs(665).

When compared to existing literature, the performance of our Model 3 stands out as highly competitive, and in many aspects, superior. For instance, a notable MLR study dedicated to Chlorophyll-a estimation in Quan Son Reservoir, utilizing Sentinel-2B imagery, reported an $R^2$ of 0.95 during the training phase but a lower value of 0.87 during validation (Thao et al., 2024), thereby indicating potential overfitting issues in their model. Our Model 3's impressive testing $R^2$ of 0.944 not only aligns with the training performance of this aforementioned benchmark study but significantly surpasses its reported validation performance, thereby underscoring the strong generalizability of our model. Similarly, research focusing on Chlorophyll-a estimation in turbid Yellow Sea waters, also employing MLR, reported a correlation coefficient (r) of 0.94 (Baek et al., 2019). Notably, the testing r of 0.972 achieved by our Model 3 demonstrates an even more robust correlation, indicating a superior predictive capability. Beyond these strong correlation metrics, Model 3 also demonstrated exceptional performance in terms of error assessment. It achieved the minimum RMSE values: 9.534 mg/m$^3$ for the training phase and 9.366 mg/m$^3$ for the independent testing phase. Furthermore, the associated bias in the testing phase was notably low at 3.028 mg/m$^3$. While direct numerical comparisons of RMSE can often be challenging due to potential inconsistencies in units and concentration ranges across diverse studies, our RMSE values are indicative of high accuracy. For comparative context, certain studies employing advanced machine learning algorithms for Chlorophyll-a retrieval, such as an LGBM model by Kim et al. (2022), reported higher RMSE values (e.g., 15.15 mg/m$^3$). Consequently, the robust RMSE demonstrated by Model 3 further reinforces its precision and reliability in estimating Chlorophyll-a concentrations.

### *Spatial distribution of the chlorophyll-a and TSI*

*Figure 9a* presents the spatial pattern of Chl-a concentrations across the study area. Chl-a serves as a crucial indicator of phytoplankton biomass and primary productivity, providing insights into the ecological status of aquatic environments. The concentration values, ranging from 0 to 250 mg/m$^3$, are visually represented by a progressive color gradient, with deep purple indicating the lowest concentrations and bright red signifying the highest. Analysis of the map reveals a heterogeneous distribution of Chl-a. Low Chl-a concentrations are observed in limited, often narrower, riverine segments and less impacted peripheral areas. Conversely, significant portions of the water bodies, particularly the broader river sections and potentially larger interconnected water systems, exhibit elevated Chl-a concentrations. Specifically, areas depicted in yellow, orange, and red, corresponding to concentrations exceeding 60 mg/m$^3$, indicate regions of high to very high phytoplankton biomass. These elevated Chl-a levels are predominantly observed in the central and southern parts of the mapped region, suggesting localized or widespread algal proliferation. Such hotspots underscore areas experiencing heightened primary productivity, which is often indicative of nutrient enrichment and potential water quality degradation.

The spatial pattern of the TSI within the study region is illustrated in *Figure 9b*, providing a comprehensive assessment of eutrophication levels. The TSI, classified into


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

seven distinct categories ranging from oligotrophic (0–20) to extremely hypereutrophic (80–100), offers a standardized metric for evaluating the trophic status. The color scale on the map corresponds to these categories, progressing from dark blue (oligotrophic) to dark red (extremely hypereutrophic). Quantitative analysis of the spatial coverage indicates that the "eutrophic" category (50–70) dominates the study area, encompassing 56.69% of the water body surface. This widespread prevalence of eutrophic conditions highlights a significant challenge in managing nutrient inputs. "Upper mesotrophic" areas (40–50) account for 14.71%, while "hypereutrophic" conditions (70–80) are observed in 10.52% of the area, indicating severe nutrient over-enrichment in substantial localized regions. The less impacted categories, "mesotrophic" (30–40), "lower mesotrophic" (20–30), and "oligotrophic" (0–20), represent 9.49%, 4.74%, and 2.73% of the area, respectively, primarily concentrated in certain upstream or less disturbed segments. "Extremely hypereutrophic" conditions (80–100), though representing a smaller proportion (1.12%), signify critical areas of severe ecological imbalance. The aggregated distribution of TSI values, particularly the dominance of eutrophic and hypereutrophic conditions, strongly correlates with regions exhibiting elevated Chl-a concentrations (as shown in *Fig. 9a*), collectively pointing to a pervasive issue of anthropogenic nutrient loading impacting the aquatic ecosystems within the study domain.

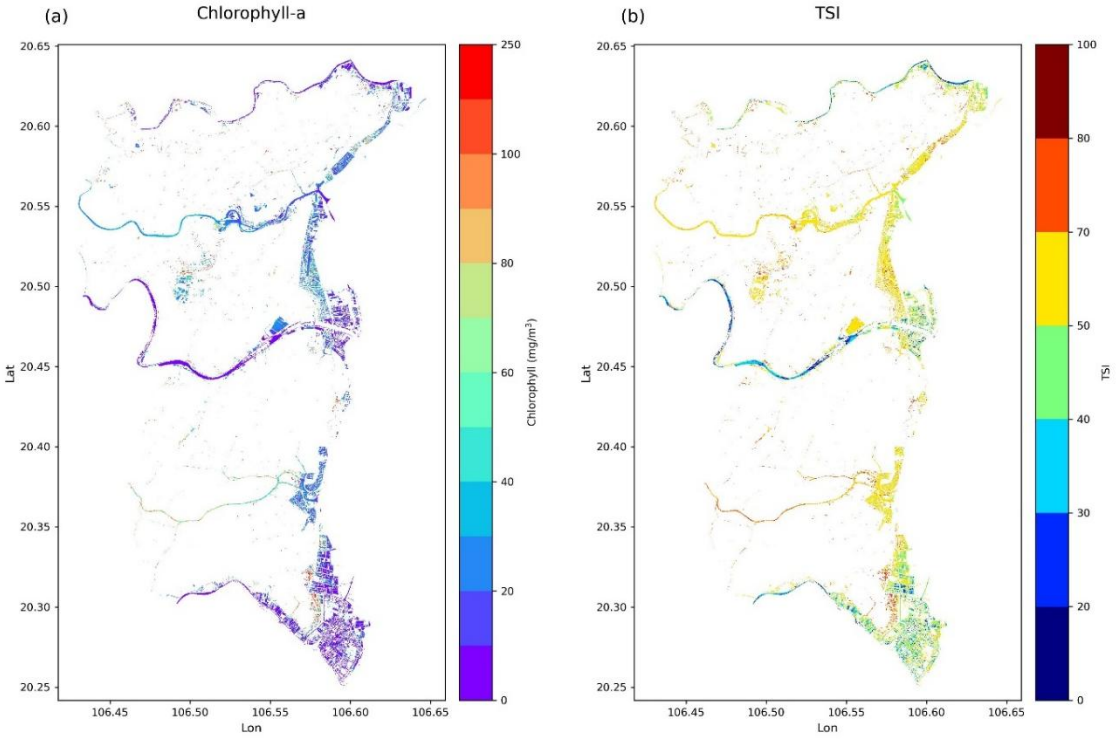


***Figure 9.*** *Spatial distribution of (a) Chl-a and (b) TSI values in the study area*

The provided histograms (*Fig. 10*) illustrate the distributions of Chlorophyll-a (Chl-a) and Trophic State Index (TSI) values within the study area. Chl-a concentrations (*Fig. 10a*) exhibit a highly right-skewed distribution, peaking at low values but extending significantly towards high concentrations (Max: 249.96 mg/m$^3$, Mean: 28.76 mg/m$^3$, Median: 18.46 mg/m$^3$, Std: 33.48 mg/m$^3$). This pronounced skewness suggests that while much of the area maintains low primary productivity, localized zones experience


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

considerable algal growth, indicative of potential eutrophication. Conversely, TSI values (*Fig. 10b*) present a more symmetrical, though still slightly right-skewed, distribution, with a prominent mode around 60 (Max: 84.86, Mean: 53.99, Median: 56.5, Std: 14.63). The close proximity of the mean and median, along with a lower standard deviation compared to Chl-a, signifies a more concentrated spread of trophic states. The prevalence of TSI values in the eutrophic range suggests that a substantial portion of the study area is characterized by elevated nutrient enrichment, reinforcing the implications drawn from the Chl-a distribution regarding the overall ecological health of the system.

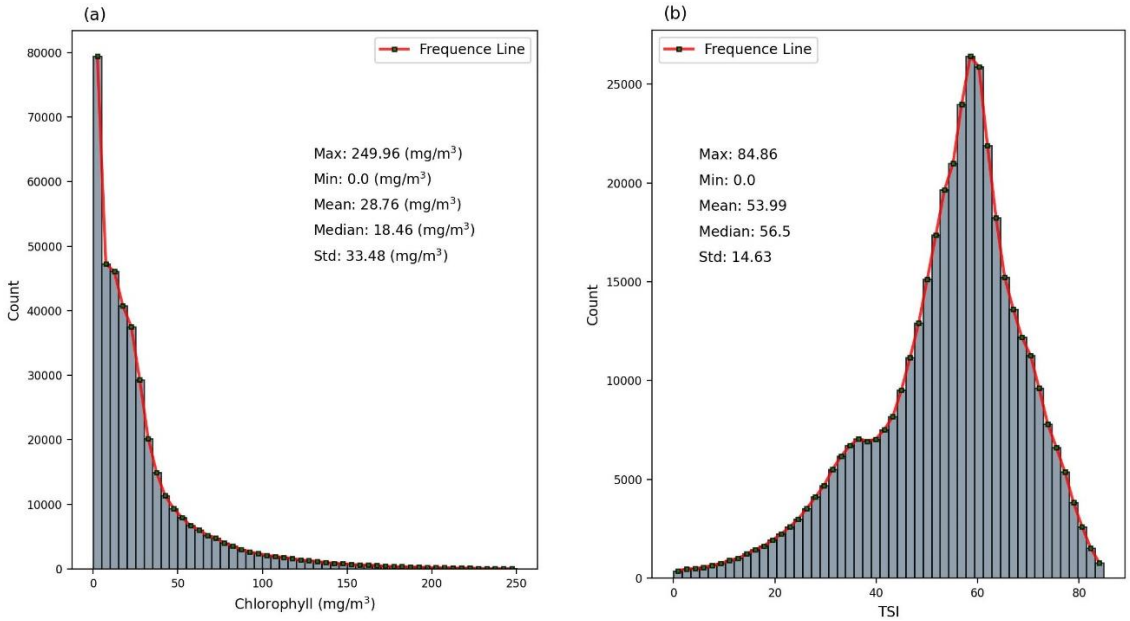


***Figure 10.*** *Histogram plots of (a) Chl-a and (b) TSI values in the study area*

## *Limitation*

Beyond the promising results achieved, this study also has several limitations that should be noted. Firstly, the field sampling dates (September 30 – October 2, 2024) do not perfectly align with the satellite imagery acquisition date (October 6, 2024), resulting in a temporal mismatch of 4–6 days. Therefore, this temporal gap means the estimated Chlorophyll a concentrations may not perfectly reflect the actual conditions at the time the satellite image was captured. Water quality parameters, especially Chlorophyll a concentration, are dynamic and can change due to wind, currents, tidal cycles, or biological processes within a few days. In this work, we proceeded based on the assumption that the overall concentration and spatial pattern of Chlorophyll a remained relatively stable across the study area during this short period. While this assumption is reasonable for a large-scale mapping objective, the mismatch introduces potential uncertainty or error into the empirical model, which could slightly reduce the overall accuracy of the chlorophyll a estimates compared to using a perfectly aligned image. Secondly, although the optimized MLR model demonstrated high predictive performance, achieving an $R^2$ of 0.950 and an RMSE of 9.534 mg/m$^3$ in the training phase, and an $R^2$ of 0.944 and an RMSE of 9.366 mg/m$^3$ in the testing phase, a notable limitation of this work is the constrained sample size of ground truth measurements utilized for both training (N = 27) and testing (N = 14). Moving forward, we plan to collect more samples


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

in the upcoming project to increase our data size. At that time, we can adopt a more robust evaluation method, such as the cross-validation approach. Furthermore, the MLR model assumes a linear and additive relationship between explanatory variables. This assumption can be problematic given that Chl-a concentrations exhibit non-linear and unstable characteristics, influenced by numerous anthropogenic and hydro-meteorological factors. This mismatch can lead to poor or unreliable predictive performance when applied to complex real-world data. Additionally, MLR is less effective in handling complex data patterns and non-linear relationships. More sophisticated machine learning algorithms, such as Random Forest, Support Vector Machine, Neural Networks, and Deep Learning, are generally considered superior to traditional linear regression in addressing complex data patterns and non-linear relationships. Moreover, MLR is sensitive to variations across different regions or environmental conditions. Specifically, MLR models assume that training and testing data originate from the same distribution. However, data from diverse environmental conditions can impact the performance of empirical models and reduce accuracy when applied to other study areas. Finally, the performance of the MLR model is also dependent on atmospheric correction methods. If these methods are not robust enough to handle the complexity of in-water optical constituents, it could lead to uncertainties in reflectance data, consequently affecting the accuracy of Chl-a estimations.

## Conclusion

This study successfully developed and optimized a Multiple Linear Regression (MLR) model for estimating and mapping Chlorophyll-a (Chl-a) concentrations and the Trophic State Index (TSI) in coastal aquaculture areas of Thai Binh province, Vietnam. Leveraging in situ Chl-a data, field spectral measurements, and Sentinel-2 satellite imagery, our methodology involved a comprehensive Pearson correlation analysis to identify optimal band ratios and combinations from synthetic Sentinel-2 surface reflectances. This rigorous selection process ensured that only variables with the highest correlation coefficients were utilized as inputs for the subsequent MLR model development.

The optimized MLR model demonstrated high predictive performance, achieving an $R^2$ of 0.950 and a root mean square error (RMSE) of 9.534 mg/m$^3$ in the training phase, and an $R^2$ of 0.944 and an RMSE of 9.366 mg/m$^3$ in the testing phase. These robust results underscore the significant potential of Sentinel-2 data as a freely accessible, high-spatial-resolution resource for Chl-a research and monitoring. The application of the optimized model to a cloud-minimal Sentinel-2 image (acquired on October 06, 2024) enabled the effective generation of spatial distribution maps for Chl-a and TSI across the study region.

The resulting Chl-a and TSI maps provide crucial insights into the spatial heterogeneity and trophic states of coastal aquaculture zones, facilitating the identification of areas susceptible to eutrophication and supporting proactive strategies for mitigating the risks associated with harmful algal blooms. This research significantly enhances our understanding of aquatic environmental dynamics and contributes to the promotion of sustainable aquaculture practices. While the proposed method demonstrates high applicability within the study region, future research should focus on validating its broader applicability to water bodies with diverse optical properties and geographical conditions.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

**Acknowledgements.** Financial support for this research was provided by the Vietnam Academy of Science and Technology through project KHCBBI.02/24-25. The authors extend their sincere gratitude to the European Space Agency (ESA) for the provision of satellite data. Sincere thanks are extended to anonymous reviewers for their valuable comments and suggestions that helped improve the clarity of the manuscript.

## REFERENCES

[1] Aizaki, M., Otsuki, A., Fukushima, T., Hosomi, M., Muraoka, K. (1981): Application of Carlson's trophic state index to Japanese lakes and relationships between the index and other parameters: with 2 figures and 4 tables in the text. – Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen 21(1): 675-681.

[2] Akbarnejad Nesheli, S., Quackenbush, L. J., McCaffrey, L. (2024): Estimating chlorophyll-a and phycocyanin concentrations in inland temperate lakes across New York State using Sentinel-2 images: application of Google Earth engine for efficient satellite image processing. – Remote Sensing 16(18): 3504.

[3] Alawadi, F. (2010): Detection of surface algal blooms using the newly developed algorithm surface algal bloom index (SABI). – Paper presented at the Remote Sensing of the Ocean, Sea Ice, and Large Water Regions. 20-23 September 2010, Toulouse, France.

[4] Baek, J.-Y., Jo, Y.-H., Kim, W., Lee, J.-S., Jung, D., Kim, D.-W., Nam, J. (2019): A new algorithm to estimate chlorophyll-a concentrations in turbid yellow sea water using a multispectral sensor in a low-altitude remote sensing system. – Remote Sensing 11(19): 2257.

[5] Bertone, E., Ajmar, A., Tonolo, F. G., Dunn, R. J., Doriean, N. J., Bennett, W. W., Purandare, J. (2024): Satellite-based estimation of total suspended solids and chlorophyll-a concentrations for the Gold Coast Broadwater, Australia. – Marine Pollution Bulletin 201: 116217.

[6] Cairo, C., Barbosa, C., Lobo, F., Novo, E., Carlos, F., Maciel, D., Flores Júnior, R., Silva, E., Curtarelli, V. (2019): Hybrid chlorophyll-a algorithm for assessing trophic states of a tropical Brazilian reservoir based on msi/sentinel-2 data. – Remote Sensing 12(1): 40.

[7] Cao, Z., Ma, R., Duan, H., Pahlevan, N., Melack, J., Shen, M., Xue, K. (2020): A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. – Remote Sensing of Environment 248: 111974.

[8] Carlson, R. E. (1977): A trophic state index for lakes 1. – Limnology and Oceanography 22(2): 361-369.

[9] Carstensen, J., Klais, R., Cloern, J. E. (2015): Phytoplankton blooms in estuarine and coastal waters: seasonal patterns and key species. – Estuarine, Coastal and Shelf Science 162: 98-109.

[10] Chen, F., Li, S., Song, K. (2024): Remote sensing of lake chlorophyll-a in Qinghai-Tibet Plateau responding to climate factors: implications for oligotrophic lakes. – Ecological Indicators 159: 111674.

[11] Chislock, M. F., Doster, E., Zitomer, R. A., Wilson, A. E. (2013): Eutrophication: causes, consequences, and controls in aquatic ecosystems. – Nature Education Knowledge 4(4): 10.

[12] Du, Y., Zhang, X., Ma, S., Yao, N. (2024): Chlorophyll-a concentration variations in Bohai sea: impacts of environmental complexity and human activities based on remote sensing technologies. – Big Data Research 36: 100440.

[13] Duan, H., Zhang, Y., Zhang, B., Song, K., Wang, Z. (2007): Assessment of chlorophyll-a concentration and trophic state for Lake Chagan using Landsat TM and field spectral data. – Environmental Monitoring and Assessment 129: 295-308.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

[14] Duan, H., Ma, R., Xu, X., Kong, F., Zhang, S., Kong, W., Hao, J., Shang, L. (2009): Two-decade reconstruction of algal blooms in China's Lake Taihu. – Environmental Science & Technology 43(10): 3522-3528.
[15] Duan, H., Ma, R., Xu, J., Zhang, Y., Zhang, B. (2010): Comparison of different semi-empirical algorithms to estimate chlorophyll-a concentration in inland lake water. – Environmental monitoring and assessment 170: 231-244.
[16] Duan, H., Ma, R., Hu, C. (2012): Evaluation of remote sensing algorithms for cyanobacterial pigment retrievals during spring bloom formation in several lakes of East China. – Remote Sensing of Environment 126: 126-135.
[17] Eaton, A. D., Clesceri, L. S., Greenberg, A. E., Franson, M. A. H. (1995): Standard Methods for the Examination of Water and Wastewater. – American Public Health Association, Washington, DC.
[18] Gilerson, A. A., Gitelson, A. A., Zhou, J., Gurlin, D., Moses, W., Ioannou, I., Ahmed, S. A. (2010): Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. – Optics Express 18(23): 24109-24125.
[19] Gitelson, A. (1992): The peak near 700 nm on radiance spectra of algae and water: relationships of its magnitude and position with chlorophyll concentration. – International Journal of Remote Sensing 13(17): 3367-3373.
[20] Gitelson, A., Kondratyev, K. Y. (1991): Optical models of mesotrophic and eutrophic water bodies. – International Journal of Remote Sensing 12(3): 373-385.
[21] Gitelson, A., Garbuzov, G., Szilagyi, F., Mittenzwey, K., Karnieli, A., Kaiser, A. (1993): Quantitative remote sensing methods for real-time monitoring of inland waters quality. – International Journal of Remote Sensing 14(7): 1269-1295.
[22] Gitelson, A. A., Dall'Olmo, G., Moses, W., Rundquist, D. C., Barrow, T., Fisher, T. R., Gurlin, D., Holz, J. (2008): A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: validation. – Remote Sensing of Environment 112(9): 3582-3593.
[23] Gitelson, A. A., Gao, B.-C., Li, R.-R., Berdnikov, S., Saprygin, V. (2011): Estimation of chlorophyll-a concentration in productive turbid waters using a hyperspectral imager for the Coastal Ocean—the Azov Sea case study. – Environmental Research Letters 6(2): 024023.
[24] Gower, J., King, S., Borstad, G., Brown, L. (2005): Detection of intense plankton blooms using the 709 nm band of the MERIS imaging spectrometer. – International Journal of Remote Sensing 26(9): 2005-2012.
[25] Guo, L., Xiu, P., Chai, F., Xue, H., Wang, D., Sun, J. (2017): Enhanced chlorophyll concentrations induced by Kuroshio intrusion fronts in the northern South China Sea. – Geophysical Research Letters 44(22): 11,565-511,572.
[26] Gurlin, D., Gitelson, A. A., Moses, W. J. (2011): Remote estimation of chl-a concentration in turbid productive waters—return to a simple two-band NIR-red model? – Remote Sensing of Environment 115(12): 3479-3490.
[27] Ha, N. T. T., Thao, N. T. P., Koike, K., Nhuan, M. T. (2017): Selecting the best band ratio to estimate chlorophyll-a concentration in a tropical freshwater lake using sentinel 2A images from a case study of Lake Ba Be (Northern Vietnam). – ISPRS International Journal of Geo-Information 6(9): 290.
[28] Hestir, E. L., Brando, V. E., Bresciani, M., Giardino, C., Matta, E., Villa, P., Dekker, A. G. (2015): Measuring freshwater aquatic ecosystems: the need for a hyperspectral global mapping satellite mission. – Remote Sensing of Environment 167: 181-195.
[29] Hu, C., Lee, Z., Franz, B. (2012): Chlorophyll a algorithms for oligotrophic oceans: a novel approach based on three-band reflectance difference. – Journal of Geophysical Research: Oceans 117(C1).
[30] Jang, W., Kim, J., Kim, J. H., Shin, J.-K., Chon, K., Kang, E. T., Park, Y., Kim, S. (2024): Evaluation of Sentinel-2 based chlorophyll-a estimation in a small-scale reservoir: assessing accuracy and availability. – Remote Sensing 16(2): 315.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

[31] Johansen, R. A., Reif, M. K., Saltus, C. L., Pokrzywinski, K. L. (2022): A Review of Empirical Algorithms for the Detection and Quantification of Harmful Algal Blooms Using Satellite-Borne Remote Sensing. – The U.S. Army Engineer Research and Development Center, Vicksburg, MS.
[32] Johansen, R. A., Reif, M. K., Saltus, C. L., Pokrzywinski, K. L. (2024): A broadscale assessment of sentinel-2 imagery and the google earth engine for the nationwide mapping of chlorophyll a. – Sustainability 16(5): 2090.
[33] Kakade, A., Salama, E.-S., Han, H., Zheng, Y., Kulshrestha, S., Jalalah, M., Harraz, F. A., Alsareii, S. A., Li, X. (2021): World eutrophic pollution of lake and river: biotreatment potential and future perspectives. – Environmental Technology & Innovation 23: 101604.
[34] Khan, F. A., Naushin, F., Rehman, F., Masoodi, A., Irfan, M., Hashmi, F., Ansari, A. A. (2014): Eutrophication: Global Scenario and Local Threat to Dynamics of Aquatic Ecosystems. – In: Ansari, A. A., Gill, S. S. (eds.) – Eutrophication: Causes, Consequences and Control. Vol. 2. Springer, Dordrecht, pp. 17-27.
[35] Kim, H. H., Ko, B. C., Nam, J. Y. (2016): Predicting chlorophyll-a using Landsat 8 OLI sensor data and the non-linear RANSAC method–a case study of Nakdong River, South Korea. – International Journal of Remote Sensing 37(14): 3255-3271.
[36] Kim, S. W., Jung, D., Choung, Y.-J. (2020): Development of a multiple linear regression model for meteorological drought index estimation based on Landsat satellite imagery. – Water 12(12): 3393.
[37] Kim, Y. W., Kim, T., Shin, J., Lee, D.-S., Park, Y.-S., Kim, Y., Cha, Y. (2022): Validity evaluation of a machine-learning model for chlorophyll a retrieval using Sentinel-2 from inland and coastal waters. – Ecological Indicators 137: 108737.
[38] Kuhn, C., de Matos Valerio, A., Ward, N., Loken, L., Sawakuchi, H. O., Kampel, M., Richey, J., Stadler, P., Crawford, J., Striegl, R. (2019): Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-a and turbidity. – Remote Sensing of Environment 224: 104-118.
[39] Lao, Q., Liu, S., Ling, Z., Jin, G., Chen, F., Chen, C., Zhu, Q. (2023a): External dynamic mechanisms controlling the periodic offshore blooms in Beibu Gulf. – Journal of Geophysical Research: Oceans 128(6): e2023JC019689.
[40] Lao, Q., Liu, S., Wang, C., Chen, F. (2023b): Global warming weakens the ocean front and phytoplankton blooms in the Luzon Strait over the past 40 years. – Journal of Geophysical Research: Biogeosciences 128(12): e2023JG007726.
[41] Li, S., Song, K., Wang, S., Liu, G., Wen, Z., Shang, Y., Lyu, L., Chen, F., Xu, S., Tao, H. (2021): Quantification of chlorophyll-a in typical lakes across China using Sentinel-2 MSI imagery with machine learning algorithm. – Science of the Total Environment 778: 146271.
[42] Li, Z., Yang, X., Zhou, T., Cai, S., Zhang, W., Mao, K., Ou, H., Ran, L., Yang, Q., Wang, Y. (2024): Monitoring chlorophyll-a concentration variation in fish ponds from 2013 to 2022 in the Guangdong-Hong Kong-Macao Greater Bay Area, China. – Remote Sensing 16(11): 2033.
[43] Martins, V. S., Barbosa, C. C. F., De Carvalho, L. A. S., Jorge, D. S. F., Lobo, F. d. L., Novo, E. M. L. d. M. (2017): Assessment of atmospheric correction methods for Sentinel-2 MSI images applied to Amazon floodplain lakes. – Remote Sensing 9(4): 322.
[44] Matthews, M. W. (2011): A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. – International Journal of Remote Sensing 32(21): 6855-6899.
[45] Matthews, M. W., Bernard, S., Robertson, L. (2012): An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. – Remote Sensing of Environment 124: 637-652.
[46] Matus-Hernández, M. Á., Hernández-Saavedra, N. Y., Martínez-Rincón, R. O. (2018): Predictive performance of regression models to estimate Chlorophyll-a concentration based on Landsat imagery. – PLoS One 13(10): e0205682.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

[47] Mishra, P., Naik, S., Babu, P. V., Pradhan, U., Begum, M., Kaviarasan, T., Vashi, A., Bandyopadhyay, D., Ezhilarasan, P., Panda, U. S. (2022): Algal bloom, hypoxia, and mass fish kill events in the backwaters of Puducherry, Southeast coast of India. – Oceanologia 64(2): 396-403.
[48] Mishra, S., Mishra, D. R. (2012): Normalized difference chlorophyll index: a novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. – Remote Sensing of Environment 117: 394-406.
[49] Moses, W. J., Gitelson, A. A., Berdnikov, S., Povazhnyy, V. (2009): Estimation of chlorophyll-a concentration in case II waters using MODIS and MERISdata—successes and challenges. – Environmental Research Letters 4(4): 045005.
[50] Ogashawara, I., Kiel, C., Jechow, A., Kohnert, K., Ruhtz, T., Grossart, H.-P., Hölker, F., Nejstgaard, J. C., Berger, S. A., Wollrab, S. (2021): The use of Sentinel-2 for chlorophyll-a spatial dynamics assessment: a comparative study on different lakes in Northern Germany. – Remote Sensing 13(8): 1542.
[51] Pan, W., Yu, F., Li, J., Li, C., Ye, M. (2025): Quantification of chlorophyll-a in inland waters by remote sensing algorithm based on modified equivalent spectra of Sentinel-2. – Ecological Informatics 87: 103061.
[52] Prepas, E., Charette, T. (2003): Worldwide eutrophication of water bodies: causes, concerns, controls. – Treatise on Geochemistry 9: 612.
[53] Prieto, A. J., Silva, A., de Brito, J., Macías-Bernal, J. M., Alejandre, F. J. (2017): Multiple linear regression and fuzzy logic models applied to the functional service life prediction of cultural heritage. – Journal of Cultural Heritage 27: 20-35.
[54] Qin, Z., Ruan, B., Yang, J., Wei, Z., Song, W., Sun, Q. (2022): Long-term dynamics of chlorophyll-a concentration and its response to human and natural factors in Lake Taihu based on MODIS data. – Sustainability 14(24): 16874.
[55] Quang Vinh, P., Thi Thu Ha, N., Thao, N. T. P., Ha Linh, P., Hien, T. T., Parsons, M. (2024): Investigating the spatial-temporal pattern of the trophic state index in Hanoi's largest urban lake (Vietnam) using Sentinel-2 images. – Urban Water Journal 21(6): 774-789.
[56] Salem, S. I., Higa, H., Kim, H., Kobayashi, H., Oki, K., Oki, T. (2017): Assessment of chlorophyll-a algorithms considering different trophic statuses and optimal bands. – Sensors 17(8): 1746.
[57] Salls, W. B., Schaeffer, B. A., Pahlevan, N., Coffer, M. M., Seegers, B. N., Werdell, P. J., Ferriby, H., Stumpf, R. P., Binding, C. E., Keith, D. J. (2024): Expanding the application of Sentinel-2 chlorophyll monitoring across United States lakes. – Remote Sensing 16(11): 1977.
[58] Sivonen, K. (1996): Cyanobacterial toxins and toxin production. – Phycologia 35(sup6): 12-24.
[59] Smith, G. M., Milton, E. J. (1999): The use of the empirical line method to calibrate remotely sensed data to reflectance. – International Journal of Remote Sensing 20(13): 2653-2662.
[60] Smith, M. E., Lain, L. R., Bernard, S. (2018): An optimized chlorophyll a switching algorithm for MERIS and OLCI in phytoplankton-dominated waters. – Remote Sensing of Environment 215: 217-227.
[61] Smith, V. H. (2003): Eutrophication of freshwater and coastal marine ecosystems a global problem. – Environmental Science and Pollution Research 10: 126-139.
[62] Sousa, S., Martins, F. G., Alvim-Ferraz, M. C., Pereira, M. C. (2007): Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. – Environmental Modelling & Software 22(1): 97-103.
[63] Thao, N. T. P., Ha, N. T. T., Vinh, P. Q., Hien, T. T., Thanh, D. X. (2024): A multivariate linear regression model for estimating chlorophyll-a concentration in Quan Son Reservoir (Hanoi, Vietnam) using Sentinel-2B Imagery. – Vietnam Journal of Earth Sciences 46(3): 360-380.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569

[64] Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., Nõges, T. (2016): First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery. – Remote Sensing 8(8): 640.
[65] Tóth, V., Ladányi, M., Jung, A. (2021): Adaptation and validation of a Sentinel-based chlorophyll-a retrieval software for the central European freshwater lake, Balaton. PFG–J. – Photogrammetry, Remote Sens. Geoinform. Scie. 89: 335-344.
[66] Tran, D.-P., Liou, Y.-A. (2022): Estimation of monthly air temperature using Random Forest algorithm. – Paper presented at the 2022 IET International Conference on Engineering Technologies and Applications (IET-ICETA). October 14–16, Changhua, Taiwan.
[67] Tran, D.-P., Liou, Y.-A. (2024): Creating a spatially continuous air temperature dataset for Taiwan using thermal remote-sensing data and machine learning algorithms. – Ecological Indicators 158: 111469.
[68] Tran, T. T. V., Hoang, L. T. T., Le, B. B. (2017): Sinh khí hậu và phát triển rừng ngập mặn ven biển tỉnh Thái Bình. – VNU Journal of Science: Earth and Environmental Sciences 33(1).
[69] Vanhellemont, Q., Ruddick, K. (2016): Acolite for Sentinel-2: aquatic applications of MSI imagery. – Paper presented at the 2016 ESA Living Planet Symposium, Prague, Czech Republic.
[70] Vinh, P. Q., Ha, N. T. T., Thao, N. T. P., Linh, N. T., Oanh, L. T., Phuong, L. T., Huyen, N. T. T. (2022): Monitoring the trophic state of shallow urban lakes using Landsat 8/OLI data: a case study of lakes in Hanoi (Vietnam). – Frontiers of Earth Science: 1-16.
[71] Wang, J., Zhang, Y., Yang, F., Cao, X., Bai, Z., Zhu, J., Chen, E., Li, Y., Ran, Y. (2015): Spatial and temporal variations of chlorophyll-a concentration from 2009 to 2012 in Poyang Lake, China. – Environmental Earth Sciences 73: 4063-4075.
[72] Wang, S., Li, J., Zhang, B., Spyrakos, E., Tyler, A. N., Shen, Q., Zhang, F., Kuster, T., Lehmann, M. K., Wu, Y. (2018): Trophic state assessment of global inland waters using a MODIS-derived Forel-Ule index. – Remote Sensing of Environment 217: 444-460.
[73] Woźniak, M., Bradtke, K. M., Krężel, A. (2014): Comparison of satellite chlorophyll a algorithms for the Baltic Sea. – Journal of Applied Remote Sensing 8(1): 083605-083605.
[74] Wynne, T., Stumpf, R., Tomlinson, M., Warner, R., Tester, P., Dyble, J., Fahnenstiel, G. (2008): Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. – International Journal of Remote Sensing 29(12): 3665-3672.
[75] Xing, K., Guo, H., Sun, Y., Huang, Y. (2005): Assessment of the spatial-temporal eutrophic character in the Lake Dianchi. – Journal of Geographical Sciences 15(1): 37-43.
[76] Yang, W., Matsushita, B., Chen, J., Fukushima, T., Ma, R. (2010): An enhanced three-band index for estimating chlorophyll-a in turbid case-II waters: case studies of Lake Kasumigaura, Japan, and Lake Dianchi, China. – IEEE Geoscience and Remote Sensing Letters 7(4): 655-659.
[77] Zhang, Y., Pulliainen, J., Koponen, S., Hallikainen, M. (2003): Empirical algorithms for Secchi disk depth using optical and microwave remote sensing data from the Gulf of Finland and the Archipelago Sea. – Boreal Environment Research 8(3): 251-261.
[78] Zhang, Y., Ma, R., Duan, H., Loiselle, S., Xu, J. (2014): A spectral decomposition algorithm for estimating chlorophyll-a concentrations in Lake Taihu, China. – Remote Sensing 6(6): 5090-5106.
[79] Zhao, M., Bai, Y., Li, H., He, X., Gong, F., Li, T. (2022): Fluorescence line height extraction algorithm for the geostationary ocean color imager. – Remote Sensing 14(11): 2511.


http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2306_1154111569