**TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT**

**KHOA CÔNG NGHỆ THÔNG TIN**

# BÁO CÁO SINH HOẠT HỌC THUẬT

***Mô hình thực nghiệm trong chiết suất thông tin về Chlorophyll-a từ ảnh hài hòa Landsat – Sentinel đa thời gian***

**Người báo cáo: TS. Trần Thị Hòa**

**Bộ môn: Tin học – Trắc địa**

**HÀ NỘI, 01 -2026**

# MỤC LỤC

MỞ ĐẦU.....3
NỘI DUNG.....4
1. Chlorophyll – a và đặc trưng về phổ.....4
1.1. Khái niệm chung và ý nghĩa.....4
1.2. Đặc trưng phản xạ phổ của Chlorophyll -a.....5
1.3. Một số nghiên cứu về Chrophyll-a khu vực cửa sông.....6
2. Dữ liệu và phương pháp.....8
2.1. Dữ liệu.....8
2.2. Phương pháp.....10
3. Kết quả sơ bộ.....12
KẾT LUẬN.....17

# MỞ ĐẦU

Trong những thập kỷ gần đây, sự gia tăng nhanh chóng của các hoạt động kinh tế - xã hội và biến đổi khí hậu đã gây áp lực lớn lên hệ thống thủy văn toàn cầu. Sự suy giảm chất lượng nước, đặc biệt là hiện tượng phú dưỡng, đang trở thành vấn đề cấp bách tại các lưu vực sông, hồ nội địa và vùng nước ven bờ. Chlorophyll-a (Chl-a) là sắc tố quang hợp chính có trong tảo và thực vật phù du, đóng vai trò là chỉ thị quan trọng nhất để đánh giá nồng độ sinh khối thực vật, trạng thái dinh dưỡng và sức khỏe tổng thể của hệ sinh thái nước.

Phương pháp quan trắc chất lượng nước truyền thống chủ yếu dựa trên việc lấy mẫu trực tiếp tại hiện trường và phân tích trong phòng thí nghiệm. Mặc dù có độ chính xác cao, nhưng phương pháp này bộc lộ nhiều hạn chế như: Khó có thể thực hiện định kỳ trên diện rộng; Các điểm lấy mẫu đơn lẻ không thể phản ánh được sự biến đổi không gian phức tạp và liên tục của nồng độ Chl-a trên toàn bộ bề mặt vực nước.

Sự phát triển của công nghệ viễn thám (Remote Sensing) đã mở ra một hướng đi mới hiệu quả và kinh tế hơn. Với khả năng quan sát diện rộng, tính lặp lại theo thời gian và kho dữ liệu vệ tinh đa dạng (như Sentinel-2, Landsat 8/9), viễn thám cho phép theo dõi diễn biến chất lượng nước một cách trực quan và liên tục.

Cơ sở của phương pháp này dựa trên đặc tính phản xạ phổ đặc trưng của nước chứa diệp lục. Các mô hình thực nghiệm thiết lập mối quan hệ thống kê giữa giá trị phản xạ phổ ghi nhận từ vệ tinh và dữ liệu thực đo tại hiện trường, từ đó cho phép ước tính và xây dựng bản đồ phân bố nồng độ Chl-a trên quy mô lớn.

Bài viết này tập trung vào việc trình bày quy trình xây dựng và áp dụng mô hình thực nghiệm tính toán nồng độ Chlorophyll-a từ dữ liệu ảnh vệ tinh. Nội dung sẽ bao gồm việc tìm hiểu và thử nghiệm một số mô hình thực nghiệm khôi phục và tính toán nồng độ Chrorophyll – a, đối sánh và lựa chọn mô hình phù hợp với khu vực thực nghiệm, dữ liệu và thời điểm thu nhận thông tin.

# NỘI DUNG

## 1. Chlorophyll – a và đặc trưng về phổ

### 1.1. Khái niệm chung và ý nghĩa

Chlorophyll-a (Chl-a) là sắc tố quang hợp màu xanh lá cây có mặt trong hầu hết các loài tảo, thực vật phù du (phytoplankton) và vi khuẩn lam trong môi trường nước. Nó đóng vai trò hấp thụ năng lượng ánh sáng mặt trời để chuyển hóa nước và $CO_2$ thành chất hữu cơ. Trong nghiên cứu thủy sinh học, nồng độ Chl-a được coi là một thước đo gián tiếp để xác định sinh khối thực vật phù du hiện diện trong vực nước.

Chl-a là một trong những thông số quan trọng nhất để đánh giá sức khỏe hệ sinh thái nước vì những lý do sau:

- Chỉ thị trạng thái dinh dưỡng: Nồng độ Chl-a cao thường là dấu hiệu của hiện tượng phú dưỡng (eutrophication). Dựa vào chỉ số này, các nhà quản lý có thể phân loại vực nước là nghèo dinh dưỡng (oligotrophic), trung dưỡng (mesotrophic), hay phú dưỡng (eutrophic).

*Hình 1. Tảo nở hoa ở ven biển Nam Trung Bộ (https://antoanmiennam.com/hien-tuong-phu-duong)*

- Cảnh báo hiện tượng tảo nở hoa: Việc theo dõi Chl-a giúp phát hiện sớm sự bùng phát của tảo, điều có thể dẫn đến tình trạng suy giảm oxy hòa tan (DO), gây chết tôm cá và ảnh hưởng đến nước sinh hoạt.

- Đánh giá chất lượng môi trường: Chl-a phản ánh tác động của các nguồn thải nhân tạo (như phân bón nông nghiệp hoặc nước thải sinh hoạt chứa nhiều Nitơ và Phốt pho) chảy vào nguồn nước.

**1.2. Đặc trưng phản xạ phổ của Chlorophyll -a**

Từ ảnh vệ tinh, thông tin về Chl-a có thể được chiết tách thông qua sự thay đổi về phổ của nước trong dải sáng từ xanh dương đến hồng ngoại ngắn. Hình 2 mô tả đặc tính phản xạ phổ của nước khi có sự xuất hiện của Chl-a Nước tinh khiết hấp thụ mạnh năng lượng ở vùng bước sóng dài (đỏ và cận hồng ngoại). Tuy nhiên, khi có sự xuất hiện của Chl-a, đặc tính phản xạ phổ của nước thay đổi đáng kể:

- Các vùng hấp thụ (Absorption Peaks): Sắc tố Chl-a hấp thụ năng lượng mạnh nhất tại hai "cửa sổ" phổ chính:

- Vùng Blue (khoảng 430 - 450 nm): Hấp thụ cực đại. Tuy nhiên, vùng này thường bị nhiễu bởi chất hữu cơ hòa tan ($CDOM$) và các hạt lơ lửng khác.

- Vùng Red (khoảng 660 - 675 nm): Đây là vùng hấp thụ đặc trưng của Chl-a. Nồng độ tảo càng cao thì giá trị phản xạ tại vùng này càng giảm.

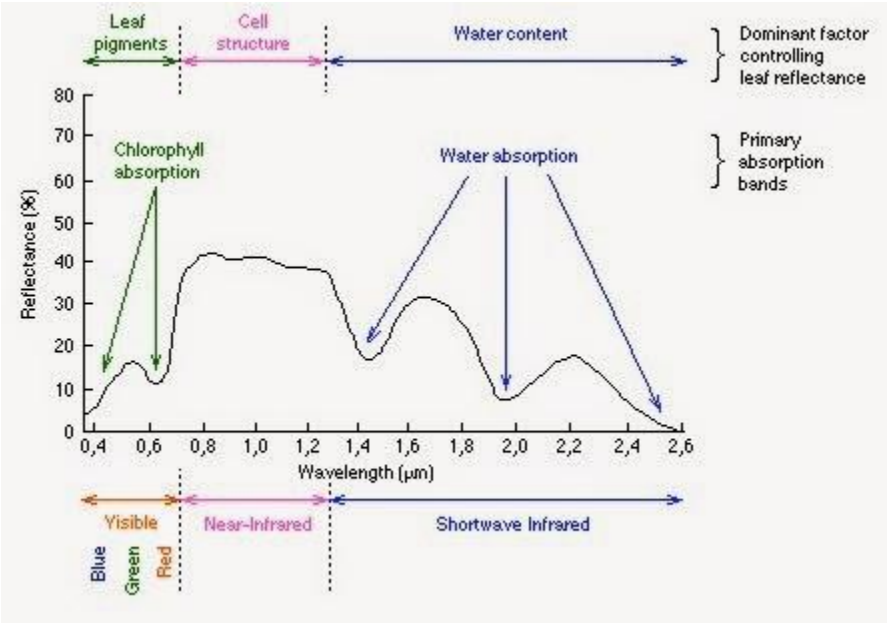
- Các vùng phản xạ (Reflectance Peaks)

- Vùng Green (khoảng 550 nm): Do Chl-a hấp thụ ít ánh sáng xanh lá cây nên nó phản xạ mạnh tại đây, làm cho nước có nhiều tảo thường có màu xanh lục bằng mắt thường.

- Vùng Cận hồng ngoại (NIR) và Red-edge (khoảng 700 - 720 nm): Đây là đặc trưng quan trọng nhất đối với nước đục hoặc nước giàu dinh dưỡng. Khi nồng độ Chl-a rất cao, một đỉnh phản xạ mạnh xuất hiện ở vùng giáp ranh giữa đỏ và cận hồng ngoại (Red-edge) do sự tán xạ ngược từ tế bào tảo.

Như vậy, với đặc trưng của sự thay đổi trong phản xạ của nước, ta có thể ứng dụng trong chọn kênh vệ tinh để ước tính sự xuất hiện của Chl-a, cụ thể:

- Với nước trong (Đại dương): Người ta thường dùng tỷ số giữa kênh Blue và Green.
- Với nước đục/nội địa (Sông, Hồ Việt Nam): Người ta ưu tiên dùng tỷ số giữa kênh Red-edge và Red (ví dụ: Band 5 và Band 4 trên Sentinel-2) vì vùng này ít bị ảnh hưởng bởi các tạp chất khác trong nước.




*Hình 2. Đặc trưng phản xạ phổ nước khi có Chlorophyll -a (https://ladchadapon-rs.blogspot.com/p/blog-page.html)*

### 1.3. Một số nghiên cứu về Chrophyll-a khu vực cửa sông

Cửa sông là nơi có hệ sinh thái có năng suất sinh học cao và có ý nghĩa quan trọng về mặt sinh thái, đóng vai trò là mắt xích quan trọng kết nối giữa môi trường đất liền và môi trường biển. Tuy nhiên, các khu vực này cũng đặc biệt nhạy cảm trước các áp lực từ hoạt động của con người, dẫn đến sự suy thoái chất lượng nước. Việc giám sát các chỉ số chính, chẳng hạn như nồng độ Chlorophyll-a (Chl-a), là vô cùng thiết yếu để đánh giá và quản lý sức khỏe của hệ sinh thái cửa sông, bởi nồng độ Chl-a cao có thể là dấu hiệu của hiện tượng phú dưỡng và tảo độc nở hoa.

Tuy nhiên, việc giám sát chất lượng nước một cách chính xác và nhất quán trong môi trường cửa sông tương đối khó khăn, cụ thể là ước tính nồng độ Chl-a. Các cửa sông là những hệ thống có đặc điểm quang học phức tạp, thường được gọi là vùng nước ven bờ. Điều này có nghĩa là các đặc tính quang học của chúng không chỉ bị ảnh hưởng bởi thực vật phù du (chứa Chl-a) mà còn bởi các chất khác như trầm tích lơ lửng, chất hữu cơ hòa tan có màu (CDOM) và sự phản xạ từ đáy ở các vùng nước nông. Sự phức tạp này khiến việc tách biệt tín hiệu của Chl-a bằng dữ liệu vệ tinh trở nên khó khăn.

Mặc dù viễn thám là một công cụ giám sát mạnh mẽ, các phương pháp hiện nay vẫn đối mặt với những hạn chế đáng kể:

- Thứ nhất, độ phân giải thời gian: Các nhiệm vụ vệ tinh truyền thống như Landsat có chu kỳ lặp lại tương đối dài (16 ngày), thường là quá thưa để ghi lại những thay đổi nhanh chóng và năng động diễn ra tại cửa sông do chu kỳ thủy triều, các đợt xả lũ của sông hoặc các đợt tảo nở hoa phát triển nhanh. Ta có thể bỏ lỡ hoàn toàn các sự kiện quan trọng.
- Thứ hai, độ phân giải không gian: Trong khi một số vệ tinh cung cấp độ phân giải thời gian cao, độ phân giải không gian của chúng lại quá thô, không đủ để phân tách các đặc điểm quy mô nhỏ trong lòng cửa sông.
- Thứ ba, dữ liệu thực địa: Phương pháp truyền thống là thu thập mẫu nước và phân tích tại phòng thí nghiệm tuy chính xác nhưng lại thâm dụng lao động, tốn kém và bị hạn chế về phạm vi không gian lẫn thời gian. Việc lấy mẫu toàn bộ cửa sông nhiều lần trong ngày hoặc trong tuần là không khả thi.

Báo cáo này hướng tới là sự thiếu hụt một phương pháp viễn thám tần suất cao, độ phân giải cao có khả năng ước tính nồng độ Chl-a một cách nhất quán và chính xác trong môi trường cửa sông động và phức tạp về quang học. Mặc dù các nghiên cứu trước đây đã sử dụng riêng lẻ Landsat hoặc Sentinel-2, nhưng họ chưa khai thác đầy đủ sức mạnh kết hợp của cả hai nhiệm vụ này để tạo ra một chuỗi thời gian liên tục và mạnh mẽ. Giải pháp đề xuất: Sử dụng dữ liệu ảnh Landsat-Sentinel đã được đồng nhất (Harmonized Landsat-Sentinel - HLS) để vượt qua các hạn chế nêu trên. Bằng cách kết hợp dữ liệu từ cả hai nhiệm vụ thành một tập dữ liệu duy nhất sẵn sàng cho phân tích (Analysis-Ready Data), có thể:

- Tăng độ phân giải thời gian: Đạt được tần suất quan sát cao hơn nhiều (2-3 ngày), cho phép ghi lại những thay đổi nhanh chóng của nồng độ Chl-a.
- Duy trì độ phân giải không gian: Tập dữ liệu đồng nhất cung cấp độ phân giải 30 mét nhất quán, phù hợp để phân tách các đặc điểm của cửa sông.
- Cho phép giám sát liên tục: Cách tiếp cận này cho phép tạo ra một chuỗi thời gian dài hạn và liền mạch để hiểu toàn diện hơn về động thái chất lượng nước cửa sông.

Tuy nhiên, cần lưu ý rằng việc tích hợp dữ liệu Landsat 8/9 OLI và Sentinel-2 MSI đòi hỏi quá trình đồng nhất phổ (spectral harmonization) do sự khác biệt trong cấu hình kênh phổ và hàm đáp ứng phổ. Mặc dù cả hai cảm biến đều cung cấp dữ liệu phản

xạ bề mặt trong dải sóng nhìn thấy và cận hồng ngoại, sự biến thiên về độ rộng băng thông và bước sóng trung tâm có thể gây ra sự không nhất quán khi kết hợp các tập dữ liệu để giám sát dài hạn (Roy et al., 2016; Claverie et al., 2018).

## 2. Dữ liệu và phương pháp

### 2.1. Dữ liệu

Các ảnh vệ tinh Landsat 8/9 và Sentinel-2 bao phủ khu vực nghiên cứu được thu thập từ các cổng dữ liệu của USGS (Cục Khảo sát Địa chất Hoa Kỳ) và ESA (Cơ quan Vũ trụ Châu Âu) với độ phủ mây dưới 30%. Các kênh phổ cụ thể được sử dụng từ ảnh Landsat 8/9 và Sentinel-2 bao gồm: Blue_coastal (443nm - Xanh dương ven biển), Blue (490nm - Xanh dương), Green (560nm - Xanh lá), Red (665nm - Đỏ) và NIR (865nm - Cận hồng ngoại). Số lượng ảnh cụ thể được liệt kê dưới đây.

*Bảng 1. Dữ liệu ảnh vệ tinh*

| Loại ảnh | Năm | Số lượng kênh |
|---|---|---|
| Landsat 8/9 (124053) | 2020 | 12 |
| | 2021 | 10 |
| | 2022 | 19 |
| | 2023 | 14 |
| | 2024 | 19 |
| Sentinel -2 (T48PYS) | 2020 | 12 |
| | 2021 | 11 |
| | 2022 | 11 |
| | 2023 | 9 |
| | 2024 | 12 |

Mặc dù lượng dữ liệu thu thập được là tương đối lớn, tuy nhiên do ảnh hưởng của các điều kiện khí tượng, sự gián đoạn của các nguồn dữ liệu đơn lẻ này không thể đảm bảo cho việc giám sát chất lượng nước một cách liên tục.

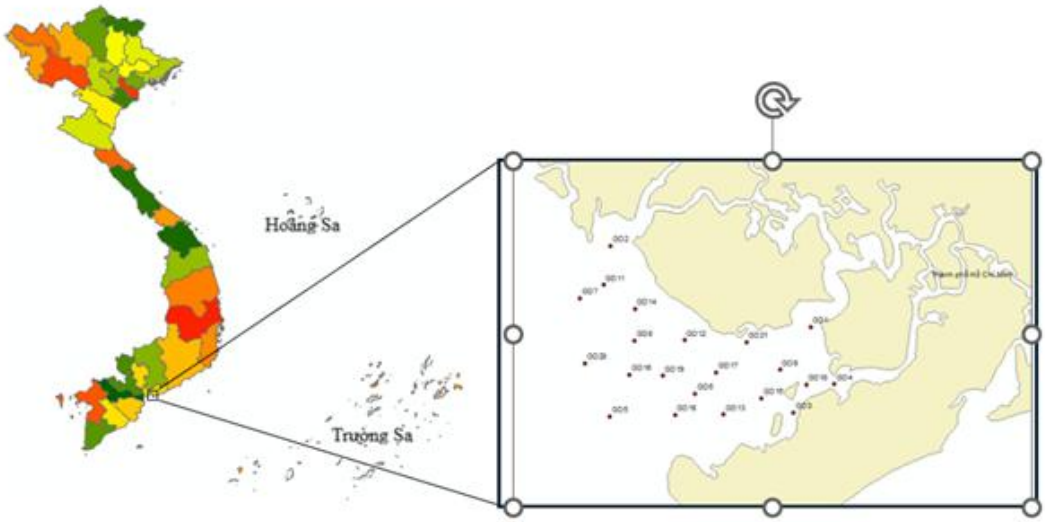
Do đó, các ảnh đồng nhất đã được sử dụng. Dự án Đồng nhất dữ liệu Landsat và Sentinel-2 (HLS) là một sáng kiến của NASA phối hợp với USGS nhằm tạo ra dữ liệu phản xạ bề mặt tương thích vệ tinh, bao gồm cảm biến OLI (Operational Land Imager) trên vệ tinh Landsat-8 và MSI (Multi-Spectral Instrument) trên vệ tinh Sentinel-2. Sự

kết hợp này cho phép quan sát bề mặt trái đất trên phạm vi toàn cầu với tần suất 2-3 ngày một lần ở độ phân giải không gian trung bình (30 m). Dữ liệu HLS sử dụng một bộ thuật toán để tạo ra các sản phẩm liền mạch từ OLI và MSI, bao gồm: hiệu chỉnh khí quyển, lọc mây và bóng mây, đồng đăng ký không gian và đưa về lưới tọa độ chung, chuẩn hóa góc nhìn và hiệu chỉnh dải thông phổ) (Junchang Ju, 2025).

Sản phẩm dữ liệu HLS có thể được coi là các khối cấu trúc để xây dựng một "khối dữ liệu" (data cube), cho phép người dùng kiểm tra bất kỳ điểm ảnh (pixel) nào theo thời gian và xử lý chuỗi thời gian phản xạ với tần suất gần như hàng ngày như thể chúng đến từ một cảm biến duy nhất. Các dữ liệu này hiện có sẵn trên Earthdata Search cũng như trung tâm LP DAAC của NASA.

Dữ liệu thực địa được sử dụng trong nghiên cứu này được thu thập thông qua các trạm đo đạc trong khu vực. Vị trí của các trạm được thể hiện trong Hình 3. Dữ liệu thu thập bao gồm 21 trạm, trong đó 14 trạm được sử dụng để xây dựng mô hình và 7 trạm được sử dụng để kiểm định. Hình 1 thể hiện khu vực nghiên cứu và vị trí của các mẫu dữ liệu đã thu thập.

Vịnh Gành Rái là một đầm phá nước lợ, nơi có nhiều con sông đổ vào lưu vực từ phía Bắc như sông Ngã Bảy (Lòng Tàu), sông Đồng Tranh, sông Thị Vải và sông Dinh. Phía Đông là bán đảo Vũng Tàu với mũi Nghinh Phong; phía Tây là mũi Cần Thạnh (Cần Giờ) cùng với mũi Đồng Tranh, tạo thành thế bao bọc lấy vịnh từ ba phía.




*Hình 3. Khu vực thực nghiệm và vị trí các trạm thu thập dữ liệu*

Vịnh Gành Rái có chiều rộng khoảng 12 km và chiều dài khoảng 21 km, tương ứng với tọa độ từ 10º20'B đến 10º30'B vĩ độ Bắc và từ 106º50'Đ đến 107º08'Đ kinh độ Đông. Chế độ thủy triều tại vịnh Gành Rái là chế độ bán nhật triều không đều, với biên độ dao động mực nước khoảng ±1,5 m. Hệ thống dòng chảy phụ thuộc chặt chẽ vào chế độ thủy triều và chế độ gió. Bên cạnh đó, độ sâu trong vịnh cũng thay đổi một cách phức tạp, gây ảnh hưởng đáng kể đến động lực dòng chảy trên toàn vịnh.

Ngoài ra, nhờ vị trí địa lý thuận lợi, vịnh Gành Rái còn có ý nghĩa to lớn trong giao thương hàng hóa đường thủy của vùng Đông Nam Bộ, đặc biệt là TP. Hồ Chí Minh. Đây là nơi nhiều con sông lớn đổ ra Biển Đông, đồng thời cũng là một ngư trường khai thác thủy hải sản lớn với nhiều trang trại nuôi trồng thủy sản như tôm, cá, nghêu và hàu.

**2.2. Phương pháp**

Đầu tiên, các dữ liệu Sentinel-2 đã đồng nhất sẽ được thu thập từ NASA LP DAAC thông qua nền tảng Google Earth Engine với mã truy cập (Earth Engine Snippet) là: ee.ImageCollection("NASA/HLS/HLSS30/v002"). Trong nghiên cứu này, sự nhất quán giữa các giá trị phản xạ của ảnh Sentinel-2 và Landsat 8/9 đối với từng kênh phổ cụ thể đã được đánh giá thông qua chỉ số sai lệch tuyệt đối trung bình (Mean Absolute Difference - MAD).

$$MAD = \frac{1}{n}\sum_{i=1}^{k}\left|R_{\lambda i}^{L} - R_{\lambda i}^{S}\right|$$

Trong đó, $R_{\lambda i}^{S}$ là giá trị phản xạ của Sentinel-2 tại vị trí *i* trong kênh phổ *λ*.

$R_{\lambda i}^{L}$ là giá trị phản xạ của Landsat 8/9 tại vị trí *i* trong kênh phổ *λ*.

*n* là tổng số điểm khảo sát.

Sau khi đồng nhất, các đặc trưng phổ được trích xuất từ giá trị phản xạ mặt đất $R_{rs}$ nhằm tăng độ nhạy của mô hình đối với sự thay đổi nồng độ Chl-a, đồng thời giảm thiểu các tác động gây nhiễu từ tổng chất rắn lơ lửng (TSS) và chất hữu cơ hòa tan có màu (CDOM) vốn chiếm ưu thế trong các vùng nước cửa sông như vịnh Gành Rái. Việc thu thập đặc trưng tuân theo các phương pháp đã được thiết lập trong lĩnh vực quang học sinh học đại dương, bao gồm:

- Giá trị phản xạ của các kênh phổ chung (có sẵn trên cả ảnh Landsat 8/9 và Sentinel-2), bao gồm: Blue_coastal (443nm), Blue (490nm), Green (560nm), Red (665nm) và

NIR (865nm). Các kênh này được sử dụng trực tiếp do mối liên hệ mật thiết của chúng với đặc tính hấp thụ và tán xạ ngược của thực vật phù du.

- Tỷ số giữa các kênh phổ (Band ratios): Giúp giảm thiểu ảnh hưởng của điều kiện chiếu sáng và làm nổi bật các đặc tính hấp thụ cụ thể. Các tỷ số được sử dụng là Blue/Green và Red/Green, vốn là những tổ hợp mang lại hiệu quả chính xác trong việc tách biệt tín hiệu Chl-a khỏi TSS tại các vùng nước đục.

Phương pháp được sử dụng để đánh giá tính độc lập của các biến khi ước tính là hệ số tương quan Pearson. Trong đó, giá trị tương quan *r* được phân chia thành các mức độ như trong bảng 2 sau:

*Bảng 2. Phân loại hệ số tương quan (Schober et al, 2018)*

| ***r*** | ***Diễn giải*** |
|---|---|
| 0.00 – 0.09 | Tương quan không đáng kể |
| 0.10 – 0.39 | Tương quan yếu |
| 0.40 – 0.69 | Tương quan trung bình |
| 0.70 – 0.89 | Tương quan cao |
| 0.90 – 1.00 | Tương quan rất cao |

Trong nghiên cứu này, bốn dạng biến đổi logarit cho các mô hình hồi quy đa biến được sử dụng để ước tính nồng độ Chl-a từ các dữ liệu đầu vào như giá trị phản xạ, tỷ số kênh phổ và sai khác giữa các kênh phổ, bao gồm:

Hồi quy tuyến tính : $y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k$

Hồi quy ln: $\ln y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k$

Hồi quy ln – tuyến tính : $y = b_0 + b_1\ln x_1 + b_2\ln x_2 + \ldots + b_k\ln x_k$

Hồi quy ln-ln: $\ln y = b_0 + b_1\ln x_1 + b_2\ln x_2 + \ldots + b_k\ln x_k$

Việc lựa chọn mô hình phù hợp nhất cho khu vực nghiên cứu dựa trên giá trị hệ số xác định $R^2$, p-value và để tăng độ tin cậy của nghiên cứu, sai số tiêu chuẩn SE cũng sẽ được xem xét. Trong đó, giá trị hệ số xác định $R^2$được phân chia thành các mức độ như trình bày trong Bảng 3 dưới đây:

*Bảng 3. Phân loại hệ số xác đinh (Moriasi và cộng sự, 2015)*

| ***$R^2$*** | ***Interpretation*** |
|---|---|
| ≤ 0.5 | Không thỏa mãn |

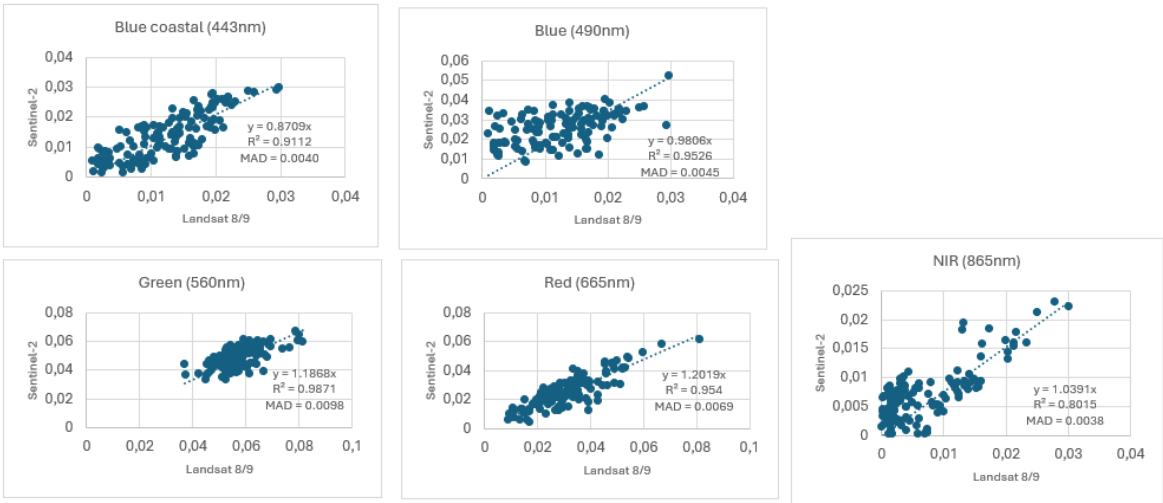
| | |
|---|---|
| 0.5 – 0.7 | Thỏa mãn |
| 0.7 – 0.8 | Tốt |
| ≥ 0.8 | Rất tốt |

## 3. Kết quả sơ bộ

### 3.1 Đánh giá tính nhất quán của dữ liệu

Trước khi sử dụng dữ liệu Sentinel-2 đã được đồng nhất để ước tính nồng độ Chl-a nhằm bổ trợ cho dữ liệu Landsat, nhóm tác giả đã đánh giá tính nhất quán của hai loại dữ liệu này tại các kênh Blue (Xanh dương), Green (Xanh lá), Red (Đỏ) và NIR (Cận hồng ngoại) trong cùng một ngày tại khu vực nghiên cứu. Vị trí trích xuất dữ liệu trùng khớp với các vị trí khảo sát nồng độ Chl-a thực địa (xem Hình 3).

Để đánh giá tính đồng nhất, các cặp ảnh Sentinel-2 và Landsat 8/9 được chọn phải được chụp trong cùng một ngày với độ phủ mây dưới 10%, nhằm đảm bảo các vị trí lấy mẫu không bị ảnh hưởng bởi mây và bóng mây. Trong số các dữ liệu thu thập từ năm 2020 đến 2024, chỉ có các ngày sau đây đáp ứng được yêu cầu: ngày 17 tháng 1 năm 2021; ngày 7 tháng 4 năm 2021; ngày 21 tháng 6 năm 2022; ngày 11 tháng 2 năm 2024; ngày 22 tháng 3 năm 2024; và ngày 17 tháng 11 năm 2024.




*Hình 4. Biểu đồ phân tán giá trị phản xạ với Landsat 8/9 trên trục hoành (x) và Sentinel-2 trên trục tung (y), thể hiện tính nhất quán của dữ liệu tại các kênh phổ Blue_coastal, Blue, Green, Red và NIR.*

Sự tương quan về giá trị phản xạ giữa hai loại dữ liệu trên năm kênh phổ là không đồng đều. Hệ số xác định $R^2$ cao nhất tập trung ở kênh Green (0,9871), tiếp theo là Red (0,954), Blue (0,9526), Blue_coastal (0,9112) và cuối cùng là kênh NIR (0,8015). Tuy nhiên, nếu xem xét tính nhất quán thông qua chỉ số MAD, thứ tự này bị đảo ngược lại, lần lượt là NIR (0,0038), Blue_coastal (0,0040), Blue (0,0045), Red (0,0069) và Green (0,0098). Điều này là do giá trị phản xạ tại kênh NIR của các điểm lấy mẫu nhỏ hơn nhiều so với các kênh khác, đặc biệt là kênh Green (xem Hình 4). Hầu hết các điểm lấy mẫu có giá trị phản xạ ở dải NIR <0.01, trong khi ở kênh Green, giá trị này chủ yếu $> >0.04. Do đó, việc độ nhạy đối với biên độ dao động khác nhau dẫn đến tình trạng nêu trên là điều dễ hiểu.

### 3.2 Ước tính nồng độ Chl-a

Để ước tính nồng độ Chl-a trong khu vực nghiên cứu này, dữ liệu thực địa đã được thu thập vào thời điểm gần với ngày thu nhận ảnh Sentinel-2 (mảnh ảnh T48PYS) vào ngày 26/06/2022. Nồng độ Chl-a được tính toán bằng cách sử dụng các hàm hồi quy cho bốn dạng biến đổi logarit.

Mối tương quan giữa dữ liệu thực địa (Chl-a) và giá trị phản xạ của ảnh được trình bày trong Bảng 3. Mặc dù hầu hết các hệ số tương quan đều mang giá trị âm, nhưng khi so sánh với bảng phân loại ở Bảng 3, phần lớn các hệ số này đều ở mức mạnh, ngoại trừ tỷ số Blue/Green ở mức trung bình. Như vậy, các biến này có thể được sử dụng để ước tính nồng độ Chl-a.

*Bảng 4. Hệ số tương quan giữa dữ liệu thực địa (Chl-a) và các biến đầu vào*

| | *Blue coastal* | *Blue* | *Green* | *Red* | *NIR* | *Red/Green* | *Blue/Green* | *Chl-a* |
|---|---|---|---|---|---|---|---|---|
| *Blue coastal* | 1 | | | | | | | |
| *Blue* | 0.97 | 1 | | | | | | |
| *Green* | 0.84 | 0.88 | 1 | | | | | |
| *Red* | 0.95 | 0.91 | 0.76 | 1 | | | | |
| *NIR* | 0.89 | 0.87 | 0.61 | 0.96 | 1 | | | |
| *Red/Green* | 0.89 | 0.84 | 0.63 | 0.98 | 0.96 | 1 | | |
| *Blue/Green* | 0.87 | 0.88 | 0.56 | 0.84 | 0.91 | 0.83 | 1 | |
| *Chl-a* | -0.76 | -0.76 | -0.72 | -0.79 | -0.74 | -0.76 | -0.62 | 1 |

Đối với mỗi dạng biến đổi logarit theo công thức nêu trên, các hệ số $\beta_i$ được xác định tương ứng với các tham số đầu vào là giá trị phản xạ, tỷ số kênh phổ và sai khác phổ; khi đó, các công thức sẽ trở thành:

Hàm hồi quy:

$$Chl - a = -72.956 * R_{rs}^{Bc} + 5827.38 * R_{rs}^{B} - 5971.95 * R_{rs}^{G} + 1760.36 * R_{rs}^{R} - 441.513 * R_{rs}^{N} - 373.823 * R_{rs}^{B/G} - 92.6523 * R_{rs}^{R/G} + 401.07$$

Hàm ln:

$$Ln(Chl - a) = -2.9452 * R_{rs}^{Bc} + 325.813 * R_{rs}^{B} - 330.669 * R_{rs}^{G} + 90.8296 * R_{rs}^{R} - 25.1136 * R_{rs}^{N} - 20.9008 * R_{rs}^{B/G} - 4.6998 * R_{rs}^{R/G} + 24.0554$$

Hàm kn hồi quy:

$$Chl - a = 10.7612 * \ln(R_{rs}^{Bc}) + 6.4520 * \ln(R_{rs}^{B}) - 32.5949 * \ln(R_{rs}^{G}) + 9.7873 * \ln(R_{rs}^{R}) - 7.62859 * R_{rs}^{N} - 18.0678$$

Hàm ln -ln:

$$Ln(Chl - a) = 0.6321 * \ln(R_{rs}^{Bc}) + 0.2265 * \ln(R_{rs}^{B}) - 1.6569 * \ln(R_{rs}^{G}) + 0.5079 * \ln(R_{rs}^{R}) - 0.4215 * \ln(R_{rs}^{N}) + 0.9137$$

Với dữ liệu ảnh hài hòa, các công thực này sẽ tiếp tục được chuyển đổi thành:

Hàm hồi quy:

$$Chl - a = 296.2607 * R_{rs}^{Bc} - 2192.25 * R_{rs}^{B} + 969.5326 * R_{rs}^{G} - 870.416 * R_{rs}^{R} - 839.785 * R_{rs}^{N} + 84.8161 * R_{rs}^{B/G} + 71.2239 * R_{rs}^{R/G} - 32.3359$$

Hàm ln:

$$Ln(Chl - a) = 16.6612 * R_{rs}^{Bc} - 131.132 * R_{rs}^{B} + 57.8338 * R_{rs}^{G} - 44.0288 * R_{rs}^{R} - 49.1933 * R_{rs}^{N} + 5.0416 * R_{rs}^{B/G} + 3.8336 * R_{rs}^{R/G} - 0.0549$$

Hàm kn hồi quy:

$$Chl - a = 15.1415 * \ln(R_{rs}^{Bc}) - 7.8266 * \ln(R_{rs}^{B}) - 4.8316 * \ln(R_{rs}^{R}) + 1.2231 * R_{rs}^{N} - 6.1564 * \ln(R_{rs}^{B/G}) + 32.6751$$

Hàm ln -ln:

$$ln(Chl - a) = 0.891 * \ln(R_{rs}^{Bc}) - 0.3458 * \ln(R_{rs}^{B}) - 0.3151 * \ln(R_{rs}^{R}) + 0.0738 * \ln(R_{rs}^{N}) - 0.5153 * \ln(R_{rs}^{B/G}) + 3.9717$$

Trong đó, $R_{rs}^{Bc}$, $R_{rs}^{B}$, $R_{rs}^{G}$, $R_{rs}^{R}$, $R_{rs}^{N}$ tương ứng là phản xạ của các kênh Blue coastal, Blue, Green, Red and NIR.

$R_{rs}^{B/G}$ tỷ số của kênh Blue và Green.

$R_{rs}^{R/G}$ tỷ số của kênh Red và Green.

Việc lựa chọn mô hình phù hợp nhất cho từng loại dữ liệu (dữ liệu gốc và dữ liệu đã đồng nhất) đối với mỗi dạng biến đổi logarit được dựa trên ba tham số: hệ số xác định $R^2$, p-value, và sai số tiêu chuẩn SE. Kết quả chi tiết được trình bày trong bảng 5 dưới đây.

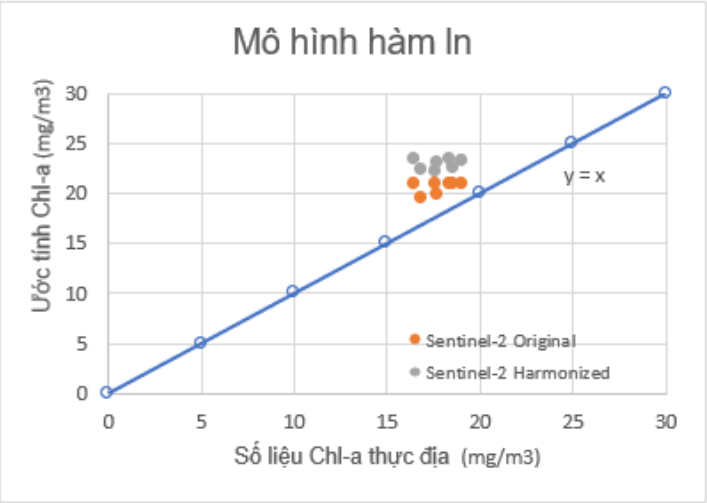
*Bảng 5. So sánh 4 mô hình thực nghiệm dựa vào kết quả tính dành cho sảnh Sentinel gốc và ảnh hải hòa*

| ***Mô hình*** | ***Sentinel-2 gốc*** | | | ***Sentinel-2 hài hòa*** | | |
|---|---|---|---|---|---|---|
| | ***$R^2$*** | ***p-value*** | ***SE*** | ***$R^2$*** | ***p-value*** | ***SE*** |
| Hàm hồi quy | 0.86 | 0.028 | 1.06 | 0.87 | 0.023 | 1.02 |
| Hàm ln | *0.88* | *0.021* | *0.06* | *0.88* | *0.020* | *0.06* |
| Hàm ln hồi quy | 0.77 | 0.114 | 1.38 | 0.86 | 0.027 | 1.05 |
| Hàm ln – ln | 0.78 | 0.094 | 0.08 | 0.87 | 0.021 | 0.06 |

Giá trị $R^2$ của các mô hình đều ở mức cao, nằm trong khoảng Tốt (0,7 - 0,8) và Rất tốt (> 0,8), cho thấy sự tương quan chặt chẽ với dữ liệu thực địa. Tuy nhiên, đối với hai mô hình *ln-ln* và *ln hồi quy* sử dụng ảnh Sentinel-2 gốc, p-values cao hơn 0,05 (lần lượt là 0,094 và 0,114), điều này đồng nghĩa với việc hai mô hình này không có giá trị thống kê. Trong các mô hình còn lại, p-values đều đạt mức ý nghĩa, nhưng khi xem xét sai số tiêu chuẩn SE, duy nhất mô hình *ln* có giá trị SE nhỏ nhất. Điều này khẳng định rằng mô hình *ln* đại diện tốt nhất cho nồng độ Chl-a tổng thể.

Mô hình *ln* sau đó đã được sử dụng để ước tính hàm lượng Chl-a, được kiểm định bằng 7 điểm lấy mẫu và sử dụng giá trị RMSE để đánh giá độ chính xác. Chỉ số RMSE đối với ảnh Sentinel-2 gốc và ảnh đồng nhất lần lượt là 2,32 và 2,73. Các kết quả minh họa trong Hình 5 cho thấy, mặc dù mô hình *log-level* của cả hai trường hợp ảnh đều tốt với các giá trị $R^2$, p-value, SE tương đương nhau, nhưng khi tính toán, giá

trị Chl-a từ ảnh đồng nhất cao hơn khoảng 5% so với ảnh gốc. Nguyên nhân được giải thích là do ảnh đồng nhất đã được tính toán lại với giá trị phổ tiệm cận với giá trị phản xạ phổ của ảnh Landsat 8/9.




*Hình 5. Biểu đồ tương quan ước tính Chl-a từ mô hình lựa chọn so với dữ liệu thực địa.*

Bước tiếp theo của nghiên cứu sẽ áp dụng hàm đã lựa chọn để ước tính Chl-a trên các dữ liệu ảnh còn lại và phân tích sự phân bố của Chl- a ở khu vực thực nghiệm.

# KẾT LUẬN

Dựa trên dữ liệu Chl-a từ các phép đo thực địa và giá trị $R_{rs}$ thu được từ ảnh Sentinel-2 (gốc) và Sentinel-2 đồng nhất (HLS), nghiên cứu đã xây dựng một thuật toán địa phương để giám sát nồng độ Chl-a tại vịnh Gành Rái. Nồng độ Chl-a được ước tính theo mô hình biến đổi logarit dạng *ln-level* với $R^2$= 0,88$ và giá trị RMSE cho ảnh Sentinel-2 gốc và đồng nhất lần lượt là 2,32 và 2,73.

Nghiên cứu tiếp theo sẽ làm rõ sự phân bố không gian của Chl-a tại vịnh Gành Rái để đánh giá việc : (1) phân bố khu vực chịu ảnh hưởng trực tiếp bởi lượng nước xả từ sông Chà Và và sông Dinh; (2) xác định xu hướng nồng độ Chl-a giữa các mùa trong năm; (3) đánh giá sự khác biệt chủ yếu đến từ khu vực ven bờ, nơi chịu tác động mạnh mẽ bởi các hoạt động của con người và chế độ bán nhật triều trong khu vực hay không.

Dữ liệu Sentinel-2 đồng nhất hoàn toàn có thể được sử dụng để đánh giá sự phân bố không gian - thời gian của nồng độ Chl-a. Phương pháp nghiên cứu trong bài báo cáo này có tiềm năng ứng dụng cao cho việc giám sát chất lượng nước tại các vùng ven biển tương tự.

Dữ liệu đồng nhất mang lại nhiều ưu thế, chẳng hạn như cho phép phân tích chuỗi thời gian đa cảm biến một cách nhất quán; giảm thiểu nhiễu từ TSS và CDOM thông qua các tỷ số và chỉ số kênh phổ; đồng thời cung cấp bộ đặc trưng đa dạng cho các mô hình học máy (như Hồi quy Ridge, Random Forest) để lựa chọn các biến dự báo phù hợp nhất. Mặc dù vẫn còn những hạn chế, ví dụ như khả năng xảy ra hiện tượng đa cộng tuyến, đặc biệt là ở các tỷ số kênh phổ, dẫn đến việc giảm hiệu suất của các thuật toán OC (Ocean Color) cổ điển trong môi trường nước cửa sông, nhưng những hạn chế này sẽ được giải quyết thông qua việc áp dụng Hồi quy Ridge hoặc Random Forest,... trong nghiên cứu tiếp theo.

Báo cáo được tiến hành dựa trên một phần của kết quả nghiên cứu đề tài mã số B2025 MDA05.

## TÀI LIỆU THAM KHẢO

P. Schober, C. Boer, and L. A. Schwart (2018). Correlation coefficients: Appropriate use and interpretation. Anesth. Analg., vol. 126, no. 5, pp. 1763–1768. doi: 10.1213/ANE.0000000000002864;

D. N. Moriasi, M. W. Gitau, N. Pai, and P. Daggupati (2015). Hydrologic and water quality models: performance measures and evaluation criteria. Am. Soc. Agric. Biol. Eng., vol. 58, no. 6, pp. 1763–1785. doi: 10.13031/trans.58.10715;

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J. C., Skakun, S. V., & Justice, C. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sensing of Environment, 219, 145–161. https://doi.org/10.1016/j.rse.2018.09.002

J. Ju, Qiang Zhou, Brian Freitag, David P. Roy, Hankui K. Zhang, Madhu Sridhar, John Mandel, Saeed Arab, Gail Schmidt, Christopher J. Crawford, Ferran Gascon, Peter A. Strobl, Jeffrey G. Masek, Christopher S.R. Neigh (2025). The Harmonized Landsat and Sentinel-2 version 2.0 surface reflectance dataset. Remote Sensing of Environment, Volume 324. https://doi.org/10.1016/j.rse.2025.114723;

Blondeau-Patissier, D., Brando, V. E., Oubelkheir, K., Dekker, A. G., Clementson, L. A., & Daniel, P. (2014). A review of ocean color remote sensing methods and statistical techniques for the detection of phytoplankton blooms in coastal and open oceans. Progress in Oceanography, 123, 123–144. https://doi.org/10.1016/j.pocean.2013.12.008;

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. https://doi.org/10.1080/01431169208904125;

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). https://doi.org/10.1029/2011JC007395;

O'Reilly, J. E., et al. (1998). Ocean color chlorophyll algorithms for SeaWiFS. Journal of Geophysical Research: Oceans, 103(C11), 24937–24953. https://doi.org/10.1029/98JC02160;

Roy, D. P., Li, J., Zhang, H. K., Yan, L., Huang, H., Vermote, E., & Skakun, S. (2016). Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and comparison with Landsat-8 OLI. Remote Sensing of Environment, 176, 19–34. https://doi.org/10.1016/j.rse.2016.01.028;

Wang, Q., et al. (2020). Consistency assessment of Sentinel-2 MSI and Landsat-8 OLI products: Implications for combining multi-source satellite data in vegetation monitoring. Remote Sensing of Environment, 236, 111426. https://doi.org/10.1016/j.rse.2019.111426;

IOCCG. (2000). Remote sensing of ocean colour in coastal, and other optically-complex, waters. Reports of the International Ocean-Colour Coordinating Group, No. 3. IOCCG, Dartmouth, Canada;

Morel, A. & Prieur, L. (1977). Analysis of Variation in Ocean Color. Limnology and oceanography. 22. 709-722. 10.4319/lo.1977.22.4.0709.

https://thuydacvietnam.org.vn/thuy-trieu/vung-tau-DNP280101983DIZI28/20240127.html?lang=en-GB

Ba Ria - Vung Tau Department of Science and Technology (2005), Report of project "Research on the design of the environmental monitoring network in Ba Ria-Vung Tau Province and the development of support tool on the digital map".