

Tổng quan về ứng dụng dữ liệu Sentinel-1 và Sentinel-2 trong ước tính sinh khối và các-bon rừng tại Đông Dương

Touksamphanh Maokhamphiou^{1,2}, Nguyễn Hải Hòa^{1*}, Bùi Thế Đồi^{1*}, Trần Thanh Tùng³

¹Trường Đại học Lâm nghiệp

²Bộ Nông Lâm Lào và Môi trường, Viêng Chăn, CHDCND Lào

³Quỹ Bảo vệ và Phát triển rừng tỉnh Lâm Đồng

Sentinel-1 and Sentinel-2 data capabilities for above-ground biomass and carbon stock estimation in Indochina: a systematic review

Touksamphanh Maokhamphiou^{1,2}, Nguyen Hai Hoa^{1*}, Bui The Doi^{1*}, Tran Thanh Tung³

¹Vietnam National University of Forestry

²Ministry of Agriculture, Forestry and Environment of Laos, Vientiane, Laos PDR.

³Lam Dong Forest Protection and Development Fund

*Corresponding author: hoan@vnuf.edu.vn, doibt@vnuf.edu.vn

<https://doi.org/10.55250/jo.vnuf.15.7.2026.102-111>

TÓM TẮT

Nghiên cứu tổng quan việc ứng dụng dữ liệu Sentinel-1 (SAR) và Sentinel-2 trong ước tính trữ lượng các-bon rừng tại khu vực Đông Dương - cơ sở khoa học quan trọng phục vụ quản lý rừng bền vững và phát triển các cơ chế tài chính các-bon. Dựa trên phân tích chuyên sâu 35 công trình công bố giai đoạn 2016-2023, kết quả cho thấy việc tích hợp dữ liệu radar Sentinel-1 và quang học Sentinel-2 với dữ liệu điều tra thực địa thông qua thuật toán học máy giúp cải thiện đáng kể độ chính xác của mô hình ước tính trữ lượng các-bon rừng so với việc khai thác đơn lẻ từng loại dữ liệu. Tuy nhiên, các nghiên cứu tại Đông Dương hiện còn tản mạn, thiếu các phương trình sinh khối (allometric) đặc thù và chưa hình thành hệ thống cơ sở dữ liệu giám sát rừng đồng bộ. Trên cơ sở đó, nghiên cứu đề xuất 7 nhóm giải pháp kỹ thuật nhằm tối ưu hóa hiệu suất mô hình, bao gồm: (i) chuẩn hóa thiết kế hệ thống ô điều tra thực địa; (ii) xây dựng dữ liệu cơ sở về sinh khối rừng chất lượng; (iii) tăng cường khai thác, xử lý dữ liệu viễn thám trên nền tảng điện toán đám mây; (iv) lựa chọn biến và phát triển mô hình học máy; (v) sử dụng ngoại suy không gian để lập bản đồ sinh khối và trữ lượng các-bon, (vi) đánh giá độ không chắc chắn mô hình; và (vii) phân tích biến động rừng tích hợp vào hệ thống giám sát rừng quốc gia. Đồng thời, bài viết nhấn mạnh hai định hướng ưu tiên cốt lõi: phát triển hàm allometric địa phương và xây dựng hệ thống giám sát rừng liên tục trên nền tảng điện toán đám mây.

ABSTRACT

This study provides a comprehensive review of the application of Sentinel-1 (SAR) and Sentinel-2 data in estimating forest carbon stocks in Indochina—an important scientific foundation for sustainable forest management and the development of carbon financing mechanisms. Based on an in-depth analysis of 35 studies published between 2016 and 2023, the results show that integrating Sentinel-1 radar and Sentinel-2 optical data with field survey data through machine learning algorithms significantly improves the accuracy of forest carbon estimation models compared to using each data source individually. However, research in Indochina remains fragmented, lacking region-specific allometric equations and a standardized forest monitoring database system. On this basis, the study proposes seven technical solutions to optimize model performance, including: (i) standardizing the design of field sampling plot systems; (ii) constructing high-quality baseline forest biomass

Thông tin chung:

Ngày nhận bài: 13/04/2026

Ngày phản biện: 13/05/2026

Ngày quyết định đăng: 16/07/2026

Từ khóa:

Các-bon rừng, Đông Dương, mô hình học máy, Sentinel-1, Sentinel-2, sinh khối rừng.

Keywords:

Forest biomass, forest carbon stock, Indochina, machine learning, Sentinel-1, Sentinel-2

datasets; (iii) enhancing the retrieval and processing of remote sensing data on cloud computing platforms; (iv) selecting variables and developing machine learning models; (v) applying spatial extrapolation to map biomass and carbon stocks; (vi) evaluating model uncertainty; and (vii) analyzing forest dynamics for integration into national forest monitoring systems. Furthermore, the paper emphasizes two core priority directions: developing local allometric equations and building continuous cloud-based forest monitoring systems.

1. ĐẶT VẤN ĐỀ

Áp lực từ tác động của biến đổi khí hậu, sự suy giảm đa dạng sinh học và nhu cầu phát triển bền vững đặt ra yêu cầu cấp thiết về việc định lượng chính xác, kịp thời và minh bạch các dịch vụ hệ sinh thái rừng, đặc biệt là khả năng lưu trữ các-bon. Các phương pháp ước tính sinh khối trên mặt đất (AGB) và trữ lượng các-bon truyền thống dựa vào hệ thống các ô mẫu điều tra thực địa và các hàm allometric, luôn là "tiêu chuẩn vàng" về độ chính xác tại chỗ [1]. Tuy nhiên, các phương pháp truyền thống không đáp ứng nhu cầu về độ bao phủ không gian, tần suất cập nhật liên tục và chi phí khi phạm vi nghiên cứu lớn như vùng, quốc gia hay vùng rộng lớn hơn với địa hình phức tạp.

Sự phát triển của công nghệ viễn thám đã chứng minh tính khả thi và vượt trội so với các phương pháp truyền thống. Trong số các nguồn dữ liệu, dữ liệu ảnh vệ tinh Sentinel của Cơ quan Vũ trụ châu Âu (ESA) đã tạo ra một sự thay đổi mang tính hệ thống nhờ chính sách dữ liệu mở và các đặc tính kỹ thuật đặc trưng của nó [2, 3]. Ảnh Sentinel-2 cung cấp dữ liệu quang học đa phổ với các kênh phổ "rìa đỏ- Red-edge" trên thiết bị MSI, trong khi ảnh Sentinel-1 (C-Band SAR) hoạt động với bước sóng vi ba (độ dài bước sóng 5,6 cm) có xuyên mây và không phụ thuộc vào thời tiết, mỗi loại ảnh mang lại thông tin bổ sung về thảm thực vật. Đặc biệt, các nghiên cứu gần đây có xu hướng tập trung vào việc kết hợp Sentinel-1 và Sentinel-2 (fusion data) cùng với dữ liệu điều tra thực địa (OTC) thông qua các thuật toán học máy (Machine learning, ML) tiên tiến nhằm phát triển các mô hình ước tính sinh khối và trữ lượng các-bon rừng có độ tin cậy, có khả năng khắc phục những hạn chế vốn có của từng cảm biến riêng lẻ [4].

Tuy nhiên, sự phát triển nhanh chóng xu hướng này dẫn đến một khối lượng lớn các

nghiên cứu phát triển phương pháp, các kết quả và kết luận đa dạng, đôi khi mâu thuẫn. Sự phát triển này đặt ra thách thức trong việc tổng hợp và lựa chọn phương pháp tối ưu, đặc biệt cho khu vực Đông Dương với đặc thù rừng nhiệt đới phức tạp và nguồn lực nghiên cứu còn hạn chế. Điều này gây khó khăn cho việc tổng hợp, đánh giá và rút ra các bài học kinh nghiệm, hướng đi tối ưu, đặc biệt đối với các nhà nghiên cứu và nhà quản lý tại các quốc gia đang phát triển trong khu vực, trong đó có Lào và Việt Nam, nơi nguồn lực cho nghiên cứu thử nghiệm là hạn chế. Hơn nữa, phần lớn các tổng quan hiện có mang tính toàn cầu hoặc tập trung vào các hệ sinh thái ôn đới, thiếu sự phân tích chuyên sâu về bối cảnh đặc thù của rừng nhiệt đới trong đó có Lào, nơi có tính đa dạng sinh học cao, cấu trúc rừng phức tạp, mùa mưa kéo dài, độ che phủ mây lớn với áp lực chuyển đổi sử dụng đất mạnh mẽ.

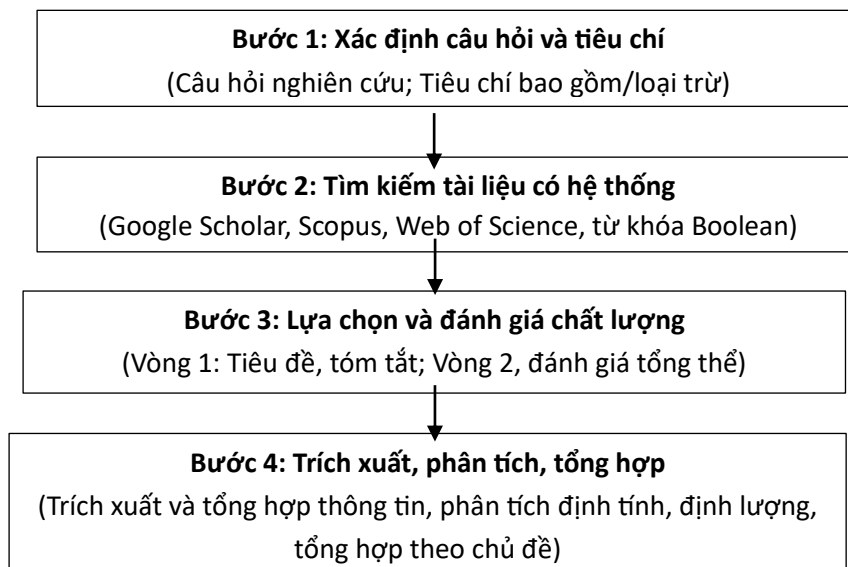
Do vậy, nghiên cứu đánh giá tổng quan về khả năng sử dụng dữ liệu viễn thám, đặc biệt là sự kết hợp giữa Sentinel-1 và Sentinel-2 trong nghiên cứu các-bon rừng là rất cần thiết. Bài báo này không chỉ đơn thuần liệt kê các nghiên cứu, mà còn áp dụng một phương pháp luận rõ ràng để tìm kiếm, lựa chọn, đánh giá và tổng hợp một cách có hệ thống các bằng chứng khoa học, từ đó đưa ra những phân tích sâu sắc và khuyến nghị thực tiễn. Các câu hỏi nghiên cứu chính của bài tổng quan này bao gồm: (1) Tính khả thi và những hạn chế chính của việc sử dụng riêng rẽ và kết hợp dữ liệu Sentinel-1 và Sentinel-2 trong ước tính sinh khối rừng trên mặt đất (AGB) và trữ lượng các-bon rừng trên mặt đất (AGC) ở khu vực nhiệt đới là gì?; (2) Các thuật toán nào được sử dụng, đặc biệt là các thuật toán học máy, trong ước tính sinh khối và trữ lượng các-bon rừng, ưu điểm và hạn chế của chúng là gì?; (3) Những khoảng trống nghiên cứu nào đang tồn tại trong việc ứng dụng các phương pháp

trên ở khu vực Đông Dương?; (4) Một khung phương pháp luận tích hợp dữ liệu và các định hướng nghiên cứu nào trong tương lai có thể áp dụng cho Việt Nam và Lào?

2. PHƯƠNG PHÁP NGHIÊN CỨU

Bài tổng quan được thực hiện theo phương

pháp tổng quan hệ thống (systematic review) có cấu trúc, tuân thủ các nguyên tắc tìm kiếm, lựa chọn và tổng hợp tài liệu một cách minh bạch và có thể lặp lại [5]. Quá trình thực hiện qua 4 bước (Hình 1).



Hình 1. Sơ đồ phương pháp nghiên cứu tài liệu tổng quan

Bước 1: Xác định câu hỏi và tiêu chí lựa chọn

Câu hỏi nghiên cứu

Tiêu chí bao gồm (Inclusion Criteria): Tài liệu được thu thập gồm các bài báo khoa học đăng trên tạp chí có phản biện độc lập, luận án tiến sĩ, và báo cáo kỹ thuật từ các tổ chức khoa học có uy tín;

Thời gian: Ưu tiên các nghiên cứu công bố từ năm 2015 trở lại đây, đó là khoảng thời gian trùng với dữ liệu Sentinel sẵn có;

Nội dung: Các nghiên cứu sử dụng dữ liệu Sentinel-1 và/hoặc Sentinel-2 để lập bản đồ rừng, phân loại trạng thái rừng, đặc biệt là ước tính sinh khối (AGB) hoặc trữ lượng các-bon rừng;

Phạm vi: Các nghiên cứu trên toàn cầu, đặc biệt các nghiên cứu tại khu vực Đông Nam Á, Đông Dương (Việt Nam, Lào, Campuchia), các khu vực có sinh thái rừng nhiệt đới tương đồng;

Tiêu chí loại trừ (Exclusion Criteria): Các bài báo chỉ sử dụng dữ liệu viễn thám khác (Landsat, MODIS) mà không có tích hợp với Sentinel; các nghiên cứu chỉ tập trung vào nông nghiệp, đô thị; các bài báo không có đầy đủ thông tin về phương pháp luận hoặc kết quả

định lượng.

Bước 2: Tìm kiếm tài liệu có hệ thống

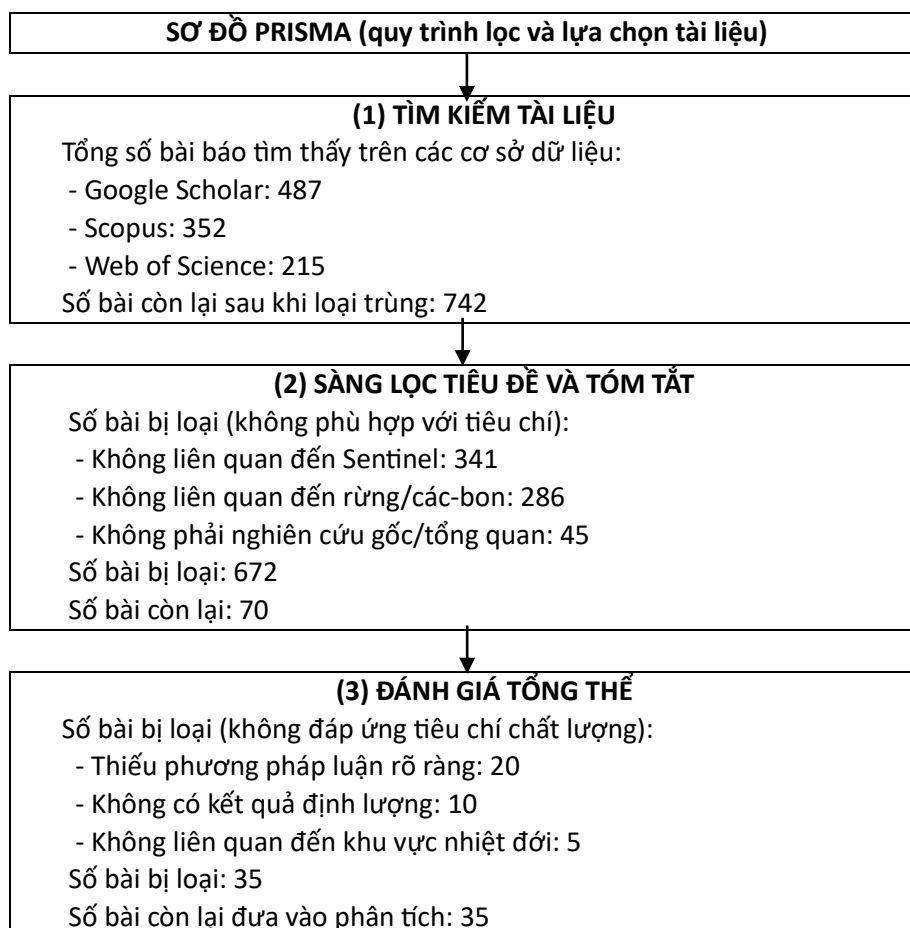
Việc tìm kiếm được thực hiện trên các cơ sở dữ liệu học thuật chính, gồm có Google Scholar, Scopus, và Web of Science. Các từ khóa được sử dụng kết hợp bằng toán tử Boolean (AND, OR): Nhóm 1 (Cảm biến- sensor): ("Sentinel-1" OR "Sentinel-2" OR "SAR" OR "Copernicus"); Nhóm 2 (Biến mục tiêu): ("Above-ground biomass" OR "AGB" OR "Above-ground carbon stocks" OR "AGC"); Nhóm 3 (Phương pháp): ("remote sensing" OR "machine learning" OR "Random Forest" OR "data fusion"); Nhóm 4 (Khu vực): ("Tropical forest" OR "Southeast Asia" OR "Vietnam" OR "Laos" OR "Cambodia" OR "Indochina"). Chẳng hạn chuỗi tìm kiếm được thiết lập như: ("Sentinel-1" OR "Sentinel-2") AND "aboveground biomass" AND "tropical forest" AND "machine learning".

Bước 3: Lựa chọn và đánh giá chất lượng tài liệu

Các kết quả tìm kiếm được sàng lọc qua hai vòng: Vòng 1 (Sàng lọc tiêu đề và tóm tắt): dựa trên tiêu chí bao gồm/loại trừ; Vòng 2 (đánh giá

tổng thể): các tài liệu qua vòng 1 được đọc tổng thể để đánh giá chi tiết về chất lượng phương pháp luận, phương pháp nghiên cứu (thiết kế điều tra thực địa, quy trình tiền xử lý dữ liệu, phương pháp xây dựng và kiểm định mô hình, đánh giá độ không đảm bảo). Các nghiên cứu

có phương pháp rõ ràng, kết quả phân tích sử dụng các chỉ số R^2 , RMSE, MAE để đo lường độ tin cậy của mô hình, có đóng góp rõ ràng vào việc trả lời các câu hỏi nghiên cứu sẽ được giữ lại để phân tích sâu.



Hình 2. Sơ đồ PRISMA (Preferred reporting items for systematic reviews and meta-analyses) mô tả quy trình tìm kiếm, sàng lọc và lựa chọn tài liệu cho bài tổng quan hệ thống

Sau quá trình sàng lọc, 35 bài báo đáp ứng đầy đủ tiêu chí và được đưa vào phân tích chuyên sâu. Quy trình lọc tài liệu được thể hiện chi tiết qua sơ đồ PRISMA (Hình 2).

Bước 4: Trích xuất, phân tích và tổng hợp thông tin

Thông tin từ các tài liệu tại bước 3 được trích xuất vào một bảng tổng hợp có cấu trúc, bao gồm: Tác giả/năm, khu vực nghiên cứu, nguồn dữ liệu viễn thám, biến đầu vào (Features), thuật toán mô hình, độ chính xác (R^2 , RMSE), các phát hiện/đóng góp chính, và hạn chế được nêu. Các thông tin này sau đó được phân tích về mặt định tính và định lượng (nếu có thể) để

rút ra các xu hướng, so sánh hiệu quả phương pháp, xác định mâu thuẫn trong kết quả và tổng hợp thành các chủ đề chính trình bày trong bài báo này.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Đánh giá tính khả thi, ưu điểm và hạn chế dữ liệu ảnh Sentinel trong nghiên cứu các-bon rừng

Dữ liệu ảnh Sentinel-2: Đây là dữ liệu ảnh vệ tinh quang học.

Ưu điểm: Các dải Red-Edge (B5, B6, B7) là bước đột phá, cung cấp thông tin cực kỳ nhạy với hàm lượng diệp lục, tình trạng sinh lý và cấu trúc lá, vượt xa khả năng của dải NIR [6]. Điều

này được minh chứng rõ nét qua nghiên cứu của Nguyễn Hải Hòa và Bùi Thế Đồi (2023) [7] tại các hệ sinh thái rừng ngập mặn, nơi các dải Red-Edge và SWIR không chỉ đóng vai trò then chốt trong việc phân biệt các trạng thái rừng mà còn là biến số quan trọng nhất trong mô hình ước tính sinh khối. Khả năng nắm bắt những biến động tinh tế về hàm lượng diệp lục và độ ẩm tán lá đã giúp các dải phổ này vượt trội hơn hẳn so với các chỉ số thực vật truyền thống trong việc phản ánh mối quan hệ với sinh khối rừng nhiệt đới.

Hạn chế: Dữ liệu ảnh Sentinel-2 có thể bị bão hòa tín hiệu khi nghiên cứu các-bon rừng. Các chỉ số như NDVI (Normalised Difference Vegetation Index), EVI (Enhanced Vegetation Index) nhanh chóng đạt ngưỡng bão hòa ở các khu rừng có LAI (chỉ số diện tích lá - Leaf Area Index) cao và tán rừng khép kín, dẫn đến mất độ nhạy ở phạm vi sinh khối cao, phổ biến ở rừng nguyên sinh nhiệt đới [8]. Ngoài ra, dữ liệu này có thể bị ảnh hưởng của khí quyển và mây. Đây là hạn chế lớn nhất của nguồn dữ liệu này khi nghiên cứu rừng nhiệt đới. Chu kỳ lặp lại của Sentinel-2 là 10 ngày đối với 1 vệ tinh riêng lẻ (Sentinel-2A, Sentinel-2B) và 5 ngày khi cả 2 vệ tinh (Sentinel-2A + Sentinel-2B) cùng hoạt động là tiềm năng lý thuyết cho nghiên cứu các-bon rừng; trên thực tế việc thu nhận được một ảnh hoàn toàn không mây cho một khu vực cụ thể vào thời điểm mong muốn là rất khó, gây gián đoạn cho giám sát liên tục.

Dữ liệu ảnh Sentinel-1: Đây là dữ liệu ảnh vệ tinh SAR.

Ưu điểm: Khả năng xuyên mây là ưu thế tuyệt đối cho vùng nhiệt đới. Tín hiệu tán xạ ngược (backscatters), đặc biệt ở phân cực chéo VH, phản ánh trực tiếp cấu trúc thể tích và độ nhám của tán rừng, những yếu tố vật lý có quan hệ trực tiếp hơn với sinh khối so với tín hiệu quang phổ. Cụ thể, nghiên cứu của Bourbia và cộng sự (2020) [9] tại khu vực rừng Địa Trung Hải đã chỉ ra rằng hệ số backscatters ở phân cực VH có tương quan chặt chẽ với sinh khối rừng (R^2 đạt 0,72), vượt trội so với phân cực VV. Điều

này được lý giải bởi cơ chế tán xạ thể tích trong tán rừng vốn liên quan trực tiếp đến khối lượng vật chất được ghi nhận tốt hơn thông qua phân cực chéo. Dữ liệu này có thể cho phép xây dựng các chuỗi thời gian liên tục để phát hiện mất rừng [10].

Hạn chế: Tín hiệu SAR kênh C bão hòa ở mức sinh khối trung bình (khoảng 100-150 tấn/ha) [11]. Điều này nghiêm trọng hơn so với bão hòa quang học, hạn chế khả năng ước tính sinh khối cho các khu rừng có trữ lượng giàu. Ngoài ra, cường độ backscatters phụ thuộc mạnh vào độ ẩm của thảm thực vật và đất. Sự biến thiên độ ẩm theo mùa có thể gây nhiễu cho mô hình nếu không được kiểm soát. Bên cạnh đó, mức độ nhiễu speckle và độ phức tạp trong xử lý đòi hỏi kiến thức chuyên sâu về xử lý SAR và các bộ lọc nhiễu phù hợp.

Từ những phân tích trên, có thể thấy rõ tính kết hợp tốt giữa hai nguồn dữ liệu (Fusion data), viễn thám quang phổ Sentinel-2 cung cấp chi tiết về thành phần loài cây và hiện trạng, trạng thái rừng (kiểu rừng), trong khi Sentinel-1 với khả năng xuyên qua tán rừng, xuyên mây vượt trội đạt hiệu quả "chiều sâu" về đặc trưng cấu trúc rừng và độ ẩm, những yếu tố vật lý mà mắt thường và cảm biến quang học khó có thể chạm tới. Khó có một nguồn dữ liệu đơn lẻ nào là tối ưu; chính sự kết hợp bổ trợ cho nhau giữa hai loại viễn thám đã tạo thành nền tảng vững chắc cho tích hợp dữ liệu đa nguồn, mở ra hướng tiếp cận toàn diện và chính xác hơn trong ước tính sinh khối và các-bon rừng nhiệt đới.

3.2. Khả năng tích hợp dữ liệu Sentinel-1 và Sentinel-2 với mô hình học máy (Machine Learning)

Từ việc tổng hợp và phân tích các nghiên cứu trước cho thấy Mô hình kết hợp (Sentinel-1 + Sentinel-2) luôn cho kết quả vượt trội so với mô hình sử dụng đơn lẻ, với sự cải thiện về R^2 và giảm RMSE có ý nghĩa thống kê [4, 12]. Tính khả thi của việc tích hợp hai dữ liệu này đã được chứng minh. Cụ thể:

Giám điểm bão hòa: Thông tin cấu trúc từ

SAR (Sentinel-1) giúp phân biệt các khu vực có chỉ số thực vật (từ Sentinel-2) tương tự nhau nhưng sinh khối thực tế khác nhau do khác biệt về mật độ thân cây, cấu trúc tán. Ngược lại, thông tin quang phổ giúp hiệu chỉnh ảnh hưởng của độ ẩm lên tín hiệu SAR.

Đảm bảo tính liên tục: Trong mùa mưa, khi ảnh Sentinel-2 bị mây che phủ, dữ liệu Sentinel-1 vẫn có thể được sử dụng để ước tính hoặc nội suy, duy trì khả năng giám sát.

Tuy nhiên, mối quan hệ giữa các biến đa nguồn dữ liệu tích hợp và AGB thực địa là phức tạp, phi tuyến và đa chiều. Các mô hình hồi quy tuyến tính truyền thống chưa làm nổi bật mối tương quan giữa dữ liệu viễn thám với sinh khối và các-bon rừng. Các nghiên cứu đã chỉ ra mô hình học máy (Machine Learning-ML) chính là chìa khóa để khai phá mối quan hệ tiềm ẩn này. Kết quả so sánh về ưu, nhược điểm của các thuật toán ML được tổng hợp tại Bảng 1.

Bảng 1. Các thuật toán học máy sử dụng để ước tính AGB từ dữ liệu viễn thám đa nguồn

Thuật toán	Ưu điểm chính	Hạn chế	Ứng dụng tiêu biểu trong rừng nhiệt đới	Khuyến nghị cho khu vực nghiên cứu	Tài liệu tham khảo
Random Forest (RF)	- Hạn chế overfitting tốt (Bagging). - Hiệu suất tính toán nhanh. - Cung cấp chỉ số Variable Importance, chỉ số về mức độ đóng góp vào mô hình. - Dễ sử dụng, ít tham số tinh chỉnh.	- Khó ngoại suy cho giá trị ngoài phạm vi huấn luyện. - Có thể kém hiệu quả với các quan hệ liên tục cực kỳ phức tạp.	Phạm Thị Lan Hương và cộng sự (2020) [12] đã ứng dụng RF để kết hợp Sentinel-1 và Sentinel-2 ước tính sinh khối rừng ngập mặn tại Cần Giờ, Việt Nam, đạt độ chính xác vượt trội ($R^2 = 0,85$), cao hơn 20% so với mô hình đơn nguồn.	Lựa chọn tối ưu ban đầu do tính ổn định, khả năng giải thích cao và phù hợp với số lượng mẫu nhỏ. Đặc biệt hữu ích khi cần xác định một số yếu tố sinh thái điển hình (cấu trúc rừng hay trạng thái thực vật...) những chỉ tiêu thể hiện hấp thụ các-bon.	[12, 13]
Gradient Boosting (XGBoost, LightGBM)	- Thường đạt độ chính xác số học cao hơn so với thuật toán khác. - Xử lý tốt các quan hệ phi tuyến phức tạp (nhiều biến số). - Có cơ chế xử lý dữ liệu khuyết.	- Rất dễ bị overfitting nếu các tham số không được không tối ưu. - Huấn luyện chậm hơn RF. - Giải thích mô hình phức tạp hơn.	Vương Văn Thành và cộng sự (2021) sử dụng LightGBM ước tính sinh khối rừng ở Trung Quốc từ Sentinel-2 và DEM, đạt độ chính xác cao ($R^2 > 0,8$) với tốc độ xử lý vượt trội.	Thuật toán này chỉ nên áp dụng khi yêu cầu chính xác cao và có đủ dữ liệu, thời gian để thực hiện (cross-validation chặt chẽ).	[14]
Support Vector Regression (SVR)	- Hiệu quả trong không gian đa chiều. - Hiệu quả cao với số lượng mẫu ít. - Khả năng phổ quát tốt.	- Kết quả phụ thuộc nhiều vào việc chọn hàm kernel và các tham số. - Kết quả thấp với tập dữ liệu rất lớn. - Khó giải thích kết quả nghiên cứu.	Mountrakis và cộng sự (2011) trong tổng quan SVM đã chỉ ra tiềm năng của SVR trong ước lượng thông số lâm học ở rừng nhiệt đới.	Khi lượng mẫu nhỏ nên lựa chọn thuật toán hoặc cần một phương pháp khác để kiểm chứng kết quả từ RF/XGBoost, nhưng hạn chế về khả năng giải thích.	[5]

Thuật toán	Ưu điểm chính	Hạn chế	Ứng dụng tiêu biểu trong rừng nhiệt đới	Khuyến nghị cho khu vực nghiên cứu	Tài liệu tham khảo
Artificial Neural Networks (ANN)	- Khả năng học các mẫu hình phi tuyến rất phức tạp. - Tiềm năng lớn với kiến trúc Deep Learning.	- Yêu cầu lượng dữ liệu huấn luyện rất lớn. - "Hộp đen" hoàn toàn, không giải thích được. - Huấn luyện không ổn định, phụ thuộc vào khởi tạo.	Watanachaturaporn và cộng sự (2008) [15] so sánh MLP với SVM và RF trong ước lượng sinh khối từ Hyperion, cho thấy MLP có kết quả tương đương, đôi khi tốt hơn.	Phù hợp khi có cơ sở dữ liệu thực địa cực kỳ lớn và đội ngũ chuyên môn sâu. Trong bối cảnh Đông Dương hiện nay, việc ứng dụng ANN còn nhiều thách thức do yêu cầu dữ liệu và tài nguyên tính toán.	[15]

Để minh họa rõ hơn hiệu quả của các phương pháp và nguồn dữ liệu, Bảng 2 tổng hợp các chỉ số độ chính xác (R^2 , RMSE) từ các nghiên cứu tiêu biểu được phân tích trong bài tổng quan này.

Bảng 2. Tổng hợp kết quả độ chính xác từ các nghiên cứu ước tính sinh khối rừng sử dụng dữ liệu Sentinel

Tác giả/Năm	Khu vực nghiên cứu	Dữ liệu sử dụng	Thuật toán	R^2	Nguồn tham khảo
Bourbia và cộng sự (2020)	Địa Trung Hải	Sentinel-1 (VH)	Hồi quy	0,72	[9]
Phạm Thị Lan Hương và cộng sự (2020)	Cần Giờ, Việt Nam	Sentinel-1 + Sentinel-2	Random Forest	0,85	[12]
Vương Văn Thành và cộng sự (2021)	Trung Quốc	Sentinel-2 + DEM	LightGBM	> 0,80	[14]
Solberg và cộng sự (2019)	Na Uy	Sentinel-1 + Sentinel-2	Random Forest	0,74	[4]
Nguyễn Hải Hòa và Bùi Thế Đồi (2023)	Đông Dương	Sentinel-1 + Sentinel-2	Random Forest	> 0,80	[7]

Kết quả tổng hợp tại Bảng 1 và 2 cho thấy mô hình kết hợp Sentinel-1 và Sentinel-2 cho độ chính xác cao hơn so với mô hình sử dụng từng nguồn dữ liệu riêng lẻ [4, 12]. Sự cải thiện này có thể được lý giải qua hai cơ chế bổ sung chính:

Thứ nhất, Sentinel-2 cung cấp thông tin về đặc tính quang phổ của tán rừng, phản ánh trạng thái sinh lý, hàm lượng diệp lục và sức khỏe của thảm thực vật [6]. Tuy nhiên, các chỉ số thực vật như NDVI, EVI nhanh chóng đạt ngưỡng bão hòa ở các khu rừng có chỉ số diện tích lá (LAI) cao, dẫn đến mất độ nhạy ở phạm vi sinh khối lớn [8]). Ngược lại, Sentinel-1 với tín hiệu Radar phân cực chéo VH có khả năng xuyên thấu tán rừng và phản ánh cấu trúc thể tích cũng như độ nhám của tán, những yếu tố vật lý có quan hệ nhân quả trực tiếp hơn với sinh khối [9]. Do đó, thông tin cấu trúc từ

Sentinel-1 giúp phân biệt các khu vực có chỉ số thực vật từ Sentinel-2 tương tự nhau nhưng sinh khối thực tế khác nhau do khác biệt về mật độ thân cây, chiều cao và cấu trúc tán.

Thứ hai, dữ liệu Sentinel-1 có khả năng xuyên mây, đảm bảo tính liên tục của dữ liệu trong mùa mưa, một lợi thế vượt trội cho khu vực nhiệt đới [10]. Trong khi đó, Sentinel-2 cung cấp thông tin chi tiết về bề mặt khi điều kiện thời tiết thuận lợi. Sự kết hợp này giúp duy trì khả năng giám sát quanh năm và giảm thiểu sai số do thiếu dữ liệu trong các giai đoạn mây che phủ dày đặc. Ngoài ra, thông tin quang phổ từ Sentinel-2 có thể hỗ trợ hiệu chỉnh ảnh hưởng của độ ẩm lên tín hiệu SAR, vốn là một yếu tố gây nhiễu quan trọng đối với dữ liệu Radar. Như vậy, chính sự bổ sung thông tin về cấu trúc (từ Sentinel-1) và quang phổ (từ Sentinel-2) đã tạo ra một mô hình ước tính bền

vững và chính xác hơn, đặc biệt hiệu quả trong việc khắc phục hiện tượng bão hòa tín hiệu ở các ngưỡng sinh khối cao và đảm bảo tính liên tục của chuỗi dữ liệu trong điều kiện thời tiết bất lợi.

3.3. Khoảng trống sử dụng dữ liệu Sentinel-1 và Sentinel-2 trong ước tính sinh khối, các-bon rừng

Sử dụng dữ liệu Sentinel ước tính sinh khối và trữ lượng các-bon rừng đã phát triển mạnh mẽ. Tuy nhiên, ngoài một số kết quả chính đạt được, chúng còn có những điểm tồn tại sau:

Mặc dù đã có một số nghiên cứu tiên phong đã chứng minh ưu điểm của việc tích hợp Sentinel-1 và Sentinel-2 [4, 12], việc ứng dụng rộng rãi và có hệ thống phương pháp tích hợp này tại khu vực Đông Dương, đặc biệt là ở Lào, vẫn còn rất hạn chế. Các nghiên cứu hiện tại chủ yếu tập trung vào từng loại dữ liệu riêng lẻ [6, 9] hoặc mới chỉ dừng lại ở việc khai thác các chỉ số thực vật và backscatters cơ bản, trong khi các chỉ số kết cấu (texture indices) từ cả Sentinel-1 và Sentinel-2, được chứng minh có hiệu quả và tiềm năng trong ước tính sinh khối, song chưa được khai thác triệt để rộng rãi [12].

Hạn chế về dữ liệu thực địa và hàm allometric đặc thù. Các nghiên cứu hiện nay trong khu vực chủ yếu dựa vào các hàm allometric toàn cầu [1] mà chưa có sự hiệu chỉnh hoặc xây dựng riêng cho các kiểu rừng, trạng thái cũng như mục đích sử dụng, cụ thể như các vườn quốc gia. Chave và cộng sự (2019) [16] nhấn mạnh rằng sai số từ hàm allometric có thể mang lại hiệu quả và nói lên độ tin cậy của mô hình viễn thám. Điều này, đặt ra yêu cầu cấp thiết về việc thiết kế các ô mẫu thực địa đại diện cho các ngưỡng ước tính sinh khối và phát triển các hàm allometric phù hợp với điều kiện rừng cho Lào.

Ứng dụng các thuật toán học máy còn thiếu tính đồng bộ và chưa giải quyết triệt để hiện tượng bão hòa. Mặc dù các thuật toán như Random Forest đã được khẳng định về hiệu suất [13], việc so sánh và tối ưu hóa giữa các thuật toán (RF, Gradient Boosting, SVR) cho từng điều kiện cụ thể, chưa được áp dụng tại khu vực. Đặc biệt, hiện tượng bão hòa tín hiệu ở khu vực rừng có sinh khối cao, một thách

thức chung khi sử dụng dữ liệu quang học và SAR kênh C [8, 11] chưa được phân tích và đề cập trong các mô hình, dẫn đến việc ước tính thiếu chính xác ở các khu rừng giàu trữ lượng.

Chưa có hệ thống giám sát biến động các-bon liên tục dựa trên chuỗi thời gian thực. Hầu hết các nghiên cứu mới chỉ dừng lại ở việc xây dựng bản đồ sinh khối cho một thời điểm. Trong khi đó, dữ liệu Sentinel cho phép xây dựng chuỗi thời gian liên tục từ năm 2015 đến nay, tạo cơ hội phát hiện sớm các vụ mất rừng và suy thoái rừng [10]. Việc chưa khai thác tiềm năng này là một khoảng trống lớn, ảnh hưởng đến khả năng đáp ứng các yêu cầu báo cáo REDD+ và tham gia thị trường tín chỉ các-bon tại khu vực nghiên cứu.

Chưa đánh giá độ không chắc chắn các mô hình ước tính (uncertainty assessment) một cách toàn diện. Các kết quả ước tính hiện nay thường được công bố dưới dạng giá trị duy nhất mà không kèm theo khoảng tin cậy. Trong khi đó, sai số có thể tích lũy từ nhiều nguồn, gồm sai số đo đạc thực địa, sai số từ hàm allometric equations, sai số mô hình viễn thám và sai số bản đồ phân loại. Chave và cộng sự (2019) [16] chỉ ra rằng việc bỏ qua các bước đánh giá độ không chắc chắn sẽ làm giảm đáng kể độ tin cậy của kết quả ước tính, đặc biệt khi sử dụng cho các báo cáo kiểm kê khí nhà kính hoặc các giao dịch tín chỉ các-bon.

Những tồn tại với các nghiên cứu được nêu trên có mối quan hệ chặt chẽ với nhau, từ dữ liệu thứ cấp (thực địa, hàm allometric equations) đến phương pháp xử lý (kết hợp dữ liệu, mô hình ML) cho đến sản phẩm đầu ra là trữ lượng các-bon. Việc nhận diện rõ các khoảng trống này là cơ sở để định hướng các nghiên cứu tiếp theo nhằm xây dựng một quy trình ước tính sinh khối và các-bon rừng có độ tin cậy cao, phù hợp với điều kiện thực tế tại khu vực Đông Dương.

3.4. Định hướng nghiên cứu các-bon rừng trong thời gian tới

Trên kết quả tổng hợp tài liệu, một số cách tiếp cận trong nghiên cứu các-bon rừng trong thời gian tới cũng như nâng cao hiệu quả trong việc ước tính sinh khối, trữ lượng các-bon rừng từ dữ liệu viễn thám đa nguồn và thực địa trên

nền tảng điện toán đám mây và các thuật toán tiên tiến, nghiên cứu đưa ra một số hàm ý (implications) và đề xuất như sau:

- Chuẩn hóa việc thiết kế hệ thống ô (OTC) mẫu và điều tra thực địa: Hệ thống OTC nghiên cứu kiểm chứng nên phân bố đại diện cho các kiểu rừng, trạng thái rừng tương ứng với từng ngưỡng sinh khối, trữ lượng các-bon, đồng thời tính đến khả năng tiếp cận thực tế. Kết hợp dữ liệu OTC với các tuyến điều tra nhanh hoặc sử dụng công nghệ UAVs để giảm chi phí và tăng độ bao phủ cho các điểm, OTC mẫu trên thực địa.

- Xây dựng bộ cơ sở dữ liệu sinh khối chất lượng: Hiệu chỉnh các hàm allometric toàn cầu [1] bằng cách thu thập kết quả nghiên cứu tỷ trọng gỗ thông qua các chỉ số đo đếm lâm học truyền thống hay xác định trọng lượng riêng, chiều cao vút ngọn các loài ưu thế tại khu vực nghiên cứu và xác định chính xác thông số hiệu chỉnh nhằm đạt độ chính xác nghiên cứu so với kết quả thực địa.

- Khai thác, xử lý và trích xuất tối đa đặc trưng mã nguồn mở trên nền tảng điện toán đám mây: sử dụng nền tảng điện toán đám mây GEE (Google Earth Engine) để xử lý dữ liệu ảnh viễn thám, như ảnh Sentinel-1 (hiệu chỉnh, lọc nhiễu, trích xuất backscatter VH/VV, tính toán chỉ số kết cấu) và ảnh Sentinel-2 (hiệu chỉnh khí quyển, tính chỉ số thực vật như NDVI, EVI, NDRE). Các bước chuẩn hóa giúp này giúp giảm khối lượng công việc đối với nguồn dữ liệu lớn (big data) và thời gian xử lý, đồng thời hạn chế những sai số hệ thống.

- Lựa chọn biến cho mô hình và xây dựng mô hình học máy riêng biệt: ưu tiên Random Forest (RF) làm thuật toán nền tảng nhờ tính ổn định, ít tham số và dễ giải thích thông qua các biến chỉ số quan trọng Variable Importance. Khi cần tối ưu độ chính xác, có thể thử nghiệm với XGBoost/LightGBM kèm tối ưu siêu tham số chặt chẽ.

- Ngoại suy và lập bản đồ sinh khối, bản đồ trữ lượng các-bon: sử dụng mô hình đã được xây dựng dựa trên kết quả kiểm chứng để ngoại suy toàn bộ khu vực nghiên cứu, tạo bản đồ có độ phân giải 10–20 m, thực hiện trên nền tảng đám mây để đảm bảo tính tái lập và khả năng xử lý dữ liệu lớn.

- Kiểm tra, đánh giá sai số cho phép: Áp dụng phương pháp bootstrapping hoặc kiểm định chéo để ước lượng khoảng tin cậy cho các ước tính, cung cấp bức tranh toàn diện về độ tin cậy, yếu tố then chốt cho báo cáo REDD+ và giao dịch tín chỉ các-bon.

- Đánh giá biến động và tích hợp vào hệ thống giám sát quản lý trữ lượng các-bon rừng: xây dựng chuỗi thời gian sinh khối/các-bon từ kho dữ liệu Sentinel (2015–nay) để phát hiện sớm về mất rừng, suy thoái rừng, thêm rừng. Sản phẩm định dạng GeoTIFF, Shapefile, sẵn sàng tích hợp vào các phần mềm GIS mã nguồn mở (QGIS) phục vụ quản lý rừng tại địa phương.

Về ứng dụng thực tiễn xác định trữ lượng các-bon tại VQG Nam Ka Đỉnh, Lào PDR, để triển khai việc nghiên cứu các-bon rừng có hiệu quả, nghiên cứu đề xuất ba nhóm ưu tiên:

- Thực hiện nghiên cứu thực nghiệm ước tính các-bon rừng tại Vườn Quốc gia Nam Ka Đỉnh, Lào PDR nhằm xây dựng mô hình ước tính sinh khối, trữ lượng các-bon rừng lá rộng thường xanh hoặc hiệu chỉnh hàm allometric cho 5–10 loài cây gỗ bản địa có phân bố chính trên các kiểu rừng và trạng thái rừng tại khu vực nghiên cứu. Đồng thời, xây dựng chuẩn hóa cấu trúc và thuật toán scripts xử lý Sentinel (Sentinel-1, Sentinel-2) trên nền tảng GEE (Google Earth Engine), tạo nền tảng chung cho cộng đồng nghiên cứu trong khu vực.

- Hình thành cơ sở dữ liệu sinh khối, trữ lượng các-bon trên thực địa (VQG Nam Ka Đỉnh) có tính chia sẻ; phát triển nền tảng WebGIS để giám sát biến động rừng, biến động trữ lượng các-bon gần thời gian thực, tích hợp với các lớp thông tin quản lý để hỗ trợ công tác kiểm lâm và chính quyền địa phương.

- Tích hợp dữ liệu Sentinel với công nghệ viễn thám tiên tiến (LiDAR không gian GEDI, ICESat-2) để nâng cao độ chính xác việc ước tính trữ lượng các-bon ở các khu rừng có sinh khối cao, cấu trúc phức tạp tương đồng với Vườn quốc gia Nam Ka Đỉnh. Thiết lập cơ chế cung cấp số liệu các-bon chính thức, có độ tin cậy cao, đáp ứng yêu cầu của các thị trường các-bon tự nguyện và cam kết quốc tế về giảm phát thải.

4. KẾT LUẬN

Bài viết tổng quan hệ thống này đã tổng hợp và phân tích về khả năng sử dụng dữ liệu viễn thám Sentinel và các thuật toán học máy (ML) trong ước tính sinh khối và trữ lượng các-bon rừng. Việc lựa chọn dữ liệu Sentinel-1, Sentinel-2, hay tích hợp hai dữ liệu cùng với dữ liệu thực địa cho phép ước tính sinh khối và trữ lượng các-bon rừng với độ tin cậy cao. Mô hình Random Forest được các nghiên cứu ứng dụng có tính khả thi và đảm bảo độ tin cậy. Tuy nhiên, ưu điểm của sự tích hợp hai nguồn dữ liệu này sẽ không phát huy được nếu không giải quyết các khoảng trống then chốt, đặc biệt là việc xây dựng cơ sở dữ liệu nền đáng tin cậy thông qua các hàm allometric địa phương và một thiết kế điều tra thực địa phù hợp với dữ liệu viễn thám sẵn có. Trong thời gian tới, cộng đồng các nhà khoa học và các cơ quan quản lý lâm nghiệp trong khu vực cần hợp tác chặt chẽ để phát triển các hệ thống giám sát các-bon liên tục, có độ tin cậy cao, dựa trên nền tảng dữ liệu mở và các bước nghiên cứu chuẩn hóa như đã đề xuất. Chỉ bằng cách đó, viễn thám mới thực sự trở thành công cụ đắc lực, không chỉ cho khoa học, mà còn cho công tác bảo tồn và phát triển các-bon rừng, quản lý tài nguyên rừng bền vững và thực hiện các cam kết khí hậu của các quốc gia Đông Dương.

Lời cảm ơn: Đây là một phần kết quả nghiên cứu của luận án tiến sĩ “*Nghiên cứu cơ sở khoa học sử dụng dữ liệu viễn thám để ước tính ước tính trữ lượng các-bon rừng tại VQG Nam Ka Định, CHDCND Lào*” của nghiên cứu sinh Touksanphanh Maokhamphiou. Nhóm tác giả trân trọng cảm ơn VQG Nam Ka Định đã tạo điều kiện thuận lợi cho nhóm trong quá trình điều tra, thu thập dữ liệu.

TÀI LIỆU THAM KHẢO

[1]. J. Chave, Réjou Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B., & Vieilledent, G (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*. 20(10): 3177–3190.

[2]. M. Drusch, Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., & Bargellini, P (2012). Sentinel 2: ESA's optical high resolution mission for GMES operational services. *Remote Sensing of Environment*. 120: 25–36.

[3]. R. Torres, Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., & Rostan, F. (2012). GMES

Sentinel 1 mission. *Remote Sensing of Environment*. 120: 9–24.

[4]. S. Solberg, Astrup, R., & Breidenbach, J (2019). Combining Sentinel 1 and Sentinel 2 for improved forest biomass mapping. *International Journal of Remote Sensing*. 40(7): 2461–2478.

[5]. G. Mountrakis, Im, J., & Ogole, C (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*. 66(3): 247–259.

[6]. M. Immitzer, Vuolo, F., & Atzberger, C (2016). First experience with Sentinel 2 data for crop and tree species classifications in central Europe. *Remote Sensing of Environment*. 8(3): 166–169.

[7]. Bùi Thế Đồi & Nguyễn Hải Hòa (2023). Nghiên cứu ứng dụng tổ hợp dữ liệu Sentinel 1 và Sentinel 2 trong ước tính sinh khối rừng ngập mặn tại khu vực Đông Dương, Hội thảo khoa học: Carbon rừng, Trường Đại học Lâm nghiệp.

[8]. O. Mutanga, & Skidmore, A. K. (2020). Saturation in remote sensing: A review and implications for estimating biomass. *International Journal of Remote Sensing*. 41(4): 1269–1284.

[9]. A. Bourbia, Boussema, M. R., & Habaieb, H (2020). Forest above ground biomass estimation using Sentinel 1 data in a Mediterranean forest. *Geocarto International*. 35(12): 1323–1343.

[10]. J. Reiche, Hamunyela, E., Verbesselt, J., Hoekman, D., & Herold, M (2018). Improving near real time deforestation monitoring in tropical dry forests by combining dense Sentinel 1 time series with Landsat and ALOS 2 PALSAR 2. *Remote Sensing of Environment*. 204: 147–161.

[11]. S. Enghart, Keuck, V., & Siegert, F (2011). Aboveground biomass retrieval in tropical forests – The potential of combined X and L band SAR data use. *Remote Sensing of Environment*. 115(5): 1260–1271.

[12]. L. T. H. Pham, Brabyn, L., & Ashraf, S (2020). Combining Sentinel 1 and Sentinel 2 data for mangrove mapping and biomass estimation in the Red River Delta, Vietnam. *GIScience & Remote Sensing*. 57(7): 638–659.

[13]. M. Belgiu, & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*. 114: 24–31.

[14]. Y. Wang, Ziv, G., Adami, M., de Almeida, C. A., Antunes, J. F. G., Coutinho, A. C., & Galbraith, D (2021). Mapping Brazil's forest restoration trajectories. *Nature Ecology & Evolution*. 5(4): 477–478.

[15]. P. Watanachaturaporn, Arora, M. K., & Varshney, P. K (2008). Multisource classification using support vector machines and decision trees. *Photogrammetric Engineering & Remote Sensing*. 74(2): 239–246.

[16]. J. Chave, Davies, S. J., Phillips, O. L., Lewis, S. L., Sist, P., Schepaschenko, D. & Zo Bi, I. C. (2019). Ground data are essential for biomass remote sensing missions. *Surveys in Geophysics*. 40(4): 863–880.