

Bài báo khoa học

Nghiên cứu ứng dụng ảnh viễn thám Sentinel-2 kết hợp hồi quy đa biến xác định diễn biến nồng độ tổng chất rắn lơ lửng sông Thị Tính, tỉnh Bình Dương

Dương Bích Tuyền¹, Đỗ Xuân Hồng², Lê Trọng Diệu Hiền^{3*}

¹ Chương trình Thạc sĩ Khoa học Môi trường, Đại học Thủ Dầu Một, 06 Trần Văn Ôn, Thành phố Thủ Dầu Một, Việt Nam; tuyenduong0965@gmail.com

² Khoa Môi trường và Tài nguyên, Đại học Nông Lâm Tp. HCM; doxuanhong@hcmuaf.edu.vn

³ Khoa tài nguyên và Môi trường, Đại học Thủ Dầu Một, 06 Trần Văn Ôn, Thành phố Thủ Dầu Một, Việt Nam; hienltd@tdmu.edu.vn

*Tác giả liên hệ: hienltd@tdmu.edu.vn; Tel.: +84-372831517

Ban biên tập nhận bài: 7/10/2024; Ngày phản biện xong: 20/11/2024; Ngày đăng bài: 25/4/2025

Tóm tắt: Ô nhiễm chất lượng nước là một vấn đề môi trường nghiêm trọng, ảnh hưởng đến hệ sinh thái, đến sức khỏe con người và đến sự phát triển nền kinh tế. Khi các chất ô nhiễm như hóa chất, kim loại nặng, chất thải hữu cơ và vi khuẩn từ các nguồn công nghiệp, nông nghiệp dẫn đến sự suy giảm chất lượng nước. Ô nhiễm nước thường được đo lường qua sự gia tăng nồng độ các thông số chất lượng nước. Nghiên cứu trích xuất chỉ số nước (AWEI, WRI, và NDWI) từ dữ liệu ảnh viễn thám Sentinel-2 kết hợp với dữ liệu đo đạc thông số TSS thực tế để xây dựng mô hình hồi quy xác định diễn biến nồng độ TSS trên phạm vi sông Thị Tính và các phụ lưu. Kết quả nghiên cứu cho thấy phương pháp nghiên cứu mang tính thực tiễn cao giúp đánh giá nhanh và hiệu quả. Kết quả dự đoán giá trị nồng độ TSS giai đoạn 2021-2024 trong nước mặt dao động trong khoảng 1,97-83,78 mg/l. Mô hình hồi quy có hệ số xác định $R^2 = 67,4\%$ và $MSE = 0,17$. Phương pháp được áp dụng trong nghiên cứu có thể được sử dụng để đánh giá chất lượng nước các khu vực khác của tỉnh Bình Dương.

Từ khóa: Hồi quy đa biến; TSS; Sông Thị Tính; Sentinel-2.

1. Mở đầu

Tổng chất rắn lơ lửng (TSS) được xem là một trong những chất gây ô nhiễm chính góp phần làm suy giảm chất lượng nước, là một trong những thông số làm tăng chi phí xử lý nước, tác động đến loài động thực vật có trong nước. Các hạt lơ lửng (SPM) ảnh hưởng trực tiếp lẫn gián tiếp đến hệ sinh thái trong môi trường nước và các hoạt động có liên quan đến con người [1]. Các phương pháp đánh giá chất lượng nước truyền thống vẫn còn tồn tại nhiều hạn chế về thiết bị, phương pháp, và quy mô nghiên cứu, đặc biệt chi phí thực hiện là một vấn đề lớn. Phương pháp truyền thống đòi hỏi số lượng mẫu lớn và thời gian thực hiện kéo dài, điều này không chỉ làm giảm độ chính xác của nghiên cứu mà có khả năng gây phát sinh chi phí [2]. So với các phương pháp truyền thống, công nghệ viễn thám có nhiều lợi thế nổi bật, chẳng hạn như khả năng thu thập lượng dữ liệu lớn, tiết kiệm chi phí và thời gian so với việc chụp ảnh từ trên không ở các khu vực rộng. Thêm vào đó, công nghệ này còn cho phép phân tích dữ liệu và theo dõi sự thay đổi của chúng theo thời gian [3]. Công nghệ viễn thám, với khả năng sử dụng dữ liệu ảnh đa phổ và đa thời gian, đem lại một công cụ hiệu quả trong việc theo dõi liên tục và toàn diện diễn biến của chất lượng nước. Những yếu tố như độ đục, *Tạp chí Khí tượng Thủy văn* **2025**, 772, 38-50; doi:10.36335/VNJHM.2025(772).38-50 <http://tapchikttv.vn/>

độ mặn, và sự hiện diện của chất rắn lơ lửng trong nước có thể được theo dõi thường xuyên mà không bị ảnh hưởng từ các trạm quan trắc cố định [4]. Để đảm bảo việc quản lý và bảo vệ môi trường nước tại các con sông đạt hiệu quả cao, việc phát triển và áp dụng các giải pháp giúp chiết tách chỉ số chất lượng nước từ các dữ liệu ảnh viễn thám một cách nhanh chóng và chính xác là vô cùng quan trọng. Việc sử dụng công nghệ viễn thám cho phép thu thập và xử lý dữ liệu chất lượng nước từ một phạm vi rộng lớn mà không gặp phải các rào cản về thời gian và không gian như các phương pháp truyền thống. Một trong những công cụ mạnh mẽ hỗ trợ cho công tác này là Google Earth Engine, một nền tảng điện toán đám mây tiên tiến, được thiết lập đặc biệt để phục vụ các phân tích dữ liệu quy mô toàn cầu. Nền tảng này không chỉ cung cấp khả năng xử lý và phân tích khối lượng lớn dữ liệu ảnh vệ tinh, mà còn giúp người sử dụng thực hiện các nghiên cứu và theo dõi các biến động môi trường một cách hiệu quả và toàn diện, hỗ trợ quá trình ra quyết định trong việc bảo vệ tài nguyên nước. Nó cung cấp quyền truy cập vào một loạt hình ảnh vệ tinh, đặc biệt là các sản phẩm Landsat và Sentinel-2, đồng thời tạo điều kiện cho điện toán hiệu suất cao để phân tích xã hội và môi trường, bao gồm cả giám sát nước [5]. Việc Cơ quan Vũ trụ Châu Âu (ESA) ra mắt chương trình Sentinel như một phần của chương trình Copernicus là một thời điểm đột phá vì lần đầu tiên nó cung cấp quyền truy cập vào các hình ảnh có độ phân giải cao, miễn phí trên quy mô lớn. Các đặc điểm băng tần và phương pháp phản xạ quang phổ của dữ liệu Sentinel-2 rất hữu ích để phát hiện các vùng nước bề mặt [6]. Do đó, trong nghiên cứu này, Google Earth Engine đã được sử dụng để ứng dụng tiềm năng của các phương pháp tính toán chỉ số nước và chọn ra các chỉ số hoạt động tốt nhất để phát hiện nồng độ ô nhiễm bề mặt nước bằng cách sử dụng hình ảnh vệ tinh đa phổ của dữ liệu ảnh viễn thám Sentinel-2. Nghiên cứu liên quan đến sử dụng các chỉ số nhận biết nước trên nền tảng Google Earth Engine từ dữ liệu ảnh viễn thám Sentinel-2, tác giả đã tiến hành nghiên cứu sử dụng ba chỉ số NDWI, MNDWI và WNDWI để phân biệt và nhận diện các khu vực có nước và không có nước trong thành phố Sa Đéc, tỉnh Đồng Tháp. Kết quả nghiên cứu chỉ ra rằng chỉ số WNDWI và MNDWI hiệu quả hơn chỉ số NDWI trong việc phân biệt và nhận diện các khu vực có nước và không có nước. Cụ thể, các chỉ số WNDWI và MNDWI cho ra kết quả chính xác hơn, với giá trị của các chỉ số OA (*Overall Accuracy*), KC (*Kappa Coefficient*), EC (*Error Commission*), và EO (*Error Omission*) lần lượt là 0,98, 0,85, 0,015 và 0,006, cho thấy mức độ phân biệt rõ ràng và độ chính xác cao hơn so với chỉ số NDWI [7]. Trong một nghiên cứu gần đây, tác giả đã áp dụng phương pháp hồi quy để phân tích mối tương quan giữa chỉ số vật chất lơ lửng (*Normalized Suspended Material Index - NSMI*) từ ảnh viễn thám Sentinel-2A và lượng phù sa thực đo để xây dựng sơ đồ phân bố không gian của nồng độ tổng chất rắn lơ lửng (TSS) trong nước mặt tại tỉnh An Giang. Các nhà nghiên cứu đã sử dụng ảnh vệ tinh quang học Sentinel-2A để tính toán giá trị chỉ số NSMI và xác định hàm lượng TSS trong nước mặt, với giá trị khoảng từ 0-100 mg/l. Áp dụng phương trình hồi quy tuyến tính để thể hiện mối quan hệ này, cho ra hệ số xác định $R^2 = 0,868$, cho thấy sự tương quan mạnh mẽ giữa chỉ số NSMI và hàm lượng TSS. Kết quả nghiên cứu chỉ ra rằng việc sử dụng dữ liệu từ vệ tinh Sentinel-2A có thể hỗ trợ hiệu quả trong việc thành lập bản đồ phân bố hàm lượng ô nhiễm TSS trong nước mặt, đặc biệt là đối với năm 2019, và có độ tin cậy cao. Công nghệ viễn thám này không chỉ giúp theo dõi chất lượng nước một cách chính xác mà còn đóng góp trong việc bảo vệ tài nguyên nước và cải thiện công tác bảo vệ môi trường của cơ quan quản lý [8]. Chất lượng nước mặt sông Hồng và suối Đuống đoạn chảy qua thành phố Hà Nội cũng được nhóm tác giả xem xét bằng phương pháp tương tự. Nhóm tác giả đã lựa chọn phương pháp đánh giá dựa trên kết quả đo phổ phản xạ của mặt nước và kết quả phân tích mẫu nước để xác định các thông số chất lượng nước, bao gồm Độ đục, COD, BOD₅ và NH₄⁺. Qua nghiên cứu, nhóm tác giả nhận thấy rằng thông số Độ đục có mối liên hệ mật thiết với quang phổ phản xạ mặt nước, với hệ số xác định $R^2 = 0,663$, cho thấy sự phù hợp cao trong các mô hình hồi quy. Điều này cho thấy khả năng sử dụng phổ phản xạ để dự báo độ đục của nước một cách hiệu quả. Tuy nhiên, các thông số còn lại như COD, BOD₅ và NH₄⁺ không có mối liên hệ mật

thiết với quang phổ phản xạ mặt nước, cho thấy rằng việc sử dụng phổ phản xạ để xác định những thông số này không đạt được hiệu quả tương tự. Kết quả nghiên cứu này cung cấp một cái nhìn sâu sắc về mối quan hệ giữa các thông số chất lượng nước và đặc tính phổ phản xạ của mặt nước, góp phần vào việc cải thiện các phương pháp giám sát chất lượng nước từ xa [9].

Sông Thị Tính là một nhánh phụ của sông Sài Gòn, bắt nguồn từ huyện Bình Long, tỉnh Bình Phước, và chảy qua các khu vực như huyện Dầu Tiếng, huyện Bàu Bàng, thành phố Bến Cát, cùng một phần của thành phố Thủ Dầu Một, tỉnh Bình Dương. Con sông này chủ yếu chảy qua vùng địa hình gò đồi thấp, trước khi hợp lưu với sông Sài Gòn [10]. Trong nước cũng đã có những nghiên cứu về sông Thị Tính đa phần tập trung vào các thông số hữu cơ, điển hình là nhóm tác giả đã áp dụng mô hình MIKE 11 để tiến hành tính toán khả năng tiếp nhận và xác định mức độ phục hồi chất lượng nước của sông Thị Tính. Nghiên cứu này tập trung vào việc xem xét ảnh hưởng của các nguồn thải phân tán và nguồn thải điểm trong mùa mưa, đặc biệt là từ tháng 5 - tháng 9/2010, đối với hai chỉ tiêu nghiên cứu là BOD₅ và NH₄⁺. Các kết quả tính toán từ các kịch bản mô phỏng cho thấy rằng nếu các cơ sở xử lý đạt nước thải theo tiêu chuẩn xử lý cho phép, chất lượng nước sông sẽ có sự cải thiện rõ rệt. Tuy nhiên, bên cạnh sự cải thiện, các giá trị của một số chỉ tiêu chất lượng nước vẫn vượt quá mức quy chuẩn cho phép, điều này cho thấy rằng cần tiếp tục nỗ lực tăng cường hiệu quả của các công nghệ xử lý nước thải để đảm bảo đạt các tiêu chuẩn môi trường trong tương lai [11]. Bên cạnh đó, nhóm tác giả trong nước đã nghiên cứu và tính toán về hiện trạng chất lượng nước sông Thị Tính, đồng thời dự báo khả năng tiếp nhận chất ô nhiễm hữu cơ của con sông này đến năm 2020. Nghiên cứu được thực hiện với hai kịch bản khác nhau: một là kịch bản quản lý không tốt và một là kịch bản quản lý tốt. Kết quả của nghiên cứu cho thấy, dưới điều kiện quản lý không tốt, sông Thị Tính có dấu hiệu ô nhiễm nghiêm trọng, đặc biệt là đối với các hợp chất hữu cơ gây ô nhiễm, khả năng tiếp nhận chất ô nhiễm hữu cơ của sông đã giảm sút nghiêm trọng, đến mức không còn khả năng đáp ứng nhu cầu sử dụng nước cho sinh hoạt. Tuy nhiên, đối với mục đích sử dụng nước cho nông nghiệp, khả năng tiếp nhận chất ô nhiễm vẫn còn, mặc dù ở mức độ hạn chế, không đủ để đảm bảo nước có thể sử dụng lâu dài cho các hoạt động nông nghiệp, không tạo nên tác động tiêu cực. Kết quả này cho thấy sự cần thiết phải cải thiện công tác quản lý chất lượng nước sông Thị Tính nhằm bảo vệ và duy trì các mục đích sử dụng nước cho cộng đồng và sản xuất [12]. Một nghiên cứu khác cũng đã tiến hành đánh giá sự phân bố tải lượng ô nhiễm hữu cơ tại các tiểu vùng trong lưu vực sông Thị Tính, sử dụng phương pháp điều tra thực tế kết hợp với việc lấy mẫu nước để phân tích các chỉ tiêu ô nhiễm hữu cơ, bao gồm BOD, COD, tổng Nitơ (N) và tổng Phốt pho (P). Các mẫu được lấy tại các điểm xả thải lớn ở ba tiểu vùng 1, 2 và 3, nơi nước thải được xả vào sông Thị Tính. Từ các kết quả phân tích này, nghiên cứu đã tính toán tải lượng ô nhiễm dự báo cho đến năm 2020. Kết quả nghiên cứu cho thấy, với ba kịch bản khác nhau, các cơ quan quản lý môi trường địa phương đã nhận được một cái nhìn tổng thể và chi tiết về sự phân bố mức độ ô nhiễm cũng như các nguồn ô nhiễm chính trong từng tiểu vùng của lưu vực sông Thị Tính. Những thông tin này đóng vai trò quan trọng trong việc xây dựng các giải pháp quản lý, nâng cao chất lượng nước và bảo vệ môi trường sống của cộng đồng tại khu vực này [13].

Nhận thấy các nghiên cứu về thông số vô cơ tại sông Thị Tính tương đối ít, từ đó tác giả lựa chọn thông số TSS và sử dụng ảnh vệ tinh có độ phân giải không gian và thời gian kết hợp với dữ liệu quan trắc thông số TSS thực tế trên bề mặt nước, được thu thập từ các vị trí quan trắc cố định. Phương pháp này cho phép tận dụng ưu thế của ảnh vệ tinh trong việc giám sát và phân tích diễn biến chất lượng nước trong không gian rộng lớn, đồng thời kết hợp chặt chẽ với các số liệu quan trắc thực tế để nâng cao độ chính xác và độ tin cậy trong việc đánh giá mức độ ô nhiễm và các xu hướng thay đổi của thông số TSS trên sông Thị Tính. Sự kết hợp này dựa trên việc tích hợp công nghệ viễn thám và hệ thống thông tin địa lý (GIS), giúp tận dụng hết những ưu điểm của cả hai công nghệ trong việc thu thập, lưu trữ,

phân tích và xử lý dữ liệu địa lý. Từ đó, việc xây dựng, cập nhật và phân tích thông tin không gian được thực hiện một cách hiệu quả hơn, cho phép theo dõi chi tiết sự biến động của các chỉ số môi trường trong thời gian thực và trên phạm vi rộng. Điều này không chỉ hỗ trợ các nhà nghiên cứu trong việc đánh giá chất lượng môi trường mà còn cung cấp một công cụ mạnh mẽ cho quản lý môi trường. Mục tiêu của nghiên cứu là theo dõi diễn biến nồng độ TSS sông Thị Tính, tỉnh Bình Dương bằng dữ liệu viễn thám Sentinel-2 kết hợp hồi quy đa biến từ đó phân vùng mức độ ô nhiễm TSS theo các mức rất thấp, thấp, cao, rất cao giúp nhằm quản lý chất lượng nước khoanh vùng ô nhiễm TSS và có biện pháp quản lý kịp thời.

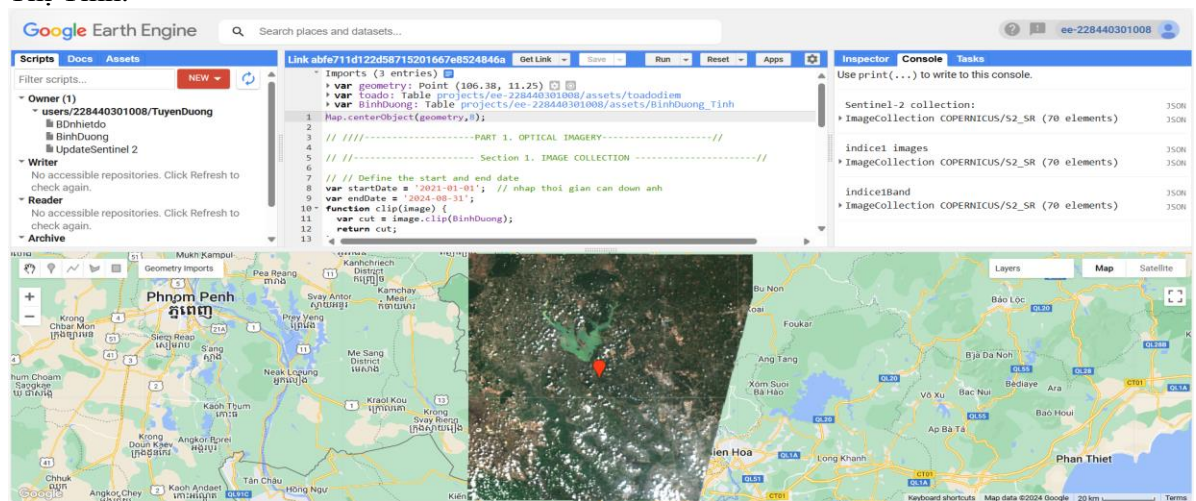
2. Phương pháp nghiên cứu

2.1. Dữ liệu

Thu thập ảnh viễn thám Sentinel-2 và chỉ số nước AWEI, WRI, và NDWI tính toán trực tiếp trên nền Google Earth Engine (GEE) tại website: <https://earthengine.google.com/> giai đoạn 2021-2024 với độ phân giải không gian 10m, độ che phủ mây dưới 30% (Hình 1). Thay vì chỉ tính toán trên một máy tính, GEE có thể phân phối công việc xử lý cho nhiều máy tính để giảm thời gian xử lý. Trong một số trường hợp, hình ảnh vệ tinh đã trải qua quá trình xử lý cơ bản. Điều này có nghĩa là xử lý bộ dữ liệu RS nhanh hơn, mang lại thời gian quay vòng nhanh hơn cho người dùng với chi phí thấp hơn [5].

Thu thập số liệu mẫu phân tích thông số TSS cùng thời điểm thu thập ảnh viễn thám hoặc chênh lệch không quá 5 ngày.

Thu thập bản đồ ranh giới hành chính huyện thuộc tỉnh Bình Dương và lớp bản đồ sông Thị Tính.



Hình 1. Giao diện website Google Earth Engine.

2.2. Khảo sát thực tế và phân tích chất lượng nước

Dữ liệu ban đầu bao gồm 82 mẫu phân tích nồng độ tổng chất rắn lơ lửng (TSS) có trong nước mặt, được kết hợp với dữ liệu tính toán các chỉ số nước như AWEI, NDWI, và WRI. Trong đó, nghiên cứu thực hiện thu thập 70 mẫu phân tích nồng độ TSS có trong nước mặt tại 5 vị trí thuộc Chương trình quan trắc chất lượng môi trường của tỉnh Bình Dương trong giai đoạn 2021-2024 và thực hiện lấy bổ sung 12 mẫu phân tích nồng độ TSS tại 6 vị trí vào 2 thời điểm (tháng 2 và tháng 6) năm 2024 (Bảng 1).

Bảng 1. Điểm quan trắc.

Vị trí	Ký hiệu mẫu	Tọa độ X	Tọa độ Y
Điểm thu thập			
Cầu Phú Bình	STT1	106.4922	11.2408
Cầu trên đường vành đai 4	STT2	106.5983	11.1008
Cầu Ông Cộ	STT3	106.6108	11.0383

Vị trí	Ký hiệu mẫu	Tọa độ X	Tọa độ Y
Suối Cắm Xe tại ngã 3 suối Bài Lãng và suối Cắm Xe	RTT1	106.4756	11.3233
Hộp lưu của suối Đông Sô và suối Đồi tại Cầu Quan	RTT2	106.5858	11.1542
Điểm lấy mẫu bổ sung			
Suối Hồ Muồng	NM1	106.5117	11.2642
Suối Đá Yêu	NM2	106.5109	11.3012
Suối Nhà Mát	NM3	106.5339	11.2240
Suối Bà Lãng	NM4	106.6022	11.2017
Suối Cầu Củi	NM5	106.5935	11.1581
Suối Tân Định	NM6	106.6125	11.0686

Bảng 2. Thời gian lấy mẫu.

STT	Ngày lấy mẫu	Ký hiệu mẫu				
		STT1	STT2	STT3	RTT1	RTT2
1	03/03/2021	24	20	18	25	25
2	06/05/2021	41	18	24	25	21
3	09/06/2021	97	29	36	68	29
4	05/11/2021	49	57	57	19	40
5	07/01/2022	27	9	15	14	22
6	03/03/2022	29	20	32	18	31
7	06/06/2022	5	8	8	6	7
8	05/12/2022	11	50	68	9	47
9	06/02/2023	27	134	58	7	20
10	03/11/2023	23	33	29	25	41
11	05/12/2023	28	33	33	24	26
12	29/02/2024	17	13	32	10	11
13	09/04/2024	8	15	17	6	14
14	13/06/2024	22	28	11	22	21

Mẫu nước được lấy ở giữa sông và ở bề mặt (cách 30 cm), lưu và bảo quản trong bình nhựa dung tích 1 lít. Các điểm lấy mẫu bổ sung được lựa chọn phân bố tại các nhánh phụ lưu chính của sông Thị Tịnh. Tọa độ vị trí lấy mẫu được xác định bằng thiết bị định vị toàn cầu GPS và kỹ thuật lấy mẫu được thể hiện trên Hình 2.



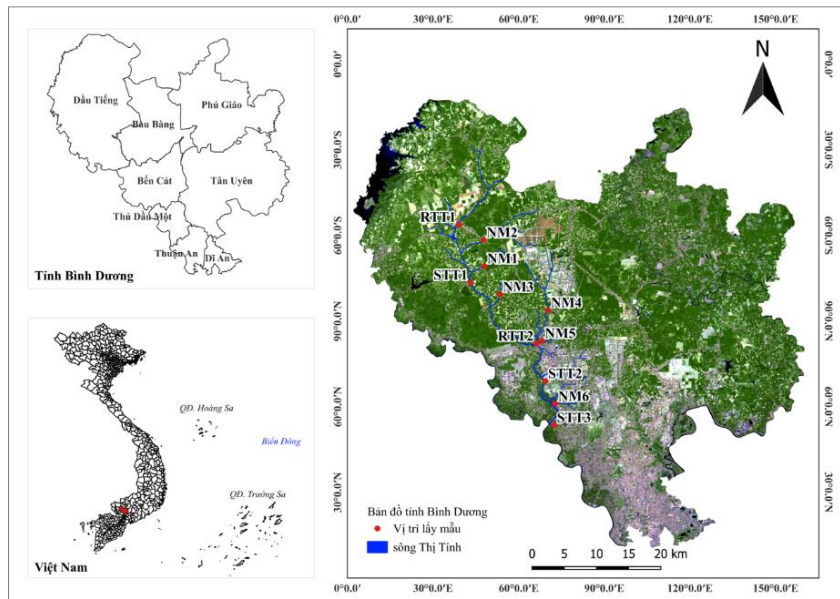
Hình 2. Minh họa thiết bị định vị toàn cầu GPS và kỹ thuật lấy mẫu.

Nghiên cứu áp dụng phương pháp phân tích theo TCVN 6625:2000 - Phương pháp phân tích khối lượng. Giá trị thông số TSS phân tích chất lượng nước trong các mẫu nước mặt có giá trị trong khoảng 4-97 mg/l. Các mẫu được chia thành hai nhóm ngẫu nhiên, 70% mẫu được dùng để tính toán và lựa chọn các chỉ số tối ưu nhất nhằm xây dựng mô hình dự đoán và 30% mẫu còn lại được sử dụng để kiểm định kết quả của mô hình dự đoán. Kết quả tích nồng độ TSS được sử dụng làm biến phụ thuộc, trong khi đó các biến độc lập là các chỉ số nước (AWEI, WRI và NDWI) được chiết tách từ dữ liệu ảnh viễn thám.

2.3. Giá trị chỉ số nước mặt từ dữ liệu ảnh viễn thám Sentinel-2

Một trong những phương pháp phổ biến trong việc chiết tách nước (*Water Indices - WIs*) là sử dụng các chỉ số toán học được tính từ các kênh hình ảnh, nhờ tính hiệu quả và tiện lợi. Dữ liệu các kênh đa phổ (Green, Red, NIR với độ phân giải không gian 10m; SWIR với độ phân giải không gian 20m) từ ảnh viễn thám Sentinel-2 [14] được lựa chọn để áp dụng nghiên

cứ. Các chỉ số nước như chỉ số khác biệt nước được chuẩn hóa (NDWI) [15] và chỉ số chiết tách nước tự động (AWEI) [16] và chỉ số tỉ lệ nước (WRI) [17] được tính toán thông qua các kênh ảnh (kênh 3), Red (kênh 4), NIR (kênh 8), SWIR1 (kênh 11) và SWIR2 (kênh 12) của Sentinel-2. Các chỉ số được thiết lập nhằm tăng cường sự khác biệt với các đối tượng nước trên ảnh quang học, bằng cách tăng sự khác biệt giữa các điểm ảnh chứa thông tin về nước và các điểm ảnh đại diện cho các đối tượng khác xung quanh, đồng thời loại bỏ các điểm ảnh không đạt.

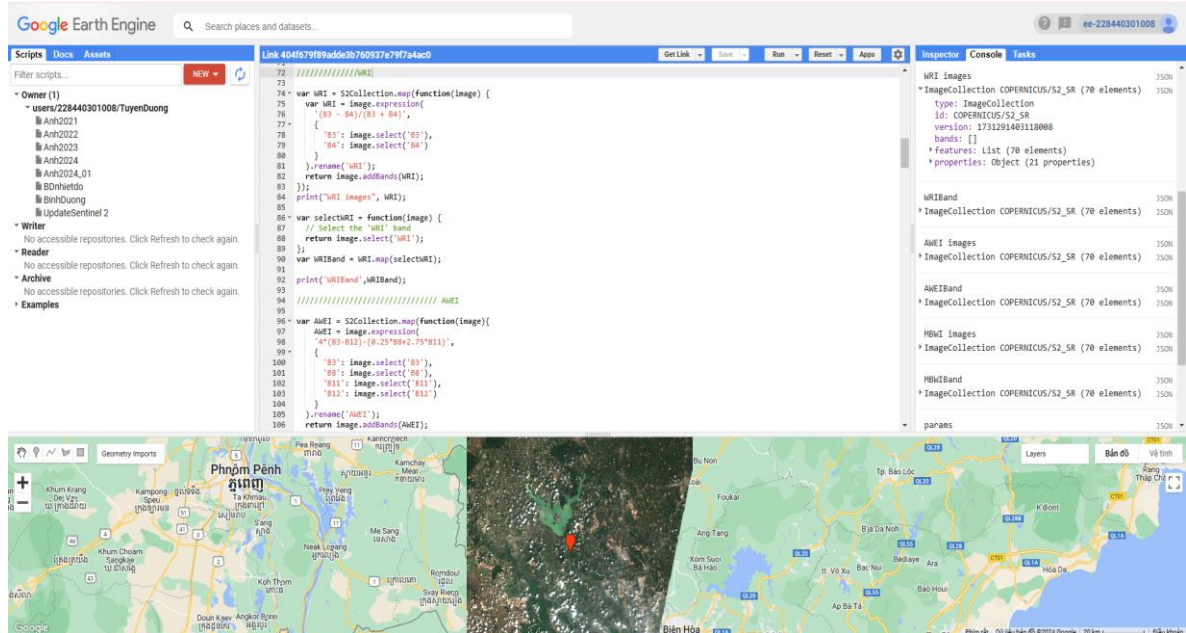


Hình 3. Vị trí lấy mẫu khu vực nghiên cứu.

Bảng 3. Chỉ số chiết tách từ dữ liệu viễn thám.

Chỉ số	Mô tả	Công thức
NDWI	Normalized Difference Water Index	$NDWI = \frac{GREEN - NIR}{GREEN + NIR}$
WRI	Water Ratio Index	$WRI = \frac{GREEN - RED}{GREEN + SWIR1}$
AWEI	Automated Water Extraction Index	$AWEI = 4 \times (GREEN - SWIR2) - (0,25 \times NIR + 2,75 \times SWIR1)$

Những chỉ số này được tính toán và trích xuất từ dữ liệu viễn thám Sentinel-2 trên nền tảng Google Earth Engine và được tải về Google Drive để dễ dàng xử lý dữ liệu và xây dựng mô hình (Hình 4).



Hình 3. Tính toán và trích xuất dữ liệu trên nền tảng Google Earth Engine.

2.4. Xây dựng mô hình quy tuyến tính

Ứng dụng mô hình hồi quy đa biến để tính toán thông số TSS thể hiện diễn biến chất lượng nước từ dữ liệu ảnh viễn thám. Nồng độ TSS được ước tính bằng cách sử dụng các biến độc lập khác nhau trong các thuật toán hồi quy đa biến. Nghiên cứu đã sử dụng 70% dữ liệu tham gia xây dựng mô hình và lựa chọn các chỉ số tối ưu nhất ước tính chất lượng nước bằng thông số TSS. Mô hình hồi quy đa biến thực hiện dựa trên các nghiên cứu đã công bố liên quan đến mô hình chất lượng nước [18–20] thực hiện phương trình (1) như sau:

$$y = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \dots + \beta_p x_p + \varepsilon \quad (1)$$

Trong đó y là biến phụ thuộc; $\beta_0, \beta_1, \beta_2, \dots$ là các tham số; x_1, x_2 là các biến độc lập; ε : sai số ngẫu nhiên. Phương trình có nhiều biến x (x_1, x_2, \dots, x_k) và mỗi biến có một thông số β_p ($k = 1, 2, \dots, p$) cần phải ước tính. Vì thế mô hình còn được gọi là mô hình hồi quy tuyến tính đa biến.

2.5. Kiểm định và đánh giá độ chính xác mô hình

Từ mô hình tính toán, kết quả của các phương trình hồi quy được kiểm định, đánh giá bằng cách sử dụng 30% dữ liệu còn lại để đối chiếu và lựa chọn mô hình tối ưu nhất. Phương trình tối ưu nhất được chọn dựa trên giá trị hệ số tương quan R^2 lớn nhất và sai số bình phương trung bình MSE giữa các giá trị được dự đoán và thực tế nhỏ nhất gần bằng 0. Tính toán giá trị hệ số tương quan R^2 để đánh giá mức độ tin cậy, nghiên cứu sử dụng công thức (2) như sau:

$$R^2 = \left[\frac{\sum_{i=1}^n (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_i)^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2}} \right]^2 \quad (2)$$

Trong đó x_i là giá trị thực đo; \bar{x}_i là giá trị thực đo trung bình; y_i là giá trị dự đoán; \bar{y}_i là giá trị dự đoán trung bình; n là số lượng giá trị tính toán.

Giá trị hệ số tương quan R^2 nằm trong khoảng từ 0-1, thể hiện mối tương quan giữa giá trị thực đo và giá trị dự đoán. Trường hợp R^2 gần bằng 0, thì kết quả có độ tin cậy kém. Ngược lại, nếu những giá trị này bằng 1, thì kết quả mô phỏng của mô hình là tối ưu nhất.

Sự khác nhau giữa giá trị đo đạc và giá trị dự báo được gọi là sai số. Khi sai số càng nhỏ thì độ chính xác sẽ càng cao. Kiểm định mô hình bằng sai số bình phương trung bình MSE, sai số càng thấp thì mô hình dự đoán càng tốt, nghiên cứu sử dụng công thức (3) như sau:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (3)$$

Trong đó \hat{y}_i là giá trị thực đo; y_i là biến dự đoán; n là biểu thị tổng giá trị.

3. Phân tích kết quả và thảo luận

3.1. Xây dựng mô hình hồi quy tương quan sử dụng nồng độ TSS và các chỉ số nước tính toán từ dữ liệu ảnh viễn thám Sentinel-2

Tiến hành xử lý dữ liệu đầu vào, thực hiện các bước kiểm tra phân phối chuẩn (histogram) và loại bỏ các giá trị ngoại lai (outliers) để đảm bảo chất lượng dữ liệu đầu vào. Những mẫu không đạt yêu cầu hoặc có khả năng gây sai lệch trong kết quả mô hình hồi quy được loại bỏ, sau cùng tổng số mẫu cuối cùng còn lại là 76 mẫu, đáp ứng tiêu chuẩn để đưa vào tính toán và kiểm định.

Thực hiện chia bộ dữ liệu ngẫu nhiên, nghiên cứu sử dụng 63 mẫu hiện trường kết hợp với dữ liệu tính toán các chỉ số nước (AWEI, NDWI, WRI) chiết tách từ ảnh Sentinel-2 để xây dựng mô hình hồi quy tuyến tính, tính toán mối quan hệ thống kê giữa nồng độ TSS trong nước mặt sông Thị Tính với 3 chỉ số nước.

Thực hiện kiểm tra mối tương quan giữa các biến độc lập, phân tích tương quan thực hiện trong phần mềm SPSS 20, kết quả thể hiện trong Bảng 4.

Bảng 4. Kết quả tương quan các biến độc lập.

		AWEI	NDWI	lnWRI
AWEI	Tương quan	1	-0,515	-0,274
	Mức ý nghĩa (2 đầu)		0,0	0,030
	Cỡ mẫu	63	63	63
NDWI	Tương quan	-0,515	1	0,55
	Mức ý nghĩa (2 đầu)	0,0		0,0
	Cỡ mẫu	63	63	63
lnWRI	Tương quan	-0,274	0,550	1
	Mức ý nghĩa (2 đầu)	0,030	0,0	
	Cỡ mẫu	63	63	63

Kết quả đầu ra phân tích về mối tương quan giữa các chỉ số nước < 0,6 và hệ số tương quan rất có ý nghĩa vì giá trị p-value là 0.00 và nhỏ hơn mức ý nghĩa ($\alpha = 5\%$). Kết quả mô hình và phân tích các giá trị hồi quy tuyến tính được trình bày ở Bảng 5 và 6.

Bảng 5. Tóm tắt mô hình hồi quy.

Mô hình	Hệ số R	Hệ số R bình phương	Hệ số R bình phương hiệu chỉnh	Sai số chuẩn của ước lượng	Trị số Durbin-Waston
1	0,821	0,675	0,685	0,434	1,015

Bảng 6. Mô hình quy tuyến tính thể hiện mối tương quan giữa các chỉ số nước và nồng độ TSS.

Mô hình với biến phụ thuộc là lnTSS						
Mô hình	Hệ số chưa chuẩn hóa		Hệ số đã chuẩn hóa		Kiểm định t	Mức ý nghĩa
	Hệ số B	Sai số chuẩn	Hệ số Beta			
Hằng số	2,962	0,055			54,012	0,00
AWEI	0,258	0,064	0,347		4,007	0,00
NDWI	-0,547	0,077	-0,708		-7,096	0,00
lnWRI	0,222	0,066	0,299		3,358	0,001

Như vậy, phương trình hồi quy để tính toán nồng độ TSS từ dữ liệu ảnh viễn thám được tính theo công thức như sau:

$$\ln TSS = 2,962 + 0,258 \times AWEI - 0,547 \times NDWI + 0,222 \times \ln WRI \quad (4)$$

Mô hình có hệ số xác định $R^2 = 0,675$, từ phương trình hồi quy tuyến tính được xây dựng có thể dự đoán giá trị của biến TSS (biến phụ thuộc) dựa trên giá trị của biến AWEI, NDWI, WRI (biến độc lập) cho thấy các biến độc lập phản ánh được 67% giá trị biến động nồng độ TSS.

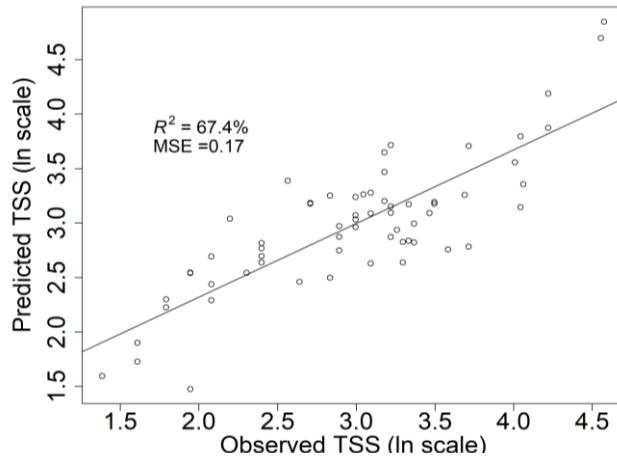
3.2. Đánh giá mức độ chính xác của mô hình

Kết quả của mô hình hồi quy tuyến tính được đánh giá độ tin cậy bằng hệ số MSE cho thấy mối tương quan giữa nồng độ TSS thực đo và dữ liệu TSS dự đoán. Mô hình hồi quy đáp ứng với hệ số xác định (R^2) và sai số bình phương trung bình (MSE). Kết quả cho thấy kiểm định mô hình có mức độ chính xác với tương quan thuận 0,17, hệ số xác định $R^2 = 67,4\%$ giá trị biến động (Hình 5).

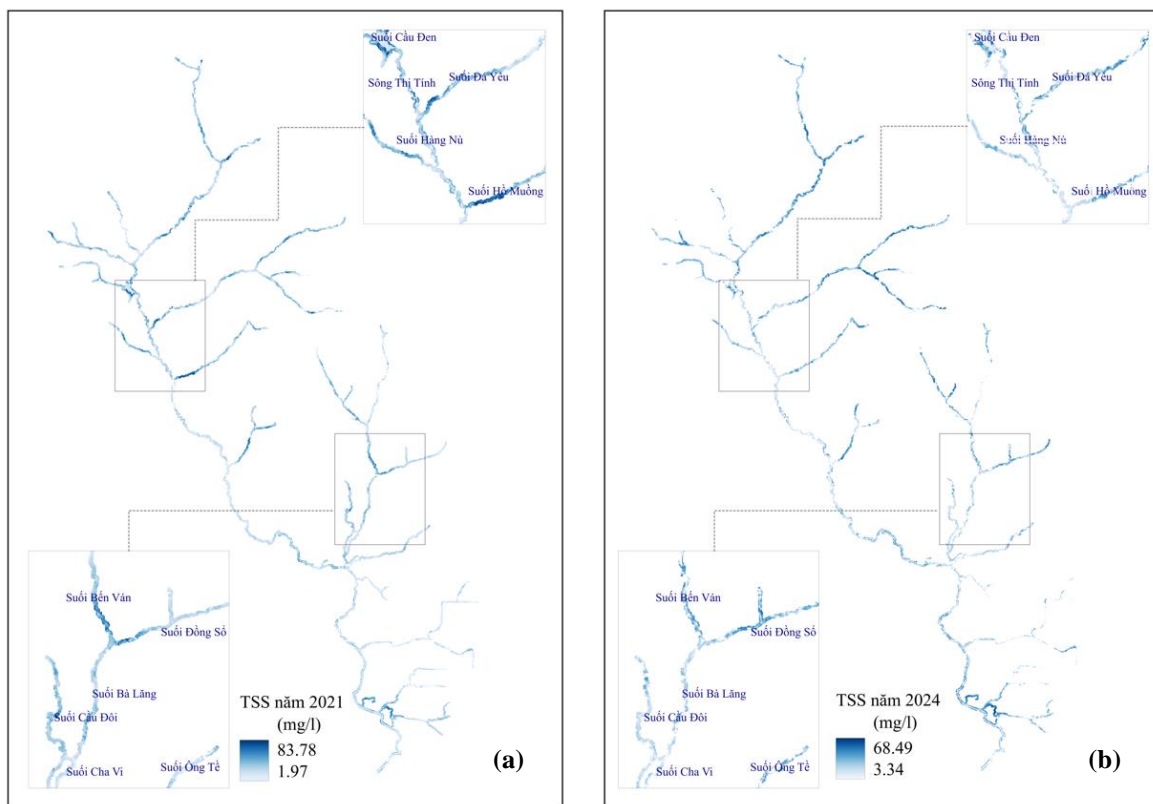
3.3. Phân bố không gian nồng độ TSS

Nồng độ TSS phân bố trong nước mặt sông Thị Tính, tỉnh Bình Dương giai đoạn 2021-2024 được trình bày trong Hình 6. Nồng độ TSS được tính toán từ dữ liệu ảnh vệ tinh Sentinel-2 theo công thức (4).

Sự phân bố không gian nồng độ TSS tương đối cao ở điểm hợp lưu suối đổ vào sông Thị Tính như suối Đá Yêu, suối Hồ Muồng, nồng độ năm 2021 dao động trong khoảng 1,9-83,7 mg/l (Hình 6a) và sự phân bố không gian nồng độ TSS cải thiện đáng kể khi nồng độ năm 2024 dao động trong khoảng 3,3-68,4 mg/l (Hình 6b).



Hình 4. Kết quả kiểm định mô hình hồi quy tuyến tính.



Hình 5. Phân bố nồng độ TSS năm 2021 (a) và năm 2024 (b).

3.4. Bản đồ phân bố không gian mức độ ô nhiễm TSS sông Thị Tính, tỉnh Bình Dương

Thông số TSS dùng để đánh giá sự thay đổi trong chất lượng nước sông Thị Tính được dự đoán từ mô hình, bản đồ phân bố nồng độ ô nhiễm nước mặt được xác định cụ thể. Kết quả thể hiện trong Hình 7 thể hiện sự phân bố theo không gian của nồng độ TSS sông Thị Tính, tỉnh Bình Dương. Mức độ ô nhiễm TSS tham khảo từ nghiên cứu [21] được phân thành 5 mức độ: (i) Ô nhiễm rất thấp, (ii) Ô nhiễm thấp, (iii) Ô nhiễm vừa, (iv) Ô nhiễm cao; (v) Ô nhiễm rất cao.

Bảng 7. Mức độ ô nhiễm khu vực nghiên cứu.

Rất thấp	Thấp	Vừa	Cao	Rất cao
< 11,3	11,3 – 34,2	34,2 – 63,9	63,9 – 104,3	>104,3

nghiên cứu đã được công bố tại [9] hệ số tương quan $R^2 = 0,663$ thì kết quả này có thể chấp nhận được. Bên cạnh đó, kết quả kiểm định mô hình khi hệ số xác định $R^2 = 67,4\%$ và $MSE = 0,17$ cho thấy nghiên cứu đã khai thác khả năng dự đoán vượt trội của phương pháp này. Một nghiên cứu tương tự trong đó tác giả chứng minh sự phù hợp mạnh mẽ giữa các giá trị dự đoán với giá trị đo đạc thực tế khi hệ số xác định $R^2 = 87\%$ [22].

4. Kết luận

Nghiên cứu từ việc sử dụng công nghệ viễn thám trong đánh giá diễn biến chất lượng nước mặt sông Thị Tính, tỉnh Bình Dương cho thấy kết quả tích cực. Lựa chọn kết quả đo đạc thực tế thông số TSS kết hợp với chỉ số nước chiết tách từ dữ liệu ảnh viễn thám Sentinel-2 là những tiêu chí đầu vào dùng để xây dựng mô hình hồi quy tuyến tính xác định nồng độ TSS trên phạm vi rộng một cách hiệu quả. Giá trị nồng độ TSS giai đoạn 2021-2024 trong nước mặt dao động trong khoảng 1,97-83,78 mg/l từ kết quả dự đoán của mô hình quy tuyến tính với hệ số xác định $R^2 = 0,675$. Kết quả thể hiện trên bản đồ cho thấy, giai đoạn 2021-2024 tại điểm hợp lưu suối Đá Yêu với sông Thị Tính và điểm hợp lưu suối Hồ Muồng với Thị Tính có nồng độ TSS ô nhiễm giảm đáng kể. Tại khu vực hạ lưu sông Thị Tính nồng độ ô nhiễm TSS có xu hướng tăng lên, sự phân bố nồng độ TSS cao ở một số phụ lưu chịu sự ảnh hưởng của hoạt động công nghiệp, nông nghiệp và đô thị.

Bên cạnh kết quả đạt được, nghiên cứu cũng có mặt hạn chế nhất định, hệ số xác định R^2 tuy ở mức chấp nhận được nhưng độ chính xác vẫn chưa cao. Sự phân bố nồng độ TSS chịu sự ảnh hưởng lớn từ các hoạt động công nghiệp, nông nghiệp và đô thị nhưng nghiên cứu chưa thể định lượng chính xác mức độ tác động của từng loại hình hoạt động và từng nguyên nhân cụ thể.

Nghiên cứu cũng có thể nhân rộng quy mô khi bổ sung thêm các mẫu thực tế để cải thiện độ chính xác của mô hình hồi quy và tăng hệ số xác định R^2 . Đồng thời, mở rộng nghiên cứu đến các sông lớn khác như sông Sài Gòn, sông Đồng Nai trên địa bàn tỉnh Bình Dương để tạo nên bức tranh toàn diện về chất lượng nước mặt của tỉnh.

Đóng góp tác giả: Đóng góp vào việc thu thập dữ liệu, trích xuất dữ liệu, phân tích dữ liệu, soạn thảo bài báo: D.B.T.; Kiểm định mô hình, giải thích dữ liệu, chỉnh sửa bản thảo: L.T.D.H.; Chỉnh sửa bài báo, giải trình số liệu: D.X.H.

Lời cảm ơn: Nhóm nghiên cứu xin chân thành cảm ơn Trung tâm Quan trắc - Kỹ thuật Tài nguyên và Môi trường tỉnh Bình Dương đã cấp cung cấp dữ liệu đo đạc thông số TSS phục vụ cho nghiên cứu này. Nghiên cứu này được tài trợ bởi Quỹ Phát triển khoa học và công nghệ Quốc gia (NAFOSTED) trong đề tài mã số: 105.06-2021.17.

Lời cam đoan: Tập thể tác giả cam đoan bài báo này là công trình nghiên cứu của tập thể tác giả, chưa được công bố ở đâu, không được sao chép từ những nghiên cứu trước đây; không có sự tranh chấp lợi ích trong nhóm tác giả.

Tài liệu tham khảo

1. Gernez, P. Toward Sentinel-2 high resolution remote sensing of suspended particulate matter in very turbid waters: SPOT4 (Take5) experiment in the Loire and Gironde estuaries. *Remote Sens.* **2015**, 7(8), 9507–9528. doi: 10.3390/rs70809507.
2. Graveline, N.; Maton, L.; Rinaudo, J.D.; Lückge, H.; Interwies, E.; Rouillard, J.; Strosser, P.; Palkaniete, K.; Taverne, D. An operational perspective on potential uses and constraints of emerging tools for monitoring water quality. *TrAC Trends Anal. Chem.* **2010**, 29(5), 378–384. doi: 10.1016/j.trac.2010.02.006.
3. Stacy, L.; Marvin, E. Satellite remote sensing of wetlands. *Wetl. Ecol. Manag.* **2002**, 10, 381–402.
4. Vũ, N.N.; Trung, LV.; Vân T.T. Đánh giá xâm nhập mặn vùng cửa sông từ dữ liệu viễn thám kết hợp quan trắc mặt đất. *Tạp chí Khoa học và Công nghệ Thủy lợi* **2018**,

- 2(2), 50–56.
5. Gorelick, N.; Hancher, M.; Dixon, M.; lyushchenko, S.; Thau, I.D.; Moore, R. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. doi: 10.1016/j.rse.2017.06.031.
 6. Jiang, W. An effective water body extraction method with new water index for Sentinel-2 imagery. *Water* **2021**, *13*(12), 1647. doi: 10.3390/w13121647.
 7. Chiến P. V. Nghiên cứu các chỉ số nhận biết nước từ ảnh Sentinel-2 trên nền Google Earth Engine: Áp dụng cho thành phố Sa Đéc, tỉnh Đồng Tháp. *Tạp chí Khoa học và Công nghệ Thủy lợi* **2020**, *60*, 1–8.
 8. Điệp, N.T.H.; Nam, T.S.; Tiễn, P.D.; Linh, T.B.; Hồ, N.; Giám, T.T.; Nhung, Đ.T.C. Phân bố hàm lượng chất rắn lơ lửng (TSS) tỉnh An Giang sử dụng ảnh viễn thám sentinel 2A. *Can Tho Univ. J. Sci.* **2021**, *57*(1), 1–7. doi: 10.22144/ctu.jvn.2021.001.
 9. Thắng, Đ.L.; Hiền, Đ.T.T.; Thức, C.V.; Hùng, T.L.; Nga, N.T.T. Xây dựng quy trình xác định một số thông số chất lượng nước từ ảnh viễn thám quang học Sentinel 2 và số liệu quan trắc. *Tạp chí Khoa học Đo đạc và bản đồ* **2022**, *53*, 57–64. doi: 10.54491/jgac.2022.53.656.
 10. Tiên, C.T.T.; Sỹ, P.C. Đánh giá hiện trạng và dự báo tải lượng các nguồn thải đổ vào sông Thị Tịch. *Tài nguyên và môi trường* **2012**, 46–47.
 11. Thắng, L.V.; Tân, L.M.; Quân, N.H.; Triết, L.M. Mô hình toán tích hợp các nguồn thải điểm và nguồn phân tán cho lưu vực sông Thị Tịch, phục vụ công tác đánh giá khả năng tiếp nhận và phục hồi chất lượng nước. *Tạp chí Khí tượng Thủy văn* **2012**, *620*, 21–26.
 12. Sỹ, P.C.; Tiên, C.T.T. Tính toán hiện trạng và dự báo khả năng tiếp nhận chất ô nhiễm hữu cơ trên sông Thị Tịch. *Tài nguyên và Môi trường* **2013**, 19–22.
 13. Tiên, C.T.T.; Hà, L.T.Q.; Sỹ, P.C. Đánh giá phân bố tải lượng ô nhiễm hữu cơ theo các tiêu vùng thuộc lưu vực sông Thị Tịch. *Tạp chí Đại học Thủ Dầu Một* **2014**, *1*, 59–66.
 14. Đỗ, T.N.; Nguyễn, T.D.M.; Trần, Q.T.; Nghiêm, V.T.; Phạm, V.M. 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* **2022**, (52), 54–64. doi: 10.54491/jgac.2022.52.598.
 15. McFeeters, S. Using the normalized difference water index (NDWI) within a geographic information system to detect swimming pools for mosquito abatement: A practical approach. *Remote Sens.* **2013**, *5*(7), 3544–3561. doi: 10.3390/rs5073544.
 16. Feyisa, G.L.; Meilby, H.; Fensholt, R.; Proud, S.R. Automated water extraction index: A new technique for surface water mapping using Landsat imagery. *Remote Sens. Environ.* **2014**, *140*, 23–35. doi: 10.1016/j.rse.2013.08.029.
 17. Fang-fang, Z.; Bing, Z.; Jun-sheng, L.; Qian, S.; Yuanfeng, W.; Yang, S. Comparative analysis of automatic water identification method based on multispectral remote sensing. *Procedia Environ. Sci.* **2011**, *11*, 1482–1487. doi: 10.1016/j.proenv.2011.12.223.
 18. Schneider, A.; Hommel, G.; Blettner, M. Linear regression analysis. *Dtsch. Arztebl. Int.* **2010**, *107*(44), 776–782. doi: 10.3238/arztebl.2010.0776.
 19. Japitana, M.V.; Burce, M.E.C. A Satellite-based remote sensing technique for surface water quality estimation. *Eng. Technol. Appl. Sci. Res.* **2019**, *9*(2), 3965–3970. doi: 10.48084/etasr.2664.
 20. Al-Mukhtar, M.; Al-Yaseen, F. Modeling water quality parameters using data-driven models, a case study Abu-Ziriq Marsh in South of Iraq. *Hydrology* **2019**, *6*(1), 24. doi: 10.3390/hydrology6010024.
 21. Đỗ, T.N.; Nguyễn, N.T.D.; Nguyễn, T.H.; Bùi, Q.T.; Lư, T.P.M.; Phạm, V. M. Ứng dụng định lượng rủi ro ô nhiễm nước mặt dựa trên dữ liệu viễn thám quang học và

phương pháp phân tích đa tiêu chí. *Tap chí khoa học Trường Đại học Sư phạm TPHCM* **2019**, 15(3), 2283–2296. doi:10.54607/hcmue.js.15.3.145(2018).

22. Thao, N.T.P.; Ha, N.T.T.; Vinh, P.Q.; Hien, T.T.; Thanh, D.X. A multivariate linear regression model for estimating chlorophyll-a concentration in Quan Son Reservoir (Hanoi, Vietnam) using Sentinel-2A Imagery. *VN J. Earth Sci.* **2024**, 360–380.

Application of Sentinel-2 Remote Sensing Imagery Combined with Multivariate Regression to Determine the Dynamics of Total Suspended Solids Concentration in Thi Tinh River, Binh Duong Province

Duong Bich Tuyen¹, Do Xuan Hong², Le Trong Dieu Hien^{3*}

¹ Master's Program in Environmental Science, Thu Dau Mot University, 06 Tran Van On Street, Thu Dau Mot City, Binh Duong Province, Vietnam; tuyenduong0965@gmail.com

² Department of Environment and Natural Resources, Nong Lam University, Ho Chi Minh City; doxuanhong@hcmuaf.edu.vn

³ Faculty of Natural Resources and Environment, Thu Dau Mot University, 06 Tran Van On Street, Thu Dau Mot City, Binh Duong Province, Vietnam; hienltd@tdmu.edu.vn

Abstract: Water quality pollution is a serious environmental issue that affects ecosystems, human health, and economic development. When pollutants such as chemicals, heavy metals, organic waste, and bacteria from industrial and agricultural sources lead to the degradation of water quality, it often results in measurable increases in concentrations of water quality indicators. In this study, we extracted water indices (AWEI, WRI, and NDWI) from Sentinel-2 satellite remote sensing data, combined with actual measured TSS parameters, to develop a regression model that identifies TSS concentration trends across the Thi Tinh River and its tributaries. The research results show that the method is highly practical, enabling quick and effective evaluation. Predicted TSS concentrations for the 2021-2024 period in surface water range from 1.97 to 83.78 mg/l. The regression model achieved an R^2 value of 67.4% and an MSE of 0.17. The method applied in this research can be utilized to assess water quality in other areas within Binh Duong Province.

Keywords: Multivariate regression; TSS; Thi Tinh River; Sentinel-2.