

Assessing the Effectiveness of a Ridge Regression Model for Estimating Chlorophyll-A Content from Sentinel-3 Olci Images in Vietnamese Coastal Water

TRAN Thanh Ha¹, NGUYEN Minh Hai^{1,*}, TRAN Thi Hoa¹, NGUYEN Minh Ngoc²,
PHAN Viet Son¹

¹ Hanoi University of Mining and Geology, Hanoi, Vietnam

² Vietnam National Space Centre, Vietnam Academy of Science and Technology, Hanoi, Vietnam

* Corresponding author: nguyenminhhai@humg.edu.vn

Abstract: Chlorophyll-a (Chl-a) is a crucial indicator of water quality, and its concentration can be determined through various methods, including in-situ measurements, laboratory analysis, and remote sensing. While in-situ measurements are highly accurate, they are often costly and time-consuming. Remote sensing techniques, on the other hand, provide an efficient way to monitor Chl-a over large areas and offer diverse options for different temporal and spatial scales. However, their accuracy is typically low in coastal regions. This study presents an advanced method for estimating Chl-a content in the Southern coastal sea of Vietnam using Sentinel-3 Ocean and Land Color Instrument (OLCI) data. The dataset consists of 28 in-situ measurements, with 22 used to build the model and 6 for testing. The proposed approach develops a Ridge Regression (RR) model to address the multicollinearity issue commonly found in multiple linear regression models. The model's inputs include Chl-a estimation from established algorithms – OC2, OC3, NDCI, and Gurlin 11 – as well as two certain accompanying Sentinel – 3 products, CHL_NN and CHL_OC4ME. The regularization parameter λ was chosen by testing values from 0.1 to 1, with the optimal value of 0.4 selected because it ensured that the Variance Inflation Factor (VIF) for all variables was less than 5, thus overcoming multicollinearity. The effectiveness of the algorithm was demonstrated by determined coefficients (R^2) of 0.811 and a root mean square error (RSME) of 0.456, indicating a relatively high estimation accuracy. Furthermore, a Chl-a content distribution map for the study was created from this algorithm, which achieved spatial agreement with the CHL_NN results both qualitatively through map analysis and quantitatively through distribution graph analysis. The results obtained from the RR model, developed with a suitable value, are promising and effective for coastal areas.

Keywords: Chlorophyll-a, Coastal Area, Ridge Regression, Sentinel-3, Remote Sensing, Water Quality

1. Introduction

Global coastal areas significantly face increasing pressure from human activities such as urban development, coastal industry, aquaculture, and the impacts of climate change (Halpern et al., 2015). In Vietnam, the coastal area of the Southern is a key economic zone with a high density of ports, industrial activities, and aquaculture. Consequently, the resulting increase in nutrient and suspended matter loads has led to a decline in water quality and serious impacts on the marine ecosystem (Vinh, 2016). Therefore, continuous, effective, and spatially extensive water quality monitoring is urgently needed.

Water quality is a complex index indicated by various physical, chemical and biological parameters. From a biological perspective, Chlorophyll-a (Chl-a) is a key pigment in algae and phytoplankton (Gholizadeh et al., 2016; Werdell et al., 2013), and its concentration serves as a vital indicator of water quality. Elevated levels often suggest poor water quality due to increased algae and cyanobacteria biomass, potentially caused by nutrient pollution from sources like agricultural runoff and wastewater. Monitoring Chl-a levels helps assess the trophic state of water bodies and identify potential issues like eutrophication, which can negatively impact aquatic ecosystems.

While in-situ measuring is the most accurate approach to assess water quality based on direct collected samples, it is costly, requiring time-consuming, temporally discontinuous, and spatially insufficient (Tyler et al., 2006). In contrast, remote sensing techniques have proven to be powerful tools for estimating and monitoring water quality indices over large areas with high repeatability (Dube et al., 2015; Matthews, 2011). Sensors like MODIS, MERIS, Landsat OLI, Sentinel-2 MSI, and more recently, Sentinel-3 OLCI, have been used to retrieve Chl-a with varying effectiveness depending on the water's optical properties (Dörnhöfer & Oppelt, 2016; Wang & Shi, 2005). The Sentinel-3 OLCI sensor, developed by the European Space Agency (ESA), stands out with 21 continuous spectral bands from 400 to 1020 nm and a

revisit time of 1–2 days. This provides high spectral resolution data and temporal frequency suitable for monitoring complex coastal waters (Binh et al., 2022; EUMETSAT, 2024). The sensor's design makes OLCI particularly sensitive to biological pigments, suspended matter, and dissolved organic matter—typical components in Case 2 waters (Doerffer & Schiller, 2007). Case 2 waters are a type of water environment where optical properties are complex due to the mixing of dissolved matter, sediment, and phytoplankton. The Sentinel-3 OLCI sensor is particularly sensitive to these components.

However, the quantitative extraction of Chl-a from the spectral reflectance data remains a challenge due to the effects of multicollinearity between spectral bands, high noise, and complex spatio-temporal variability in coastal waters (Kutser et al., 2016; Vanhellemont & Ruddick, 2021). Traditional methods, such as linear regression or spectral ratio indices, often lack flexibility and are prone to overfitting when applied in Case 2 environments (Su et al., 2021). To overcome these limitations, regularized regression methods such as Ridge Regression (L2 regularization), LASSO, and Elastic Net have been introduced to increase stability, reduce multicollinearity, and control overfitting (Friedman et al., 2010; Hoerl & Kennard, 1970). Recent studies have shown that Ridge Regression often outperforms traditional regression models in estimating Chl-a from Sentinel-2 images (Ivanda et al., 2021).

In Vietnam, recent research on water quality remote sensing has primarily focused on applying machine learning algorithms and regularized regression model to estimate parameters like Chl-a and total suspended (TSS) using Sentinel-2 MSI satellite data (Thanh Son & Duc Loc, 2024; Thien Phuong Thao et al., 2024). These studies have mainly focused on inland river systems and reservoirs, where hydro-physical conditions are less variable and less affected by complex optical factors. However, the use of Sentinel-3 OLCI data, with its advantages of high revisit frequency and suitable spectral configuration for coastal marine environments, remains significantly limited. This is particularly noteworthy in the context of Case-2 coastal waters, where optical properties are. Currently, there have been no published studies that apply regularized regression models with Sentinel-3 OLCI data for the coastal area of Ho Chi Minh City, Dong Thap, and Vinh Long Province.

This gap highlights the urgent need to expand research on applying machine learning models for coastal water remote sensing using Sentinel-3 data in the future. Based on this, the objectives of this study are to (1) develop and calibrate a Ridge Regression model to estimate Chl-a concentrations from Sentinel-3 OLCI satellite data, and (2) evaluate the model's performance by comparing it with traditional linear models and validating it with in-situ data. The research results are expected to contribute to the improvement of remote sensing methods for water quality monitoring in Vietnam, thereby supporting effective and sustainable marine environmental management.

2. Data Collection and Research Methodologies

2.1 Data Collection

2.1.1 Sentinel-3 images

The rapid advancement of remote sensing technology, particularly with modern ocean-observing satellites like Sentinel-3 with the Ocean and Land Color Instrument (OLCI), has opened new avenues for coastal water quality monitoring. Sentinel-3 OLCI provides continuous spectral information across 21 bands, ranging from 400 to 1020 nm, with a spatial resolution of 300 meters and a revisit time of 1-2 days. These features make it highly suitable for detecting rapid changes in Chlorophyll a (Chl-a), turbidity (TSS), and extreme events such as algal blooms or sediment plumes in tropical coastal waters (EUMETSAT, 2024).

However, converting the spectral reflectance from the satellite into accurate estimates of Chl-a and TSS requires developing appropriate regression models that can fully utilize the rich spectral information while remaining stable and accurate – especially when in-situ sampling is limited. For this study, Sentinel-3 OLCI Level 1 data was obtained from the Copernicus Open Access Hub of the European Space Agency. The specific characteristics of the spectral bands used in this study are summarized in Table 1.

Tab. 1. Sentinel-3 channels

Band	Central wavelength/Spectral bandwidth (nm)	Spatial resolution (m)
B3	442.5/10	300
B4	490/10	300
B5	510/10	300
B6	560/10	300
B7	620/10	300
B8	665/10	300
B9	673.75/7.5	300
B10	681.25/7.5	300
B11	708.75/10	300
B12	753.75/7.5	300

2.1.2. In-situ data

The in-situ data used in this study was collected through measurement stations located in the area, as shown in Figure 1. The dataset contains a total of consists of 28 values collected in the dry season of 2021 (6th of March). Field sample collection took place between 9:30 AM and 2:40 PM. At the water surface, spectral measurements were performed at each sampling point using a handheld PRR 2600. This is a specialized device for measuring spectrum in turbid water, capable of measuring spectral properties in the range of 340 nm to 900 nm. The time of the water surface spectral measurements coincided with the Sentinel-3 OLCI satellite overpass. The methodology for measuring surface reflectance spectra and spectral transformation followed the Mobley method (Mobley, 1999). At each field location, water samples were collected into test tubes, which were then wrapped in a black bag with ice (at a temperature of 0°C) for preservation. The samples were then filtered through a 25mm Whatman GF/F filter, submerged in 10 ml of 90% acetone solution, and analyzed using a U-2900 spectrophotometer.

Of these, 22 values were used to build the model, while the remaining 6 values were used for testing.

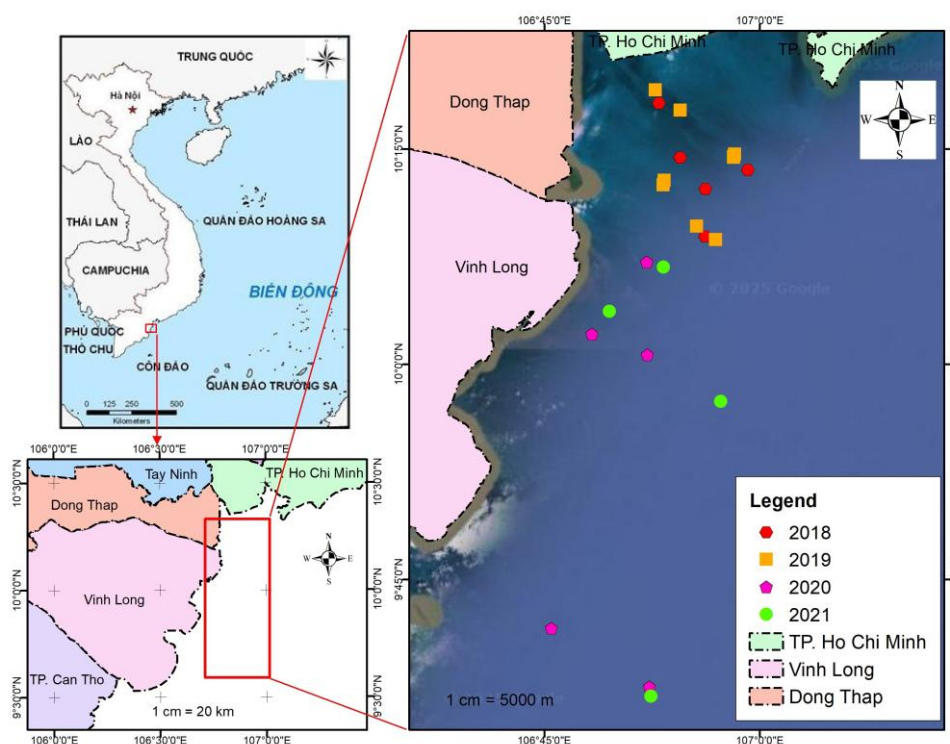


Fig. 1. Positions of samples in the study area (coastal area of Ba Ria – Vung Tau extending from Ho Chi Minh city to Vinh Long province)

2.2. Research Methodologies

2.2.1. Chl-a Inversion Algorithms

Published studies have utilized empirical bio-optical algorithms based on spectral band ratios to estimate Chl-a from remote sensing data. These models generally fall into two categories: those based on the Blue/Green ratio, which are suitable for clear to moderately turbid waters, and those based on the Red/NIR ratio, which are for turbid waters. Given the large study area, which includes major estuaries, the authors selected both types of band ratios to determine the most suitable Chl-a estimation model.

a. Blue/Green ratio-based models

- Ocean Color 2 model (OC2): The OC2 model is a widely used empirical algorithm for estimating Chl-a concentration from ocean color satellite data. It's based on the ratio of blue band (e.g., 490 nm) to green band (e.g., 560 nm). This ratio has been shown to correlate with Chl-a concentration in relatively clear or moderately turbid waters (Merder et al., 2024). This model estimates Chl-a using a polynomial function of this ratio following the formula (1) below, unit mg/m^3 :

$$Chl - a = 10^{(a_0 + a_1 * R + a_2 * R^2 + a_3 * R^3 + a_4 * R^4)} \quad (1)$$

where $R = \log_{10} \left(\frac{R_{rs}(490)}{R_{rs}(560)} \right)$

- Ocean Color 3 model (OC3): The OC3 is an empirical logarithmic algorithm. Developed by NASA, it's widely used for estimating Chl-a from optical remote sensing images, particularly from sensors like SeaWiFS, MODIS, MERIS, and Sentinel-3's OLCI. The algorithm is based on the reflectance ratio between the blue and green spectral bands. These bands are sensitive to the absorption and scattering of light by the Chl-a pigment in water (O'Reilly et al., 1998).

OC3 algorithm employs a fourth-degree polynomial formula based on the logarithm of the reflectance ratio between spectral channels. The formula, which is provided as formula (2) in the document, is as follows:

$$\log_{10}(Chl - a) = a_0 + a_1 R + a_2 R^2 + a_3 R^3 + a_4 R^4 \quad (2)$$

where $R = \log_{10} \left(\frac{R_{rs}(490)}{R_{rs}(560)} \right)$

$R_{rs}(\lambda)$: surface water reflectance at wavelength λ

$a_0, a_1, \dots, a_4$: coefficients calibrated for each specific sensor

Based on the document provided, the OC3 algorithm has been adjusted for the Sentinel-3 OLCI sensor and is referred to as OC3Me or OC3S. The spectral bands used are a blue band at either 443nm (Oa03) or 490nm (Oa05) and a green band at 560 nm (Oa06) (Eplee et al., 2019):

$$\text{with } R = \log_{10} \left(\frac{\max(R_{rs}(443), R_{rs}(490))}{R_{rs}(560)} \right)$$

It's important to note that OC3 algorithm was primarily developed for clear ocean waters (as known as Case-1 waters). In turbid coastal areas, the accuracy of the algorithm can be compromised because of the reflectance from non-chlorophyll matter (Merder et al., 2024).

b. Red/NIR ratio- based models

- Normalized Difference Chlorophyll Index (NDCI) model is an empirical model developed specifically for estuaries and coastal waters (Mishra & Mishra, 2012). This model uses NDCI as an input variable to calculate Chl-a (mg/m^3) based on the following formula (3):

$$Chl - a = a + b * NDCI + c * NDCI^2 \quad (3)$$

where $NDCI = \frac{(R_{rs}(709) - R_{rs}(665))}{(R_{rs}(709) + R_{rs}(665))}$

- Gurlin11 model is an empirical model, developed by Gurlin et al. It uses a second-degree polynomial function based on the band ratio of $R_{rs}(709)/R_{rs}(665)$. This model calculates Chl-a (mg/m^3) using the formula (4): (Gurlin et al., 2011)

$$Chl - a = a * R^2 + b * R + c \quad (4)$$

where $R_{rs}(709)/R_{rs}(665)$.

For a large study area with a complex water environment, a single model is unlikely to accurately estimate Chl-a concentration across the entire region. While Multiple Linear Regression (MLR) could be used to address this, it often encounters the issue of multicollinearity when there are many independent variables. Multicollinearity leads to unstable and unreliable estimates, particularly with incomplete data,

such as in-situ Chl-a measurements. To mitigate this issue and create a more generalized model, the authors propose using a Ridge Regression model.

2.2.2. Ridge Regression model

Traditional methods for measuring Chl-a are expensive and difficult to scale in terms of space or frequency. Additionally, traditional interpolation methods, like kriging and inverse distance weighting (IDW), can result in significant errors because they fail to capture the complex spatio-temporal nature of coastal water environments. While Multiple Linear Regression (MLR) and spectral index-based models are commonly used to estimate Chl-a and turbidity from remote sensing data, they face two key challenges when applied to Sentinel-3 OLCI imagery with many spectral bands and a limited number of in-situ measurement points (typically 10-30 points per season). These challenges are multicollinearity and the risk of overfitting. Multicollinearity causes regression coefficients to fluctuate widely, leading to an unstable model, while overfitting results in poor predictive performance on new data outside the training set (Kim et al., 2022).

Ridge Regression (L2 Regularization) has been increasingly studied and applied in remote sensing data analysis to overcome limitations of traditional methods. It is an extension of linear regression where the loss function minimizes the sum of squared prediction errors and adds a penalty term proportional to the sum of the squared regression coefficients. This prevents the regression coefficients from becoming too large, especially when there are many highly correlated independent variables or when the number of variables exceeds the number of samples. The optimization function for Ridge Regression is shown in formula (5):

$$\text{loss}_{\text{Ridge}} = \sum_{i=1}^n (y_i - x_i^T \beta)^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (5)$$

Where, y_i is the observed value (Chl-a or turbidity), x_i is the vector of spectral band values for the i -th sample, β is the vector of regression coefficients, and λ is the regularization parameter that controls the extent to which the coefficients are shrunk towards zero (0). When $\lambda=0$, Ridge Regression reverts to traditional linear regression; as λ increases, the regression coefficients are adjusted to be smaller, making the model more stable and reducing the risk of overfitting (Ehsanes Saleh et al., 2019).

Compared to Lasso Regression (L1 regularization), Ridge Regression has the advantage of not eliminating any spectral bands from the model. Instead, it reduces the influence of less important bands, making full use of the spectral information from the satellite image. Ridge Regression also has a simple tuning process for the parameter λ . This is often done using cross-validation to find the optimal value that balances accuracy and model generalization. Recent studies have demonstrated the effectiveness of Ridge Regression for estimating Chl-a and turbidity from Sentinel-3 OLCI data, particularly when in-situ samples are limited (Su et al., 2021).

2.2.3. Deploying Ridge Regression algorithms

In Case-2 waters, such as the coastal areas of Ba Ria – Vung Tau, the optical reflectance properties are influenced by multiple components, including Chl-a, colored dissolved organic matter (CDOM), inorganic sediment, and suspended impurities. The relationship between the spectral reflectance signal $R_{rs}(\lambda)$ and parameters like Chl-a and TSS is non-linear and is also affected by noise and multicollinearity between spectral bands. To address these challenges, this study developed a Ridge Regression model, a regularized linear regression model, to estimate Chl-a from Sentinel-3 OLCI satellite imagery. The model is based on the water surface spectral reflectance $R_{rs}(\lambda)$ at specific spectral bands that are highly reflective of different water components in coastal areas. The spectral bands used in the model include 443 nm, 490 nm, 510 nm, 560 nm, 620 nm, 665 nm, and 681 nm. (Su et al., 2021).

To ensure the accuracy of the water surface reflectance, Sentinel-3 OLCI image data was pre-processed using the C2RCC atmospheric correction technique. In-situ Chl-a data were collected simultaneously at 28 coastal survey stations during the dry season of 2021. This data served as the training and calibration set for the model.

The calculation process is as follows:

- Step 1: Collect Sentinel-3 OLCI images and perform atmospheric correction (using C2RCC to extract water surface reflectance $R_{rs}(\lambda)$).
- Step 2: Estimate Chl-a concentration by models mentioned in section 2.2.1.
- Step 3: The input matrix X is normalized using the z-score to bring all variables to the same scale

based on formula (6):

$$x'_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (6)$$

Where μ_j and σ_j are the mean and standard deviation of variable x_j respectively.

- Step 4: Train the Ridge Regression model by optimizing the coefficient β , solving the L2 regularized loss function minimization problem based on (7). The value of λ is chosen based on the variance inflation factor (VIF) with the condition $VIF < 5$ to avoid multicollinearity.

$$\beta = \arg \min_{\beta} \left\{ \sum_{i=1}^n (y_i - x_i^T \beta)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (7)$$

- Step 5: Evaluate the model's performance using metrics such as RMSE and R^2 .

- Step 6: Apply the model to the entire Sentinel-3 OLCI image of the study area to create a spatial distribution map of Chl-a.

3. Results and Discussion

While implementing the models mentioned, two other Chl-a estimation products from ESA, CHL_NN and CHL_OC4ME, were also included in the Ridge Regression model. These products are based on Sentinel-3 data processed with the C2RCC and OC4ME algorithms, respectively. Figure 2 illustrates the correlation of the Chl-a values between these two products, and the in-situ measured data.

The results of the regression models built for the OC2, OC3, NDCI, and Gurlin 11 algorithms to estimate Chl-a concentration in the area are described in Table 2.

The coefficient of determination (R^2) for the single algorithms used to estimate Chl-a concentration is very low. This indicates that a combined-parameter model would be a more suitable approach to improve the accuracy of the estimation.

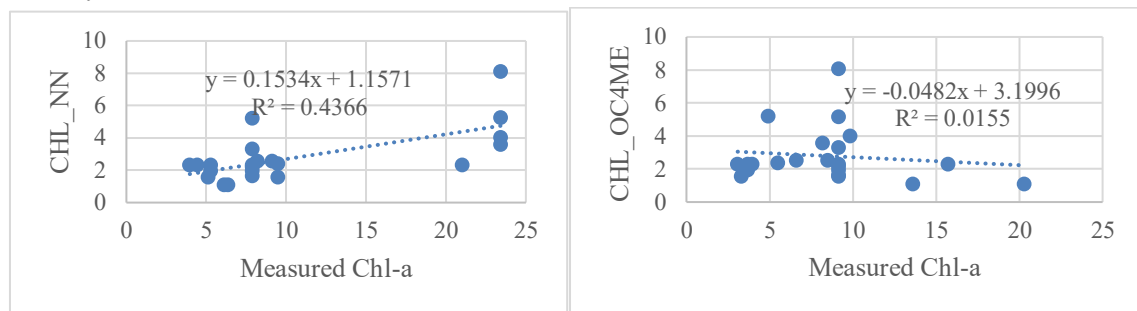


Fig. 2. Graphs show correlation between measured data and data from ESA Chlorophyll products (unit: mg/m^3)

Tab. 2. Regression equation for each model for Chl-a estimation

Model	Regression Equation	R^2
OC2	$y = -8565,2x^4 - 4883,2x^3 - 943,16x^2 - 71,355x - 1,3648$	0,22
OC3	$y = 5,0104x^2 - 5,9x + 2,0604$	0,11
NDCI	$y = -34,893x^2 - 5,5287x + 3,2149$	0,12
Gurlin	$y = -24,323x^2 + 39,885x - 12,803$	0,15

To figure out the most suitable λ value, a range of values from 0.1 to 1 were tested. The goal was to ensure that the Variance Inflation Factor (VIF) for the variables met the required criteria. The results of this testing are shown in Table 3.

Tab. 3. VIF values regarding λ

λ	VIF					
	OC2	OC3	NDCI	Gurlin11	CHL_NN	CHL_OC4ME
0,1	1,871	3,307	28,296	24,857	2,946	2,879
0,2	1,801	2,966	12,488	11,320	2,732	2,521
0,3	1,739	2,751	7,118	6,695	2,579	2,329
0,4	1,682	2,583	4,666	4,564	2,452	2,190

0,5	1,630	2,440	3,342	3,400	2,340	2,077
0,6	1,581	2,314	2,546	2,688	2,239	1,980
0,7	1,536	2,201	2,028	2,218	2,147	1,894
0,8	1,494	2,098	1,671	1,887	2,063	1,817
0,9	1,454	2,005	1,415	1,644	1,985	1,747
1,0	1,417	1,919	1,224	1,459	1,912	1,682

Beginning with a λ value of 0,4, the VIF values of all variables fall below the required threshold of 5. This confirms that the issue of multicollinearity in the input data has been resolved. However, as the λ value increases, the parameters approach zero. Therefore, to determine the most appropriate λ value, both the R^2 and RMSE must be evaluated. Table 4 represents the values of those two parameters regarding λ .

Tab. 4. Evaluating R^2 and RMSE of Ridge Regression model respects to λ

λ	R^2	RMSE
0.1	0,581	0,737
0.2	0,576	0.603
0.3	0,572	0,518
0.4	0,568	0,456
0.5	0,565	0,410
0.6	0,561	0,376
0.7	0,557	0,351
0.8	0,554	0,334
0.9	0,551	0.325
1	0,547	0,321

To be more specific, as $\lambda = 0,4$, the value of $R^2 = 0,568$ is the most appropriate. Thus, this value $\lambda = 0,4$ selected to further apply in the Ridge Regression model for Chl-a (mg/m^3) estimation for the study area following formula (8):

$$\begin{aligned}
 Chl - a = & 0,164 * Chl - a_{OC2} - 0,337 * Chl - a_{OC3} - 0,06 * Chl - a_{NDCI} \\
 & + 0,149 * Chl - a_{Gurlin} \\
 & + 0,782 * Chl - a_{CHL-NN} - 0,182 * Chl - a_{CHL-OC4ME}
 \end{aligned} \quad (8)$$

Where $Chl - a_{OC2}$, $Chl - a_{OC3}$, $Chl - a_{NDCI}$, $Chl - a_{Gurlin}$ are Chl-a values derived from OC2, OC3, NDCI and Gurlin11; $Chl - a_{CHL-NN}$ and $Chl - a_{CHL-OC4ME}$ are extracted from provided products with Sentinel-3.

The model's results were validated using test points, showing a good correlation with measured Chl-a. The determined coefficient (R^2) was 0,811, and the RMSE was 0,456, as illustrated in Figure 3.

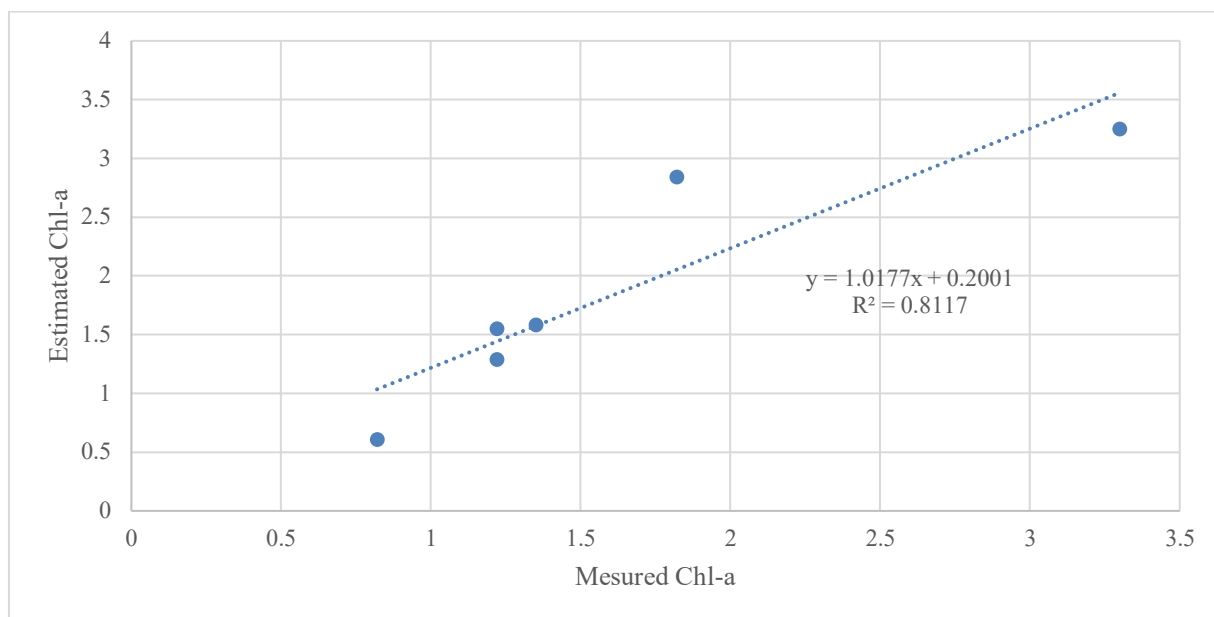


Fig. 3. Correlation between estimated Chl-a form Ridge Regression and measure Chl-a respectively (unit: mg/m^3)

Figure 4 presents a spatial distribution map of estimated Chl-a derived from the Ridge Regression model, which is then compared with Chl-a directly extracted from a Sentinel-3 image acquired on March 6, 2021, for the study area.

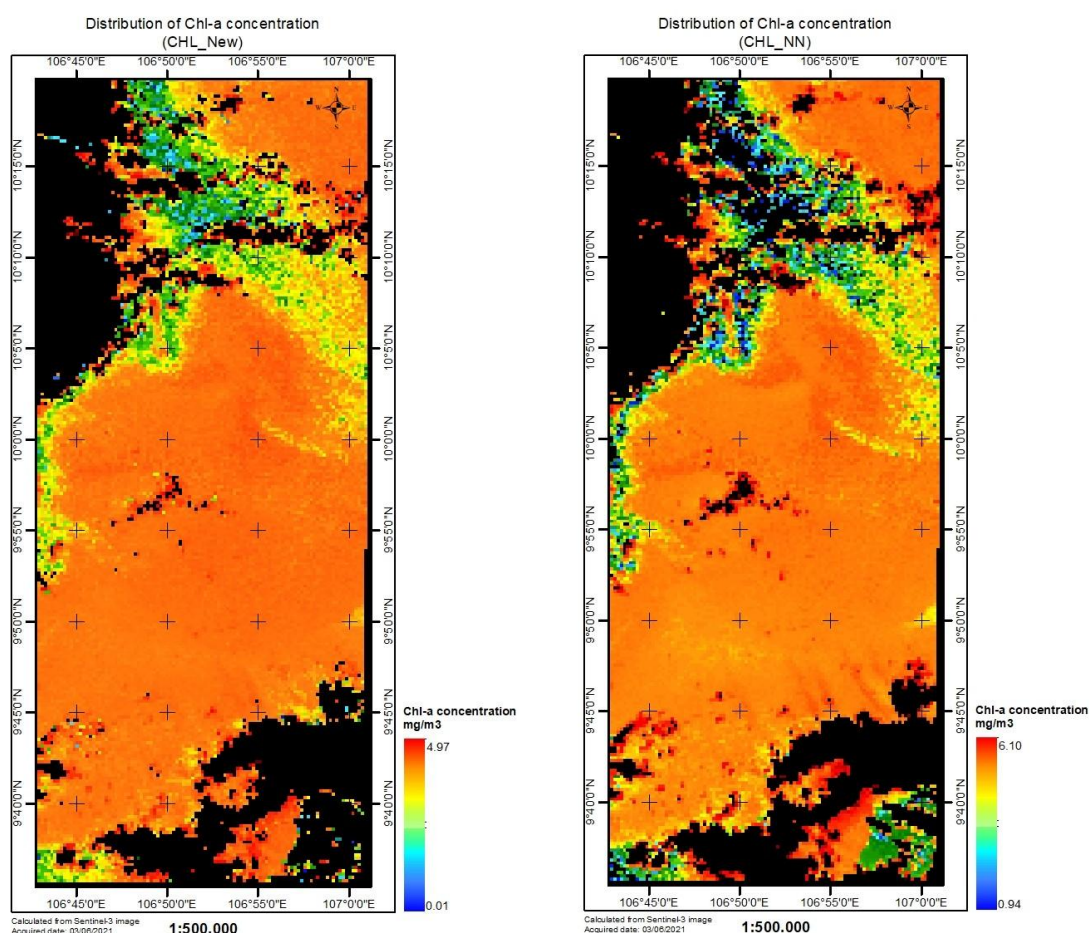


Fig. 4. Spatial distribution of Ridge Regression estimated Chl-a (left); Spatial distribution of directly extracted Chl-a from Sentinel-3 (right)

The Chl-a content estimated by the Ridge Regression model has demonstrated an improvement over

the CHL_NN data available in the Sentinel-3 imagery, specifically by reducing the overestimation of Chl-a in the northern and central parts of the study area. However, the new algorithm was unable to overcome the issue of cloud shadows in the southeastern part of the area.

Distribution comparison: The statistical distribution of Chl-a content from both estimation methods is shown in the histograms in Figure 5. A key difference is in the concentration of values:

- The Chl-a values from the Sentinel-3 data (CHL_NN) are heavily concentrated in the range of 2–4 mg/m³
- The values estimated by the Ridge Regression model are mainly concentrated in the range of 1–2 mg/m³.

This result is consistent with the actual conditions in the area and helps to limit the overestimation of Chl-a content by algorithms that are designed for coastal waters.

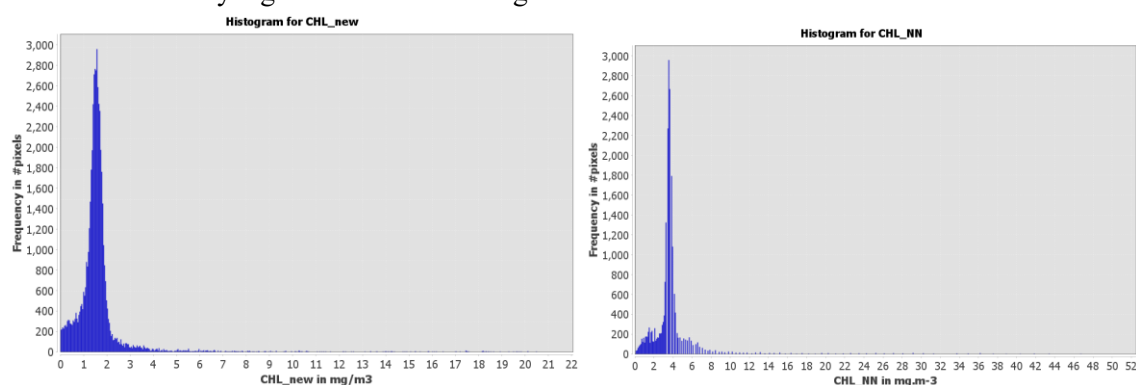


Fig. 5. Distribution chart of Chl-a content values from the Ridge Regression (left) and from the Sentinel-3 data (right) (unit: mg/m³)

The application of remote sensing in marine and ocean research is a complex process. Even with specialized software like SeaDAS, the variability of atmospheric and sea surface conditions continues to pose challenges for modeling. The preliminary results of the Ridge Regression algorithm for estimating Chl-a content are very promising, as the new algorithm shows it can optimize traditional ones. However, its accuracy still needs further experimental verification.

4. Conclusion

A Ridge Regression model was developed and tested to estimate Chl-a concentration in a complex coastal environment in Vietnam. This novel model uses input data from previous Chl-a estimation algorithms, including OC2, OC3, NDCI, and Gurlin11, as well as products provided by the European Space Agency (ESA) from the C2RCC and OC4ME models (CHL_NN and CHL_OC4ME).

The model's effectiveness was highly evaluated, with an R^2 of 0.811 and a RMSE of 0.456. These results were achieved when the regularization parameter λ was set to 0.4, demonstrating the model's suitability for use in coastal areas compared to traditional models. Although the preliminary results of this advanced method are promising and suggest it can optimize traditional algorithms, its accuracy still requires further experimental verification.

Acknowledgement

This study was supported by the T25-37 research project of Hanoi University of Mining and Geology.

References

- Binh, N.A.; Hoa, P.V.; Thao, G.T.P.; Duan, H.D.; Thu, P.M. (2022). Evaluation of Chlorophyll-a estimation using Sentinel 3 based on various algorithms in southern coastal Vietnam. *International Journal of Applied Earth Observation and Geoinformation*, 112, 102951. <https://doi.org/10.1016/j.jag.2022.102951>
- Doerffer, R.; Schiller, H. (2007). The MERIS Case 2 water algorithm. *International Journal of Remote Sensing*, 28(3–4), 517–535. <https://doi.org/10.1080/01431160600821127>
- Dörnhöfer, K.; Oppelt, N. (2016). Remote sensing for lake research and monitoring – Recent advances. *Ecological Indicators*, 64, 105–122. <https://doi.org/10.1016/j.ecolind.2015.12.009>
- Dube, T.; Mutanga, O.; Seutloali, K.; Adelabu, S.; Shoko, C. (2015). Water quality monitoring in sub-Saharan African lakes: a review of remote sensing applications. *African Journal of Aquatic Science*, 40(1), 1–7. <https://doi.org/10.2989/16085914.2015.1014994>
- Ehsanes Saleh, A.K.Md.; Arashi, M.; Golam Kibria, B.M. (2019). Theory of Ridge Regression Estimation

- with Applications. Wiley. <https://doi.org/10.1002/9781118644478>
- Eplee, R.E.; Meister, G.; Patt, F.S.; Turpie, K.R.; Bailey, S.W.; Franz, B.A. (2019). The NASA OBPB 2020 on-orbit calibration of SNPP VIIRS for ocean color applications. In Butler, J. J.; Xiong, X. (Jack); Gu, X. (Eds.), *Earth Observing Systems XXIV* (p. 46). SPIE. <https://doi.org/10.1117/12.2528769>
- EUMETSAT. (2024). Sentinel-3 OLCI level 1 data guide. <https://user.eumetsat.int/resources/user-guides/sentinel-3-olci-level-1-data-guide>
- Friedman, J.; Hastie, T.; Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1). <https://doi.org/10.18637/jss.v033.i01>
- Gholizadeh, M.; Melesse, A.; Reddi, L. (2016). A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors*, 16(8), 1298. <https://doi.org/10.3390/s16081298>
- 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. <https://doi.org/10.1016/j.rse.2011.08.011>
- Halpern, B.S.; Frazier, M.; Potapenko, J.; Casey, K.S.; Koenig, K.; Longo, C.; Lowndes, J.S.; Rockwood, R.C.; Selig, E.R.; Selkoe, K.A.; Walbridge, S. (2015). Spatial and temporal changes in cumulative human impacts on the world's ocean. *Nature Communications*, 6(1), 7615. <https://doi.org/10.1038/ncomms8615>
- Hoerl, A.E.; Kennard, R.W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>
- Ivanda, A.; Šerić, L.; Bugarić, M.; Braović, M. (2021). Mapping Chlorophyll-a Concentrations in the Kaštela Bay and Brač Channel Using Ridge Regression and Sentinel-2 Satellite Images. *Electronics*, 10(23), 3004. <https://doi.org/10.3390/electronics10233004>
- Kim, E.; Nam, S.H.; Ahn, C.H.; Lee, S.; Koo, J.W.; Hwang, T.M. (2022). Comparison of spatial interpolation methods for distribution map an unmanned surface vehicle data for chlorophyll-a monitoring in the stream. *Environmental Technology & Innovation*, 28, 102637. <https://doi.org/10.1016/j.eti.2022.102637>
- Kutser, T.; Paavel, B.; Verpoorter, C.; Ligi, M.; Soomets, T.; Toming, K.; Casal, G. (2016). Remote Sensing of Black Lakes and Using 810 nm Reflectance Peak for Retrieving Water Quality Parameters of Optically Complex Waters. *Remote Sensing*, 8(6), 497. <https://doi.org/10.3390/rs8060497>
- 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. <https://doi.org/10.1080/01431161.2010.512947>
- Merder, J.; Zhao, G.; Pahlevan, N.; Rigby, R.A.; Stasinopoulos, D.M.; Michalak, A.M. (2024). A novel algorithm for ocean chlorophyll-a concentration using MODIS Aqua data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 210, 198–211. <https://doi.org/10.1016/j.isprsjprs.2024.03.014>
- 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. <https://doi.org/10.1016/j.rse.2011.10.016>
- Mobley, C.D. (1999). Estimation of the remote-sensing reflectance from above-surface measurements. *Applied Optics*, 38(36), 7442. <https://doi.org/10.1364/AO.38.007442>
- O'Reilly, J.E.; Maritorena, S.; Mitchell, B.G.; Siegel, D.A.; Carder, K.L.; Garver, S.A.; Kahru, M.; McClain, C. (1998). Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research: Oceans*, 103(C11), 24937–24953. <https://doi.org/10.1029/98JC02160>
- Su, H.; Lu, X.; Chen, Z.; Zhang, H.; Lu, W.; Wu, W. (2021). Estimating Coastal Chlorophyll-A Concentration from Time-Series OLCI Data Based on Machine Learning. *Remote Sensing*, 13(4), 576. <https://doi.org/10.3390/rs13040576>
- Thanh Son, N.; Duc Loc, N. (2024). Establishing Optimal Machine Learning Models for Monitoring Water Quality in Vietnam's Upper Ma River. *Applied Environmental Research*. <https://doi.org/10.35762/AER.2024053>

- Thien Phuong Thao, N.; Thi Thu Ha, N.; Quang Vinh, P.; Thi Hien, T.; Xuan Thanh, D. (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*. <https://doi.org/10.15625/2615-9783/20714>
- Thien Phuong Thao, N.; Huu Toan, D.; Thi Thu Ha, N.; Quang Vinh, P.; Cong Tien, H.; Minh Tien, T.; Thi Hien, T. (2024). Modeling Spatial Distribution of Suspended Sediments in Hoa Binh Reservoir Using Sentinel-2 Image and Co-Kriging. *VNU Journal of Science: Earth and Environmental Sciences*. <https://doi.org/10.25073/2588-1094/vnuees.5123>
- Tyler, A.N.; Svab, E.; Preston, T.; Présing, M.; Kovács, W.A. (2006). Remote sensing of the water quality of shallow lakes: A mixture modelling approach to quantifying phytoplankton in water characterized by high-suspended sediment. *International Journal of Remote Sensing*, 27(8), 1521–1537. <https://doi.org/10.1080/01431160500419311>
- Vanhellemont, Q.; Ruddick, K. (2021). Atmospheric correction of Sentinel-3/OLCI data for mapping of suspended particulate matter and chlorophyll-a concentration in Belgian turbid coastal waters. *Remote Sensing of Environment*, 256, 112284. <https://doi.org/10.1016/j.rse.2021.112284>
- Vinh, L.T. (2016). Coastal water quality at Vung Tau monitoring station (1997 - 2014). *Vietnam Journal of Marine Science and Technology*, 16(1), 64–72. <https://doi.org/10.15625/1859-3097/16/1/6166>
- Wang, M.; Shi, W. (2005). Estimation of ocean contribution at the MODIS near-infrared wavelengths along the east coast of the U.S.: Two case studies. *Geophysical Research Letters*, 32(13). <https://doi.org/10.1029/2005GL022917>
- Werdell, P.J.; Franz, B.A.; Bailey, S.W.; Feldman, G.C.; Boss, E.; Brando, V.E.; Dowell, M.; Hirata, T.; Lavender, S.J.; Lee, Z.; Loisel, H.; Maritorena, S.; Mélin, F.; Moore, T.S.; Smyth, T.J.; Antoine, D.; Devred, E.; D'Andon, O.H.F.; Mangin, A. (2013). Generalized ocean color inversion model for retrieving marine inherent optical properties. *Applied Optics*, 52(10), 2019. <https://doi.org/10.1364/AO.52.002019>

