

## Tích hợp viễn thám đa nguồn và dữ liệu mặt đất để ước tính nồng độ BOD, COD và TSS trong vùng nước mặt ven biển tỉnh Cà Mau

Phạm Văn Mạnh<sup>1</sup>, Trần Quốc Tuấn<sup>2</sup>, Nguyễn Văn Việt<sup>3</sup>,  
Trần Mai Hùng<sup>4</sup>, Nguyễn Thị Thu Hằng<sup>1</sup>, Đỗ Thị Nhung<sup>1\*</sup>

<sup>1</sup>Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

<sup>2</sup>Công ty Cổ phần Công nghệ, Dịch vụ Tài nguyên và Môi trường

<sup>3</sup>Công ty TNHH tư vấn 869

<sup>4</sup>Trung tâm Kỹ thuật-Công nghệ-Quan trắc, Sở Nông nghiệp và Môi trường tỉnh Cà Mau

### Integration of multi-source remote sensing and ground data to estimating BOD, COD and TSS concentration in coastal surface waters of Ca Mau province

Pham Van Manh<sup>1</sup>, Tran Quoc Tuan<sup>2</sup>, Nguyen Van Viet<sup>3</sup>,  
Tran Mai Hung<sup>4</sup>, Nguyen Thi Thu Hang<sup>1</sup>, Do Thi Nhung<sup>1\*</sup>

<sup>1</sup>University of Science, Vietnam National University, Hanoi

<sup>2</sup>Technology JSC, Natural R & E Services

<sup>3</sup>869 Consultants co.,ltd

<sup>4</sup>Center for Engineering-Technology-Monitoring, Department of Agriculture and Environment of Ca Mau province

\*Corresponding author: nhungdo1262@gmail.com

<https://doi.org/10.55250/jo.vnuf.15.2.2026.085-095>

#### TÓM TẮT

Chất lượng nước mặt tại vùng ven biển tỉnh Cà Mau đã và đang chịu áp lực lớn từ các hoạt động sản xuất nông nghiệp, nuôi trồng và khai thác thủy sản, phát triển đô thị, cùng với tác động ngày càng gia tăng của biến đổi khí hậu, xâm nhập mặn và nước biển dâng. Những áp lực này làm gia tăng nồng độ các chất ô nhiễm hữu cơ, chất rắn lơ lửng và chất dinh dưỡng, đe dọa đến sức khỏe cộng đồng, sinh kế của người dân cũng như sự bền vững của các hệ sinh thái ven biển. Nghiên cứu này đề xuất một cách tiếp cận tích hợp viễn thám đa nguồn và dữ liệu mặt đất kết hợp mô hình học máy Cubist để ước tính ba thông số chất lượng nước quan trọng: tổng chất rắn lơ lửng (TSS), nhu cầu oxy hóa học (COD) và nhu cầu oxy sinh học (BOD). Độ chính xác mô hình được kiểm chứng thông qua các chỉ số thống kê gồm RMSE, R<sup>2</sup> và MAE, với kết quả cho thấy RMSE dao động từ 3,47–4,38 mg/l, R<sup>2</sup> đạt 0,74–0,85 và MAE từ 2,43–3,28 mg/l. Kết quả nghiên cứu khẳng định tiềm năng của việc tích hợp viễn thám đa nguồn với các mô hình học máy để giám sát nhanh chất lượng nước mặt ven biển, cung cấp cơ sở khoa học đáng tin cậy cho quản lý, giám sát và quyết sách nhằm bảo vệ nguồn nước và duy trì độ đa dạng sinh học ở các khu vực ven biển tương tự.

#### ABSTRACT

Surface water quality in the coastal areas of Ca Mau province has been subjected to substantial pressure from agricultural production, aquaculture and fisheries exploitation, and urban expansion, compounded by escalating impacts of climate change, saltwater intrusion, and sea-level rise. These pressures have increased concentrations of organic pollutants, suspended solids, and nutrients, threatening public health, local livelihoods, and the sustainability of coastal ecosystems. This study proposes an integrated approach that combines multi-source remote sensing and ground-based data with the Cubist machine learning model to estimate three key water quality parameters: total suspended solids (TSS), chemical oxygen demand (COD), and biochemical oxygen demand (BOD). Model accuracy was assessed using RMSE, R<sup>2</sup>, and MAE, with results indicating RMSE values of 3.47–4.38 mg/L, R<sup>2</sup> of 0.74–0.85, and MAE of 2.43–3.28 mg/L. The study's findings confirm the potential of integrating multi-source remote sensing with machine learning models for rapid monitoring of coastal surface water quality, providing a reliable scientific basis for management, surveillance, and decision-making to protect water resources and maintain biodiversity in similar coastal areas.

#### Thông tin chung:

Ngày nhận bài: 20/10/2025

Ngày phản biện: 28/11/2025

Ngày quyết định đăng: 06/01/2026

#### Từ khóa:

BOD, COD, mô hình lập thể, tỉnh Cà Mau, TSS, viễn thám đa nguồn.

#### Keywords:

BOD, Ca Mau province, COD, Cubist model, multi-source remote sensing, TSS.

## **1. ĐẶT VẤN ĐỀ**

Vùng nước ven biển đóng vai trò thiết yếu cả về sinh thái lẫn kinh tế-xã hội, là nơi cư trú của nhiều loài sinh vật biển, giúp điều hòa chu trình dinh dưỡng và duy trì sinh kế cho hàng triệu người trên toàn cầu [1, 2]. Trước áp lực ngày càng gia tăng từ hoạt động nhân sinh và biến đổi khí hậu, nhiều báo cáo toàn cầu đã nhiều lần nhấn mạnh tính cấp thiết của việc giám sát và cải thiện chất lượng nước trong việc đạt được các mục tiêu phát triển bền vững liên quan đến nước sạch và vệ sinh (SDG 6) cũng như bảo tồn tài nguyên biển (SDG 14). Chất lượng nước được xác định bởi các đặc trưng sinh học, hóa học và vật lý của nước và quyết định khả năng sử dụng cho các mục đích như cấp nước sinh hoạt, sản xuất nông nghiệp và các hoạt động giải trí [3, 4]. Thực tế, các phương pháp giám sát tại chỗ dựa vào việc lấy mẫu và tần suất quan trắc, tiêu tốn nhiều nhân lực và chi phí khi triển khai trên phạm vi lớn [3]. Trước những hạn chế đó, công nghệ viễn thám đã được chứng minh là công cụ hiệu quả trong giám sát chất lượng nước nhờ khả năng bao phủ không gian rộng và cập nhật liên tục [5]. Ứng dụng ưu điểm của tín hiệu phản xạ và tán xạ thu bởi cảm biến vệ tinh ở các bước sóng khác nhau có thể được liên hệ với các chỉ tiêu như tổng nitơ (TN), tổng phốt pho (TP), tổng chất rắn lơ lửng (TSS), nhu cầu oxy sinh hóa (BOD), nhu cầu oxy hóa học (COD) và oxy hòa tan (DO). Trong đó, giá trị BOD/COD chủ yếu liên quan đến quá trình phân hủy sinh học và hàm lượng chất hữu cơ hòa tan, những thành phần này không tạo tín hiệu quang học đặc trưng như TSS, chlorophyll-a hay các chất màu hòa tan (CDOM). Do đó, việc ước tính các thông số này từ dữ liệu vệ tinh quang học cần được thực hiện gián tiếp thông qua mô hình hóa các biến trung gian liên quan đến đặc tính quang học của nước [6].

Cách tiếp cận này dựa trên việc giám sát biến động của các chỉ số thông qua thay đổi phổ phản xạ/tán xạ ngược từ môi trường nước và xây dựng mô hình thực nghiệm, phân tích tương quan [6]. Khi kết hợp với số liệu đo thực

địa và các mô hình thống kê hoặc mô hình học máy, dữ liệu viễn thám không chỉ tăng cường cả độ chính xác mà còn mở rộng khả năng giám sát chất lượng nước bề mặt trên quy mô của không gian và thời gian lớn [3]. Bởi đặc tính của ảnh Radar (SAR) hoạt động ở vùng vi sóng nên không bị ảnh hưởng bởi mây che và nhạy với độ nhám bề mặt nước, hàm lượng vật chất lơ lửng, dòng chảy và động lực học trên bề mặt và độ ẩm, đặc tính pha của lớp phủ nước mặt là các yếu tố có liên quan gián tiếp đến sự phân bố và biến động của chất hữu cơ trong nước. Tại Việt Nam, các nghiên cứu chuyên sâu về chất lượng nước vùng ven biển vẫn chủ yếu dựa trên quan trắc tại chỗ. Trong khi khu vực này chịu tác động đồng thời của hoạt động nhân sinh (đô thị hóa, nuôi trồng thủy sản, du lịch) và quá trình tự nhiên (thủy triều, xói lở, bồi tụ) dẫn đến biến động chất lượng nước phức tạp theo không gian-thời gian [7]. Do đó, cần một cách tiếp cận tích hợp gồm dữ liệu quan trắc tại chỗ, dữ liệu viễn thám đa nguồn cùng với mô hình hóa liên kết đất liền và biển để phát triển các chiến lược quản lý, giám sát nhanh và liên tục.

Khu vực nghiên cứu thuộc phạm vi ven biển tỉnh Cà Mau, nơi có hệ sinh thái đặc thù và chế độ thủy văn phức tạp do ảnh hưởng bán nhật triều không đều, dẫn tới biến động lớn về mực nước, vận tốc dòng chảy và quá trình vận chuyển vật chất. Mục tiêu của nghiên cứu tập trung giải quyết những vấn đề sau: (i) xây dựng và định lượng bản đồ phân bố không gian của BOD, COD và TSS bằng cách tích hợp ảnh viễn thám đa nguồn và dữ liệu thực địa; (ii) đánh giá khả năng ứng dụng viễn thám như một khung giám sát liên tục hỗ trợ quản lý và ra quyết định về tài nguyên nước ven biển. Kết quả nghiên cứu được kỳ vọng sẽ cung cấp bằng chứng khoa học hỗ trợ việc ra quyết định quản lý tài nguyên ven biển bền vững và bảo vệ môi trường.

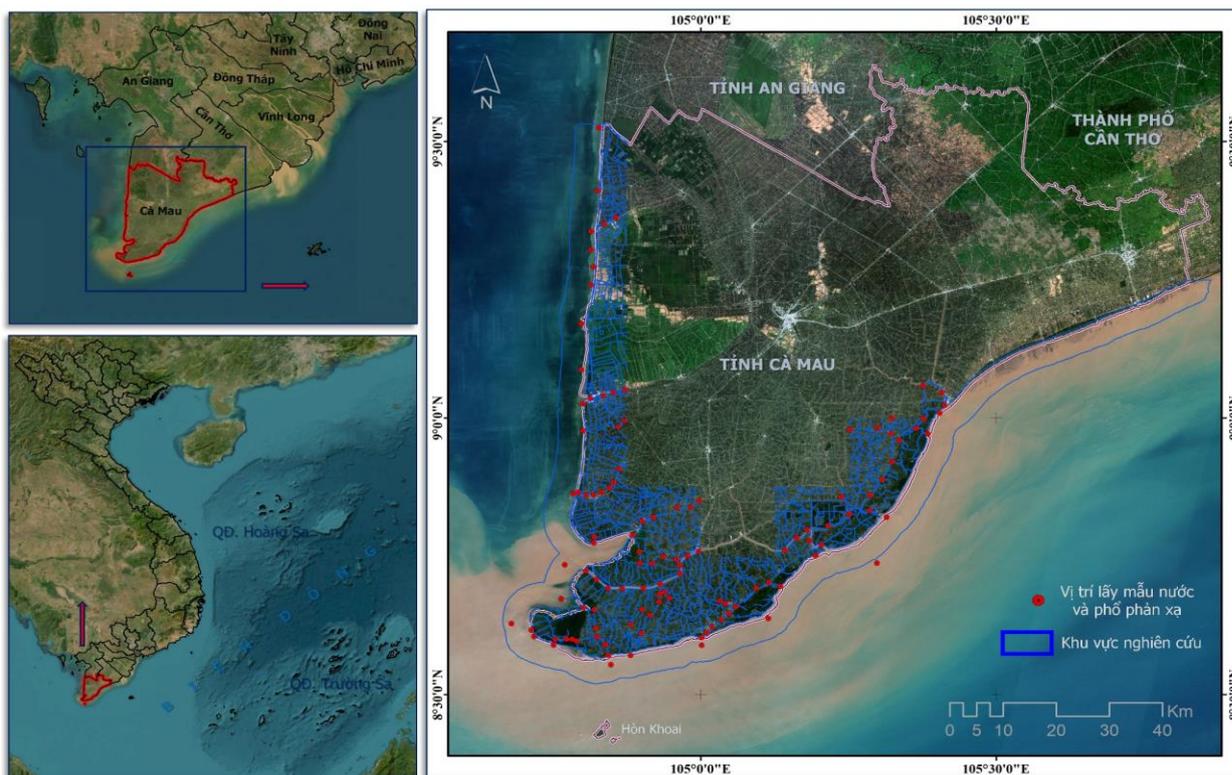
## **2. PHƯƠNG PHÁP NGHIÊN CỨU**

### **2.1. Khu vực nghiên cứu và hiện trạng chất lượng nước ven biển tỉnh Cà Mau**

Cà Mau nằm ở cực Nam Việt Nam, thuộc Đồng bằng sông Cửu Long và là một phần của vùng kinh tế trọng điểm (<https://www.camau.gov.vn/>).

Tỉnh Cà Mau có vị trí bán đảo với ba mặt giáp biển, phía Đông giáp Biển Đông (107 km bờ biển), phía Tây và phía Nam giáp Vịnh Thái Lan (147 km bờ biển), phía Bắc tiếp giáp tỉnh An Giang và thành phố Cần Thơ (Hình 1). Khu vực có địa hình tương đối thấp và bằng phẳng, mạng lưới sông rạch dày đặc với tổng chiều dài khoảng 14.000 km. Độ cao trung bình 0,5–1,5

m so với mực nước biển, có xu hướng nghiêng dần từ Bắc xuống Nam. Khí hậu mang đặc trưng nhiệt đới gió mùa cận xích đạo chia thành hai mùa rõ rệt: mùa mưa (tháng 5–11) và mùa khô (tháng 12–4). Lượng mưa trung bình năm đạt khoảng 2.360 mm, nhiệt độ trung bình 26,5°C và độ ẩm trung bình 85,6%, với dao động nhiệt độ từ 18°C đến 38°C trong những năm gần đây.



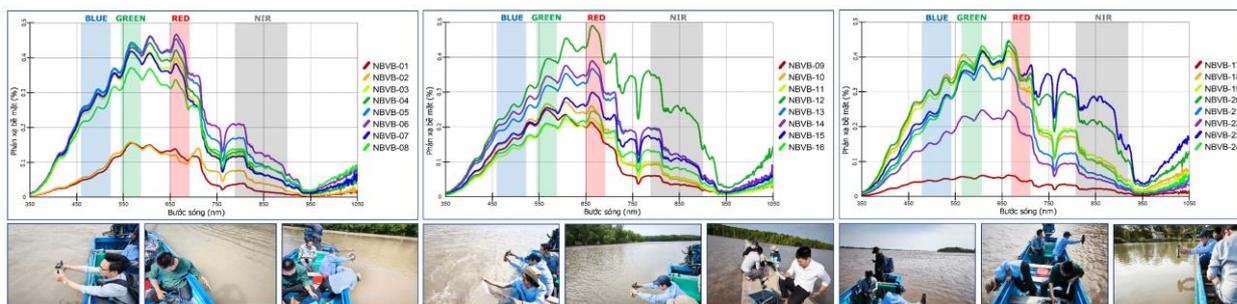
Hình 1. Vị trí các điểm lấy mẫu nước mặt và ranh giới hành chính tỉnh Cà Mau trên lãnh thổ Việt Nam

Nguồn nước sông ngòi ven biển tỉnh Cà Mau đang chịu nguy cơ ô nhiễm nghiêm trọng, tác động trực tiếp tới sản xuất nông nghiệp và nuôi trồng thủy sản. Các quan trắc cho thấy nồng độ ô nhiễm hữu cơ và tổng chất rắn lơ lửng thường xuyên cao, đồng thời hiện tượng bồi lắng và biến đổi dòng chảy diễn ra nhanh, làm suy giảm khả năng tự làm sạch của hệ thủy sinh [8]. Mặc dù đã triển khai nhiều biện pháp quản lý và kiểm soát, tình trạng ô nhiễm trên các tuyến sông, kênh rạch ven biển vẫn phức tạp và chưa được kiểm soát triệt để.

## 2.2. Cách tiếp cận và dữ liệu sử dụng

### 2.2.1. Thu thập và xử lý dữ liệu mẫu nước mặt ven biển

Tổng cộng 115 vị trí lấy mẫu nước mặt, trong đó 24 vị trí thực hiện đo phổ phản xạ; các vị trí này được bố trí dọc ven biển tỉnh Cà Mau và việc thu mẫu được tiến hành từ ngày 15-20/01/2025; mỗi mẫu được lấy ở độ sâu khoảng 0,5 đến 0,7 m dưới mặt nước, cách bờ sông/bờ biển tối thiểu 10 m bằng bình nhựa dung tích 1.000 ml và được bảo quản trong thùng đá lạnh cho đến khi phân tích; các vị trí được chọn nhằm đảm bảo tính đại diện cho chất lượng nước mặt ven biển, phân bố không gian bao phủ các kiểu khu vực ven biển khác nhau và ưu tiên những điểm có độ tĩnh của thủy động, ít bị xáo trộn bởi sóng, dòng chảy hoặc hoạt động tàu thuyền.



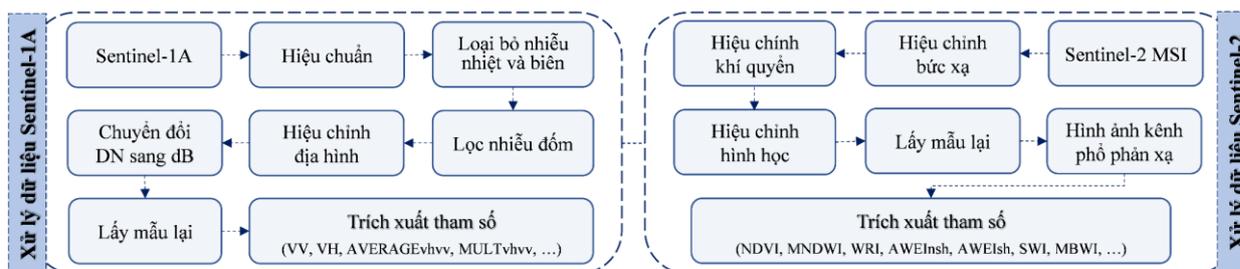
Hình 2. Thu thập mẫu nước mặt và đo phổ phản xạ tại khu vực nghiên cứu.

Các giá trị của các thông số BOD, COD và TSS là kết quả phân tích trong phòng thí nghiệm từ các chuyên gia của Trung tâm Kỹ thuật-Công nghệ-Quan trắc thuộc Sở Nông nghiệp và Môi trường tỉnh Cà Mau; kết quả cho thấy nồng độ BOD dao động 16–63 mg/L, COD 11–35 mg/L và TSS 46–173 mg/L. Trong phân tích thống kê và mô hình hóa của nghiên cứu này, các giá trị BOD, COD và TSS được coi là các biến phụ thuộc, trong khi các biến độc lập được trích xuất từ dữ liệu ảnh viễn thám đa nguồn để mô tả các yếu tố môi trường và không gian liên quan.

**2.2.2. Xử lý dữ liệu viễn thám đa nguồn và trích xuất chỉ số vật lý cho vùng ven biển**

Trong nghiên cứu này, hai nguồn dữ liệu ảnh vệ tinh được sử dụng gồm Sentinel-2 MSI và Sentinel-1 SAR được thu nhận vào ngày

30/01/2025, thời điểm gần với đợt lấy mẫu ngoài thực địa. Mặc dù tồn tại sai lệch nhỏ về thời gian giữa ngày thu nhận ảnh viễn thám và ngày đo đạc các thông số chất lượng nước, khoảng cách này được xem là chấp nhận được, cho phép tiến hành so sánh và phân tích một cách khả thi [8]. Để đơn giản hóa quy trình xử lý, cả hai loại dữ liệu được tái lấy mẫu về cùng kích thước pixel 10 m bằng kỹ thuật nội suy tuyến tính theo phương pháp song tuyến (Bilinear). Trên cơ sở tổng hợp lý thuyết và thực tiễn từ các nghiên cứu trước đây [3, 4], các chỉ số vật lý được lựa chọn trong nghiên cứu nhằm phản ánh đầy đủ những đặc trưng của môi trường nước tại khu vực ven biển tỉnh Cà Mau. Quy trình xử lý và trích xuất các chỉ số này được minh họa trong Hình 3.



Hình 3. Sơ đồ xử lý dữ liệu viễn thám đa nguồn

Dữ liệu viễn thám được tiền xử lý từ hai nguồn chính là Sentinel-2 MSI và Sentinel-1 SAR. Đối với ảnh Sentinel-2 MSI [9], hiệu chỉnh khí quyển và bức xạ được thực hiện bằng quy trình ATCOR tích hợp trong Catalyst Professional (<https://catalyst.earth/solutions/catalyst-professional>), gồm ba bước: (i) hiệu chỉnh đỉnh phản xạ ở tầng khí quyển; (ii) loại bỏ mây và sương mù; (iii) chuyển đổi giá trị bức xạ thành phản xạ bề mặt. Với dữ liệu Sentinel-1A SAR

[10], quy trình trong SNAP (<https://step.esa.int/main/download/snap-download>) bao gồm: (i) hiệu chuẩn tán xạ, chuyển DN sang hệ số tán xạ ngược (sigma naught); (ii) giảm nhiễu hạt bằng bộ lọc thích ứng Lee (cửa sổ 7x7); (iii) hiệu chỉnh địa hình bằng DEM độ phân giải 10 m để khắc phục biến dạng hình học. Cuối cùng, cả hai tập dữ liệu được hiệu chỉnh hình học và quy chiếu về hệ tọa độ VN2000, múi chiếu 48 nhằm đảm bảo

tính thống nhất và độ chính xác không gian. Nghiên cứu này lựa chọn các biến đầu vào từ ảnh viễn thám quang học và từ ảnh SAR (Bảng

1). Các tập biến này được sử dụng để mô hình hóa và ước tính nồng độ BOD, COD và TSS.

**Bảng 1. Các biến tham gia mô hình ước tính BOD, COD, TSS từ Sentinel-1 SAR và Sentinel-2 MSI.**

Biến dự đoán/ ký hiệu	Chú thích	Công thức
Sentinel-2 MSI	BLUE	Xanh lam (Blue)
	GREEN	Xanh lá (Green)
	RED	Đỏ (Red)
	NIR	Cận hồng ngoại (Near Infrared)
Các chỉ số quang phổ từ Sentinel-2	NDWI	Chỉ số Nước chênh lệch chuẩn hóa (Normalized Difference Water Index) $NDWI = \frac{GREEN - NIR}{GREEN + NIR}$
	MNDWI	Chỉ số Nước chênh lệch chuẩn hóa sửa đổi (Modification Of Normalized Difference Water Index) $MNDWI = \frac{GREEN - SWIR1}{GREEN + SWIR1}$
	WRI	Chỉ số Tỷ lệ nước (Water Ratio Index) $WRI = \frac{GREEN + RED}{NIR + SWIR1}$
	AWEI <sub>nsh</sub>	Chỉ số Khai thác nước tự động (Automated Water Extraction Index) $AWEI_{nsh} = 4 \times (GREEN - SWIR1) - (0.25 \times NIR + 2.75 \times SWIR2)$
	AWEI <sub>sh</sub>	Chỉ số Khai thác nước tự động (Automated Water Extraction Index) $AWEI_{sh} = BLUE + 2.5 * GREEN - 1.5 * (NIR + SWIR1) - 0.25 * SWIR2$
	SWI	Chỉ số nước Sentinel-2 (Sentinel-2 water index) $SWI = \frac{VRE1 - SWIR2}{VRE1 + SWIR2}$
	MBWI	Chỉ số Nước đa kênh phổ (Multi-Band Water Index) $MBWI = 2 \times (GREEN) - RED - NIR - SWIR1 - SWIR2$
Sentinel-1 SAR	VH	Phân cực truyền dọc-nhận ngang (dB) (Vertical Transmit-Horizontal Receive polarizations) $Sigma\theta_{VH}$
	VV	Phân cực truyền dọc-nhận dọc (dB) (Vertical Transmit-Vertical Receive Polarizations) $Sigma\theta_{VV}$
	ADDITION <sub>vhvv</sub>	Chỉ số phân cực tổng (dB) (Polarization addition) $VH+VV$
	AVERAGE <sub>vhvv</sub>	Chỉ số phân cực trung bình (dB) (Polarization average) $(VH+VV)/2$
	MULT <sub>vhvv</sub>	Chỉ số phân cực tích (dB) (Polarization multiply) $VH*VV$

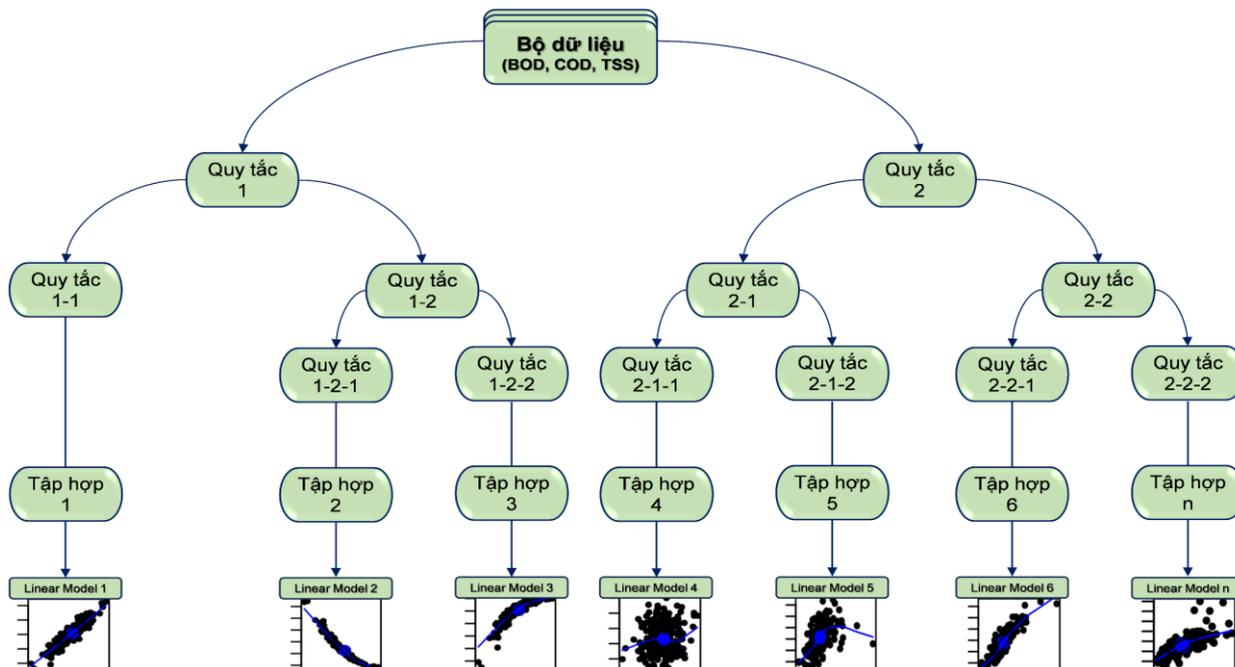
**2.2.3. Xây dựng và đánh giá mô hình dự đoán thông số chất lượng nước**

Nghiên cứu sử dụng mã nguồn mở của mô hình Cubist trong phần mềm R để dự đoán các thông số BOD, COD và TSS trong vùng nước mặt ven biển tỉnh Cà Mau. Mô hình Cubist được sử

dụng bởi sự kết hợp cấu trúc dạng cây với mô hình hồi quy đa biến tại mỗi quy tắc, trong đó mỗi nhánh kết thúc không chỉ trả về một giá trị cố định mà gắn kèm một mô hình tuyến tính đa biến áp dụng khi các điều kiện quy tắc thỏa mãn (Hình 4). Cơ chế này cho phép Cubist mô tả các

quan hệ phức tạp giữa biến đầu vào và biến mục tiêu tốt hơn các cây quyết định đơn thuần, đồng thời cung cấp khả năng diễn giải cao hơn các mô hình mạng nơ-ron phức tạp. So với rừng ngẫu nhiên (RF) hay Gradient Boosting (GBM), Cubist có khả năng ngoại suy và dự đoán vượt ra ngoài phạm vi giá trị quan sát trong bộ mẫu huấn luyện, điều quan trọng khi giá trị cực trị của thông số chất lượng nước có thể chưa được bao phủ đầy đủ trong tập dữ liệu hiện có

[11-13]. Mặc dù RF và GBM có ưu điểm về độ chính xác trong các dự báo thông thường, nhưng Cubist lại thể hiện sự vượt trội khi cần kết hợp các biến số có tính chất không tuyến tính mạnh mẽ. Do đó phương pháp này phù hợp để xử lý mối quan hệ phi tuyến và biến thiên không gian phức tạp của các thủy vực ven biển tỉnh Cà Mau, đồng thời cung cấp kết quả có tính diễn giải và độ tin cậy cao cho các thông số BOD, COD và TSS.



Hình 4. Sơ đồ khái niệm mô hình học máy Cubist dùng để dự đoán các thông số chất lượng nước

Để đánh giá độ chính xác và khả năng khái quát hóa của mô hình, nghiên cứu áp dụng phương pháp xác nhận chéo k-fold. Dữ liệu được chia ngẫu nhiên thành 10 tập con (k = 10); mỗi lần lặp, một tập con làm bộ kiểm định và chín tập còn lại làm bộ huấn luyện, tiến hành lặp lại sao cho mỗi tập con được sử dụng một lần làm kiểm định. Việc chọn k = 10 giúp cân bằng giữa độ tin cậy của ước tính và hiệu suất tính toán [3, 14]. Các chỉ số thống kê dùng để đánh giá và lựa chọn mô hình phù hợp cho ước tính BOD, COD và TSS gồm: hệ số xác định R<sup>2</sup>, sai số trung bình bình phương gốc RMSE và sai số tuyệt đối trung bình MAE. Các chỉ số này được tính cho cả giai đoạn huấn luyện (training) và giai đoạn xác thực chéo (cross-validation) để xác định mô hình Cubist phù hợp nhất với dữ liệu và vị trí nghiên cứu. Hệ số xác định R<sup>2</sup> được sử

dụng theo phương trình (1) để đánh giá mức độ phù hợp giữa giá trị quan sát và giá trị dự đoán.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i^{ob} - y_i^{pre})^2}{\sum_{i=1}^n (y_i^{ob} - \bar{y}^{ob})^2} \tag{1}$$

Trong đó:

n là một số mẫu quan sát được;

$y_i^{ob}$  là các giá trị quan sát được;

$y_i^{pre}$  là các giá trị dự đoán;

$\bar{y}^{ob}$  là mức trung bình của giá trị quan sát được.

Sai số toàn phương trung bình (RMSE) là tiêu chí khách quan dùng trong tối ưu hóa mô hình, phản ánh độ chính xác của ước tính bằng cách đo lường độ lệch trung bình theo chuẩn giữa giá trị dự đoán và giá trị quan sát. RMSE được tính theo phương trình (2):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i^{ob} - \bar{y}^{pre})^2}{n}} \quad (2)$$

Sai số tuyệt đối trung bình (MAE) biểu thị sai số dự đoán trung bình không có định hướng, đo bằng giá trị tuyệt đối của sai lệch giữa giá trị quan sát và giá trị dự đoán; MAE cung cấp phép đo trực tiếp về độ lớn trung bình của sai số. MAE được tính theo phương trình (3).

$$MAE = \frac{\sum_{i=1}^n |y_i^{ob} - y_i^{pre}|}{n} \quad (3)$$

Để tối ưu hóa ước tính các thông số chất lượng nước (BOD, COD, TSS), dữ liệu quan sát thực địa được chia ngẫu nhiên thành tập huấn luyện (70% tổng số mẫu) và tập kiểm định (30%). Sau khi hiệu chỉnh và lựa chọn mô hình tối ưu, toàn bộ mẫu thực địa sẵn có được sử dụng để so sánh trực tiếp các giá trị dự báo với giá trị quan sát để kiểm chứng độ chính xác

không gian của kết quả dự đoán các thông số BOD, COD và TSS trong vùng nước mặt ven biển tỉnh Cà Mau.

### 3. KẾT QUẢ VÀ THẢO LUẬN

#### 3.1. Độ chính xác mô hình dự đoán các thông số chất lượng nước mặt

Kết quả từ mô hình Cubist, được tối ưu bằng quy trình xác nhận chéo 10 lần, cho thấy khả năng tổng quát hóa tốt. Các chỉ số đánh giá trung bình của các mô hình tốt nhất nằm trong khoảng CV-RMSE = 3,25–3,84 mg/L, CV-R<sup>2</sup> = 0,76–0,87 và CV-MAE = 2,39–2,71 mg/L (Bảng 2). Các giá trị BOD, COD và TSS ước tính từ ảnh viễn thám có sai khác so với quan trắc thực tế nhưng mức chênh lệch này nằm trong giới hạn sai số của mô hình Cubist [3]. Điều đó chứng tỏ các ước tính BOD, COD và TSS dựa trên dữ liệu và phương pháp mô hình hóa đã chọn đạt độ chính xác chấp nhận được.

**Bảng 2. Độ chính xác của mô hình ước tính các thông số BOD, COD và TSS**

Thông số chất lượng nước	Số lượng chỉ số	Chỉ số đã chọn	Độ chính xác mô hình			Kiểm chứng		
			CV-RMSE	CV-R <sup>2</sup>	CV-MAE	RMSE	R <sup>2</sup>	MAE
<b>BOD</b> (nhu cầu oxy sinh học)	6	NDWI, WRI, MNDWI, AWEI <sub>sh</sub> , AWEI <sub>nsh</sub> , VH	3,84	0,76	2,71	4,38	0,74	3,22
<b>COD</b> (nhu cầu oxy hóa học)	8	AWEI <sub>nsh</sub> , WRI, AWEI <sub>sh</sub> , MBWI, NDWI, MNDWI, SWI, VH	3,39	0,83	2,46	4,05	0,81	3,28
<b>TSS</b> (tổng chất rắn lơ lửng)	5	MNDWI, AWEI <sub>sh</sub> , AWEI <sub>nsh</sub> , NDWI, VH	3,25	0,87	2,39	3,47	0,85	2,43

Trên cơ sở kết quả thu được, khi sử dụng các tổ hợp biến đầu vào khác nhau trong mô hình học máy Cubist cho độ chính xác ước tính tương đối tốt và đủ tin cậy để ứng dụng thực tiễn. Do đó, các phương trình ước tính BOD

(phương trình 4), COD (phương trình 5) và TSS (phương trình 6) có thể được sử dụng để định lượng không gian các thông số chất lượng nước từ dữ liệu ảnh viễn thám đa nguồn tại khu vực ven biển tỉnh Cà Mau.

$$BOD = 35.5 - 25.4 \times NDWI + 4.34 \times WRI - 4.8 \times MNDWI - 43.3 \times AWEI_{sh} - 0.12 \times VH + 15.5 \times AWEI_{nsh} \quad (4)$$

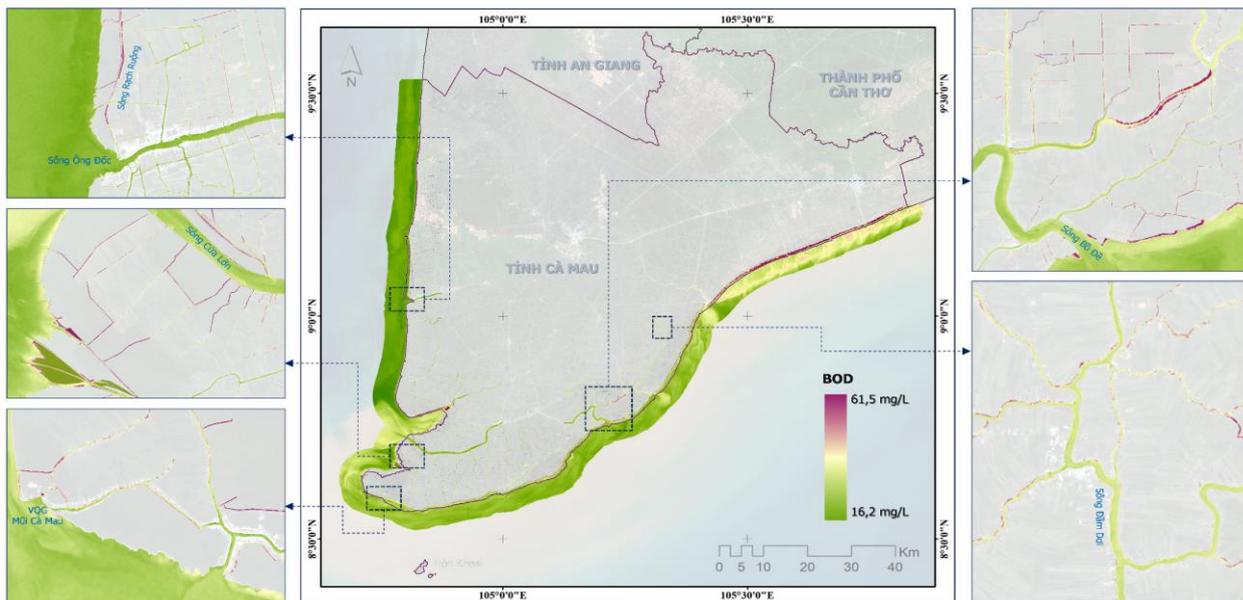
$$COD = 35.5 + 32.5 \times AWEI_{nsh} - 8.5 \times WRI - 53.5 \times AWEI_{sh} - 44 \times MBWI + 26.5 \times NDWI - 9.1 \times MNDWI + 10.6 \times SWI - 0.21 \times VH \quad (5)$$

$$TSS = 105.5 - 92.5 \times MNDWI - 43 \times AWEI_{sh} - 95 \times NDWI + 0.25 \times VH + 24.5 \times AWEI_{nsh} \quad (6)$$

#### 3.2. Phân bố không gian của BOD, COD và TSS trong vùng nước mặt ven biển tỉnh Cà Mau

Hình 5 cho thấy nồng độ BOD dao động từ 16,2 mg/L đến 61,5 mg/L và phân bố không đồng đều trên vùng nước mặt ven biển, tỉnh Cà Mau. Giá trị nồng độ BOD cao nhất (màu đỏ-tím) tập trung dọc kênh rạch ven bờ, trong các

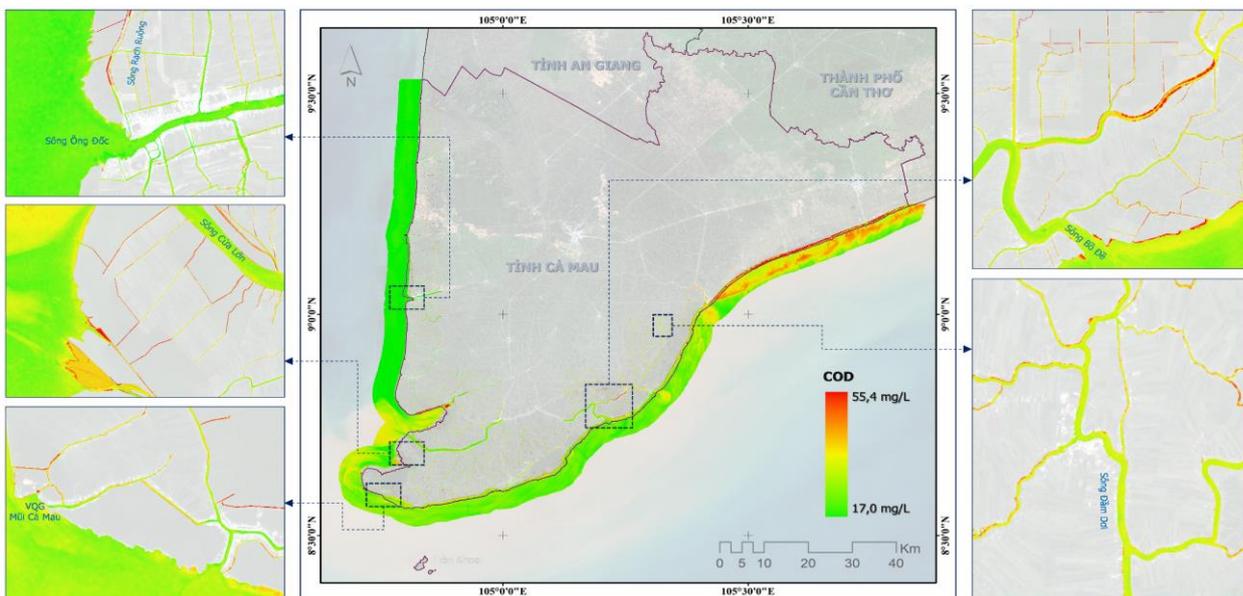
vùng nước lợ gần khu dân cư và khu nuôi trồng thủy sản. Ngược lại, các vùng cửa biển mở và khu vực có trao đổi thủy động lực mạnh có nồng độ BOD thấp hơn (màu xanh lục), phản ánh sự pha loãng và trao đổi với nước biển làm giảm nồng độ hữu cơ hòa tan trong nước.



Hình 5. Kết quả phân bố không gian đại diện được hiện thị của nồng độ BOD dự đoán

Nồng độ COD dao động từ 17,0 mg/L đến 55,4 mg/L và có sự phân bố không gian tương tự BOD nhưng lan rộng hơn do COD phản ánh tổng chất hữu cơ và các chất oxy hóa (Hình 6). Giá trị COD cao tập trung quanh cửa kênh rạch, vùng đô thị ven sông và các khu nuôi trồng thủy sản, lan dọc theo hướng dòng chảy và giảm dần

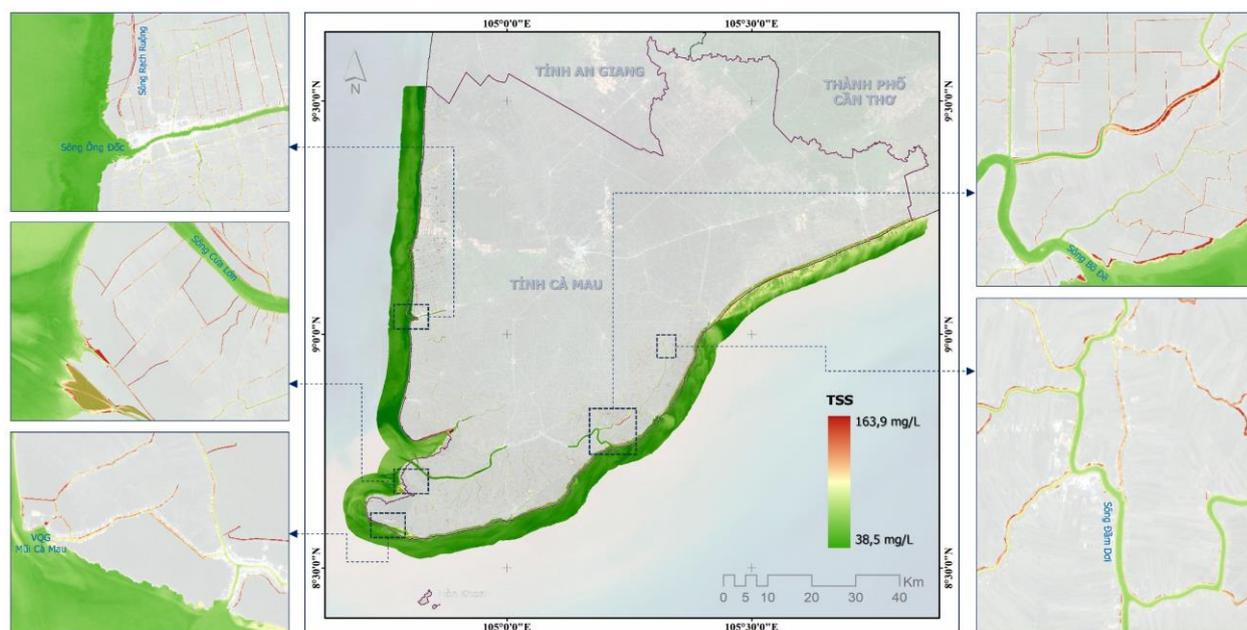
về phía biển. Các vùng COD cao (màu đỏ-cam) xuất hiện ở đoạn kênh ít chảy và khu vực tiếp nhận nước thải công nghiệp, nước thải sinh hoạt; các vùng COD thấp (màu xanh mint) tập trung ở cửa biển, nơi quá trình pha loãng, trao đổi thủy lưu và oxy hóa nhờ ánh sáng mặt trời làm giảm nồng độ các hợp chất hữu cơ.



Hình 6. Kết quả phân bố không gian đại diện được hiện thị của nồng độ COD dự đoán

Nồng độ TSS dao động mạnh từ 38,5 mg/L đến 163,9 mg/L và phân bố không đồng đều dọc ven biển, tỉnh Cà Mau (Hình 7). Hàm lượng TSS cao (màu đỏ sẫm) tập trung dọc dải ven kênh rạch và các đoạn sông gần khu đô thị; ngược lại, vùng cửa biển mở và vùng có trao đổi

thủy triều mạnh biểu hiện TSS thấp (màu xanh lá) do quá trình phân tán và lắng đọng trầm tích. Khu vực có xáo trộn đáy hoặc nơi dòng chảy thay đổi, tạo thành dải hoặc điểm có nồng độ TSS cao kéo dài từ nội thủy ra cửa biển.



Hình 7. Kết quả phân bố không gian đại diện được hiện thị của nồng độ TSS dự đoán

Kết quả mô hình Cubist cho thấy sự khác biệt rõ rệt trong phân bố không gian của ba thông số chất lượng nước BOD, COD, TSS. Sự khác biệt này phản ánh trực tiếp các tác động giữa các yếu tố nhân sinh, địa hình thấp trũng và điều kiện thủy động lực ven bờ biển cũng như tính phức tạp trong việc phân tích nước mặt vùng ven biển khu vực tỉnh Cà Mau. Cụ thể, nguồn ô nhiễm hữu cơ sẽ khuếch tán từ đất liền xuống biển theo dòng chảy ven bờ, trong khi các hạt rắn dễ bị vận chuyển và lắng đọng tại các cửa sông do thay đổi năng lượng dòng chảy. Như vậy, không chỉ hoạt động của con người mà cả điều kiện ven biển tự nhiên đều quyết định kiểu phân bố không gian của BOD, COD và TSS. Mặt khác, các cửa sông và vùng ven biển có khả năng tự làm sạch cao hơn nhờ quá trình pha loãng với nước biển và khả năng phân hủy sinh học, nhưng hiệu quả này phụ thuộc vào lưu lượng, thủy triều và tải trọng ô nhiễm đầu nguồn. Những kết quả cho thấy cần ưu tiên kiểm soát chất thải hữu cơ tại nguồn ở vùng ven sông và kênh rạch, đồng thời áp dụng biện pháp bảo vệ bờ và quản lý trầm tích tại cửa sông ven biển tỉnh Cà Mau. Giải pháp này hỗ trợ quản lý chất lượng nước mặt theo hướng phát triển bền vững, phù hợp với mục tiêu về nước sạch và bảo tồn tài nguyên biển (SDG 6 và SDG 14).

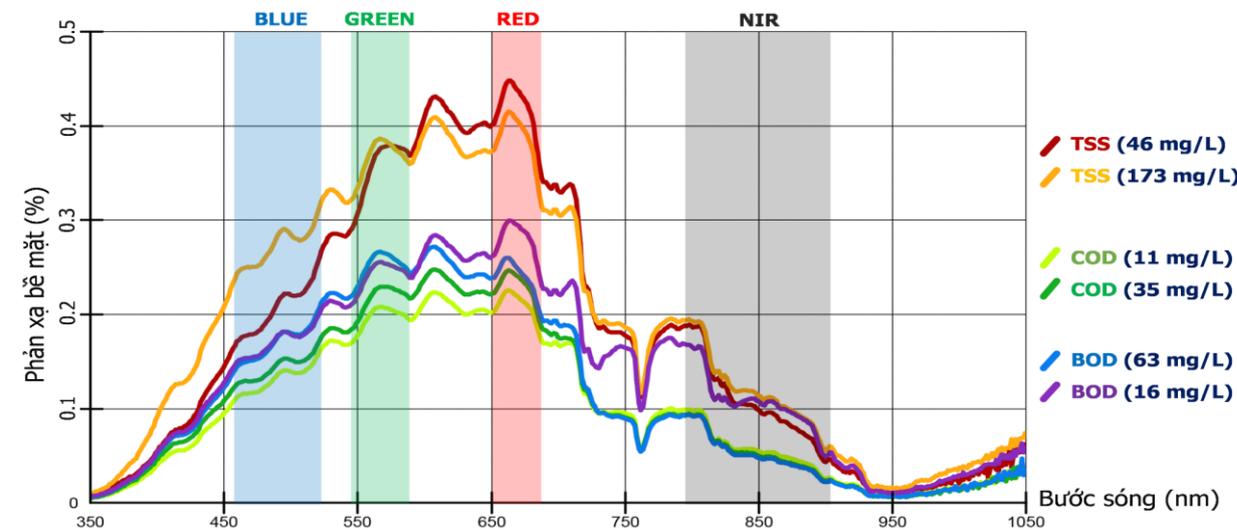
### 3.3. Khả năng ứng dụng viễn thám kết hợp mô hình Cubist để ước tính BOD, COD và TSS

Dữ liệu từ viễn thám đa nguồn (quang học

và radar) cho thấy khả năng theo dõi BOD, COD và TSS thông qua mối liên hệ giữa thành phần nước và tín hiệu phổ phản xạ (Hình 8). Chỉ số TSS và một phần hợp chất hữu cơ hòa tan làm thay đổi phản xạ ở vùng nhìn thấy (BLUE, GREEN, RED) và cận hồng ngoại (NIR), trong khi các hạt lơ lửng và độ nhám mặt nước ảnh hưởng lên hệ số tán xạ ngược trong vùng sóng vi ba. Các chỉ số BOD và COD phản ánh tải trọng chất hữu cơ tổng có thể liên kết với chất rắn hoặc làm biến đổi đặc tính quang học của cột nước, từ đó tác động gián tiếp đến tín hiệu viễn thám [3]. Trong nghiên cứu này, mô hình học máy Cubist được sử dụng nhờ khả năng kết hợp giữa tính chính xác của các mô hình dự đoán phi tuyến và tính diễn giải của các mô hình tuyến tính. Ở các khu vực kênh rạch và nhánh sông nhỏ, nơi mặt nước có kích thước tương đương với độ phân giải không gian của ảnh vệ tinh, việc phân biệt trực tiếp các vật chất lơ lửng trong nước dựa trên ảnh viễn thám quang học thường thiếu chính xác do ảnh hưởng của điều kiện chiếu sáng, nhiễu khí quyển và độ phản xạ thấp của bề mặt nước. Do đó, tích hợp dữ liệu viễn thám đa nguồn (quang học – radar) kết hợp với mô hình học máy Cubist là hướng tiếp cận hợp lý để tăng cường khả năng phát hiện nước mặt và cải thiện việc ước tính các thông số chất lượng nước. Trong nghiên cứu này đã thực hiện hiệu chỉnh nền bức xạ là phương pháp loại nhiễu và hiệu chuẩn chéo giữa dữ liệu

hiện trường (lấy mẫu tại chỗ, đo quang phổ) và ảnh sau hiệu chỉnh khí quyển là bước bắt buộc để đảm bảo tính nhất quán, giảm sai số và nâng

cao độ tin cậy của mô hình dự đoán, đặc biệt là ở vùng và ở khu vực có phản xạ rất cao (ví dụ rừng dừa, lồng nuôi, tàu thuyền).



Hình 8. Phản xạ bề mặt theo bước sóng cho các thông số TSS, COD và BOD trong nghiên cứu

Các chỉ số quang phổ và radar như  $AWEI_{nsh}$ , WRI,  $AWEI_{sh}$ , MBWI, NDWI, MNDWI, SWI và hệ số tán xạ VH có khả năng bổ trợ lẫn nhau trong ước tính các thông số chất lượng nước mặt: NDWI/MNDWI và  $AWEI$  dùng để phân lớp mặt nước cơ bản và loại bỏ nền không phải nước; SWI và WRI nhạy với tải chất rắn lơ lửng, phù hợp cho ước tính TSS;  $AWEI_{nsh}$  và  $AWEI_{sh}$  tăng nhạy với tải hữu cơ hòa tan và giảm nhiều từ thảm thực vật ven bờ; MBWI giúp phân biệt bề mặt đất trơ và bờ bùn; trong khi VH cung cấp thông tin về độ nhám bề mặt, vật thể nổi và vật chất lơ lửng khi dữ liệu quang học bị che phủ mây. Cách tiếp cận này thể hiện tiềm năng trong việc nâng cao khả năng dự đoán của mô hình bằng cách tận dụng thông tin đa chiều từ dữ liệu viễn thám, song việc tích hợp thêm các biến tùy vào bối cảnh là điều kiện cần để đảm bảo độ chính xác và khả năng áp dụng rộng rãi của mô hình trong các môi trường nước khác nhau.

#### 4. KẾT LUẬN

Nghiên cứu chỉ ra rằng tích hợp viễn thám đa nguồn (Sentinel-2 MSI và Sentinel-1 SAR) với mô hình học máy Cubist cho phép ước tính không gian nồng độ BOD, COD và TSS ở vùng ven biển Cà Mau với độ chính xác thực nghiệm đạt mức chấp nhận được (RMSE: 3,47–4,38 mg/L;  $R^2$ : 0,74–0,85; MAE: 2,43–3,22 mg/L). Các chỉ số quang học và radar đã cung cấp thông tin bổ trợ nhạy cảm với môi

trường nước mặt, chỉ số MNDWI, NDWI và  $AWEI_{sh}$  tương quan mạnh với TSS;  $AWEI_{nsh}$ ,  $AWEI_{sh}$ , MBWI, NDWI nhạy với tải trọng hữu cơ, trong khi VH cung cấp tín hiệu về độ nhám và vật chất lơ lửng. Kết quả bản đồ cho thấy BOD và COD cao ở khu vực kênh rạch, vùng nước lợng và khu vực nuôi trồng thủy sản, trong khi TSS tập trung dọc ven sông và cửa sông, phản ánh tương tác giữa nguồn thải nhân sinh và thủy động lực học ven bờ.

Công nghệ viễn thám là công cụ hiệu quả cung cấp phạm vi bao phủ rộng và dữ liệu nhất quán mà phương pháp lấy mẫu phải tốn kém nhiều hơn để có thể đạt được. Phương pháp sử dụng trong nghiên cứu này được đề xuất để góp phần làm rõ tính khả thi trong giám sát chất lượng nước mặt ở khu vực thủy động lực phức tạp, cung cấp ước tính liên tục, không xâm lấn với độ tin cậy phù hợp cho ứng dụng quản lý. Kết quả cho thấy hàm tính toán của bộ chỉ số quang phổ và chỉ số radar bổ trợ lẫn nhau, cho phép tách nguồn tín hiệu và cung cấp ước tính liên tục, không xâm lấn với độ chính xác phù hợp cho mục tiêu giám sát; tuy nhiên, hiệu quả thực tế phụ thuộc vào hiệu chỉnh ảnh, hiệu chuẩn chéo với dữ liệu hiện trường và đưa vào các điều kiện môi trường bổ sung (góc quan sát, thủy triều, gió, độ sâu). Trên cơ sở kết quả này, việc ứng dụng trong quản lý cần đi kèm các biện pháp vận hành: triển khai chương trình lấy mẫu

hiện trường đồng thời với ảnh vệ tinh để hiệu chuẩn và kiểm chứng mô hình; bổ sung điều kiện môi trường và tối ưu hóa tiền xử lý để giảm nhiễu do tảo, dầu và bề mặt bùn. Kết quả thu được nên được tích hợp vào quy hoạch quản lý nguồn thải, trầm tích và định hướng phát triển ven biển lâu dài. Các quyết định quản lý hiệu quả cần sự phối hợp giữa cơ quan quản lý, cộng đồng dân cư và khu vực tư nhân để giảm nguồn ô nhiễm và bảo đảm sử dụng tài nguyên nước ven biển theo hướng bền vững.

#### Lời cảm ơn

Nghiên cứu này được tài trợ bởi dự án “Ứng dụng công nghệ viễn thám trong quản lý đất đai, môi trường (giám sát ô nhiễm môi trường nước mặt), giám sát biến động đường bờ, quản lý biển và hải đảo tỉnh Cà Mau”. Nhóm tác giả xin cảm ơn Cơ quan Vũ trụ châu Âu (ESA) đã cung cấp quyền truy cập mở dữ liệu ảnh vệ tinh Sentinel-2 MSI và Sentinel-1 SAR.

#### TÀI LIỆU THAM KHẢO

- [1]. Katie K. Arkema, Laurel Field, Laura K. Nelson, Natalie C. Ban, Cailene Gunn & Sarah E. Lester (2024). Advancing the design and management of marine protected areas by quantifying the benefits of coastal ecosystems for communities. *One Earth*. 7(6): 989–1006. DOI: 10.1016/j.oneear.2024.04.019
- [2]. Megan Gomes, Timothy J. Ralph, Marc S. Humphries, Bradley P. Graves, Tsuyoshi Kobayashi & Damian B. Gore (2025). Waterborne contaminants in high intensity agriculture and plant production: A review of on-site and downstream impacts. *Science of the Total Environment*. 958: 178084. DOI: 10.1016/j.scitotenv.2024.178084.
- [3]. Thi-Nhung Do, Diem-My Thi Nguyen, Jiwnath Ghimire, Kim-Chi Vu, Lam-Phuong Do Dang, Sy-Liem Pham & Van-Manh Pham (2023). Assessing surface water pollution in Hanoi, Vietnam, using remote sensing and machine learning algorithms. *Environmental Science and Pollution Research*. 30(34): 82230–82247. DOI: 10.1007/s11356-023-28127-2.
- [4]. Đỗ Thị Nhung, Nguyễn Thị Diễm My, Trần Quốc Tuấn, Nghiêm Văn Tuấn & Phạm Văn Mạnh. (2022). Định lượng ô nhiễm nước mặt tại thành phố Hội An dựa trên dữ liệu viễn thám và mô hình học máy. *Tạp chí Khoa học Đo đạc và Bản đồ*. 52: 54–64. DOI: 10.54491/jgac.2022.52.598.
- [5]. Analy Baltodano, Afnan Agramont, Katoria Lekarkar, Evangelos Spyrakos, IIs Reusen & Ann van Griensven (2024). Exploring global remote sensing products for water quality assessment: Lake Nicaragua case study. *Remote Sensing Applications: Society and Environment*. 36: 101331. DOI: 10.1016/j.rsase.2024.101331.
- [6]. Hongran Li, Hui Zhao, Chao Wei, Min Cao, Jian Zhang, Heng Zhang & Dongqing Yuan (2024). Assessing water quality environmental grades using hyperspectral images and a deep learning model: A case study in Jiangsu, China. *Ecological Informatics*. 84: 102854. DOI: 10.1016/j.ecoinf.2024.102854.
- [7]. Jun Huang, Xiaojuan Liu, Yan Lin & Lipeng Ge (2024). Spatio-temporal pattern analysis of coastal zone in Nansha based on remote sensing technology. *Remote Sensing Applications: Society and Environment*. 35: 101222. DOI: 10.1016/j.rsase.2024.101222.
- [8]. Le Van Muoi, Chotpantararat Srilert, Van Pham Dang Tri & Toan Pham Van (2022). Spatial and temporal variabilities of surface water and sediment pollution at the main tidal-influenced river in Ca Mau Peninsular, Vietnamese Mekong Delta. *Journal of Hydrology: Regional Studies*. 41: 101082. DOI: 10.1016/j.ejrh.2022.101082.
- [9]. Nguyễn Thị Thu Hằng, Đỗ Diệu Anh, Đặng Đỗ Lâm Phương, Đỗ Thị Nhung, Nguyễn Thị Diễm My, Nguyễn Hữu Việt Hiệu, Nguyễn Viết Dương & Phạm Văn Mạnh (2025). Ứng dụng công nghệ địa không gian và thuật toán học máy lập bản đồ nguy cơ cháy rừng nhiệt đới theo mùa khu vực Tây Nguyên, Việt Nam. *Tạp chí Khoa học và Công nghệ Lâm nghiệp*. 14(4): 107–118. DOI: 10.55250/Jo.vnuf.14.4.2025.107-118
- [10]. Phan Lê Hoàng, Đặng Đỗ Lâm Phương, Đỗ Thị Nhung, Nguyễn Thị Thu Hằng, Nguyễn Thị Diễm My, Hứa Hoàng Huế & Phạm Văn Mạnh (2025). Tích hợp viễn thám và giải tích số định lượng xâm nhập mặn tại hạ du Đồng bằng sông Cửu Long, Việt Nam. *Tạp chí Khoa học và Công nghệ Lâm nghiệp*. 14(5): 087–098. DOI: 10.55250/Jo.vnuf.14.5.2025.087-098
- [11]. Haixia Guo, Rongshuo Cai & Hongjian Tan (2023). Projected harmful algal bloom frequency in the Yangtze River Estuary and adjacent waters. *Marine Environmental Research*. 183: 105832. DOI: 10.1016/j.marenvres.2022.105832
- [12]. M. M. Mahbubul Syeed, Md. Shakhawat Hossain, Md. Rajaul Karim, Mohammad Faisal Uddin, Mahady Hasan & Razib Hayat Khan (2023). Surface water quality profiling using the water quality index, pollution index and statistical methods: A critical review. *Environmental Sustainability Indicators*. 18: 100247. DOI: 10.1016/j.indic.2023.100247
- [13]. Ziying Wu, Jingjia Pang, Jinyu Li, Yuwen Wang, Jingyi Ruan, Xueling Zhang, Linshu Yang, Yuxuan Pang & Ying Gao (2025). A review of remote sensing-based water quality monitoring in turbid coastal waters. *Intelligent Marine Technology Systems*. 3(1): 24. DOI: 10.1007/s44295-025-00075-2.
- [14]. Rasmus Houborg & Matthew F. McCabe (2018). A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*. 135: 173–188. DOI: 10.1016/j.isprsjprs.2017.10.004.